

AI die rendeert in het HBO

Van onderwijsdoel naar
verantwoorde opschaling

Onze reis van 4 jaar op weg naar
daadwerkelijk gebruik van AI in
het HBO-onderwijs



Mark van der Veen
Wiemer Kuik

AI die rendeert in het hbo

Van onderwijsdoel naar verantwoorde opschaling

Wiemer Kuik

Inhoud

1	Eerst vaststellen wat er werkelijk speelt	1
2	Begrip ontstaat niet door meer antwoorden	14
3	Meer oefenen waar het ertoe doet	27
4	Van weten naar professioneel handelen	40
5	Feedback die de volgende stap mogelijk maakt	53
6	Reflectie vraagt om frictie en vertrouwen	66
7	Samenwerken met zicht op ieders bijdrage	79
8	Toetsen als produceren goedkoop wordt	89
9	Kunnen handelen buiten de oefenomgeving	102
10	Eerst opnieuw diagnosticeren, dan besluiten	115
11	De kennisbron bepaalt de grens	129
12	Interactievormen doelgericht vergelijken	143
13	De context bepaalt of een toepassing past	153
14	Menselijke begeleiding doelbewust inrichten	167
15	Periodiek herbeoordelen en samen leren	179
16	Bekwaamheid groeit geleidelijk	193
17	Opschalen met blijvende menselijke regie	206
18	Van onderwijsbot naar HBO-opleiding	217
19	Praktijkcatalogus voor ontwerp en besluitvorming	222
20	Addenda	227

1 Eerst vaststellen wat er werkelijk speelt

AI maakt het verleidelijk om met de oplossing te beginnen. Een toepassing kan uitleg geven, feedback versnellen of een gesprek simuleren, vaak binnen enkele seconden. Dat is bestuurlijk riskant: wat studenten gemak biedt, kan docenten met extra controlewerk belasten. Wat tijd lijkt te besparen, kan tegelijk nieuwe kosten, afhankelijkheden en beheertaken veroorzaken.

Een overtuigende demonstratie zegt dan ook nog weinig over de waarde voor het onderwijs. Zelfs zichtbaar enthousiasme bewijst niet dat studenten beter leren of medewerkers werkelijk professionele ruimte winnen. Zonder heldere diagnose dreigt een instelling vooral een toepassing te zoeken voor beschikbare technologie.

Een verantwoorde AI-keuze begint daarom met nauwkeurig kijken naar de bestaande situatie. Wat gaat er niet goed genoeg, voor wie heeft dat gevolgen en welke verandering moet zichtbaar worden? Pas als daarop een scherp antwoord is, valt te beoordelen of AI een passende bijdrage kan leveren – en welk bewijs nodig is voordat experimenteren verantwoord wordt.

Extra controlewerk voor de docent



1.1 Van technologische kans naar onderwijsdiagnose

De eerste bestuurlijke vraag over AI hoort niet te zijn: *Wat kunnen we met deze technologie?* De eerste vraag is: *Welk probleem proberen we op te lossen?* Dat verschil lijkt klein, maar bepaalt in hoge mate of een instelling investeert in duurzame onderwijsverbetering of in een toepassing waarvoor achteraf nog een bestemming moet worden gezocht.

Technologische mogelijkheden trekken vanzelf de aandacht. Een toepassing kan binnen enkele seconden uitleg geven, oefenvragen maken, feedback formuleren of een gesprek simuleren. Dat nodigt uit tot experimenteren. Zo'n experiment kan nuttig zijn, maar het vormt nog geen onderbouwing voor invoering. Snelheid en veelvuldig gebruik door studenten krijgen pas betekenis wanneer ze gepaard gaan met een beter leerproces. Daarvoor moet eerst duidelijk zijn wat er in de bestaande situatie niet goed genoeg gaat.

Een bestuurlijk bruikbare diagnose beschrijft daarom vier zaken:

1. **welk leer- of organisatieprobleem optreedt;**
2. **voor wie dat probleem gevolgen heeft;**
3. **waardoor het probleem ontstaat of blijft bestaan;**
4. **welke verandering zichtbaar moet worden.**

Pas daarna komt de vraag of AI aan die verandering kan bijdragen.

Beschrijf het probleem als waarneembaar verschil

Een probleem als 'studenten krijgen te weinig feedback' klinkt herkenbaar, maar is nog te algemeen. Het kan betekenen dat studenten lang moeten wachten, dat de feedback te laat komt om nog te worden verwerkt, dat zij alleen feedback op eindproducten krijgen of dat docenten onvoldoende tijd hebben om meerdere versies te bekijken. Elk van deze situaties vraagt om een andere aanpak.

Een scherpere diagnose luidt bijvoorbeeld:

Eerstejaarsstudenten leveren hun beroepsproducten meestal maar één keer in voordat deze worden beoordeeld. Docenten kunnen door de groepsgrootte niet op meerdere conceptversies reageren. Studenten oefenen daardoor nauwelijks met het verwerken van feedback, terwijl juist die vaardigheid nodig is om hun werk zelfstandig te verbeteren.

Deze formulering maakt zichtbaar waar het knelpunt zit. Het gaat niet uitsluitend om een tekort aan docenttijd. Het onderwijsprobleem is dat studenten te weinig verbetercycli doorlopen. De gezochte waarde is dan ook niet primair dat een systeem sneller opmerkingen produceert, maar dat studenten vaker gericht verbeteren en gaandeweg minder ondersteuning nodig hebben.

Dat onderscheid voorkomt een veelgemaakte denkfout. Wanneer AI een docent twintig minuten werk uit handen neemt, is er mogelijk sprake van tijdwinst. Pas als die tijd

beschikbaar komt voor bijvoorbeeld persoonlijke begeleiding, onderwijsontwerp of beoordeling van complexe prestaties, ontstaat er professionele ruimte. En pas als studenten door de nieuwe werkwijze aantoonbaar beter leren handelen, is er onderwijsrendement.

Maak duidelijk voor wie het probleem bestaat

Eenzelfde situatie kan voor verschillende betrokkenen iets anders betekenen. Studenten ervaren misschien dat hulp niet beschikbaar is op het moment waarop zij vastlopen. Docenten zien dat een kleine groep veel begeleiding vraagt, terwijl andere studenten uit beeld raken. Een opleidingsmanager ziet oplopende werkdruk, studievertraging of grote verschillen tussen groepen. De instelling kan te maken krijgen met kwaliteitsrisico's, hoge uitvoeringskosten of een afnemende toegankelijkheid van het onderwijs.

Daarom hoort in een diagnose altijd een concrete doelgroep te staan. 'Onze studenten' is doorgaans te breed. Gaat het om beginnende studenten die de vaktaal nog niet beheersen? Om studenten die buiten de lesuren oefenen? Om studenten in een afstudeerfase die zelfstandig professionele afwegingen moeten maken? Of juist om docenten die bij grote groepen steeds hetzelfde type uitleg geven?

Hoe nauwkeuriger de doelgroep is omschreven, hoe beter kan worden vastgesteld of een toepassing daadwerkelijk waarde toevoegt. Een hulpmiddel dat ervaren studenten ondersteunt bij complexe afwegingen hoeft niet geschikt te zijn voor eerstejaarsstudenten die nog onvoldoende voorkennis hebben om een overtuigend maar onjuist antwoord te herkennen.

Onderzoek waarom de bestaande aanpak tekortschiet

Niet ieder knelpunt vraagt om AI. Soms ontbreekt er heldere instructie, is het curriculum overvol, zijn verantwoordelijkheden onduidelijk of sluit de toetsing niet aan op wat studenten tijdens de opleiding oefenen. Een digitaal hulpmiddel kan zo'n ontwerpprobleem zelfs verhullen. Studenten krijgen dan bijvoorbeeld meer uitleg, terwijl het werkelijke probleem is dat opdrachten tegenstrijdige aanwijzingen bevatten.

De diagnose moet daarom verder gaan dan het zichtbare symptoom. Bij achterblijvende studievoortgang kunnen onder meer voorkennis, taalvaardigheid, motivatie, planning, begeleiding, roostering en toetsontwerp een rol spelen. AI kan bij sommige oorzaken helpen, maar bij andere nauwelijks. Zonder oorzaakanalyse bestaat het risico dat de instelling een efficiënte oplossing ontwikkelt voor het verkeerde probleem.

Dat betekent niet dat vooraf alles bewezen moet zijn. Wel moet duidelijk zijn welke aannames worden onderzocht. Als een opleiding vermoedt dat studenten te weinig oefenen doordat docenten niet voldoende oefenmateriaal en feedback kunnen bieden, kan een beperkte proef uitwijzen of extra oefenmogelijkheden werkelijk worden gebruikt. Vervolgens moet worden onderzocht of het gebruik leidt tot betere prestaties, meer zelfstandigheid of een kortere feedbackcyclus. De proef toetst dan een onderwijsaannname, niet alleen de werking van de technologie.

Leg de gewenste verandering vooraf vast

Een diagnose is pas compleet wanneer de uitgangssituatie en de gewenste verandering naast elkaar staan. Zonder uitgangsmeting wordt iedere uitkomst vatbaar voor wensdenken. Een enthousiast docententeam, positieve reacties van studenten of een groot aantal gesprekken met een onderwijsbot kan dan al snel als succes worden gepresenteerd.

De gewenste verandering moet zo concreet mogelijk zijn. Denk aan:

- studenten doorlopen vóór de beoordeling vaker een volledige verbetercyclus;
- studenten herkennen en herstellen een groter deel van hun eigen fouten;
- de wachttijd voor eerste feedback neemt af;
- docenten besteden minder tijd aan herhaalde standaarduitleg en meer tijd aan complexe begeleiding;
- verschillen in toegang tot oefening en ondersteuning worden kleiner;
- de kwaliteit blijft ten minste gelijk terwijl meer studenten onderwijs krijgen.

Niet alles hoeft direct in cijfers te worden uitgedrukt. Professionele bekwaamheden vragen vaak om kwalitatieve beoordeling. Wel moet vooraf vaststaan welk gedrag, welk product of welke prestatie als aanwijzing voor verbetering geldt, wie dat beoordeelt en op welk moment.

Houd de diagnose technologie-neutraal

Een goede diagnose noemt nog geen voorkeursoplossing. Zij beschrijft bijvoorbeeld dat studenten onvoldoende gelegenheid krijgen om lastige beroepsgesprekken te oefenen, niet dat de opleiding een gesprekssimulator nodig heeft. Misschien blijkt oefenen met medestudenten, inzet van trainingsacteurs, aanpassing van het rooster of betere voorbereiding doeltreffender. AI is pas een serieuze optie wanneer aannemelijk is dat zij extra waarde biedt ten opzichte van zulke alternatieven.

Die technologie-neutraliteit is bestuurlijk belangrijk. Ze maakt het mogelijk om kosten, baten en risico's eerlijk te vergelijken. Ook voorkomt ze dat een tijdelijke proef ongemerkt verandert in een structurele afhankelijkheid van een leverancier of systeem waarvoor het probleem nooit scherp is vastgesteld.

Aan het einde van de diagnose moet een manager daarom in één alinea kunnen lezen wat er speelt, wie daarvan last heeft, waarom de huidige aanpak onvoldoende werkt en welke verbetering wordt gezocht. Lukt dat niet, dan is een besluit over een AI-toepassing te vroeg.

Vervolgens komt de vraag waar die verbetering terecht moet komen. Waarde voor studenten, ruimte voor medewerkers en resultaat voor de instelling hangen weliswaar met elkaar samen, maar zijn niet hetzelfde. Juist door die drie niveaus afzonderlijk te beoordelen, wordt zichtbaar voor wie een toepassing werkelijk rendeert – en wie vooral de kosten, risico's of het nieuwe werk draagt.

1.2 Drie niveaus van potentiële waarde

Een toepassing kan op één niveau aantrekkelijk lijken en op een ander niveau juist lasten veroorzaken. Een student krijgt bijvoorbeeld sneller feedback, terwijl docenten extra tijd kwijt zijn aan controle en correctie. Een dienst kan een proces goedkoper uitvoeren, terwijl opleidingen minder grip krijgen op gegevens of afhankelijk worden van een leverancier. En tijdwinst voor medewerkers heeft pas waarde als die tijd werkelijk beschikbaar komt én zinvol wordt besteed.

Daarom beoordeelt het batenmodel iedere toepassing op drie niveaus:

1. waarde voor studenten;
2. professionele speelruimte voor medewerkers;
3. waarde voor de instelling.

Op elk niveau worden de verwachte opbrengsten afgezet tegen de kosten, risico's, beheerlast en voorwaarden. Het gaat dus niet om drie losse lijstjes met voordelen. Het model maakt zichtbaar waar de waarde ontstaat, wie het nieuwe werk doet en welke afweging een manager uiteindelijk maakt.

Waarde voor studenten

Voor studenten ligt de belangrijkste waarde niet in het gemak waarmee AI een tekst, antwoord of product oplevert. De relevante vraag is wat studenten hierdoor beter, vaker of zelfstandiger leren doen.

Dat kan bijvoorbeeld betekenen dat zij:

- vaker met lastige beroepssituaties oefenen;
- sneller gerichte feedback krijgen;
- uitleg ontvangen die aansluit bij hun voorkennis;
- hun redenering kunnen toetsen voordat zij een product inleveren;
- verschillende handelingsmogelijkheden leren vergelijken;
- beter leren verantwoorden welke bronnen en argumenten zij gebruiken;
- toegankelijker kunnen deelnemen aan het onderwijs.

Deze opbrengsten moeten zo concreet mogelijk worden omschreven. 'Meer gepersonaliseerd leren' is daarvoor te algemeen. 'Eerstejaarsstudenten krijgen na iedere oefencasus aanwijzingen over ontbrekende stappen in hun analyse en kunnen de casus daarna opnieuw uitvoeren' is beter. Zo wordt duidelijk welke leerhandeling verandert en welk gedrag zichtbaar moet worden.

Meer gebruik is daarbij nog geen bewijs van beter leren. Studenten kunnen een toepassing vaak raadplegen omdat die snel antwoorden geeft, zonder dat hun begrip of handelingsvermogen groeit. Het gebruiksgemak kan zelfs een risico vormen wanneer studenten denkwerk overslaan, onjuiste uitleg overnemen of afhankelijk worden van voortdurende hulp.

Bij de beoordeling moet daarom niet alleen worden gekeken naar tevredenheid en gebruik, maar ook naar de kwaliteit van het handelen. Worden analyses beter onderbouwd? Herkennen studenten fouten eerder? Kunnen zij een taak later zonder hulp uitvoeren? En kunnen zij het geleerde toepassen in een nieuwe situatie? Pas zulke aanwijzingen maken aannemelijk dat er onderwijswaarde ontstaat.

Professionele speelruimte voor medewerkers

Voor medewerkers is tijdwinst vaak de eerste genoemde baat. AI kan conceptfeedback opstellen, veelgestelde vragen afhandelen, oefenmateriaal variëren of informatie uit documenten toegankelijk maken. Toch is 'de docent bespaart tijd' geen volledige batenclaim.

Er moeten minstens vier vervolgvragen worden gesteld:

- Welke medewerker wint tijd bij welke taak?
- Hoe vaak komt die taak voor en hoeveel tijd kost zij nu?
- Welk nieuw controle-, herstel- of beheerwerk ontstaat?
- Waarvoor wordt de vrijgekomen tijd daadwerkelijk gebruikt?

Dat laatste is wezenlijk. Vijf minuten minder nakijkwerk per student levert niet automatisch professionele speelruimte op. De tijd kan versnipperd raken, opgaan aan het controleren van AI-uitkomsten of eenvoudigweg worden gevuld met extra productie. Er ontstaat pas bruikbare ruimte wanneer de opleiding het werk anders organiseert en medewerkers de gewonnen tijd kunnen inzetten voor taken waarin hun deskundigheid het meeste verschil maakt.

Denk aan persoonlijke begeleiding, het voeren van moeilijke voortgangsgesprekken, het ontwerpen van betekenisvolle opdrachten, het beoordelen van professionele afwegingen en het ondersteunen van studenten die dreigen uit te vallen. AI neemt de professionele verantwoordelijkheid daarbij niet over. Zij kan routinematige delen van het werk ondersteunen, zodat menselijke aandacht gericht wordt ingezet.

Ook de verdeling van baten en lasten verdient aandacht. De docent die met een toepassing werkt, is niet altijd degene die haar onderhoudt. Bronmateriaal moet actueel blijven, onjuiste antwoorden moeten worden onderzocht, instructies moeten worden aangepast en studenten hebben uitleg nodig over verantwoord gebruik. Soms verschuift het werk van docenten naar onderwijsontwikkelaars, functioneel beheerders, informatiemanagers of privacy- en beveiligingsprofessionals. Die verschuiving hoort bij de businesscase.

Waarde voor de instelling

Op instellingsniveau gaat het om resultaten die meerdere studenten, teams of opleidingen raken en die passen bij de publieke opdracht van het hbo. Mogelijke baten zijn een betere onderwijskwaliteit, grotere toegankelijkheid, minder uitval, een doelmatiger inzet van schaarse deskundigheid of een kortere doorlooptijd van

ondersteunende processen.

Maar bij opschaling nemen niet alleen de opbrengsten toe. Ook fouten, afhankelijkheden en beheertaken nemen toe. Een toepassing die in één cursus met een betrokken docent goed werkt, kan bij instellingsbrede invoering duur blijken door koppelingen, ondersteuning, scholing, contractbeheer, gegevensbescherming en voortdurende kwaliteitscontrole. Bovendien kan standaardisering botsen met verschillen tussen opleidingen en beroepscontexten.

Daarom vraagt waarde voor de instelling om aantoonbaarheid én duurzaamheid. Een bestuurder moet kunnen vaststellen:

- welk instellingsdoel wordt ondersteund;
- hoeveel studenten, medewerkers of processen daadwerkelijk worden bereikt;
- of de opbrengst na de proefperiode standhoudt;
- welke structurele kosten en afhankelijkheden ontstaan;
- of de toepassing past bij publieke waarden en institutionele verantwoordelijkheden;
- wie eigenaar is van kwaliteit, beheer en besluitvorming.

Een succesvolle proef is in dit opzicht slechts een eerste aanwijzing. Instellingswaarde ontstaat pas wanneer de toepassing ook buiten de oorspronkelijke groep verantwoord werkt, bestuurbaar blijft en financieel en organisatorisch vol te houden is.

De drie niveaus in één afweging

De drie niveaus versterken elkaar soms. Meer oefenmogelijkheden kunnen studenten beter voorbereiden, waardoor docenten hun begeleiding gericht inzetten en de opleiding met dezelfde capaciteit meer kwaliteit biedt. Maar zulke samenhang mag niet worden verondersteld. Zij moet worden aangetoond.

Niveau	Centrale vraag	Mogelijke opbrengst	Veelvoorkomende verborgen last
Student	Wat leert of kan de student aantoonbaar beter?	Meer oefening, snellere feedback, betere toegankelijkheid	Afhankelijkheid, oppervlakkig leren, onjuiste informatie
Medewerker	Welke bruikbare professionele ruimte ontstaat?	Minder routinewerk, gerichtere begeleiding, beter onderwijsontwerp	Controle, correctie, scholing en onderhoud
Instelling	Welk duurzaam resultaat wordt bereikt?	Kwaliteit, toegankelijkheid, doelmatigheid of continuïteit	Beheer, contracten, gegevensrisico's en leveranciersafhankelijkheid

Een evenwichtige beslissing vereist niet dat iedere toepassing op alle drie de niveaus onmiddellijk winst oplevert. Een voorziening voor een kleine groep studenten kan bijvoorbeeld waardevol zijn zonder een financiële besparing voor de instelling. Andersom is een goedkoper proces niet vanzelf aanvaardbaar wanneer de onderwijskwaliteit afneemt of medewerkers de gevolgen niet kunnen overzien.

De kern is dat de afweging expliciet wordt gemaakt. Wie ontvangt de baat? Wie draagt de lasten? Welke voorwaarden moeten eerst worden geregeld? En welke aanwijzingen zouden de aanvankelijke verwachting kunnen weerleggen?

Juist bij een diagnose zijn die vragen belangrijk. Een hulpmiddel dat studenten snel een oplossing geeft, kan efficiënt en aantrekkelijk ogen, maar weinig zichtbaar maken van hun uitgangssituatie. Soms ontstaat de meeste waarde dan niet door een antwoord toe te voegen, maar door betere vragen te stellen.

1.3 Uit mijn onderwijspraktijk: de bot die te snel een antwoord gaf

Mijn eerste diagnosebot deed precies wat veel AI-toepassingen van nature lijken te doen: hij probeerde behulpzaam te zijn. Een student beschreef een vraagstuk, de bot ordende de informatie en kwam vrijwel direct met mogelijke oorzaken, oplossingen en vervolgstappen. De antwoorden waren vlot, overzichtelijk en op het eerste gezicht bruikbaar.

Toch ging er iets wezenlijks mis. De bot hielp de student vooral om snel verder te komen, terwijl ik juist wilde vaststellen waar de student stond.

Dat verschil bleek belangrijker dan ik aanvankelijk dacht. Wanneer een student bijvoorbeeld aangaf moeite te hebben met een programmeeropdracht, stelde de bot al snel voor om de code anders op te bouwen. Bij een projectvraag kwam er meteen een stappenplan. En wanneer een student onzeker was over een beroepsproduct, verscheen binnen enkele seconden een keurige lijst met verbeterpunten.

Daarmee leek het probleem opgelost. Alleen wist ik nog steeds niet wat de student zelf begreep, welke aanpak al was geprobeerd en waar de redenering vastliep. Ook bleef onduidelijk of er sprake was van een kennistekort, een verkeerde aanname, gebrekkige planning of simpelweg onzekerheid. De bot produceerde dus wel een antwoord, maar leverde nauwelijks een diagnose.

Van oplossen naar onderzoeken

In de eerste versie had ik de bot feitelijk de verkeerde opdracht gegeven. Ik had vooral beschreven bij welke inhoud hij studenten moest helpen en welke uitkomsten bruikbaar waren. Veel minder aandacht had ik besteed aan de vragen die hij eerst moest stellen.

Dat veranderde ik. De bot mocht niet langer meteen een oplossing voorstellen, maar moest de uitgangssituatie stap voor stap verkennen. Hij vroeg voortaan onder meer:

- Wat probeert u precies te bereiken?
- Wat lukt al wel?
- Waar loopt u voor het eerst vast?
- Welke aanpak hebt u zelf geprobeerd?
- Op grond waarvan koos u die aanpak?
- Welke informatie of kennis denkt u nog te missen?
- Hoe zou u kunnen controleren of uw redenering klopt?

Niet iedere vraag hoefde altijd te worden gesteld. De bedoeling was juist dat de bot doorvroeg op wat de student antwoordde. Een algemene opmerking als 'ik snap de opdracht niet' moest worden teruggebracht tot iets waarmee de student en de docent verder konden. Ging het om onbekende begrippen, onduidelijke beoordelingscriteria, moeite met het plannen van de taak of onzekerheid over de kwaliteit van het eigen werk?

De gesprekken werden daardoor niet alleen langer, maar vooral informatiever. Studenten moesten hun eigen aanpak onder woorden brengen voordat zij hulp kregen. Dat maakte zichtbaar waar hun begrip ophield en waar een aanname of werkwijze moest worden onderzocht.

Ik zag bovendien dat het probleem soms anders lag dan de eerste vraag deed vermoeden. Een student die om uitleg vroeg, bleek de theorie best te kennen, maar kon die nog niet toepassen in een nieuwe situatie. Een ander vroeg om inhoudelijke feedback, terwijl vooral de opdrachtinterpretatie niet klopte. Weer een ander had geen extra uitleg nodig, maar een kleiner en overzichtelijker eerste stapje.

Dat zijn didactisch verschillende situaties. Wie ze allemaal met dezelfde uitleg of hetzelfde stappenplan beantwoordt, verleent misschien een dienst, maar biedt geen gerichte ondersteuning.

De bot werd minder indrukwekkend en meer bruikbaar

De aangepaste bot voelde aanvankelijk minder spectaculair. Hij gaf niet meteen een compleet antwoord en kwam niet binnen enkele seconden met een ogenschijnlijk sluitende oplossing. Soms vonden studenten dat zelfs wat omslachtig. Zij wilden immers verder met hun opdracht en verwachtten van een digitaal hulpmiddel vooral snelheid.

Toch zat juist in die vertraging de onderwijswaarde. De student moest eerst laten zien wat hij al wist, welke keuze hij had gemaakt en waarom. De bot nam het denken niet onmiddellijk over, maar hielp dat denken zichtbaar te maken.

Dat betekende overigens niet dat iedere interactie goed verliep. Soms stelde de bot te veel vragen achter elkaar. Dan kreeg het gesprek het karakter van een vragenlijst en haakten studenten af. In andere gevallen bleef hij doorvragen terwijl het probleem al voldoende duidelijk was. Ook kwam het voor dat een student met korte antwoorden probeerde zo snel mogelijk bij de oplossing uit te komen.

Ik heb de opzet daarom opnieuw aangepast. De bot stelde voortaan één gerichte vraag per keer, vatte tussendoor samen wat er volgens hem speelde en vroeg de student die samenvatting te controleren. Pas wanneer de uitgangssituatie voldoende scherp was, mocht hij een volgende leerhandeling voorstellen: een begrip opzoeken, een redenering testen, een klein deelprobleem oplossen, een voorbeeld vergelijken of de hulp van een docent inroepen.

Daarmee werd ook de grens van de bot duidelijker. Hij kon een eerste verkenning ondersteunen, maar niet zelfstandig bepalen wat een student nodig had. Een plausibele samenvatting kon immers nog steeds verkeerd zijn. Bovendien zijn aarzeling, frustratie en onzekerheid niet altijd goed uit tekst af te leiden. De student moest de diagnose kunnen corrigeren en de docent bleef nodig wanneer het vraagstuk complex, hardnekkig of persoonlijk werd.

De bestuurlijke les

Deze eerste praktijkstap leverde mij geen bewijs op dat studenten hierdoor aantoonbaar beter leerden. Daarvoor waren systematische gebruiksgegevens, evaluatie en vergelijking nodig. Wel werd een ontwerpverschil zichtbaar dat later veel bepalender bleek dan de keuze voor een bepaald AI-model: probeert de toepassing het vraagstuk snel af te handelen, of helpt zij eerst vast te stellen wat er werkelijk aan de hand is?

Voor managers is dat geen detail in de technische inrichting. Het bepaalt welke waarde een toepassing überhaupt kan leveren. Een bot die direct antwoorden geeft, kan studenten het gevoel geven sneller geholpen te worden. Een bot die zorgvuldig doorvraagt, kan informatie opleveren over begripsproblemen, gekozen strategieën en benodigde ondersteuning. Die tweede opbrengst is minder opvallend, maar kan voor het onderwijs veel waardevoller zijn.

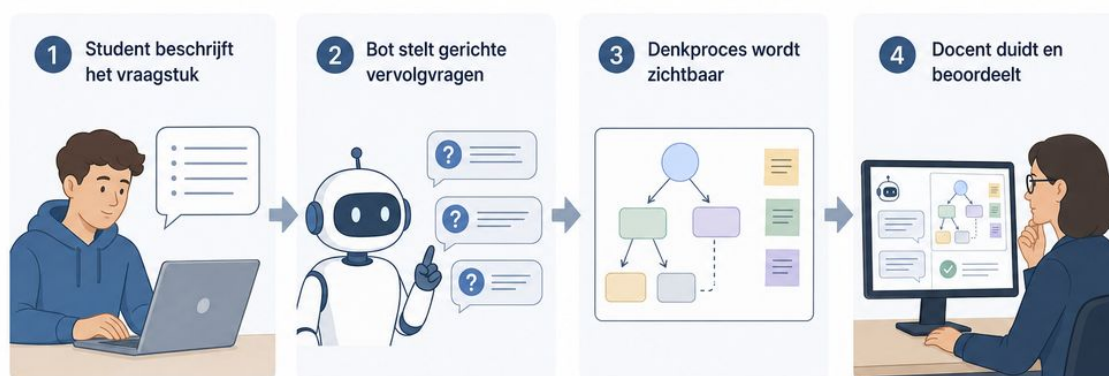
Daar staat nieuw werk tegenover. De vragen moeten didactisch worden ontworpen, gesprekken moeten worden gevolgd en onjuiste of ongepaste diagnoses moeten kunnen worden gecorrigeerd. Ook moet duidelijk zijn welke gegevens worden vastgelegd, wie die mag inzien en of gespreksinformatie wel geschikt is voor onderwijsbeslissingen. Een diagnosegesprek dat bedoeld is om te helpen, mag niet ongemerkt veranderen in een formele beoordeling.

De belangrijkste les was dan ook eenvoudig: begin niet met de vraag welk antwoord de bot moet geven. Bepaal eerst welke uitgangssituatie zichtbaar moet worden, welke vragen daarvoor nodig zijn en wie uiteindelijk beoordeelt wat de antwoorden betekenen. Pas daarna ontstaat er een grondslag voor gerichte uitleg, oefening, feedback of begeleiding.

Zo werd de eerste diagnosebot geen los hulpmiddel dat vóór het onderwijsproces stond, maar het begin van een samenhangender onderwijsconcept. Niet het snelle antwoord vormde het vertrekpunt, maar een beter onderbouwde keuze voor wat de student vervolgens moest doen.

De bot die doorvraagt

Van antwoord naar diagnose



1.4 Het bestuurlijke kader in één oogopslag

Die keuze kan alleen goed worden gemaakt als de belangrijkste vragen in samenhang worden beantwoord. Een veelvoorkomende fout is dat een instelling afzonderlijk kijkt naar onderwijskwaliteit, techniek, privacy, kosten en personele inzet. Daardoor kan een toepassing op ieder onderdeel afzonderlijk aanvaardbaar lijken, terwijl de totale afweging toch negatief uitvalt. Andersom kan een beheersbaar risico een kansrijk proefproject onnodig blokkeren wanneer de mogelijke onderwijswaarde buiten beeld blijft.

Het bestuurlijke kader brengt daarom acht vragen bijeen. Ze hoeven in deze fase nog niet allemaal definitief te worden beantwoord. Wel moet duidelijk zijn wat al bekend is, wat nog wordt verondersteld en welke informatie ontbreekt.

Onderdeel	Kernvraag	Wat minimaal duidelijk moet zijn
Onderwijsdoel	Welk onderwijs- of organisatievraagstuk willen we verbeteren?	Het concrete probleem en de gewenste verandering
Doelgroep	Voor wie moet waarde ontstaan?	De betrokken studenten, medewerkers en organisatieonderdelen

Uitgangssituatie	Wat gebeurt er nu en waar loopt dat vast?	Werkwijze, prestaties, verschillen en knelpunten vóór invoering
Verwachte baten	Welke verbetering verwachten we?	De beoogde opbrengst voor student, medewerker en instelling
Bewijsniveau	Hoe stellen we vast of die verbetering optreedt?	Indicatoren, gegevensbronnen en een passende vergelijking
Lasten	Wat kost ontwikkeling, invoering en gebruik werkelijk?	Geld, tijd, deskundigheid, controle, onderhoud en ondersteuning
Risico's en voorwaarden	Wat kan er misgaan en wat moet vooraf zijn geregeld?	Onderwijskundige, juridische, organisatorische en publieke waarborgen
Menselijke verantwoordelijkheid	Wie beoordeelt, grijpt in en blijft aanspreekbaar?	Rollen, bevoegdheden en grenzen van het AI-gebruik

Drie voorlopige besluiten

Met het kader hoeft u nog geen definitief invoeringsbesluit te nemen. Aan het einde van de eerste diagnose zijn drie voorlopige uitkomsten mogelijk.

1. Niet beginnen

Begin niet wanneer het vraagstuk onvoldoende belangrijk of concreet is, de verwachte baat niet aannemelijk kan worden gemaakt of een eenvoudiger maatregel waarschijnlijk beter werkt. Hetzelfde geldt wanneer fundamentele voorwaarden ontbreken, bijvoorbeeld omdat gevoelige gegevens niet verantwoord kunnen worden verwerkt of omdat niemand de inhoudelijke verantwoordelijkheid kan dragen.

2. Nader onderzoek doen

Onderzoek verder wanneer het vraagstuk relevant is, maar de uitgangssituatie, doelgroep of bewijsbehoefte nog onvoldoende bekend is. Verzamel dan eerst gegevens, observeer het bestaande proces of spreek met studenten en medewerkers. Deze fase moet een duidelijke onderzoeksvraag en einddatum hebben; anders wordt nader onderzoek gemakkelijk uitstel zonder besluit.

3. Een afgebakend proefproject voorbereiden

Een proefproject is passend wanneer het vraagstuk helder is, AI daar aantoonbaar aan kan bijdragen en de belangrijkste risico's beheersbaar lijken. Baken doelgroep, looptijd, toepassing en onderwijscontext vooraf af. Formuleer bovendien de beoogde baten, de volledige lasten, de meetwijze, de menselijke verantwoordelijkheden en de criteria om te stoppen, aan te passen of door te gaan.

De beslisroute is daarmee eenvoudig: toets eerst of het vraagstuk belangrijk, concreet en besluitbaar is. Bepaal vervolgens welke informatie nog ontbreekt en of de voorwaarden voor verantwoord gebruik aanwezig zijn. Besluit daarna om niet te beginnen, gericht nader onderzoek te doen of een afgebakend proefproject voor te bereiden. Dat is de eerste bestuurlijke drempel: niet bewijzen dat AI iets kan, maar aannemelijk maken dat het onderzochte vraagstuk een verantwoorde proef verdient.

2 Begrip ontstaat niet door meer antwoorden

Bij begripsontwikkeling ligt die drempel verrassend hoog. Een AI-toepassing kan binnen enkele seconden een heldere uitleg geven, voorbeelden aanpassen en moeilijke begrippen vereenvoudigen. Studenten waarderen die directe hulp vaak. Toch bewijst een goed antwoord nog niet dat er begrip is ontstaan.

Dat onderscheid is bestuurlijk relevant. Als studenten vooral sneller bij een bruikbaar antwoord komen, kan de toepassing doelmatig lijken terwijl het denkwerk verschuift van de student naar de techniek. De tijdwinst is dan zichtbaar, maar de onderwijswaarde blijft onzeker. Juist wanneer een antwoord overtuigend klinkt, is het moeilijk vast te stellen of de student verbanden doorziet, een redenering kan reconstrueren en het geleerde zelfstandig kan toepassen.

De kernvraag is daarom niet hoeveel uitleg AI kan aanbieden, maar welk denken die uitleg bij de student uitlokt. Daarvoor moet eerst duidelijk zijn waar het begrip vastloopt. Pas dan valt te beoordelen of AI een passende volgende stap ondersteunt – en waaraan te zien is dat de student die stap werkelijk zelf heeft gezet.

Een antwoord zonder begrip



2.1 Waar begrip vastloopt

Begrip kan vastlopen door een misvatting, ontbrekende voorkennis of een schijnbaar correcte redenering. Elk van deze oorzaken vraagt om een andere aanpak.

In de onderwijspraktijk blijven deze verschillen gemakkelijk buiten beeld. Een ingeleverde opdracht toont het eindantwoord, maar meestal niet hoe dat antwoord tot stand kwam. Een docent ziet dat een berekening fout is, een advies onvoldoende onderbouwd blijft of een beroepsproduct niet aan de eisen voldoet. De oorzaak kan echter op verschillende plaatsen liggen: in de kennisbasis, in het verband tussen begrippen, in een redeneerstap of in de toepassing op een nieuwe situatie.

Zonder nadere diagnose krijgt iedere student al snel dezelfde uitleg. Dat is doelmatig in de betekenis van *één keer aanbieden*, maar niet noodzakelijk doelmatig in de betekenis van *begrip tot stand brengen*.

Drie oorzaken die om een andere aanpak vragen

Een eerste oorzaak is een **misvatting**. De student beschikt dan wel over een mentaal model, maar dat model klopt niet. Zo kan een student bedrijfskunde denken dat een hogere omzet vanzelf betekent dat een onderneming financieel gezonder is. Een student verpleegkunde kan aannemen dat een protocol in iedere situatie dezelfde handeling voorschrijft. Extra informatie toevoegen helpt dan maar beperkt. Nieuwe uitleg wordt namelijk geïnterpreteerd vanuit het bestaande, onjuiste model.

Bij een misvatting moet de student eerst ervaren dat de eigen verklaring niet in alle gevallen standhoudt. Daarvoor zijn een tegenvoorbeeld, een vergelijking of een gerichte vraag vaak geschikter dan een langere toelichting. Wat gebeurt er bijvoorbeeld met de financiële gezondheid wanneer de omzet stijgt, maar de kosten nog sneller toenemen? En wanneer wijkt een professional gemotiveerd af van een protocol? Zulke vragen maken de spanning in het bestaande denkmodel zichtbaar.

Een tweede oorzaak is **ontbrekende voorkennis**. De student kan een nieuwe uitleg dan niet goed verbinden met wat al bekend is. Een technisch correcte toelichting over statistische betrouwbaarheid heeft weinig waarde wanneer begrippen als steekproef, spreiding en meetfout nog onvoldoende zijn ontwikkeld. Het probleem zit dan niet in de aangeboden uitleg, maar in het ontbrekende fundament waarop die uitleg moet rusten.

Hier is teruggaan geen vertraging, maar een voorwaarde om verder te komen. De docent of toepassing moet achterhalen welke bouwstenen ontbreken en die in een bruikbare volgorde aanbieden. Dat hoeft niet te betekenen dat een heel vorig hoofdstuk opnieuw moet worden behandeld. Vaak gaat het om één cruciaal begrip of verband dat nooit stevig genoeg is geworden.

Een derde oorzaak is een **schijnbaar correcte redenering**. Dit is bestuurlijk en didactisch misschien wel de lastigste categorie. De student gebruikt passende termen, volgt een herkenbare structuur en komt soms zelfs bij het juiste antwoord uit. Toch kan de onderliggende redenering ondeugdelijk zijn. Misschien is een relevante uitzondering

genegeerd, wordt samenhang verward met oorzaak en gevolg of is een conclusie vooral gebaseerd op een patroon uit eerdere opdrachten.

Juist taalmodellen kunnen dit probleem minder zichtbaar maken. Zij produceren gemakkelijk een vloeiende en overtuigende formulering, ook wanneer de redenering gaten bevat. Een verzorgd antwoord mag daarom niet worden verward met aangetoond begrip. Om dat onderscheid te kunnen maken, moet de student tussenstappen duidelijk maken, alternatieven afwegen en uitleggen waarom een conclusie in deze situatie gerechtvaardigd is.

Waar het begrip vastloopt	Wat zichtbaar wordt	Passende eerste reactie
Misvatting	De student gebruikt een onjuist denkmodel consequent	Laat het model botsen met een tegenvoorbeeld of afwijkende situatie
Ontbrekende voorkennis	De student kan nieuwe informatie niet verbinden met noodzakelijke basisbegrippen	Breng de ontbrekende bouwsteen in beeld en herstel eerst die basis
Schijnbaar correcte redenering	Het antwoord klinkt aannemelijk, maar tussenstappen of onderbouwing ontbreken	Vraag naar bewijs, aannames, alternatieven en de gevolgde redenering

Niet het antwoord, maar de denkroute onderzoeken

Een gerichte analyse begint daarom niet met de vraag: *Wat weet de student niet?* Een bruikbaarere vraag is: *Welke denkstap kan de student nog niet zelfstandig en verantwoord zetten?* Daarmee verschuift de aandacht van hoeveel kennis iemand kan reproduceren naar de manier waarop die kennis wordt gebruikt.

Dat vraagt om vragen die diagnostische waarde hebben. Een controlevraag waarop slechts één term hoeft te worden ingevuld, laat weinig zien. Meer informatie ontstaat wanneer de student een begrip in eigen woorden uitlegt, twee situaties vergelijkt, een fout in een redenering aanwijst of voorspelt wat er verandert wanneer één voorwaarde wegvalt. Ook een bewust onvolledig voorbeeld kan veel onthullen: welke aanvullende informatie vraagt de student op voordat er een conclusie wordt getrokken?

Voor docenten is dat geen nieuwe didactische opdracht. Zij stellen zulke vragen al tijdens werkcolleges, voortgangsgesprekken en het bespreken van beroepsproducten. Het knelpunt is vooral de beschikbare tijd. In een groep van tientallen studenten kan een docent niet voortdurend bij ieder individu onderzoeken waar het begrip precies stukt. Daardoor komt gerichte diagnose vaak pas op gang nadat fouten zich hebben opgestapeld, bijvoorbeeld bij een tussentoets of een eerste grote opdracht.

AI kan helpen om dit onderzoek eerder en vaker uit te voeren. Dat betekent niet dat een toepassing zelfstandig en definitief vaststelt wat een student wel of niet begrijpt. Een bruikbare rol is bescheidener: antwoorden laten toelichten, vervolgvragen stellen,

tegenvoorbeelden aanbieden en patronen in terugkerende fouten zichtbaar maken. De uitkomst is dan geen oordeel over de student, maar een aanwijzing voor de volgende leerhandeling of voor menselijke begeleiding.

De bestuurlijke waarde van een betere diagnose

Voor een manager zit de mogelijke opbrengst niet in het aantal beantwoorde studentvragen. Dat is hoogstens een gebruiksgegeven. De relevante baten ontstaan wanneer studenten eerder ontdekken waar hun begrip tekortschiet, docenten gerichter kunnen begeleiden en opleidingen beter zicht krijgen op structurele knelpunten in het onderwijsontwerp.

Die opbrengsten moeten afzonderlijk worden vastgesteld. Bij studenten kan worden onderzocht of zij misvattingen eerder corrigeren, betere verklaringen geven en kennis in nieuwe situaties toepassen. Bij medewerkers gaat het erom of signalen uit de toepassing werkelijk helpen de begeleiding gerichter vorm te geven – en of de tijd die daarmee wordt gewonnen niet volledig terugkomt als controle- en herstelwerk. Voor de opleiding is van belang of terugkerende patronen aanleiding zijn om instructie, leermiddelen of de volgorde van het programma aan te passen.

Daarbij past terughoudendheid. Een systeem dat studenten snel in categorieën indeelt, kan ten onrechte zekerheid suggereren. Onduidelijke vragen, taalvaardigheid, spanning of een ongebruikelijke maar verdedigbare redenering kunnen als kennisprobleem worden geïnterpreteerd. De diagnose moet daarom controleerbaar blijven: de student moet kunnen zien waarop een vervolgvraag berust en een docent moet signalen kunnen beoordelen, corrigeren of negeren.

Pas wanneer helder is waar het begrip vastloopt, heeft aangepaste uitleg werkelijk kans van slagen. Anders krijgt de student misschien een eenvoudiger antwoord, een langer voorbeeld of nóg een samenvatting, terwijl het eigenlijke probleem ongemoeid blijft. De eerste waarde van AI ligt hier dan ook niet in beter uitleggen, maar in beter bepalen welke uitleg, vraag of oefening op dat moment nodig is.

2.2 Uitleg op maat, zelf blijven nadenken

Daarmee verschuift de ontwerp vraag van *welke uitleg is het beste?* naar *welke uitleg helpt deze student om de volgende denkstap zelf te zetten?* Juist daar kan AI iets toevoegen. Een docent kan tijdens een bijeenkomst niet voor iedere student tegelijk een andere formulering, een passend voorbeeld en een reeks controlevragen maken. Een AI-toepassing kan dat in beginsel wel, mits zij werkt binnen duidelijke inhoudelijke en didactische grenzen.

De ene student heeft genoeg aan een concreet voorbeeld, terwijl een ander eerst een onderliggend begrip moet opfrissen. Een derde kent de begrippen afzonderlijk, maar ziet het verband ertussen niet. AI kan de uitleg daarop aanpassen: eenvoudiger verwoorden zonder de inhoud te versimpelen, een vergelijking kiezen uit de

beroepspraktijk van de opleiding of een abstract beginsel stap voor stap laten toepassen.

Dat maatwerk is vooral waardevol wanneer studenten niet alleen een andere uitleg krijgen, maar ook iets met die uitleg moeten doen. Laat hen bijvoorbeeld:

- een begrip in eigen woorden omschrijven;
- twee begrippen met elkaar vergelijken;
- voorspellen wat er in een nieuwe situatie gebeurt;
- een voorbeeld én een tegenvoorbeeld formuleren;
- aanwijzen welke stap in een redenering niet klopt;
- hun antwoord onderbouwen met cursusmateriaal of een andere aangewezen bron;
- uitleggen welke eerdere aanname zij hebben bijgesteld.

Zo ontstaat een wisselwerking tussen uitleg en activiteit. De AI geeft niet meteen het eindantwoord, maar biedt precies genoeg steun om de student verder te laten denken. Na een eerste aanwijzing kan een controlevraag volgen. Blijkt de student het nog onvoldoende te begrijpen, dan kan de toepassing een concreter voorbeeld geven of teruggaan naar noodzakelijke voorkennis. Begrijpt de student het wel, dan moet de volgende vraag juist wat verder afstaan van het oorspronkelijke voorbeeld.

Maak de denkstappen zichtbaar

Het risico van uitleg op maat is dat de student een vloeiend gesprek voert zonder dat duidelijk wordt wat hij werkelijk begrijpt. Een bevestigend antwoord als “ja, nu snap ik het” zegt weinig. Ook een juist antwoord kan zijn overgenomen uit de uitleg, door de formulering van de vraag zijn ingegeven of toevallig goed zijn gekozen.

Daarom moet de student verbanden zichtbaar maken. Dat kan door een redenering af te maken, een keuze te verdedigen of hetzelfde beginsel toe te passen in een situatie waarin oppervlakkige herkenning niet volstaat. Bij bedrijfseconomie kan een student bijvoorbeeld niet alleen een kengetal berekenen, maar ook verklaren waardoor het verandert en welke bestuurlijke conclusie daar wel of niet uit mag worden getrokken. In de verpleegkunde is het niet genoeg om een protocol te noemen; de student moet kunnen toelichten welke observaties in deze situatie relevant zijn en wanneer afwijking of overleg nodig is.

De toepassing ondersteunt dan niet het produceren van antwoorden, maar begripsvorming. Dat verschil is bestuurlijk belangrijk. Sneller een correcte tekst opleveren is immers nog geen bewijs dat een student beter kan analyseren, beoordelen of handelen. Het onderwijsrendement ontstaat pas wanneer de student het geleerde zelfstandig kan gebruiken, ook zonder de oorspronkelijke uitleg en bij voorkeur in een gewijzigde context.

Een bruikbaar ontwerp hanteert daarom een eenvoudig beginsel: **eerst denken, dan hulp, daarna opnieuw denken**. De student formuleert een eerste verklaring of aanpak. Vervolgens reageert de AI met een vraag, aanwijzing of beperkt voorbeeld. Daarna

moet de student zijn redenering herzien en toelichten wat er is veranderd. Pas als dat onvoldoende blijkt, volgt uitgebreidere uitleg.

Zo'n opbouw voorkomt niet alle vormen van afhankelijkheid, maar maakt passief gebruik wel minder aantrekkelijk. Bovendien levert zij informatie op die een docent kan benutten. Niet alleen het eindantwoord is zichtbaar, maar ook de punten waarop studenten aarzelen, terugvallen op een misvatting of een aanwijzing nodig hebben.

Variatie vraagt om begrenzing

Uitleg op maat betekent niet dat iedere denkbare uitleg geschikt is. Een taalmodel kan overtuigend klinkende voorbeelden verzinnen die inhoudelijk niet kloppen, te veel uitzonderingen aanvoeren of ongemerkt afwijken van de gekozen vakdidactiek. Naarmate de toepassing vrijer mag antwoorden, neemt de variatie toe, maar wordt ook de controle lastiger.

De opleiding moet daarom bepalen waarop de uitleg is gebaseerd en hoeveel vrijheid de toepassing krijgt. Bij fundamentele begrippen ligt het voor de hand dat de AI werkt met vastgesteld cursusmateriaal, goedgekeurde voorbeelden en expliciete leerdoelen. Ook moet de toepassing onzekerheid kunnen tonen en de student naar de oorspronkelijke bron terugverwijzen. Bronverwijzing is daarbij geen versiering. Zij stelt de student en docent in staat uitleg te controleren en verschillende formuleringen naast de formele vakinhoud te leggen.

Daarnaast verdient de toon aandacht. Een toepassing die voortdurend bevestigt dat een antwoord "uitstekend" of "helemaal correct" is, kan schijnzekerheid versterken. Gerichtte terugkoppeling is nuttiger: welk deel klopt, waar ontbreekt nog een verband en welke vraag moet de student vervolgens beantwoorden? Dat is minder spectaculair, maar didactisch waardevoller.

Beoordeel de opbrengst op begrip

Voor de student ligt de mogelijke baat in beter passende uitleg, meer gelegenheid om vragen te stellen en snellere terugkoppeling tijdens het studeren. Voor de docent kan zichtbaar worden welke begrippen herhaaldelijk problemen veroorzaken en welke studenten aanvullende begeleiding nodig hebben. Op opleidingsniveau kunnen patronen aanleiding geven om voorbeelden, leermiddelen of de opbouw van een onderwijseenheid te verbeteren.

Die baten moeten wel worden aangetoond. Het aantal gesprekken met een uitlegbot zegt hooguit iets over gebruik. Ook studenttevredenheid is onvoldoende: een toepassing kan prettig worden gevonden juist omdat zij veel denkwerk wegneemt. Relevanter is of studenten na het gebruik betere verklaringen geven, hardnekkige misvattingen corrigeren en het begrip zelfstandig toepassen in nieuwe opgaven of beroepssituaties.

Daar horen ook lasten bij. Vakinhoudelijke begrenzing, controle van antwoorden, actualisering van bronnen, toezicht op gebruik en ondersteuning van docenten kosten

tijd. Als iedere docent afzonderlijk voorbeelden moet nakijken en fouten moet herstellen, kan het vermeende schaalvoordeel snel verdwijnen. De vraag is dus niet alleen hoeveel uitleg de AI verzorgt, maar ook hoeveel nieuw ontwerp- en controlewerk daardoor ontstaat.

Voor een besluit over invoering zijn uiteindelijk vier vragen doorslaggevend:

1. Welke begripsproblemen moet de toepassing helpen oplossen?
2. Welke denkhandeling blijft aantoonbaar bij de student?
3. Hoe stellen we vast dat het begrip toeneemt en niet alleen het antwoord verbetert?
4. Wie controleert de inhoud, bronnen en terugkerende fouten?

Wanneer daarop geen concreet antwoord bestaat, is de toepassing waarschijnlijk nog een algemene vraagbaak en geen doelgericht onderwijsmiddel. Dat onderscheid bleek in de praktijk bepaald niet vanzelf te ontstaan. Een uitlegbot kan namelijk zo behulpzaam worden dat zij precies wegneemt wat het onderwijs probeert op te bouwen: het vermogen van studenten om zelf verbanden te leggen, twijfel te onderzoeken en een redenering op te bouwen.

2.3 Uit mijn onderwijspraktijk: uitleg die te behulpzaam werd

Dat gebeurde ook bij een van mijn eerste uitlegbots. Het uitgangspunt leek verstandig: studenten moesten buiten de les snel uitleg kunnen krijgen over begrippen die zij nog niet beheersten. De bot beschikte over cursusmateriaal, kon definities vereenvoudigen en gaf op verzoek voorbeelden. Een student die vastliep, hoefde dus niet te wachten tot het volgende contactmoment.

Technisch werkte dat behoorlijk goed. Misschien wel te goed.

Wanneer een student een vraag stelde, volgde meestal een volledig antwoord: eerst een omschrijving van het begrip, daarna een voorbeeld en soms zelfs een uitgewerkte toepassing. Bij een vraag over het verschil tussen twee begrippen zette de bot de kenmerken overzichtelijk naast elkaar. Bij een casus legde hij uit welke theorie van toepassing was en waarom. Het antwoord oogde helder, volledig en bruikbaar.

Aanvankelijk zag ik dat als een voordeel. Studenten konden verder, de bot was voortdurend beschikbaar en dezelfde uitleg kon op verschillende manieren worden gegeven. Ook studenten reageerden positief. Zij vonden het prettig dat zij direct antwoord kregen en niet hoefden te zoeken in lange teksten of presentaties.

Toch begon mij iets op te vallen. Studenten namen de uitleg gemakkelijk over, maar konden niet altijd reconstrueren hoe het antwoord tot stand was gekomen. Wanneer ik doorvroeg, bleken begrippen soms minder goed begrepen dan hun eerste reactie deed vermoeden. Zij herkenden de redenering zodra de bot die presenteerde, maar konden

haar niet zonder hulp opnieuw opbouwen. De bot had het obstakel verwijderd, maar daarmee soms ook de denkhandeling.

Dat onderscheid is belangrijk. Herkenning kan aanvoelen als begrip. Een duidelijke uitleg geeft de student al snel het gevoel: ja, natuurlijk, nu snap ik het. Pas wanneer de student het begrip zelf moet uitleggen, vergelijken of toepassen, wordt zichtbaar of dat gevoel terecht is.

Ik zag ook dat de vraagstelling veranderde. In plaats van eerst zelf een poging te doen, legden sommige studenten hun hele opdracht aan de bot voor. De vraag verschoof van 'Kun je mij helpen begrijpen waarom dit begrip hier van toepassing is?' naar 'Wat is het goede antwoord op deze casus?' Omdat de bot graag behulpzaam wilde zijn, leverde hij precies wat werd gevraagd. Daarmee werd hij ongemerkt minder een hulpmiddel bij het leren en meer een producent van afgeronde antwoorden.

Het probleem zat dus niet alleen in het gedrag van studenten. Het zat ook in mijn ontwerp. Ik had de bot vooral ingericht op inhoudelijke juistheid en gebruiksgemak. De didactische functie was onvoldoende begrensd. Zolang het antwoord correct en begrijpelijk was, leek de toepassing te doen wat zij moest doen. Maar ik had nog niet scherp vastgelegd welke denkhandeling de student zelf moest uitvoeren.

Daarom veranderde ik de interactie. De bot mocht niet langer standaard beginnen met een volledige uitleg. Eerst vroeg hij wat de student al wist, welk onderdeel onduidelijk was en hoe de student zelf het begrip zou omschrijven. Daarna volgde geen afgerond antwoord, maar een tussenstap: een aanwijzing, een tegenvoorbeeld of een vraag waarmee de student een ontbrekend verband kon onderzoeken.

Ook bouwde ik verplichte eigen toelichting in. Na een uitleg moest de student het begrip in eigen woorden samenvatten, twee situaties vergelijken of beargumenteren waarom een voorbeeld wel of niet bij het begrip paste. Een antwoord als 'ik begrijp het' was niet genoeg. De student moest iets laten zien waaruit dat begrip bleek.

De bot kreeg daarmee een andere rol. Niet: uitleg geven totdat er geen vraag meer overblijft. Wel: de student helpen de volgende denkstap te zetten en vervolgens controleren of die stap daadwerkelijk is gezet. Soms betekende dit dat de bot juist minder vertelde. Een goede wedervraag bleek dan waardevoller dan een extra alinea uitleg.

Dat ging niet zonder wrijving. Studenten die gewend waren geraakt aan directe antwoorden, vonden de nieuwe versie aanvankelijk omslachtiger. Waarom stelde de bot steeds wedervragen? Waarom gaf hij niet gewoon de oplossing? Die reactie was begrijpelijk. Voor het afronden van de taak was de aangepaste bot minder efficiënt. Voor het leren was die vertraging juist de bedoeling.

Tegelijk moest ik voorkomen dat de interactie in een star verhoor veranderde. Een bot die bij iedere vraag mechanisch reageert met 'Wat denk je zelf?' helpt evenmin. Studenten kunnen werkelijk vastzitten, voorkennis missen of een verkeerde voorstelling

hebben opgebouwd. Dan is uitleg wel degelijk nodig. De kunst was dus niet om antwoorden te verbieden, maar om de hoeveelheid hulp af te stemmen op wat de student al kon laten zien.

Zo ontstond een eenvoudige opbouw. Eerst activeerde de bot de voorkennis. Daarna vroeg hij om een eerste poging. Vervolgens gaf hij zo nodig een kleine aanwijzing, een gerichte uitleg of een voorbeeld. Ten slotte moest de student het begrip opnieuw verwoorden of in een andere situatie toepassen. Wie bleef vastlopen, kreeg meer steun of het advies de vraag met een docent te bespreken.

Ik zag daarna inhoudelijk rijkere gesprekken ontstaan. Studenten stelden vaker vragen over verbanden en uitzonderingen, in plaats van alleen om definities of oplossingen te vragen. Hun toelichtingen maakten bovendien beter zichtbaar waar het misging. Soms bleek niet het nieuwe begrip het probleem, maar een ontbrekend basisbegrip. In andere gevallen kende een student de definitie wel, maar kon hij niet bepalen wanneer die in een beroepssituatie bruikbaar was.

Dat waren waardevolle waarnemingen, maar nog geen bewijs dat studenten door de aangepaste bot structureel beter leerden. Langere gesprekken zijn niet automatisch betere gesprekken, en meer activiteit is nog geen aangetoonde begripsontwikkeling. Daarvoor moest ik de antwoorden van studenten vóór en na de interactie vergelijken en nagaan of zij het begrip later zelfstandig konden toepassen.

De belangrijkste les lag daarom niet in één geslaagde aanpassing, maar in een andere manier van ontwerpen. Een onderwijsbot moet niet worden beoordeeld op de volledigheid van zijn antwoorden. Eerst moet duidelijk zijn welke denkhandeling van de student wordt verwacht. Pas daarna kan worden bepaald wanneer de bot een vraag stelt, een aanwijzing geeft, uitleg aanbiedt of doorverwijst naar een docent.

Zo ontstond de stap van een algemene uitlegbot naar doelgericht ontworpen onderwijsbots. De techniek bleef grotendeels hetzelfde, maar de onderwijslogica veranderde. Niet maximale behulpzaamheid, maar passende behulpzaamheid werd het uitgangspunt. En juist daardoor werd ook de bestuurlijke vraag scherper: niet of studenten de bot prettig vinden of er sneller mee klaar zijn, maar waaraan de opleiding kan zien dat zij beter leren verklaren, vergelijken, toepassen en beredeneren.

Antwoorden vóór en na de bot



2.4 Besluiten op basis van aantoonbaar begrip

Stoppen, aanpassen of verder beproeven

Na een proef zijn er grofweg drie verdedigbare besluiten.

Besluit	Wanneer passend?	Vervolg
Stoppen	Er is geen aannemelijke verbetering van begrip, studenten worden afhankelijker of de risico's en beheerlast wegen niet op tegen de baten.	Beëindig de toepassing en leg vast wat niet werkte, zodat hetzelfde experiment niet elders wordt herhaald.
Aanpassen	De toepassing wordt gebruikt en lijkt waardevol, maar neemt te veel denkwerk over, werkt slechts voor een deel van de studenten of levert nog onvoldoende bewijs op.	Verander de onderwijslogica, begrensd de hulp, verbeter de kennisbron of ontwerp een betere evaluatie.
Verder beproeven	Studenten laten een relevante verbetering zien in zelfstandig verklaren, toepassen, vergelijken of beredeneren en de lasten en risico's blijven beheersbaar.	Test in een andere groep of cursus en onderzoek of het effect onder minder gunstige omstandigheden standhoudt.

Voor dit besluit zijn andere maatstaven nodig dan gebruiksgemak, snelheid en tevredenheid. Die gegevens zijn wel nuttig, maar ze vertellen vooral of studenten de toepassing kunnen en willen gebruiken. Ze laten niet zien of studenten een begrip zelfstandig kunnen uitleggen, verbanden herkennen of hun kennis in een nieuwe situatie toepassen. Een onderwijsbot kan technisch goed werken, intensief worden gebruikt en hoge waarderingscijfers krijgen, terwijl het begrip nauwelijks toeneemt.

De beoogde leerhandeling moet daarom vóór de proef worden vastgelegd. Wat moet de student na de interactie beter kunnen dan daarvoor? Bij begripsontwikkeling gaat het doorgaans om vier waarneembare handelingen:

- **verklaren:** de student kan een begrip in eigen woorden uitleggen en oorzaken of mechanismen benoemen;
- **toepassen:** de student kan het begrip gebruiken in een opgave, casus of beroepssituatie die niet letterlijk in de uitleg voorkwam;
- **vergelijken:** de student kan begrippen, modellen of handelwijzen van elkaar onderscheiden en relevante overeenkomsten en verschillen aanwijzen;
- **beredeneren:** de student kan een conclusie opbouwen, aannames zichtbaar maken en uitleggen waarom een keuze in deze situatie passend is.

Deze vier handelingen bieden meer houvast dan de algemene vraag of studenten 'de stof begrijpen'. Ze maken namelijk zichtbaar welk gedrag als bewijs van begrip geldt.

Bovendien kunnen docenten ermee bepalen welke rol de bot krijgt. Soms is een korte uitleg passend. In andere gevallen moet de bot juist een tegenvoorbeeld geven, een wedervraag stellen of de student dwingen eerst zelf een redenering op te bouwen.

Meet wat de student zelfstandig kan

Een eerste beoordeling vindt bij voorkeur plaats zonder hulp van de bot. Daarmee wordt voorkomen dat de kwaliteit van het antwoord van de toepassing wordt aangezien voor de bekwaamheid van de student. Laat studenten bijvoorbeeld vóór en na het gebruik een begrip verklaren, een nieuwe casus analyseren of twee mogelijke oplossingen vergelijken. Gebruik daarbij dezelfde beoordelingscriteria, maar niet exact dezelfde opdracht.

Alleen direct na de interactie meten is evenmin voldoende. Een student kan een zojuist gegeven uitleg nog herkennen of reproduceren zonder dat er duurzaam begrip is ontstaan. Een korte vervolgmeting, enkele dagen of weken later, geeft daarom aanvullende informatie. Kan de student de redenering dan nog zelfstandig reconstrueren? En lukt dat ook wanneer de context verandert?

Voor een eerste proef hoeft deze evaluatie niet omvangrijk te zijn. Een opleiding kan al veel leren door een beperkte groep studenten te volgen en drie bronnen naast elkaar te leggen:

1. de kwaliteit van hun werk vóór en na het gebruik;
2. het gebruiksgedrag tijdens de interactie;
3. het oordeel van docent en student over wat wel en niet hielp.

De eerste bron gaat over het leerresultaat. De tweede laat zien hoe de toepassing feitelijk wordt gebruikt. De derde helpt verklaren waarom een resultaat optreedt. Juist de combinatie is van belang. Meer interacties kunnen wijzen op intensiever oefenen, maar ook op verwarring. Een kortere doorlooptijd kan duiden op efficiënter leren, maar evengoed op het overslaan van denkstappen.

Waardering is een voorwaarde, geen eindresultaat

Een toepassing die studenten onbegrijpelijk, omslachtig of onveilig vinden, zal weinig onderwijswaarde opleveren. Gebruiksgemak en waardering doen er dus wel degelijk toe. Ze zijn alleen geen vervanging voor leerresultaten.

Dat onderscheid voorkomt twee veelvoorkomende bestuurlijke vergissingen. De eerste is dat een hoge tevredenheidsscore wordt gepresenteerd als bewijs van onderwijskwaliteit. De tweede is dat veelvuldig gebruik automatisch als succes geldt. Beide gegevens zeggen iets over de aanvaarding van de toepassing, niet noodzakelijk over beter begrip.

Ook tijdswinst vraagt om nadere duiding. Als studenten sneller tot een juist antwoord komen, moet worden nagegaan waardoor dat komt. Hebben zij een denkstap beter

leren uitvoeren, of heeft de bot die stap voor hen gezet? En wanneer een docent minder uitlegvragen krijgt, ontstaat er dan werkelijk ruimte voor gerichte begeleiding, of verschuift het werk naar controle van onjuiste of oppervlakkige botantwoorden?

Een geloofwaardige batenclaim benoemt daarom steeds wie voordeel heeft, bij welke taak, hoe dat voordeel wordt vastgesteld en welke nieuwe werkzaamheden ontstaan. Bij een uitlegbot kan de opbrengst bijvoorbeeld bestaan uit meer mogelijkheden om buiten de les uitleg te krijgen. Daartegenover staan het ontwerpen van de interactie, het controleren van de inhoud, het analyseren van problematische gesprekken en het actualiseren van de kennisbron.

Verder beproeven is nog niet hetzelfde als opschalen. Een proef met een gemotiveerde docent en een kleine studentengroep kan resultaten opleveren die niet vanzelf standhouden wanneer tientallen docenten en honderden studenten met de toepassing gaan werken. Dan nemen variatie, ondersteuningsvragen, onderhoud en controlewerk toe. Ook moet duidelijk zijn wie verantwoordelijk is voor de inhoud, wie fouten corrigeert en hoe studenten zonder passend apparaat of met een ondersteuningsbehoefte volwaardig kunnen deelnemen.

De bewijslast moet daarbij in verhouding staan tot de gevolgen. Voor een vrijwillige oefenvoorziening naast bestaand onderwijs kan een beperkte evaluatie volstaan. Wordt de bot een belangrijke of verplichte route naar uitleg, dan zijn steviger gegevens nodig over toegankelijkheid, betrouwbaarheid en leerresultaten. Hoe groter de afhankelijkheid, des te minder een instelling mag varen op enthousiasme en losse ervaringen.

De bestuurlijke kernvragen

Voor een besluit over vervolggebruik zijn uiteindelijk vijf vragen doorslaggevend:

- 1. Welke vorm van begrip moest aantoonbaar verbeteren?**
- 2. Wat kunnen studenten na gebruik zelfstandiger of beter dan daarvoor?**
- 3. Waarop is die conclusie gebaseerd: ervaring, gebruiksgegevens, beoordeling of vergelijking?**
- 4. Welke kosten, risico's en nieuwe beheer- en controletaken zijn zichtbaar geworden?**
- 5. Rechtvaardigt het gevonden resultaat een volgende proef, en zo ja, onder welke voorwaarden?**

Wie deze vragen niet kan beantwoorden, beschikt nog niet over een businesscase, maar over een ervaring. Die ervaring kan veelbelovend zijn en aanleiding geven tot verder onderzoek. Zij rechtvaardigt alleen nog geen structurele investering.

Bij begrip ligt de grens bovendien helder: de kwaliteit van het antwoord van de bot is nooit het uiteindelijke resultaat. Het resultaat is wat de student daarna zonder die bot kan verklaren, toepassen, vergelijken en beredeneren. Pas wanneer daarin verbetering zichtbaar wordt, komt verdere beproeving in beeld. Tot die tijd is terughoudendheid

geen rem op innovatie, maar gewoon goed onderwijsbestuur.

3 Meer oefenen waar het ertoe doet

Begrip krijgt pas betekenis wanneer een student ermee oefent. Niet één keer in een voorspelbare opdracht, maar herhaaldelijk, in wisselende situaties en met ruimte om fouten te herstellen. Juist daar schiet het onderwijs soms tekort. Begeleiding is schaars, goede casussen kosten tijd en feedback komt geregeld pas wanneer het volgende leermoment alweer voorbij is.

AI kan die oefenruimte vergroten. Een student kan op ieder moment een nieuwe opgave krijgen, een gesprek opnieuw voeren of oefenen met een andere beroepssituatie. Dat klinkt aantrekkelijk, zeker waar docentcapaciteit de hoeveelheid oefening begrenst. Toch is beschikbaarheid nog geen leerwinst. Wie steeds dezelfde handeling herhaalt, antwoorden opvraagt of zich zonder gerichte opbouw door willekeurige casussen heen werkt, is vooral druk bezig.

De bestuurlijke vraag is daarom niet hoeveel oefeningen AI kan leveren, maar waar extra oefening werkelijk nodig is en wat studenten daarbij moeten leren doen. Pas als duidelijk is welke professionele bekwaamheid achterblijft, welke variatie ertoe doet en hoe voortgang zichtbaar wordt, valt te beoordelen of meer oefenruimte de investering rechtvaardigt. Anders dreigt een bekend patroon: hoge gebruikscijfers, enthousiaste reacties en weinig bewijs dat studenten beter voorbereid raken op de beroepspraktijk.

Een beroepsgesprek opnieuw oefenen



3.1 Van oefentekort naar oefenprofiel

Dezelfde bestuurlijke nuchterheid is nodig als het om oefenen gaat. Meer oefenactiviteit klinkt al snel als winst, maar het aantal gemaakte opdrachten zegt op zichzelf weinig. Studenten kunnen twintig keer hetzelfde kunstje herhalen zonder beter voorbereid te raken op een beroepssituatie waarin omstandigheden veranderen, belangen botsen en een standaardantwoord ontbreekt.

De eerste vraag is daarom niet hoeveel extra oefeningen AI kan produceren. De eerste vraag is **waar studenten nu te weinig, te eenzijdig of te laat oefenen om een professionele bekwaamheid goed te ontwikkelen.**

Niet het programma, maar de bekwaamheid als vertrekpunt

In veel opleidingen is het oefenaanbod historisch gegroeid. Een docent ontwikkelt een casus, een projectgroep krijgt een opdracht en in een later studiejaar volgt een praktijksimulatie. Elk onderdeel kan op zichzelf waardevol zijn, terwijl toch onduidelijk blijft hoe vaak studenten een bepaalde bekwaamheid daadwerkelijk oefenen.

Neem bijvoorbeeld het voeren van een professioneel adviesgesprek. Studenten lezen over gesprekstechnieken, bekijken een voorbeeld en voeren eenmaal een gesprek met een medestudent. Daarna worden zij beoordeeld. Op papier is het onderwerp behandeld. In de praktijk heeft een student misschien maar één volledige oefenpoging gedaan, met één type gesprekspartner en zonder gelegenheid om gerichte feedback te verwerken in een tweede poging.

Dat is geen kennistekort, maar een oefentekort.

Zo'n tekort blijft gemakkelijk buiten beeld wanneer een opleiding vooral kijkt naar lessen, studiepunten en toetsmomenten. Om het zichtbaar te maken, moet zij per professionele bekwaamheid nagaan welke leerhandelingen studenten werkelijk uitvoeren. Niet: *Waar geven we uitleg over adviseren?* Wel: *Hoe vaak formuleert een student een advies, verdedigt hij dat tegenover een kritische opdrachtgever en past hij het vervolgens aan op basis van feedback?*

Hetzelfde geldt voor bekwaamheden als analyseren, ontwerpen, diagnosticeren, samenwerken, programmeren, rapporteren en professioneel oordelen. De vraag is steeds wat de student zelf doet, onder welke omstandigheden en met welke terugkoppeling.

Vier kenmerken van een oefenprofiel

Een bruikbaar oefenprofiel brengt ten minste vier kenmerken in kaart:

1. **Frequentie** – Hoe vaak oefent een student de betreffende handeling?
2. **Variatie** – Welke verschillende situaties, doelgroepen, belangen en moeilijkheidsgraden komt de student tegen?
3. **Feedbacktijd** – Hoe lang duurt het voordat de student bruikbare terugkoppeling ontvangt?

4. Herhalingsmogelijkheid – Kan de student de feedback toepassen in een volgende poging?

Daarbij gaat het niet om schijnprecisie. Een opleiding hoeft niet van iedere student elke oefenhandeling tot achter de komma te registreren. Een eerste inventarisatie met docenten en studenten maakt vaak al duidelijk waar de grootste beperkingen zitten.

Professionele bekwaamheid	Huidige oefenfrequentie	Variatie	Tijd tot feedback	Mogelijkheid tot herhaling	Waarschijnlijkheid van herhaling
Een klantvraag analyseren	Enkele keren per periode	Beperkt aantal casussen	Enkele dagen	Soms	Beschikbaar casuïstiek
Een adviesgesprek voeren	Eén of twee keer	Vooraf met medestudenten	Direct of na de les	Beperkt	Tijd en beschikbaar van gesprekspartner
Een beroepsproduct verbeteren	Meerdere concepten	Verschilt per project	Een tot twee weken	Ja, maar laat	Nakijkcapaciteit
Beslissen onder tijdsdruk	Nauwelijks	Eén simulatie	Directe nabespreking	Zelden	Organisatie van simulaties

Zo'n overzicht levert nog geen oordeel over AI op. Het maakt eerst zichtbaar waar de opleiding een onderwijsprobleem heeft. Dat onderscheid is belangrijk. Als studenten al vaak, gevarieerd en met snelle feedback oefenen, hoeft een digitale oefentoepassing weinig extra waarde te bieden. De investering kan dan beter naar een ander knelpunt gaan.

Wachttijd is verloren leertijd

Een oefentekort ontstaat niet altijd doordat er te weinig opdrachten zijn. Regelmatig ligt het knelpunt bij de terugkoppeling. Een student levert een concept in, wacht tien dagen op feedback en is intussen met een ander onderwerp bezig. Wanneer de reactie komt, is de oorspronkelijke redenering al naar de achtergrond verdwenen. De student verwerkt enkele opmerkingen, maar krijgt binnen hetzelfde onderdeel geen nieuwe gelegenheid om te laten zien dat hij ze begrijpt.

Daarmee wordt de feedbackcyclus feitelijk afgebroken. De docent heeft werk verricht en de student heeft feedback ontvangen, maar er is niet vastgesteld of de prestatie daarna verbetert.

Bij grote groepen wordt dit probleem scherper. Een docent kan niet onbeperkt extra gesprekken beluisteren, analyses beoordelen of concepten van commentaar voorzien. Ook medestudenten kunnen een deel van die taak uitvoeren, maar hun feedback vraagt voorbereiding en kwaliteitsbewaking. De beschikbare docentcapaciteit bepaalt daardoor mede hoe vaak studenten verantwoord kunnen oefenen.

Juist op zulke punten kan AI van betekenis zijn. Een oefentoepassing kan een extra casus aanbieden, een antwoord bevragen, een gesprekspartner nabootsen of onmiddellijk een eerste terugkoppeling geven. De mogelijke opbrengst zit dan niet in het produceren van meer tekst, maar in het verkorten van de tijd tussen poging, feedback en nieuwe poging.

Toch verdwijnt het capaciteitsvraagstuk daarmee niet vanzelf. Er ontstaat ook nieuw werk: docenten moeten oefeningen ontwerpen, grenzen aangeven, uitkomsten controleren, onjuiste terugkoppeling opsporen en studenten leren hoe zij ermee omgaan. De relevante vraag is dus niet of AI docenttijd kost of bespaart, maar **waar menselijk werk verschuift en of die verschuiving onderwijskundig gunstig is.**

Kijk ook naar verschillen tussen studenten

Een gemiddeld oefenprofiel kan bovendien verschillen verhullen. Mondige studenten organiseren soms zelf extra oefenkansen, vragen gericht om feedback en gebruiken hun netwerk om beroepssituaties te verkennen. Studenten die minder zeker zijn, wachten eerder tot een docent een oefenmoment aanbiedt. Een schaars aanbod raakt hen daardoor harder.

Breng daarom niet alleen het formele programma in kaart, maar ook het feitelijke gebruik. Hoeveel studenten benutten vrijwillige oefenmogelijkheden? Wie haakt vroeg af? Welke studenten wachten lang op hulp? En wie krijgt door een stageplaats toevallig veel meer variatie dan een ander?

Een AI-toepassing kan de toegang tot oefenen verbreden, bijvoorbeeld buiten lestijd of zonder dat een student eerst een afspraak hoeft te maken. Dat kan de toegankelijkheid verbeteren, maar is geen vanzelfsprekendheid. Studenten verschillen immers ook in digitale vaardigheid, taalbeheersing en vermogen om de kwaliteit van terugkoppeling te beoordelen. Zonder passende begeleiding kan extra beschikbaarheid bestaande verschillen juist vergroten.

Van inventarisatie naar gerichte keuze

Een opleiding hoeft niet voor alle professionele bekwaamheden tegelijk een volledig oefenprofiel op te stellen. Begin met de bekwaamheden die zwaar wegen in het beroepsprofiel en waarvan docenten of werkveldpartners vermoeden dat studenten ze onvoldoende beheersen. Kies vervolgens één of twee concrete handelingen waarvoor de huidige oefenruimte aantoonbaar beperkt is.

Leg vóór een proef ten minste vast:

- hoe vaak studenten de handeling nu uitvoeren;
- welke soorten situaties zij tegenkomen;
- hoe lang zij op terugkoppeling wachten;
- hoeveel studenten een verbeterde poging kunnen doen;
- hoeveel docent- en organisatietijd daarmee gemoeid is;

- welke kwaliteit studenten aan het einde zelfstandig moeten laten zien.

Deze uitgangsmeting hoeft niet omvangrijk te zijn, maar zij moet wel bruikbaar zijn. Zonder zo'n vertrekpunt blijft achteraf alleen de constatering over dat studenten de toepassing veel of graag hebben gebruikt. Dat zegt iets over gebruik en waardering, maar nog niet of een relevant oefentekort is verkleind.

Pas op basis van het oefenprofiel wordt een gerichte keuze mogelijk. De opleiding kan dan bepalen waar extra oefening nodig is, voor welke studenten, bij welke beroepshandeling en op welk moment in het leerproces. Vervolgens komt de volgende ontwerp vraag in beeld: welke variatie moet die oefening bieden? Want een toepassing die eindeloos nieuwe opdrachten maakt, levert pas waarde wanneer die varianten binnen doordachte inhoudelijke grenzen blijven en samen een zinvolle moeilijkheidsopbouw vormen.

3.2 Variatie zonder willekeur

Meer variatie is niet automatisch betere oefening. Een oefenbot kan in korte tijd tientallen opdrachten, cliëntsituaties of beroepscasussen maken, maar de onderwijskundige waarde daarvan hangt af van de verschillen tussen die varianten. Als alleen de namen, getallen en formuleringen veranderen, blijft de oefening inhoudelijk hetzelfde. Te veel gelijktijdige veranderingen ontnemen de student juist het zicht op een herkenbare ontwikkellijn.

De kernvraag is daarom niet hoeveel varianten AI kan produceren, maar **welke variatie een student nodig heeft om een professionele bekwaamheid op te bouwen.**

Variëren op relevante kenmerken

Goede variatie ontstaat wanneer een opleiding vooraf bepaalt welke kenmerken van een situatie ertoe doen. Bij een commercieel gesprek kan dat bijvoorbeeld gaan om het type klant, de complexiteit van de behoefte, de hoeveelheid beschikbare informatie, de mate van weerstand en de belangen van andere betrokkenen. Bij een verpleegkundige handeling kunnen de gezondheidstoestand, de voorgeschiedenis, de urgentie en het vermogen van de patiënt om zelf informatie te geven van belang zijn.

Zulke kenmerken zijn geen versiering van een casus. Ze bepalen welke afweging de student moet maken en welke professionele handeling passend is. De oefenbot mag binnen die kenmerken variëren, maar daarbuiten niet improviseren.

Dat vraagt om een inhoudelijk raamwerk waarin ten minste vier zaken zijn vastgelegd:

1. **De leerhandeling:** wat moet de student in iedere variant daadwerkelijk doen?
2. **De inhoudelijke variabelen:** welke kenmerken van de situatie mogen veranderen?
3. **De kwaliteitsgrenzen:** waaraan moet iedere opdracht of casus voldoen?

4. De **moeilijkheidsopbouw**: waardoor wordt een volgende variant aantoonbaar complexer?

Door deze keuzes vooraf te maken, verschuift de rol van AI. De technologie bepaalt niet wat studenten moeten leren, maar helpt om een door docenten ontworpen oefenruimte op grotere schaal te vullen.

Moeilijker is meer dan langer

Een veelvoorkomende fout is dat moeilijkheid wordt gelijkgesteld aan meer tekst, meer gegevens of ingewikkelder taalgebruik. Daarmee neemt de belasting voor de student wel toe, maar niet noodzakelijk de professionele complexiteit.

Een oefening wordt bijvoorbeeld betekenisvol moeilijker wanneer de student:

- minder aanwijzingen krijgt;
- uit onvolledige of tegenstrijdige informatie moet kiezen;
- meerdere belangen moet afwegen;
- een afwijking van de standaardsituatie moet herkennen;
- onder tijdsdruk moet handelen;
- zijn beslissing moet onderbouwen;
- moet reageren op de gevolgen van een eerdere keuze.

Zo'n opbouw kan beginnen met afgebakende situaties waarin het juiste handelingspatroon goed herkenbaar is. Daarna volgen varianten met afleidende informatie, uitzonderingen en concurrerende belangen. Uiteindelijk oefent de student situaties waarin niet één vanzelfsprekend antwoord bestaat en professioneel oordeel nodig is.

AI kan deze reeks ondersteunen, maar de opleiding moet bepalen wat een hoger niveau inhoudt. Dat kan niet worden overgelaten aan een algemene opdracht als: 'Maak de volgende casus iets moeilijker.' Zonder nadere criteria zal het systeem vooral oppervlakkige complexiteit toevoegen. Het maakt een tekst langer, voert extra personen op of gebruikt specialistischer woorden, terwijl de vereiste denk- of handelingsstap gelijk blijft.

Aanpassen zonder studenten op te sluiten

Een oefenbot kan varianten ook afstemmen op het niveau of het foutenpatroon van een student. Wie een basisbegrip nog onvoldoende beheerst, krijgt dan eerst overzichtelijkere situaties. Wie een handeling al beheerst, kan sneller door naar uitzonderingen en dilemma's. Daarmee kan de beschikbare oefentijd gericht worden gebruikt.

Toch vraagt deze aanpassing om terughoudendheid. Een student die enkele fouten maakt, mag niet langdurig op een laag niveau blijven oefenen. Evenmin moet iedere fout meteen leiden tot een eenvoudiger opdracht. Fouten kunnen immers ook ontstaan

door vermoeidheid, een onduidelijke formulering of een verkeerde aanname van het systeem.

Een verantwoorde werkwijze combineert daarom drie soorten informatie:

- eerdere antwoorden en gemaakte keuzes;
- de aard van de fouten, niet alleen het aantal;
- vaste momenten waarop de student opnieuw op een hoger niveau oefent.

De student moet bovendien kunnen zien waarom een bepaalde oefening wordt aangeboden en zo nodig een andere route kunnen kiezen. Adaptiviteit mag geen ondoorzichtige selectie worden waarin de toepassing stilzwijgend bepaalt wat iemand vermoedelijk wel of niet aankan.

Voor docenten blijft een overzicht nodig van de gevolgde oefenroutes. Niet om iedere poging afzonderlijk te controleren, maar om te zien of bepaalde groepen studenten steeds dezelfde soorten opdrachten krijgen, waar zij vastlopen en of de beoogde moeilijkheidsopbouw werkelijk wordt doorlopen.

De docent ontwerpt de speelruimte

Een schaalbare oefentoeepassing vraagt dus niet om minder onderwijsontwerp, maar om een andere inzet daarvan. Docenten hoeven niet iedere variant zelf te schrijven. Zij moeten wel de speelruimte ontwerpen waarbinnen varianten verantwoord kunnen worden gemaakt.

Daarbij horen bijvoorbeeld voorbeeldcasussen, uitsluitingscriteria, beoordelingspunten en regels voor de opbouw. Ook moet duidelijk zijn welke inhoud niet automatisch mag worden gegenereerd, bijvoorbeeld omdat fouten ernstige gevolgen kunnen hebben, omdat actuele wet- en regelgeving doorslaggevend is of omdat een situatie menselijke gevoeligheid vereist.

Een praktische kwaliteitscontrole kan op drie niveaus plaatsvinden:

Niveau	Te controleren vraag	Passende controle
Afzonderlijke variant	Is de opdracht inhoudelijk juist, begrijpelijk en uitvoerbaar?	Steekproef door een vakdocent
Oefenreeks	Verschillen de varianten op relevante kenmerken en loopt de moeilijkheid op?	Periodieke beoordeling van volledige leerroutes
Studentgroep	Krijgen studenten voldoende breedte en komen zij niet vast te zitten in één type oefening?	Analyse van gebruiks- en voortgangsgegevens

Die controle brengt nieuw werk met zich mee. Het is dus te eenvoudig om het aantal automatisch gemaakte opdrachten als tijdswinst te presenteren. Tegenover minder handmatig schrijfwerk staan ontwerp, steekproeven, correcties, actualisering en het

volgen van oefenroutes. Rendement ontstaat pas wanneer deze nieuwe taken minder capaciteit vragen dan de oude werkwijze én de oefenkwiteit ten minste gelijk blijft.

Beoordeel de reeks, niet de voorraad

Voor een proef is het verleidelijk om te tellen hoeveel opdrachten de bot heeft gemaakt en hoe vaak studenten ermee hebben geoefend. Bestuurlijk interessanter is of de toepassing een samenhangende reeks heeft opgeleverd waarin studenten relevante variatie tegenkwamen en aantoonbaar verder kwamen.

Beoordeel daarom niet alleen het aantal pogingen, maar ook:

- welke inhoudelijke varianten studenten hebben doorlopen;
- of zij steeds complexere situaties aankonden;
- of terugkerende fouten afnamen;
- of studenten het geleerde in een onbekende situatie konden toepassen;
- hoeveel docenttijd nodig was voor ontwerp, controle en herstel;
- welke fouten of ongewenste patronen in de gegenereerde opdrachten voorkwamen.

Pas dan wordt zichtbaar of variatie werkelijk gericht was. Een toepassing kan immers intensief worden gebruikt en toch weinig ontwikkeling opleveren. Veel pogingen, hoge waarderingcijfers en een grote voorraad casussen zijn op zichzelf geen bewijs dat studenten beter leren handelen. Daarvoor moet de oefenreeks reageren op wat zij nog niet beheersen en hen vervolgens herkenbaar naar een volgende moeilijkheidsgraad brengen.

3.3 Uit mijn onderwijspraktijk: veel gebruik, weinig voortgang

De eerste versie van mijn oefenbot voldeed daar niet aan. De bot was populair, studenten gebruikten hem veel en in de gesprekken hoorde ik vooral positieve reacties. Zij waardeerden dat zij op ieder moment een nieuwe vraag konden krijgen, direct een antwoord kregen en zonder beoordeling opnieuw konden proberen. Op het eerste gezicht leek de toepassing dus precies te doen wat ik ervan verwachtte: de drempel om te oefenen werd lager en het aantal oefenmomenten nam toe.

De gebruiksgegevens bevestigden het eerste deel van dat beeld. Ik kon zien hoeveel sessies er waren, of studenten binnen een sessie vervolgvragen stelden en of zij later terugkeerden. Die gegevens lieten bereik, activiteit en herhaald gebruik zien. Ze maakten niet zichtbaar welke fouten studenten maakten, of die fouten bij een volgende poging afnamen en of studenten daarna zonder hulp een moeilijkere opgave aankonden. Als ik alleen naar bereik en activiteit had gekeken, had ik de proef geslaagd kunnen noemen.

Toch zag ik in ingeleverd werk en tijdens begeleidingsgesprekken nauwelijks vooruitgang die aan dat intensieve gebruik kon worden gekoppeld. Dat waren geen systematische effectmetingen, maar ze legden wel een duidelijke beperking van de gebruikscijfers bloot. Sommige studenten maakten na veel oefenen nog steeds dezelfde fouten. Anderen konden een bekende vraag beantwoorden, maar liepen vast zodra de formulering of context veranderde. Ook bleken studenten de bot heel verschillend te gebruiken. De één werkte een vraag zorgvuldig uit en vroeg om toelichting. De ander vroeg vrijwel meteen om het antwoord, las de uitleg vluchtig en ging door naar de volgende opgave.

Veel interactie bleek dus nog geen gerichte oefening.

De bot hield geen rekening met het foutenpatroon

Het belangrijkste tekort zat niet in de hoeveelheid beschikbare opdrachten, maar in de manier waarop de bot vervolgvragen koos. Iedere nieuwe opgave stond min of meer op zichzelf. Wie een denkfout maakte, kon bij de volgende beurt een heel ander onderwerp krijgen. Wie een eenvoudige vraag goed beantwoordde, kreeg niet vanzelf een complexere variant. De bot genereerde wel variatie, maar bouwde geen leerroute op.

Daarmee ontbrak precies wat een docent tijdens goede oefenbegeleiding probeert te doen: herkennen waar een student vastloopt, bepalen wat een passende volgende stap is en vervolgens nagaan of de student die stap ook werkelijk beheerst.

Ook de terugkoppeling was aanvankelijk te algemeen. Bij een onjuist antwoord legde de bot uit hoe het wel moest, maar hij onderzocht niet altijd *waarom* het antwoord onjuist was. Een begripsfout, een rekenfout, een verkeerde aanname en een slordige lezing kregen daardoor ongeveer dezelfde behandeling. Voor de student kon de uitleg overtuigend klinken, terwijl het onderliggende probleem bleef bestaan.

Ik had de toepassing ontworpen als een leverancier van oefenvragen en antwoorden. Wat ontbrak, was een didactisch mechanisme dat de oefening aan de ontwikkeling van de student koppelde.

Van losse vragen naar een ontwikkellijn

In een volgende versie liet ik de bot eerst het antwoord van de student analyseren aan de hand van een beperkt aantal vooraf bepaalde foutenpatronen. De bedoeling was niet dat de bot een definitief oordeel over de student zou geven. Hij moest herkennen welke vervolgstap waarschijnlijk behulpzaam was.

Bij een begripsfout volgde bijvoorbeeld eerst een eenvoudiger onderscheidingsvraag. Bij een verkeerde redenering vroeg de bot de student om de tussenstappen expliciet te maken. Bij een goed antwoord kwam niet meteen meer van hetzelfde, maar een variant met extra informatie, een tegenstrijdig belang of een minder duidelijke probleemsituatie. De moeilijkheid nam pas toe wanneer de student het onderliggende principe meer dan één keer correct kon toepassen.

Daarnaast kreeg de student minder snel een volledig uitgewerkt antwoord. De bot gaf eerst een aanwijzing, daarna eventueel een gerichte vraag en pas in laatste instantie een toelichting. Daardoor bleef de student langer zelf aan het werk. De gesprekken telden daardoor soms minder beurten, maar waren inhoudelijk rijker. Juist daardoor werd duidelijk hoe misleidend een eenvoudige gebruiksmaat kan zijn: minder berichten konden wijzen op beter oefenen, terwijl veel berichten ook het gevolg konden zijn van gokken, antwoorden opvragen of doelloos blijven proberen.

Na de aanpassing konden we behalve activiteit ook volgen welk foutenpatroon was herkend, welke vervolgstap daarop volgde en of dezelfde fout later opnieuw optrad. Ik zag vervolgens dat studenten vaker terugkwamen op een fout die zij eerder hadden gemaakt en dat zij bewuster konden benoemen wat er in hun aanpak misging. Daarmee veranderde vooral de kwaliteit van de beschikbare informatie: we kregen zicht op het oefenproces in plaats van alleen op het aantal interacties. Dat was een bemoedigende waarneming, maar nog geen hard bewijs van leerwinst. Daarvoor moest de verbetering ook zichtbaar zijn in zelfstandig werk en in nieuwe situaties waarin de bot niet stap voor stap hielp.

We voegden daarom enkele controlemomenten toe. Na een reeks oefeningen kreeg de student een nieuwe casus zonder aanwijzingen. Zo konden we vastleggen of een eerder foutenpatroon terugkeerde, welk moeilijkheidsniveau de student zelfstandig aankon en of de geboden ondersteuning kon worden verminderd. De uitkomst daarvan werd niet alleen gebruikt om een volgende oefenreeks te kiezen, maar ook om te bepalen of de eerdere ondersteuning werkelijk iets had opgeleverd. Bij twijfel bleef beoordeling door een docent nodig, zeker wanneer meerdere redeneringen verdedigbaar waren of de kwaliteit van een antwoord van de beroepscontext afhing.

Wat ik hiervan leerde

Mijn eerste ontwerp was onbedoeld gericht op activiteit. De bot maakte oefenen gemakkelijk, maar stuurde niet op ontwikkeling. Daardoor ontstonden aantrekkelijke gebruikscijfers zonder een even overtuigend beeld van wat studenten beter leerden doen.

De aanpassing veranderde ook het werk van docenten. Zij hoefden niet iedere afzonderlijke oefenpoging te begeleiden, maar moesten vooraf wel foutenpatronen beschrijven, moeilijkheidsniveaus ontwerpen, voorbeeldroutes controleren en nagaan waar de bot studenten verkeerd indeelde. Dat werk was inhoudelijk waardevol, maar het was geen gratis tijdwinst. Een deel van de begeleiding verschoof van reageren op losse studentvragen naar het ontwerpen en bewaken van de oefenstructuur.

Voor mij werd daarmee een belangrijk onderscheid zichtbaar. Extra oefenruimte heeft pas waarde wanneer drie elementen met elkaar verbonden zijn:

1. de bot herkent wat de student nog niet beheerst;
2. de volgende oefening sluit daarop aan en verhoogt de complexiteit geleidelijk;

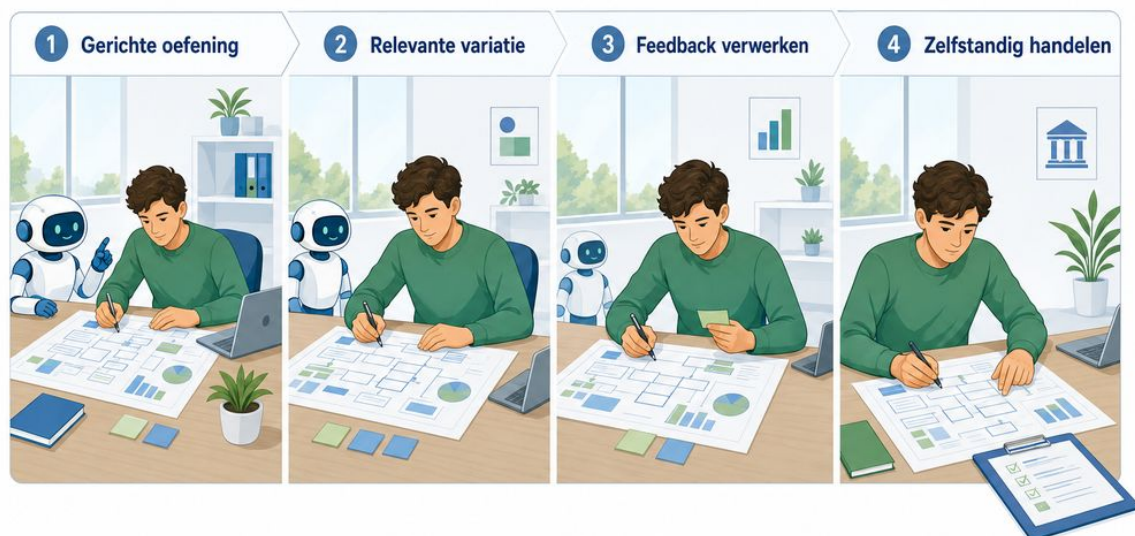
3. de opleiding controleert buiten de bot of de student daadwerkelijk zelfstandiger en beter handelt.

Ontbreekt een van deze elementen, dan bestaat het risico dat studenten vooral vaardiger worden in het voeren van een gesprek met de bot. Dat kan prettig en motiverend zijn, maar het is niet hetzelfde als groei in professionele bekwaamheid.

Sindsdien beschouw ik hoge gebruikscijfers hooguit als een aanwijzing dat studenten de oefenvoorziening weten te vinden. De bestuurlijk relevante vraag begint pas daarna: nemen terugkerende fouten af, kunnen studenten complexere situaties aan en blijft die vooruitgang zichtbaar wanneer de ondersteuning wordt verminderd? Alleen als daarop een geloofwaardig antwoord kan worden gegeven, rechtvaardigt populariteit een verdere investering.

Zelfstandig handelen zonder bot

Van ondersteuning naar zelfstandigheid



3.4 Besluiten over extra oefenruimte

Dat antwoord vraagt om meer dan tevredenheidsmetingen en gebruikscijfers. Wie extra oefenruimte wil financieren, moet aannemelijk maken dat studenten daardoor vaker de juiste leerhandelingen uitvoeren, gericht aan hun tekortkomingen werken en uiteindelijk zelfstandiger handelen. Tegelijk moet duidelijk zijn welke inzet van docenten, beheerders en ondersteunende diensten daarvoor nodig is.

De zakelijke onderbouwing begint daarom niet met de prijs van de toepassing, maar

met een vergelijking tussen de bestaande en de gewenste situatie. Pas daarna komt de vraag of AI een passend middel is.

De businesscase

Onderdeel	Bestuurlijke vraag	Stop- of aanpassingscriterium
Leerwinst	Welke professionele bekwaamheid verbetert en hoe stellen we dat buiten de bot vast?	Zelfstandige prestaties verbeteren niet of terugkerende fouten nemen niet af.
Oefenintensiteit	Krijgen studenten aantoonbaar meer relevante oefenkansen dan voorheen?	Er is vooral meer activiteit, zonder aantoonbare ontwikkeling.
Docenteninzet	Welke werkzaamheden vervallen, veranderen of komen erbij?	Controle, begeleiding en herstelwerk kosten structureel meer tijd dan voorzien.
Technische kosten	Wat kosten aanschaf, koppelingen, beveiliging, ondersteuning en onderhoud?	De totale kosten staan niet in verhouding tot de vastgestelde leerwinst.
Toegankelijkheid	Kunnen alle beoogde studenten de voorziening daadwerkelijk gebruiken en ervan profiteren?	Studentgroepen haken af, blijven achter of hebben geen bruikbaar alternatief.
Kwaliteitsbewaking	Wie controleert inhoud, feedback, moeilijkheid en ongewenste uitkomsten?	Onjuiste informatie of verkeerde werkwijzen worden onvoldoende herkend en hersteld.
Continuïteit	Kan de opleiding doorgaan als de leverancier, prijs of techniek verandert?	De toepassing blijkt zonder één leverancier of onevenredig veel beheer niet uitvoerbaar.

Een bruikbare meting hoeft niet meteen grootschalig of wetenschappelijk te zijn. Een opleiding kan beginnen met een nulmeting, een beperkt aantal duidelijke leerindicatoren en een aparte beoordeling buiten de oefenvoorziening. Denk aan het aantal terugkerende fouten, de kwaliteit van een beroepsproduct of de mate waarin een student tijdens een simulatie nog ondersteuning nodig heeft. Daarmee ontstaat nog geen hard bewijs van oorzaak en gevolg, maar wel een beter beslisbeeld dan met gebruikscijfers alleen.

Reken daarbij met de volledige inzet. Als een bot de eerste feedback op honderd oefeningen verzorgt, betekent dat niet automatisch dat de docent evenredig minder werk heeft. Misschien verschuift de inzet naar het beoordelen van uitzonderingen, het volgen van voortgang en het herstellen van misvattingen. Ook technisch schaalvoordeel is nog geen organisatorisch schaalvoordeel: meer gebruik leidt doorgaans tot meer

vragen, incidenten, controles en aanpassingen.

Het besluit

Doorgaan is verantwoord wanneer de extra oefenruimte aan vier voorwaarden voldoet:

1. studenten voeren aantoonbaar meer relevante leerhandelingen uit;
2. hun prestaties verbeteren ook buiten de ondersteuning van de bot;
3. de volledige inzet van mensen, techniek en beheer blijft in verhouding tot die verbetering;
4. toegankelijkheid, kwaliteitsbewaking en menselijke verantwoordelijkheid zijn uitvoerbaar geregeld.

Ontbreekt bewijs van zelfstandige groei, dan is voortzetting hooguit gerechtvaardigd als gerichte pilot met een duidelijke meetperiode. Zijn de baten aannemelijk maar de beheerlast of ongelijkheid te groot, dan ligt aanpassing voor de hand. Blijft vooral veel activiteit over zonder aantoonbare ontwikkeling, dan moet de toepassing stoppen.

Financier dus geen extra interactie, maar doelgerichte oefening die zichtbaar bijdraagt aan professionele bekwaamheid. Pas wanneer de opleiding weet **wie beter leert, wat er verbetert, tegen welke totale kosten en onder welk toezicht**, ontstaat er een geloofwaardige reden om de oefenruimte uit te breiden.

4 Van weten naar professioneel handelen

Meer oefenen is dus geen doel op zich. De volgende vraag is wat er gebeurt wanneer een student zijn kennis niet alleen moet reproduceren, maar onder tijdsdruk, met onvolledige informatie en in het contact met anderen moet toepassen. In zulke situaties ontstaat professionele bekwaamheid: waarnemen wat ertoe doet, belangen en risico's afwegen, een besluit nemen en daarvoor verantwoordelijkheid dragen.

Juist die samenhang is in het onderwijs lastig te oefenen. Echte beroepssituaties zijn vaak schaars, kostbaar of te riskant om onbeperkt te herhalen. Een patiënt, cliënt of opdrachtgever is nu eenmaal geen oefenmateriaal. Simulaties kunnen hier ruimte scheppen. Ze kunnen studenten laten ervaren dat informatie ontbreekt, omstandigheden veranderen en keuzes gevolgen hebben.

Maar geloofwaardigheid is nog geen kwaliteit. Een vloeiend gesprek kan vakinhoudelijk ondeugdelijk zijn, een realistisch scenario kan de verkeerde handelingen belonen en intensief gebruik kan zelfstandigheid juist verhullen. Voor managers en bestuurders gaat het daarom om meer dan de aanschaf van een overtuigende simulatie. Zij moeten kunnen beoordelen welk professioneel handelen wordt geoefend, hoe zichtbaar wordt dat studenten daarin groeien en welk menselijk toezicht nodig blijft wanneer AI de beroepspraktijk nabootst.

Beslissen met onvolledige informatie



4.1 De taak als een reeks professionele beslissingen

Meer oefenen krijgt pas betekenis wanneer de student leert handelen in een taak die op de beroepspraktijk lijkt. Zo'n taak bestaat zelden uit één afgebakende handeling met één juist antwoord. De professional moet informatie verzamelen, signalen duiden, belangen wegen, een besluit nemen, dat besluit uitvoeren en daar achteraf rekenschap van afleggen. Juist in die samenhang wordt zichtbaar of kennis ook leidt tot professioneel handelingsvermogen.

Neem een verpleegkundige die een verandering in de toestand van een patiënt opmerkt, een accountmanager die moet reageren op een ontevreden klant of een bouwkundige die afwijkingen in een ontwerp beoordeelt. Geen van hen kan volstaan met het reproduceren van vakkennis. Zij moeten bepalen wat relevant is, omgaan met onvolledige informatie en kiezen wat de situatie op dat moment vraagt.

Een beroepshandeling kan daarom worden ontleed in vijf opeenvolgende, maar onderling verbonden stappen:

1. **waarnemen:** relevante informatie herkennen en ontbrekende informatie achterhalen;
2. **afwegen:** mogelijke verklaringen, belangen, normen en risico's tegen elkaar afzetten;
3. **besluiten:** een verdedigbare handelwijze kiezen;
4. **uitvoeren:** het besluit professioneel en passend bij de situatie in de praktijk brengen;
5. **verantwoorden:** uitleggen waarop de keuze en uitvoering zijn gebaseerd en wat daarvan is geleerd.

Deze indeling is geen beschrijving van een keurig lineair proces. In de praktijk keert een professional geregeld terug naar een eerdere stap. Nieuwe informatie tijdens de uitvoering kan aanleiding zijn om de oorspronkelijke afweging te herzien. Een gesprek kan bijvoorbeeld anders verlopen dan verwacht, waardoor een eerder besluit niet langer passend is. Professioneel handelen vraagt dus niet alleen om een goede eerste keuze, maar ook om het vermogen die keuze tijdig bij te stellen.

Waarnemen is al een vorm van oordelen

Waarnemen klinkt misschien als een neutrale eerste stap, maar dat is het niet. Studenten moeten leren waarop zij letten, wat zij nog niet weten en welke signalen dringend zijn. Een beginnende student ziet vaak losse feiten. Een ervaren professional herkent patronen, afwijkingen en mogelijke risico's.

AI kan hier helpen door een rijke praktijksituatie aan te bieden en aanvullende informatie beschikbaar te maken wanneer de student daar gericht naar vraagt. Een student krijgt bijvoorbeeld een dossier, een klantbeschrijving, meetwaarden of een projectgeschiedenis en moet bepalen welke informatie ontbreekt. De toepassing kan

vervolgens reageren op de gestelde vragen.

Daarmee wordt niet alleen kennis getoetst. Zichtbaar wordt ook wat de student níét vraagt. Dat kan didactisch veelzeggend zijn. Wie een gesprek over financiële problemen voert zonder naar de thuissituatie te vragen, of een technisch incident onderzoekt zonder de omstandigheden van eerder onderhoud te bekijken, laat een blinde vlek zien die in een gewone schriftelijke opdracht gemakkelijk verborgen blijft.

De AI mag die ontbrekende waarneming niet meteen overnemen. Geeft de toepassing uit zichzelf alle relevante informatie, dan oefent de student vooral met het verwerken van een voorgesorteerd probleem. De beroepspraktijk vraagt juist dat hij zelf ontdekt wat ertoe doet.

Afwegen zonder de onzekerheid weg te poetsen

Na het waarnemen volgt de afweging. Daarbij gaat het niet alleen om de vraag welke oplossing technisch mogelijk is. Studenten moeten ook rekening houden met professionele normen, menselijke gevolgen, beschikbare tijd, tegenstrijdige belangen en de kans dat hun interpretatie onjuist is.

AI kan alternatieven helpen verkennen. De toepassing kan doorvragen naar aannames, wijzen op een belang dat nog niet is meegewogen of vragen wat er gebeurt als een gekozen verklaring niet klopt. Zo ontstaat een inhoudelijk gesprek waarin de student zijn redenering moet aanscherpen.

Er schuilt wel een belangrijk risico in die ondersteuning. Een taalmodel kan een afweging overtuigend formuleren, ook wanneer die onvolledig of onjuist is. Wie de AI vraagt wat de beste keuze is, ontvangt vaak een afgerond antwoord dat de onzekerheid kleiner doet lijken dan zij werkelijk is. De student hoeft dan nauwelijks nog zelf te wegen.

Daarom moet het ontwerp de denктаak bij de student houden. Laat hem eerst een afweging maken, alternatieven benoemen en onzekerheden aangeven. Pas daarna kan AI tegenargumenten, aanvullende perspectieven of mogelijke gevolgen aandragen. De toepassing fungeert dan als oefenpartner, niet als vervanger van professioneel oordeel.

Besluiten betekent verantwoordelijkheid nemen

Een afweging krijgt pas beroepsmatige betekenis wanneer de student een besluit durft te nemen. Dat besluit hoeft niet altijd de enige juiste keuze te zijn. In veel beroepen bestaan meerdere verdedigbare handelwijzen, zolang de gekozen aanpak past bij de feiten, de professionele normen en de risico's.

AI kan de consistentie van een besluit beproeven. Sluit de keuze aan bij de analyse? Zijn belangrijke randvoorwaarden genegeerd? Is duidelijk wanneer de student hulp moet inschakelen of de beslissing moet voorleggen aan een bevoegde professional? Vooral dat laatste is van belang. Professionele bekwaamheid blijkt immers ook uit het herkennen van de grens van de eigen verantwoordelijkheid.

De uiteindelijke norm kan niet zonder meer aan het model worden overgelaten. Wat verantwoord handelen is, wordt bepaald door het beroep, de onderwijsdoelen en de context. Vakdocenten en vertegenwoordigers uit de beroepspraktijk moeten daarom vaststellen welke beslissingen aanvaardbaar zijn, welke signalen tot overleg moeten leiden en welke fouten te risicovol zijn om ongecorrigeerd te laten.

Uitvoeren maakt zichtbaar wat papier verbergt

Een student kan op papier een verstandig besluit formuleren en toch moeite hebben met de uitvoering. Een slecht-nieuwsgesprek vraagt meer dan weten welke gesprekstechniek passend is. De student moet luisteren, reageren op emotie, begrijpelijke taal gebruiken, grenzen stellen en ondertussen het doel van het gesprek bewaken.

Bij de uitvoering komen kennis, timing en gedrag samen. Daar kan een interactieve toepassing meer zichtbaar maken dan een traditionele casusbeschrijving. De reactie van een gesimuleerde cliënt, patiënt, collega of opdrachtgever verandert namelijk afhankelijk van wat de student zegt of doet. Een onduidelijke uitleg roept nieuwe vragen op. Een te stellige houding kan weerstand vergroten. Een gemiste veiligheidscontrole kan gevolgen hebben voor het verdere verloop.

Toch is een geloofwaardige reactie nog geen bewijs dat de uitvoering beroepsmatig juist is. AI bootst gedrag na en kan daarbij inconsistent, te meegaand of juist overdreven scherp reageren. Menselijke observatie en nabespreking blijven nodig, zeker wanneer communicatie, ethiek of veiligheid zwaar wegen.

Verantwoorden sluit de handeling af én opent een nieuwe leercyclus

Na afloop moet de student kunnen uitleggen wat hij heeft waargenomen, welke alternatieven hij heeft overwogen, waarom hij voor deze aanpak koos en wat hij bij een volgende gelegenheid anders zou doen. Die verantwoording maakt het denkproces bespreekbaar. Zonder die stap ziet een docent vooral de uitkomst, niet hoe die tot stand kwam.

AI kan de student helpen zijn redenering te reconstrueren en tegenstrijdigheden op te sporen. Maar ook hier geldt dat een vloeiende toelichting niet hetzelfde is als een eigen, doorleefde verantwoording. De student moet daarom kunnen reageren op nieuwe vragen, bewijs uit de situatie aanwijzen en uitleggen waarom andere handelwijzen minder passend waren.

Voor een opleiding biedt deze ontleding een praktisch ontwerp- en besliskader. Niet de vraag *waar kunnen we een bot inzetten?* staat voorop, maar: **welke stap in het professionele handelen krijgt nu onvoldoende oefening of feedback?** Misschien herkennen studenten de relevante signalen niet. Misschien blijven hun afwegingen oppervlakkig, vermijden zij beslissingen of kunnen zij een juist besluit niet zorgvuldig uitvoeren. Elke tekortkoming vraagt om een andere oefenvorm en een ander soort begeleiding.

AI kan daarbij situaties variëren, informatie gedoseerd aanbieden, doorvragen en gevolgen laten zien. De opleiding bepaalt wat goed handelen is, bewaakt risicovolle grenzen en beoordeelt of de student ook zonder ondersteuning professioneel kan optreden. Zo ontstaat de basis voor simulatie: een veilige omgeving waarin studenten moeten handelen voordat de beroepspraktijk hun die ruimte geeft.

4.2 Simuleren voordat de praktijk het toelaat

Een simulatie heeft onderwijswaarde wanneer zij studenten laat handelen onder beperkingen die ertoe doen. AI kan daarvoor een oefensituatie laten reageren op vragen en beslissingen, nieuwe informatie geven en omstandigheden veranderen. De student moet met ontbrekende informatie leren omgaan, prioriteiten stellen, rekening houden met tijd of middelen en ervaren dat keuzes gevolgen hebben. Zo wordt zichtbaar of kennis paraat is wanneer professioneel handelen wordt gevraagd – juist in situaties die in de echte praktijk schaars, kostbaar of riskant zijn.

Veilig oefenen met situaties die schaars of riskant zijn

Niet iedere beroepssituatie is gemakkelijk te organiseren. Sommige situaties komen weinig voor, andere zijn te kostbaar, belastend of riskant om zonder voorbereiding te oefenen. Denk aan een slecht-nieuwsgesprek, een escalatie met een klant, een storing in een kritisch proces of een besluit waarbij veiligheid, geld en verschillende belangen tegelijk een rol spelen.

In een simulatie kan de student zo'n situatie herhaaldelijk doorlopen zonder dat een echte cliënt, patiënt, burger, collega of organisatie de gevolgen draagt. Dat verlaagt de praktische drempel om te oefenen. Bovendien kan een opleiding verschillende varianten aanbieden: eerst een overzichtelijke situatie, daarna een scenario met onvolledige informatie, tegengestelde belangen of toenemende tijdsdruk.

Die herhaalbaarheid is een belangrijke mogelijke baten. In het reguliere onderwijs krijgt een student soms maar één gelegenheid om een bepaalde beroepssituatie te oefenen. Bij een grote groep is het immers moeilijk om voor iedere student voldoende rollenspelers, praktijkbegeleiders en passende casussen beschikbaar te stellen. Een AI-gestuurde simulatie kan extra oefenronden mogelijk maken tussen bijeenkomsten met een docent of trainingsacteur.

Dat betekent niet dat de simulatie het menselijke oefengesprek vervangt. Zij kan vooral de schaarse menselijke begeleiding richter maken. Studenten komen beter voorbereid naar een begeleide oefening en hebben al ervaren waar hun aanpak vastloopt. De docent kan de beschikbare tijd vervolgens gebruiken voor observatie, verdieping en professioneel oordeel in plaats van voor de eerste kennismaking met de situatie.

Realistische beperkingen maken het verschil

Een simulatie wordt niet realistisch door veel details of een levendige schrijfstijl.

Doorslaggevend is of de student te maken krijgt met beperkingen die ook in het beroep het handelen bepalen. Zonder zulke beperkingen verandert de oefening al snel in een vriendelijk vraaggesprek waarin ieder redelijk antwoord wordt geaccepteerd.

Relevante beperkingen kunnen onder meer bestaan uit:

- onvolledige, dubbelzinnige of tegenstrijdige informatie;
- beperkte tijd, weinig geld of onvoldoende personele capaciteit;
- wettelijke, ethische of organisatorische grenzen;
- emoties, weerstand of uiteenlopende belangen;
- afhankelijkheid van andere professionals;
- beslissingen die niet zonder meer kunnen worden teruggedraaid.

Welke beperkingen nodig zijn, hangt af van de beroepshandeling die de opleiding wil oefenen. Bij een adviesgesprek kan het gaan om doorvragen en omgaan met weerstand. In een technisch scenario kunnen veiligheidseisen en de volgorde van handelingen bepalend zijn. Bij een managementbesluit draait het wellicht om het afwegen van kwaliteit, continuïteit, kosten en draagvlak.

De opleiding moet deze beperkingen daarom vooraf ontwerpen. Wanneer de AI ze ter plekke zelf verzint, kan een scenario wel spannend lijken, maar toch vakinhoudelijk ontsporen. De geloofwaardigheid van een antwoord zegt immers weinig over de juistheid ervan. Zeker wanneer studenten nog weinig ervaring hebben, herkennen zij niet altijd dat een voorwaarde onrealistisch is of dat een zogenaamd professioneel advies in strijd is met wetgeving, richtlijnen of gangbare beroepsnormen.

Gevolgen moeten betekenisvol zijn

Keuzes krijgen pas gewicht wanneer zij het verdere scenario beïnvloeden. Als iedere aanpak uiteindelijk tot dezelfde reactie leidt, leert de student vooral dat beslissingen er niet werkelijk toe doen. Een bruikbare simulatie laat daarom zien wat een keuze veroorzaakt, zonder dat zij elk handelen onmiddellijk als goed of fout bestempelt.

Een student die te snel een diagnose stelt, mist bijvoorbeeld informatie die later van belang blijkt. Wie een bezwaar negeert, kan vertrouwen verliezen. Wie geen veiligheidscontrole uitvoert, krijgt met een storing of incident te maken. Wie zorgvuldig doorvraagt, ontdekt misschien een achterliggend probleem, maar verbruikt ook tijd. Zo ontstaat een keten van keuzes en gevolgen waarop de student achteraf kan terugkijken.

Die gevolgen hoeven niet spectaculair te zijn. In veel beroepen is het effect van professioneel handelen juist subtiel: een gesprek loopt stroever, een collega neemt minder verantwoordelijkheid, een risico blijft langer onzichtbaar of een besluit wordt moeilijker uitvoerbaar. Zulke gevolgen zijn didactisch waardevol wanneer zij rechtstreeks samenhangen met het beoogde professionele handelen.

Wel moet de opleiding voorkomen dat de simulatie één ideale route suggereert waarlangs iedere student hetzelfde eindpunt bereikt. Professionele situaties kennen

vaak meerdere verdedigbare handelwijzen. Het gaat er dan niet alleen om *welke* keuze de student maakt, maar vooral welke signalen hij zag, welke belangen hij afwoog en hoe hij zijn besluit kan verantwoorden.

De nabespreking draagt het leren

Een afgerond scenario is nog geen afgerond leerproces. Zonder nabespreking kan de student een mislukte aanpak aan toeval wijten, een toevallig succesvolle aanpak overschatten of verkeerde feedback voor waar aannemen. De belangrijkste opbrengst ontstaat daarom vaak nadat de simulatie is gestopt.

In de nabespreking reconstrueert de student zijn handelen. Welke signalen waren beschikbaar? Welke aannames deed hij? Wanneer veranderde zijn beeld van de situatie? Welke mogelijkheid zag hij over het hoofd? Hoe verhouden zijn keuzes zich tot beroepsnormen en vakinhoudelijke kennis? En wat zou hij bij een volgende poging anders doen?

AI kan daarbij helpen door het verloop vast te leggen, beslismomenten te markeren en de student gerichte reflectievragen voor te leggen. Het deskundige oordeel blijft echter bij de docent of praktijkbegeleider. Die kan onderscheiden of een ongewenste uitkomst voortkwam uit een zwakke analyse, een onzorgvuldige uitvoering, een discutabele ontwerpkeuze in het scenario of eenvoudigweg een fout van het AI-systeem.

Voor managers ligt hier een wezenlijke ontwerpvoorwaarde. Wanneer de businesscase alleen rekent met goedkope, zelfstandige oefeningen, blijft het belangrijkste begeleidingswerk buiten beeld. Scenario's moeten worden ontworpen, vakinhoudelijk gecontroleerd, onderhouden en nabesproken. Ook moet iemand afwijkende of onjuiste reacties onderzoeken. Daar staat tegenover dat menselijke begeleiding doelgerichter kan worden ingezet en dat studenten vaker kunnen oefenen dan met uitsluitend docentgeleide werkvormen haalbaar is.

De bestuurlijke vraag luidt dus niet of een simulatie overtuigend aanvoelt, maar of zij aantoonbaar het gewenste handelingsvermogen oefent. Dat vraagt om vaste leerdoelen, realistische beperkingen, betekenisvolle gevolgen en een deskundige nabespreking. Ontbreekt een van die elementen, dan kan een simulatie een indrukwekkende ervaring opleveren zonder dat vaststaat wat de student ervan leert. Erger nog: een geloofwaardig scenario kan ook verkeerde kennis, onmogelijke omstandigheden of ondeugdelijk professioneel gedrag bevestigen.

4.3 Uit mijn onderwijspraktijk: een geloofwaardige maar verkeerde casus

Dat gebeurde bij een simulatie waarin studenten een adviesgesprek voerden met een fictieve opdrachtgever. De student moest niet alleen vragen stellen, maar ook belangen afwegen, risico's herkennen en een eerste advies formuleren. De bot speelde de opdrachtgever en reageerde op de keuzes van de student. Op het eerste gezicht werkte

dat uitstekend: het gesprek verliep natuurlijk, de opdrachtgever gaf concrete antwoorden en studenten moesten wel doorvragen.

Juist die overtuigingskracht maakte het probleem verraderlijk.

In een van de gesprekken vertelde de bot dat de organisatie alle klantgegevens versleuteld opsloeg, dat toestemming voor hergebruik correct was geregeld en dat een externe leverancier aan specifieke beveiligingsnormen voldeed. Die gegevens stonden nergens in het scenario. De bot had ze zelf toegevoegd om het gesprek samenhangend te laten verlopen. De formuleringen klonken professioneel en pasten bij de rest van de casus, maar vakinhoudelijk waren ze niet te verantwoorden.

Daarmee veranderde ook de opdracht zonder dat iemand daarvoor had gekozen. Een student die de verzonnen informatie accepteerde, kon tot een geruststellende risicoanalyse komen. Een student die kritischer doorvroeg, kreeg soms nóg meer niet-bestaande details als antwoord. De simulatie beloonde dan niet het beste professionele handelen, maar het vermogen om mee te bewegen met een geloofwaardig verteld verhaal.

Ik merkte bovendien dat studenten de fout niet vanzelf herkenden. Dat was niet vreemd. Zij zaten midden in een realistisch gesprek, wilden reageren op de opdrachtgever en mochten aannemen dat de casusinformatie doelbewust was opgenomen. Waar een docent bij een slordig antwoord misschien nog zou denken: hier klopt iets niet, zag een student vooral een zelfverzekerde gesprekspartner die concrete feiten presenteerde.

Na afloop konden studenten hun aanpak goed toelichten. Sommigen hadden keurig gevraagd naar gegevensbeheer, toestemming en leveranciersafspraken. Toch kon ik hun handelen niet zuiver beoordelen, omdat het scenario tijdens het gesprek ongemerkt was veranderd. De ene student had andere feiten gekregen dan de andere. Daardoor oefenden zij niet langer onder vergelijkbare omstandigheden en was onduidelijk of een goed advies voortkwam uit bekwaam handelen of uit toevallig gunstige informatie van de bot.

Mijn eerste reactie was de instructie voor de bot aan te scherpen: gebruik alleen de informatie uit de casus en verzin niets. Dat hielp enigszins, maar niet voldoende. Zodra studenten een vraag stelden waarop het scenario geen antwoord bevatte, probeerde de bot toch behulpzaam te zijn. Soms vulde hij een ontbrekend gegeven in, soms gaf hij een algemeen antwoord alsof dat voor deze organisatie gold. Een verbod op verzinnen bleek dus geen betrouwbare begrenzing.

Daarom heb ik de simulatie anders opgebouwd. Eerst legde ik vast welke scenario-elementen gedurende het gesprek niet mochten veranderen: het type organisatie, de beschikbare gegevens, de belangen van betrokkenen, de technische omgeving, de geldende beperkingen en de informatie die de opdrachtgever wel en niet kende. Ook bepaalde ik vooraf welke onzekerheden bewust in de casus bleven zitten. Niet iedere vraag hoefde immers een antwoord op te leveren. In de beroepspraktijk is 'dat weten

we nog niet' vaak een belangrijker gegeven dan een snel ingevuld detail.

Vervolgens kreeg de bot niet meer de ruimte om ontbrekende informatie aannemelijk aan te vullen. Bij vragen buiten de vastgelegde casus moest hij onzekerheid uitspreken, aangeven dat de opdrachtgever het antwoord niet kende of de student vragen hoe die het ontbrekende gegeven wilde onderzoeken. Daarmee werd het ontbreken van informatie onderdeel van de oefening. Studenten moesten nu beslissen of zij nader onderzoek adviseerden, een voorbehoud maakten of hun advies uitstelden.

Ook begrensde ik de kennisbron. De bot mocht voor vakinhoudelijke uitspraken alleen putten uit geselecteerd cursusmateriaal en uit de vaste beschrijving van het scenario. Die bronbegrenzing maakte de simulatie niet onfeilbaar, maar verkleinde wel de ruimte voor willekeurige toevoegingen. Bovendien werd beter controleerbaar waarop een reactie was gebaseerd.

De derde aanpassing was menselijke controle vóór en na gebruik. Vooraf doorliep ik het scenario met enkele uiteenlopende gespreksstrategieën. Ik stelde voor de hand liggende vragen, maar probeerde de bot ook uit zijn rol te lokken en vroeg bewust naar ontbrekende gegevens. Zo kwamen zwakke plekken boven voordat studenten ermee werkten. Na de oefening bekeken we niet alleen wat de student had gedaan, maar ook of de bot zich aan de casus had gehouden. Een vreemd antwoord werd daarmee geen fout die de student maar moest zien te verwerken, maar een signaal dat het ontwerp onderhoud nodig had.

Deze werkwijze kostte meer tijd dan ik aanvankelijk had voorzien. Een simulatiebot bouwen is betrekkelijk eenvoudig; een didactisch betrouwbare simulatie onderhouden is dat niet. Scenario's moeten worden getest, reacties moeten steekproefsgewijs worden gecontroleerd en fouten moeten leiden tot aanpassing. Wie alleen de oefeningen van studenten telt, ziet dat werk gemakkelijk over het hoofd.

Daar stond een duidelijke verbetering tegenover. De gesprekken werden misschien iets minder vloeiend, maar vakinhoudelijk bruikbaar. Studenten leerden dat professionele onzekerheid niet met een aannemelijk antwoord mag worden weggepoetst. Zij moesten expliciet maken welke informatie ontbrak, wat zij wel en niet konden concluderen en wanneer deskundige of bestuurlijke besluitvorming nodig was. De simulatie oefende daardoor niet alleen gespreksvaardigheid, maar ook het begrenzen van een advies.

Voor mij was dit een belangrijke ontwikkelstap. De waarde van een simulatie zit niet in de vrijheid waarmee de bot een levendige werkelijkheid oproept. Voor handelingsgericht onderwijs is juist gecontroleerde vrijheid nodig: voldoende variatie om studenten te laten kiezen, maar voldoende vaste grond om hun keuzes inhoudelijk te kunnen beoordelen.

Een geloofwaardige reactie is daarbij geen kwaliteitskenmerk op zichzelf. Soms is een stroef maar eerlijk antwoord – 'de opdrachtgever beschikt niet over die informatie' – didactisch waardevoller dan een fraai uitgewerkte onwaarheid. Pas toen

bronbegrenzing, vaste scenario-elementen en docentcontrole samenkwamen, kon ik de simulatie verantwoord als volgende oefenstep inzetten.

Ook toen bleef voorzichtigheid nodig. De aangepaste simulatie liet zien hoe studenten handelden binnen een begrensde oefenomgeving. Zij bewees nog niet dat studenten in een nieuwe, minder voorspelbare beroepssituatie hetzelfde niveau zouden tonen. Daarvoor moest ik niet alleen naar het verloop van het gesprek kijken, maar naar de kwaliteit van hun beslissingen, hun onderbouwing en hun vermogen om onzekerheid te herkennen.

Eerlijke grens in de simulatie



4.4 Besluiten over handelingsvermogen

Juist daar ligt de kern van de bestuurlijke beoordeling. Een simulatie rendeert niet omdat studenten er actief mee bezig zijn, omdat het gesprek realistisch aanvoelt of omdat de toepassing veel wordt gebruikt. De relevante vraag is of studenten daardoor aantoonbaar beter leren handelen – ook wanneer de situatie verandert, informatie onvolledig is en niet meteen duidelijk is welke aanpak de juiste is.

Dat vraagt om meer dan een tevredenheidsmeting. Studenten kunnen een simulatie leerzaam vinden zonder dat hun professionele handelen verbetert. Ook een hoge gebruiksfrequentie bewijst weinig: die kan wijzen op intensief oefenen, maar net zo goed op een onduidelijke opdracht, herhaald vastlopen of vrijblijvend gebruik. Bestuurders moeten daarom vooraf laten vastleggen welke professionele

bekwaamheden de simulatie moet helpen ontwikkelen en waaraan verbetering zichtbaar wordt. Het vermogen om die bekwaamheden zelfstandig en verantwoord in wisselende situaties in te zetten, noemen we hier handelingsvermogen.

Maak professioneel handelen waarneembaar

Professioneel handelen bestaat zelden uit één losse vaardigheid. In een beroepssituatie moet een student doorgaans informatie verzamelen, belangen wegen, een beslissing nemen, daarover communiceren en de gevolgen van die beslissing beoordelen. Bovendien moet de student herkennen wanneer de eigen kennis tekortschiet en hulp van een mens of specialist nodig is.

Een simulatie kan zulke handelingen zichtbaar maken, mits het ontwerp daarvoor ruimte biedt. Een toepassing die studenten vooral naar het juiste antwoord leidt, meet eerder patroonherkenning dan zelfstandig handelen. Een bruikbare simulatie stelt de student juist voor keuzes, laat verschillende verdedigbare routes toe en bevat gevolgen die uit die keuzes voortvloeien.

Leg daarom per toepassing ten minste vast:

- welke professionele handelingen studenten moeten oefenen;
- welk beginniveau wordt verwacht;
- welke beslissingen of gedragingen als bewijs worden gebruikt;
- welke variatie de simulatie aanbiedt;
- welke fouten leerzaam en aanvaardbaar zijn;
- wanneer een docent moet ingrijpen;
- hoe wordt vastgesteld of de student het geleerde ook buiten het geoefende scenario toepast.

Dat laatste punt is doorslaggevend. Wie één herkenbaar gesprek goed doorloopt, kan een aangeleerd stramien hebben gevolgd. Pas wanneer de student in een nieuwe situatie opnieuw relevante informatie selecteert, belangen afweegt en keuzes onderbouwt, ontstaat een aannemelijke aanwijzing voor overdracht.

Bouw bewijs in stappen op

Niet iedere simulatie hoeft meteen met een uitgebreid vergelijkend onderzoek te worden beoordeeld. Het bewijs moet wel passen bij de bestuurlijke claim. Hoe groter de investering, het beroepsrisico en het voorgenomen bereik, hoe sterker het bewijs moet zijn.

Een eerste proef kan beginnen met observaties van enkele docenten en een analyse van studentreacties. Daarmee wordt zichtbaar of het scenario begrijpelijk is, of de interactie voldoende realistisch verloopt en waar studenten vastlopen. Dat is vooral ontwerpinformatie. Het zegt nog weinig over het bereikte leereffect.

De volgende stap is een vergelijking van prestaties voor en na het oefenen. Studenten

voeren bijvoorbeeld eerst een beroepsgerichte opdracht uit, oefenen daarna met verschillende simulaties en krijgen vervolgens een nieuwe casus. Beoordelaars kijken niet alleen naar het eindresultaat, maar ook naar de gevolgde aanpak: welke informatie gebruikte de student, welke signalen herkende hij, hoe woog hij alternatieven en hoe verantwoordde hij zijn beslissing?

Sterker bewijs ontstaat wanneer prestaties worden vergeleken met een passende uitgangssituatie of een andere oefenvorm. Daarbij hoeft de vraag niet te zijn of AI beter is dan een docent. Veel relevanter is of de combinatie van simulatie en menselijke begeleiding meer of betere oefening mogelijk maakt dan binnen de bestaande inzet haalbaar was.

Bewijs	Wat het aannemelijk maakt	Wat het nog niet bewijst
Studenten gebruiken de simulatie regelmatig	De toepassing is beschikbaar en wordt benut	Dat studenten beter handelen
Studenten en docenten ervaren de simulatie als leerzaam	De oefenvorm wordt herkend als bruikbaar	Dat prestaties duurzaam verbeteren
Studenten handelen beter in een vergelijkbaar vervolgsценario	Er is vooruitgang binnen een verwante context	Dat overdracht naar andere situaties optreedt
Studenten handelen beter in een nieuwe, minder voorspelbare situatie	Overdracht is aannemelijk	Dat dit resultaat in de beroepspraktijk standhoudt
Verbetering blijft zichtbaar over meerdere cohorten	Het effect is mogelijk bestendig	Dat de toepassing zonder aanpassingen kan worden opgeschaald

Reken het nieuwe werk mee

Simulaties worden soms gepresenteerd als een manier om onbeperkt en tegen lage kosten te oefenen. Dat beeld laat een belangrijk deel van het werk buiten beschouwing. Een verantwoorde simulatie moet worden ontworpen, inhoudelijk gecontroleerd, getest, onderhouden en geëvalueerd. Scenario's verouderen, beroepsnormen veranderen en een aanpassing van het onderliggende model kan onverwacht ander gedrag veroorzaken.

Ook tijdens het onderwijs ontstaan nieuwe taken. Docenten moeten afwijkende interacties kunnen beoordelen, fouten in scenario's melden, studenten helpen reacties van de toepassing kritisch te interpreteren en nagaan of veelvuldig oefenen niet ontaardt in het aanleren van kunstmatige gesprekspatronen. Bij spraaktoepassingen komen daar technische ondersteuning, toegankelijkheidsvragen en mogelijk extra gegevensverwerking bij.

De businesscase moet daarom niet alleen vermelden hoeveel docenttijd per

oefenmoment wordt bespaard. Zij moet ook laten zien:

- hoeveel ontwikkeltijd nodig is voor een betrouwbaar scenario;
- wie inhoudelijk eigenaar is en correcties goedkeurt;
- hoe vaak controle en actualisering plaatsvinden;
- hoeveel begeleiding studenten nodig hebben;
- welke technische en functionele ondersteuning vereist is;
- welke kosten ontstaan door licenties, gegevensopslag en koppelingen;
- waarvoor eventueel vrijgekomen docenttijd daadwerkelijk wordt ingezet.

Tijdwinst krijgt immers pas onderwijswaarde wanneer die tijd beschikbaar komt én bewust wordt gebruikt voor taken die menselijk oordeel vragen, zoals verdiepend nabespreken, gerichte coaching of het beoordelen van complexe prestaties.

Kies tussen stoppen, aanpassen, voortzetten en uitbreiden

Na een proef zijn er vier verdedigbare besluiten. **Stoppen** ligt voor de hand wanneer scenario's inhoudelijk onbetrouwbaar blijven, risico's niet beheersbaar zijn of er geen aanwijzing is dat studenten beter handelen. **Aanpassen** past wanneer de didactische waarde aannemelijk is, maar scenario-opbouw, bronbegrenzing, toegankelijkheid of begeleiding tekortschiet.

Voortzetten is verantwoord wanneer studenten aantoonbaar meer of gericht oefenen, hun prestaties in verwante situaties verbeteren en de ontwikkel- en begeleidingslast in verhouding staat tot die opbrengst. **Uitbreiden** vraagt meer. Dan moet verbetering ook buiten het oorspronkelijke scenario aannemelijk zijn, moeten verantwoordelijkheden structureel zijn belegd en moet de instelling de toepassing voor meerdere cohorten kunnen onderhouden.

Voor simulaties met ingrijpende professionele gevolgen – bijvoorbeeld rond veiligheid, zorg, juridische advisering of kwetsbare personen – ligt de bewijslast hoger. Daar kan een toepassing waardevol zijn als oefenmiddel, terwijl zelfstandig gebruik zonder docentcontrole toch onverantwoord blijft. Opschaling betekent dan niet minder menselijke betrokkenheid, maar een nauwkeuriger georganiseerde vorm ervan.

De uiteindelijke beslissing rust daarmee op drie samenhangende vragen. **Leren studenten beter handelen? Kunnen zij dat ook in een nieuwe situatie? En zijn de volledige kosten en risico's aanvaardbaar?** Alleen wanneer alle drie overtuigend kunnen worden beantwoord, verdient een simulatie een structurele plaats in het onderwijs. Een levendige oefenervaring is mooi meegenomen; aantoonbaar handelingsvermogen is waar het om gaat.

5 Feedback die de volgende stap mogelijk maakt

Handelen wordt pas leren wanneer de student zicht krijgt op wat goed ging, wat beter kan en welke volgende stap nodig is. Juist daar wringt het vaak. De terugkoppeling komt wanneer de opdracht al is afgerond, de situatie uit beeld is geraakt en de aandacht alweer bij een volgend onderwerp ligt. Goede feedback verliest dan een deel van haar werking doordat zij te laat beschikbaar komt.

AI kan die wachttijd sterk verkorten. Toch is snelheid op zichzelf geen onderwijsopbrengst. Een student heeft weinig aan een onmiddellijke stroom opmerkingen die niet aansluit bij het leerdoel, geen onderscheid maakt tussen hoofd- en bijzaken of vooral uitnodigt tot gedachteloos overnemen. Snelle feedback kan het leren zelfs overslaan wanneer het systeem verbeteringen aandraagt die de student niet hoeft te begrijpen of af te wegen.

De bestuurlijke vraag is daarom niet hoeveel feedback AI kan produceren, maar of studenten er aantoonbaar beter door leren handelen. Daarvoor moet de terugkoppeling gericht, begrijpelijk en tijdig genoeg zijn om er nog iets mee te doen. Tegelijk moet helder blijven waar menselijke deskundigheid nodig is – bij normering, complexe afwegingen en beslissingen met gevolgen. Pas dan kan een kortere feedbackcyclus zowel het leren versterken als professionele ruimte creëren.

Feedback terwijl het nog telt



5.1 Wachtijd als onderwijsprobleem

Diezelfde eis geldt voor feedback. Niet de snelheid van het antwoord, maar wat de student er daarna mee kan doen, bepaalt de onderwijswaarde. Terugkoppeling die binnen enkele seconden verschijnt maar geen gerichte vervolgstap mogelijk maakt, is vooral snelle tekst. Feedback rendeert pas wanneer de student een bevinding kan begrijpen, verwerken en opnieuw kan toepassen.

In veel opleidingen zit tussen een eerste uitwerking en de reactie daarop al snel een week of langer. Dat is organisatorisch verklaarbaar. Docenten combineren het nakijken met lessen, begeleiding, vergaderingen en andere taken, terwijl opdrachten vaak rond hetzelfde moment worden ingeleverd. Voor de student heeft die wachttijd wel degelijk gevolgen. Tegen de tijd dat de feedback komt, is de oorspronkelijke redenering naar de achtergrond verdwenen, is de groep aan een volgend onderwerp begonnen of resteert er nauwelijks tijd om het werk nog wezenlijk te verbeteren.

Daarmee wordt wachttijd een onderwijsprobleem. De student krijgt misschien goede feedback, maar op een moment waarop de werking ervan afneemt. Een aanwijzing als ‘onderbouw deze keuze beter’ is immers gemakkelijker te verwerken wanneer nog helder is welke bronnen zijn geraadpleegd, welke alternatieven zijn overwogen en waarom een bepaalde redenering logisch leek. Twee weken later moet de student eerst de eigen gedachtegang reconstrueren. De terugkoppeling komt dan niet alleen laat, maar vraagt ook extra herstelwerk.

Dat effect wordt versterkt wanneer een opdracht na de feedback feitelijk is afgerond. De student leest de opmerkingen, bekijkt het resultaat en gaat verder. Er volgt geen nieuwe versie, geen vergelijkbare opgave en soms zelfs geen gesprek over de gekozen aanpak. In zo’n opzet heeft snellere feedback weinig betekenis, omdat de leercyclus na de terugkoppeling stopt.

De eerste bestuurlijke vraag is daarom niet: **hoe kunnen we sneller feedback geven?** De betere vraag luidt: **welke volgende leerhandeling willen we met die feedback mogelijk maken?**

Dat kan bijvoorbeeld zijn dat een student:

- een redenering aanscherpt en opnieuw indient;
- ontbrekende bronnen zoekt en de onderbouwing aanvult;
- een berekening controleert en een fout herstelt;
- een professioneel gesprek nogmaals oefent;
- twee oplossingsrichtingen vergelijkt;
- uitlegt waarom feedback wel of niet wordt overgenomen;
- dezelfde aanpak toepast in een nieuwe situatie.

Pas wanneer zo’n vervolgstap in het onderwijsontwerp is opgenomen, krijgt een kortere wachttijd mogelijke waarde. AI kan dan een eerste reactie geven terwijl de student nog met de taak bezig is. Daardoor hoeft een oefencyclus niet te wachten tot een docent tijd

heeft om alle uitwerkingen te bekijken. Een student kan meerdere versies maken, verschillende aanpakken proberen en sneller ontdekken waar de eigen redenering tekortschiet.

Dat betekent niet dat iedere reactie onmiddellijk moet komen. Bij sommige opdrachten is vertraging juist nuttig. Studenten moeten soms eerst zelf onzekerheid verdragen, hun werk controleren of met medestudenten overleggen. Als bij iedere aarzeling meteen een aanwijzing verschijnt, kan feedback het denken overnemen in plaats van ondersteunen. De student leert dan vooral reageren op aanwijzingen en ontwikkelt onvoldoende vermogen om het eigen werk te beoordelen.

Sneller is dus niet altijd beter. Feedback op het juiste moment is beter.

Daarvoor moeten opleidingsteams onderscheid maken tussen verschillende momenten. Tijdens een oefening kan een korte aanwijzing voorkomen dat een student langdurig op een verkeerd spoor doorgaat. Na een eerste versie kan uitgebreidere terugkoppeling helpen om keuzes te herzien. Na afronding kan feedback vooral dienen als voorbereiding op een volgende, verwante taak. Het gewenste moment hangt af van de leerhandeling, de ervaring van de student en de gevolgen van een fout.

Bij beginnende studenten kan vroege ondersteuning nodig zijn om frustratie en herhaling van basisfouten te voorkomen. Gevorderde studenten hebben vaker baat bij ruimte om eerst zelf tot een oordeel te komen. Naarmate de professionele complexiteit toeneemt, moet feedback bovendien minder het karakter krijgen van een correctie en meer van een onderzoekende vraag: Welke afweging hebt u gemaakt? Welk risico accepteert u hiermee? Welke informatie zou uw besluit kunnen veranderen?

Ook voor medewerkers is snelheid niet vanzelf een bate. Als AI veel tussentijdse feedback mogelijk maakt, kan dat nieuwe vragen oproepen. Studenten willen weten waarom twee reacties verschillen, vragen de docent om bevestiging of leveren vaker nieuwe versies in. De nakijklast kan afnemen, terwijl de begeleidingslast juist stijgt. Dat hoeft geen nadeel te zijn: een inhoudelijk gesprek over afwegingen kan waardevoller zijn dan het herhaaldelijk aanwijzen van dezelfde vormfout. Maar die verschuiving moet wel zichtbaar worden gemaakt.

Een geloofwaardige batenclaim vermeldt daarom niet alleen hoeveel sneller studenten een reactie krijgen. Daaruit blijkt ook:

- of studenten de feedback daadwerkelijk raadplegen;
- welke aanpassingen zij vervolgens maken;
- of volgende versies inhoudelijk beter worden;
- of verbetering ook optreedt bij een nieuwe opdracht;
- hoeveel docenttijd nodig blijft voor controle en uitleg;
- welke nieuwe begeleidings- en beheertaken ontstaan.

Zonder die gegevens blijft de constatering dat een systeem binnen tien seconden feedback geeft vooral een technische prestatie. Onderwijsrendement ontstaat pas als

studenten dankzij die terugkoppeling beter leren analyseren, schrijven, ontwerpen, handelen of besluiten.

Voor managers heeft dit een rechtstreeks gevolg. Een toepassing voor snelle feedback moet niet als los hulpmiddel worden ingekocht en pas daarna in het curriculum worden geplaatst. Eerst moet duidelijk zijn waar de huidige cyclus vastloopt. Wachten studenten zo lang dat verwerking niet meer plaatsvindt? Ontvangen zij te weinig oefengerichte terugkoppeling? Besteedt de docent veel tijd aan terugkerende opmerkingen die ook eerder in het proces gegeven hadden kunnen worden? Of ligt het werkelijke probleem niet bij de snelheid, maar bij de vaagheid van de opdracht, de criteria of de begeleiding?

Dat onderscheid voorkomt dat technologie een didactisch tekort slechts sneller reproduceert. Een onduidelijke reactie wordt niet waardevoller doordat zij onmiddellijk verschijnt. Een algemene opmerking als ‘meer diepgang nodig’ helpt de student nauwelijks verder, ongeacht of die na tien seconden of tien dagen komt. Bruikbare feedback moet zichtbaar maken waarop de reactie betrekking heeft en welke denk- of handelingsstap nu nodig is.

Daarmee verschuift de aandacht vanzelf van snelheid naar inhoud. Goede terugkoppeling kan zich richten op wat de student heeft gemaakt, op de werkwijze die daartoe leidde of op de professionele afweging achter een keuze. AI kan daarbij patronen herkennen, inconsistenties aanwijzen en vragen stellen. Welke feedback passend is, hangt echter af van wat de student moet leren – en van welk oordeel nadrukkelijk bij de docent moet blijven.

5.2 Feedback op product, proces en afweging

Het eerste onderscheid is dat tussen feedback op het product, het proces en de afweging. Die drie kunnen in één opdracht samenkomen, maar ze vragen niet om dezelfde reactie. Wie ze op één hoop gooit, krijgt al snel een lange lijst opmerkingen waaruit de student zelf maar moet afleiden wat ertoe doet. AI kan dan veel tekst produceren zonder een bruikbare volgende stap mogelijk te maken.

Feedback op het product

Productfeedback richt zich op wat er op dat moment ligt: een advies, ontwerp, analyse, berekening, presentatie, rapportage of ander beroepsproduct. De reactie gaat over waarneembare kenmerken daarvan. Is de hoofdvraag beantwoord? Sluiten conclusie en onderbouwing op elkaar aan? Ontbreekt er relevante informatie? Zijn begrippen consequent gebruikt? Kan de beoogde gebruiker met het product uit de voeten?

Op dit niveau kan AI relatief eenvoudig patronen signaleren. Het systeem kan bijvoorbeeld vaststellen dat een advies drie opties noemt, terwijl slechts één daarvan met argumenten wordt afgewogen. Ook kan het inconsistenties aanwijzen, controleren of verplichte onderdelen aanwezig zijn of passages markeren waarin een bewering geen

zichtbare onderbouwing heeft.

Dat is vooral bruikbaar tijdens het maken. De student hoeft dan niet te wachten tot het product af is, maar kan een tussenversie controleren en gericht verbeteren. De docent ontvangt later mogelijk een product waarin basale onvolkomenheden al zijn aangepakt en kan de beschikbare tijd gebruiken voor moeilijkere inhoudelijke en professionele vragen.

Toch kent productfeedback een duidelijke grens. Dat een onderdeel aanwezig is, wil nog niet zeggen dat het inhoudelijk deugt. Een tekst kan alle gevraagde kopjes bevatten en toch een zwakke analyse geven. Een bron kan correct zijn vermeld, maar ongeschikt zijn voor de conclusie die ermee wordt onderbouwd. AI kan zulke problemen soms herkennen, maar niet betrouwbaar genoeg om daar zonder meer een formeel oordeel aan te verbinden.

Feedback op het proces

Procesfeedback gaat niet alleen over wat de student heeft gemaakt, maar vooral over hoe die tot het resultaat is gekomen. Welke stappen zijn gezet? Welke informatie is gezocht en geselecteerd? Welke alternatieven zijn onderzocht? Wanneer heeft de student een aanname gecontroleerd? Hoe is eerdere feedback verwerkt?

Deze vorm van terugkoppeling is belangrijk omdat twee vergelijkbare producten uit heel verschillende leerprocessen kunnen voortkomen. De ene student heeft zelfstandig onderzocht, afgewogen en bijgesteld. De andere heeft een plausibele tekst samengesteld zonder de onderliggende redenering te doorgronden. Alleen het eindproduct laat dat verschil lang niet altijd zien, zeker nu studenten met generatieve AI snel verzorgde uitwerkingen kunnen maken.

AI kan het proces ondersteunen door gerichte vragen te stellen:

- Welke informatie heeft uw keuze veranderd?
- Welk alternatief hebt u overwogen en waarom hebt u dat verworpen?
- Op welk punt liep uw aanpak vast?
- Welke eerdere feedback hebt u verwerkt en waar is dat zichtbaar?
- Wat zou u bij een volgende uitvoering anders doen?

De waarde zit hier niet in het antwoord dat AI geeft, maar in het denkwerk dat de vraag bij de student uitlokt. Dat vraagt wel om een zorgvuldig ontwerp. Wanneer een systeem na iedere stap een uitgebreide aanwijzing geeft, neemt het de student al gauw bij de hand. Het proces verloopt dan misschien soepeler, maar de student oefent minder met zelf vaststellen waar het misgaat, keuzes maken en omgaan met onzekerheid.

Procesfeedback moet daarom passen bij het ontwikkelingsniveau. Een beginnende student kan baat hebben bij een concrete aanwijzing over een overgeslagen stap. Van een gevorderde student mag eerder worden gevraagd zelf te bepalen waar de aanpak tekortschiet. Naarmate de professionele bekwaamheid groeit, verschuift de

terugkoppeling idealiter van uitleg en aanwijzingen naar vragen, tegenspraak en verantwoording.

Feedback op de professionele afweging

Het derde niveau betreft de redenering achter een keuze. Juist daar wordt zichtbaar of een student professionele normen, belangen, risico's en context met elkaar kan verbinden. Waarom is in deze situatie voor deze aanpak gekozen? Wie ondervindt daarvan voordeel of nadeel? Welke onzekerheid blijft bestaan? Welk risico is aanvaardbaar en wie mag daarover beslissen?

AI kan hierbij een waardevolle tegenspeler zijn. Het systeem kan een ontbrekend belang benoemen, een tegenargument formuleren of vragen wat er verandert als een belangrijke aanname niet klopt. Bij een advies over gegevensgebruik kan het bijvoorbeeld niet alleen vragen of de gekozen werkwijze juridisch is toegestaan, maar ook of zij uitlegbaar, proportioneel en professioneel wenselijk is. Zo wordt feedback minder een controle achteraf en meer een uitnodiging om het oordeel te verdiepen.

Precies hier neemt ook het risico toe. Professionele afwegingen zijn zelden volledig uit criteria af te leiden. Ze hangen samen met omstandigheden, ervaringskennis, waarden en de mogelijke gevolgen voor echte mensen. Een overtuigend geformuleerde reactie van AI kan de indruk wekken dat er één juist antwoord bestaat, terwijl er in werkelijkheid verschillende verdedigbare keuzes zijn. Bovendien kan het systeem relevante context missen of risico's verkeerd wegen.

De docent blijft daarom verantwoordelijk voor de normering. Die bepaalt niet alleen of een beroepsproduct aan de eisen voldoet, maar ook hoeveel ruimte er binnen die eisen bestaat voor verschillende oplossingen. De docent herkent uitzonderingen, bespreekt twijfelgevallen en grijpt in wanneer een beslissing grote gevolgen heeft. AI kan argumenten spiegelen en onvolledigheden signaleren, maar mag het professionele oordeel niet ongemerkt vervangen.

Van signalering naar een volgende stap

Voor een bruikbare toepassing moet vooraf duidelijk zijn waarop de AI reageert. Bij productfeedback zijn dat bijvoorbeeld kwaliteitscriteria voor het beroepsproduct. Bij procesfeedback gaat het om een gewenste werkwijze of leerstrategie. Bij feedback op afwegingen zijn beroepsnormen, belangen en verantwoordingsvragen het vertrekpunt. Zonder zo'n grondslag ontstaat algemene terugkoppeling die wel verzorgd klinkt, maar weinig richting geeft.

Daarbij is meer feedback niet vanzelf beter. Een student die twaalf aandachtspunten tegelijk ontvangt, kan moeite hebben om te bepalen waar te beginnen. Een lange reactie vergroot bovendien de kans dat de student de voorgestelde verbeteringen klakkeloos overneemt of de tekst in haar geheel negeert. Vaak is één goed gekozen vraag of één concreet verbeterpunt didactisch waardevoller dan een volledige analyse.

Voor managers betekent dit dat een feedbacktoepassing niet alleen beoordeeld moet worden op snelheid of het aantal gegenereerde reacties. Relevanter is of de feedback aantoonbaar aansluit bij het leerdoel, of studenten haar begrijpen en verwerken en of zij daarna zelfstandiger en beter handelen. Ook moet zichtbaar zijn welk nieuw werk ontstaat: criteria onderhouden, reacties controleren, uitzonderingen behandelen, studenten leren hoe zij met de feedback omgaan en bewaken dat automatische terugkoppeling niet alsnog als beoordeling wordt gebruikt.

De ontwerpkeuze is uiteindelijk scherp: AI signaleert, bevraagt en helpt een vervolgstap te kiezen; de docent stelt de norm, duidt de context en beslist waar de gevolgen groot zijn. Binnen die taakverdeling kan snelle feedback de leercyclus werkelijk verkorten. Zonder die begrenzing ontstaat vooral méér terugkoppeling – en is nog allerm minst zeker dat die ook gericht is.

5.3 Uit mijn onderwijspraktijk: snel was nog niet gericht

Nadat ik de feedbackreactie had teruggebracht tot één leerdoel, één herkenbaar criterium en één haalbare vervolgstap, verschoof mijn aandacht van de kwaliteit van één reactie naar de kwaliteit van de cyclus. Studenten leverden een concept in, ontvingen één gericht aandachtspunt, pasten hun werk aan en vroegen daarna opnieuw feedback. Pas dan kwam zo nodig een volgend criterium aan bod. Daarmee ontstond een kort ritme van handelen, terugkoppeling en verbeteren.

Om te kunnen beoordelen wat binnen zo'n cyclus veranderde, legde ik de studentproducten vóór en na de feedback naast elkaar. Daarbij keek ik niet alleen of de tweede versie verzorgder was, maar vooral of de student het inhoudelijke leerdoel beter zichtbaar maakte. Als een student moest leren een keuze te onderbouwen, was een aangepaste indeling of formulering immers onvoldoende. In de nieuwe versie moest blijken waarom de gekozen oplossing in deze praktijksituatie passender was dan een alternatief.

Ik zag dat studenten binnen zo'n cyclus gemakkelijker konden uitleggen wat zij hadden aangepast. Ook werd beter zichtbaar of zij de feedback werkelijk hadden verwerkt. Een verandering in het product was immers niet genoeg; de student moest kunnen toelichten waarom die verandering paste bij het leerdoel en het criterium. Die toelichting gaf de docent meer informatie dan de vraag of alle opmerkingen van de bot waren 'afgevinkt'.

Ook het moment waarop een student vastliep, werd eerder zichtbaar. Bij een lange feedbackreactie kon een student meerdere oppervlakkige aanpassingen doen zonder het onderliggende probleem in de redenering op te lossen. Binnen de korte cyclus bleef dat probleem niet verscholen tussen opmerkingen over structuur, volledigheid, formulering en brongebruik. Als de student de gekozen vervolgstap niet kon uitvoeren of verklaren, was duidelijk dat aanvullende uitleg of begeleiding nodig was.

Soms bleek een vraag daarbij bruikbaar dan een voorgestelde verbetering. Bij twijfel

over de bedoeling van een passage vroeg de bot bijvoorbeeld welke keuze de student wilde verdedigen of welk bewijs daarvoor het zwaarst woog. Het antwoord maakte zichtbaar hoe de student redeneerde en voorkwam dat de bot die redenering zelf invulde. De student moest opnieuw naar het eigen werk kijken voordat een volgende bewerking mogelijk was.

Door de herhaling veranderde bovendien de functie van de feedback. De terugkoppeling was niet langer een afsluitend oordeel over een ingeleverd product, maar een aanleiding voor een volgende leerhandeling. De student moest iets vergelijken, aanvullen, herformuleren of opnieuw onderbouwen en kon daarna laten zien wat die bewerking had opgeleverd. Voor de docent ontstond zo een beter waarneembaar spoor van concept, aandachtspunt, aanpassing en toelichting.

Dat spoor maakte nog niet automatisch duidelijk dat de student zelfstandiger leerde handelen. Een student kon de voorgestelde stap immers uitvoeren zonder het criterium werkelijk te begrijpen. Daarom bleef de toelichting belangrijk: wat hebt u veranderd, waarom was dat nodig en hoe past die verandering bij het leerdoel? Juist de combinatie van een aangepast product en een onderbouwde uitleg gaf zicht op de verwerking van de feedback.

Niet alles kon aan de bot worden overgelaten. De keuze van het belangrijkste aandachtspunt bleek lastig wanneer criteria onderling samenhangen, wanneer een student een onverwachte maar verdedigbare aanpak koos of wanneer de opdracht ruimte liet voor meerdere professionele afwegingen. Juist daar bleef docentcontrole nodig. De bot kon een patroon signaleren en een vervolgstap voorstellen, maar de docent moest bewaken of die stap inhoudelijk juist en didactisch passend was.

De kortere cyclus maakte die controle wel gericht. De docent hoefde niet iedere formulering in een uitvoerige reactie te beoordelen, maar kon nagaan of het gekozen aandachtspunt aansloot bij het leerdoel, of de voorgestelde stap uitvoerbaar was en of de student daarna inhoudelijk verder kwam. Tegelijk ontstond nieuw werk: de docent moest criteria scherp houden, uitzonderingen herkennen en ingrijpen wanneer de bot een professionele afweging te eenvoudig voorstelde.

Deze werkwijze leverde mij geen bewijs dat studenten door de bot vanzelf beter leerden. Daarvoor waren een systematische vergelijking en een langere evaluatie nodig. Wel werd in de praktijk een duidelijker mechanisme zichtbaar: studenten ontvingen een aandachtspunt, handelden opnieuw en moesten de gemaakte verandering toelichten. Daardoor kon ik beter waarnemen waar zij het criterium begrepen, waar zij slechts een aanwijzing volgden en waar menselijke begeleiding nodig bleef.

Ook snelheid kreeg hierdoor een andere betekenis. De winst zat niet in het aantal reacties dat de bot binnen een uur kon produceren, maar in de mogelijkheid om kort na een eerste versie opnieuw te handelen. Alleen wanneer studenten die gelegenheid daadwerkelijk benutten en de volgende versie inhoudelijk verbeterden, droeg de snelle terugkoppeling bij aan het leerproces. Een snelle reactie die niet tot een bewerking

leidde, verkortte geen feedbackcyclus.

De bestuurlijke vraag wordt daarmee concreter. Niet: hoeveel reacties kan de bot per uur genereren? Wel: hoeveel studenten doorlopen daadwerkelijk een volgende verbeterstap, wat verandert er in hun product en toelichting, waar blijft docentcontrole nodig en wat doet de docent met de ruimte die daardoor ontstaat? Pas als die veranderingen zichtbaar worden gemaakt, wordt snelheid meer dan een technisch kunstje.

De korte feedbackcyclus



5.4 Besluiten over professionele ruimte

Tijdwinst bij feedback ontstaat alleen als er daadwerkelijk werk vervalt. Dat klinkt vanzelfsprekend, maar in veel pilots wordt de rekensom te vroeg gemaakt. De toepassing produceert binnen enkele seconden een reactie en die snelheid wordt vervolgens vergeleken met de tijd die een docent nodig heeft om feedback te formuleren. Daarmee blijft buiten beeld of de docent de reactie nog moet controleren, corrigeren of toelichten, of studenten er zelfstandig mee verder kunnen en of dezelfde feedback later alsnog tijdens een gesprek aan de orde komt.

Professionele ruimte vraagt daarom om een bredere beoordeling. Het gaat niet alleen om minuten die bij een taak vrijkomen, maar ook om de mate waarin de vrijgekomen tijd voor ander werk kan worden benut. Een docent die verspreid over de week telkens drie minuten wint, heeft daar minder aan dan een docent voor wie een aaneengesloten

dagdeel vrijkomt. Bovendien ontstaat er pas onderwijswaarde wanneer de vrijgekomen tijd bewust wordt ingezet voor taken waarbij menselijke deskundigheid van doorslaggevend belang is: begeleiding van studenten die vastlopen, bespreking van complexe afwegingen, onderwijsontwerp, beoordeling van professioneel handelen of afstemming met de beroepspraktijk.

Maak zichtbaar welk werk verandert

Een geloofwaardige batenclaim begint op taakniveau. De uitspraak dat ‘docenten tijd besparen’ is daarvoor te algemeen. Bij feedback moet ten minste duidelijk zijn op welke onderdelen de inzet van medewerkers verandert:

Onderdeel van de feedbackcyclus	Mogelijke verandering door AI	Nieuw of resterend mensenwerk
Eerste reactie op een concept	De student krijgt direct een reactie op een afgebakend criterium	De docent bepaalt criteria en controleert steekproefsgewijs de kwaliteit
Signaleren van veelvoorkomende fouten	Patronen worden eerder zichtbaar	De docent beoordeelt of het patroon didactisch betekenisvol is
Formuleren van een vervolgstap	De toepassing doet een voorstel	De student voert de stap uit en de docent bespreekt complexe gevallen
Individuele voortgang volgen	Gebruiksgegevens tonen aantallen pogingen en terugkerende problemen	De docent interpreteert de gegevens in samenhang met ander studentwerk
Beoordelen van het eindniveau	Geen automatische overname	De bevoegde beoordelaar verzamelt en weegt passend bewijs

Zo'n overzicht voorkomt dat een volledige docenttaak ten onrechte als besparing wordt ingeboekt. In de praktijk verschuift werk vaak eerder dan dat het verdwijnt. Het schrijven van veel vergelijkbare eerste reacties kan afnemen, terwijl er tijd nodig is voor het ontwerpen van goede criteria, het controleren van de toepassing en het begeleiden van studenten die de terugkoppeling verkeerd interpreteren.

Dat nieuwe werk hoeft de inzet niet onrendabel te maken. Het kan juist waardevol zijn wanneer routinematige terugkoppeling plaatsmaakt voor deskundige aandacht. Maar dan moet die verschuiving wel expliciet onderdeel zijn van de businesscase.

Bereken de nettowinst

Voor een eerste raming kunnen managers vier tijdsstromen onderscheiden:

1. **vervallen tijd** – werk dat niet meer door een medewerker hoeft te worden uitgevoerd;

2. **verkorte tijd** – werk dat blijft bestaan, maar aantoonbaar minder tijd kost;
3. **nieuwe tijd** – controle, inrichting, onderhoud, ondersteuning en scholing die door de toepassing ontstaan;
4. **herinvesteerde tijd** – de nettowinst die medewerkers gebruiken voor een vooraf gekozen onderwijsdoel.

De nettowinst is de vervallen en verkorte tijd, verminderd met het nieuwe werk. Daarbij telt niet alleen het gemiddelde. Ook de verdeling over medewerkers en perioden is van belang. Een toepassing kan op instellingsniveau veel uren lijken te besparen, terwijl de controlelast terechtkomt bij een kleine groep docenten of functioneel beheerders. Ook kan de opbrengst vooral tijdens piekweken ontstaan, terwijl inrichting en onderhoud het hele jaar door aandacht vragen.

Maak daarom zichtbaar wie welke tijd wint en wie er werk bij krijgt. Neem naast docenten ook onderwijsontwikkelaars, coördinatoren, informatiemanagers en ondersteunende diensten mee. Juist bij opschaling worden taken als kwaliteitscontrole, versiebeheer, afhandeling van meldingen en aanpassing van bronnen structureel. Wat in een pilot nog door een enthousiaste docent tussendoor wordt gedaan, moet later gewoon worden gepland en bekostigd.

Leg vooraf vast waaraan de ruimte wordt besteed

Vrijgekomen tijd vindt niet vanzelf een betere bestemming. In een drukke onderwijsorganisatie wordt die ruimte al snel gevuld met achterstallige taken, extra administratie of een hogere student-docentverhouding. Daarmee kan er organisatorisch wel een efficiëntiewinst zijn, maar is nog niet aangetoond dat de kwaliteit van feedback of begeleiding verbetert.

Het beoogde gebruik van de ruimte hoort daarom al vóór de pilot te worden benoemd. Bijvoorbeeld:

- studenten met terugkerende denkfouten krijgen eerder een persoonlijk gesprek;
- docenten bespreken vaker complexe beroepsproducten in plaats van basisfouten afzonderlijk te corrigeren;
- een onderwijsteam analyseert periodiek welke criteria studenten onvoldoende begrijpen;
- medewerkers ontwikkelen betere oefenopdrachten op basis van zichtbare leerproblemen;
- studenten krijgen meer gelegenheid om mondeling toe te lichten hoe zij feedback hebben verwerkt.

Vervolgens moet worden nagegaan of dit ook werkelijk gebeurt. Dat hoeft niet onmiddellijk met een zwaar onderzoek. Roostergegevens, taakregistraties, steekproeven, gebruiksgegevens en korte evaluaties kunnen al laten zien of medewerkers ander werk zijn gaan doen. Voor een besluit over structurele opschaling is wel steviger bewijs nodig, zeker wanneer de toepassing gepaard gaat met personele

taakstellingen of investeringen in licenties en beheer.

Bewaak de kwaliteit van het resterende oordeel

Bij feedback is de verleiding groot om behalve het formuleren ook het beoordelen geleidelijk aan de toepassing over te laten. Dan lijkt de tijdwinst groter, maar verschuift ook de verantwoordelijkheid. Een automatische reactie kan een student helpen een volgende stap te zetten; die kan niet zonder meer vaststellen of iemand een professionele bekwaamheid beheerst.

Menselijke controle moet daarom aansluiten bij het risico van de feedback. Bij een laagdrempelige oefening kan een steekproef voldoende zijn. Gaat het om een complex beroepsproduct, gevoelige informatie of feedback die zwaar meeweegt in de voortgang van een student, dan is intensievere controle nodig. Naarmate de gevolgen groter zijn, neemt dus ook het resterende mensenwerk toe. Die controle is geen hinderlijke kostenpost die uit de businesscase moet worden weggepoetst, maar een voorwaarde waaronder de toepassing verantwoord kan functioneren.

De bestuurlijke beslissing

Doorgaan is verantwoord wanneer vier zaken aannemelijk zijn. Ten eerste vervalt of verkort er aantoonbaar een concrete taak. Ten tweede blijft er na controle, beheer en ondersteuning een bruikbare nettowinst over. Ten derde wordt die ruimte daadwerkelijk besteed aan werk dat voor studenten of het onderwijs meer waarde heeft. Ten vierde blijft het professionele oordeel belegd bij medewerkers die daarvoor bevoegd en bekwaam zijn.

Voor de besluitvorming helpen de volgende vragen:

- Bij welke feedbacktaak komt tijd vrij en hoe is dat vastgesteld?
- Voor welke medewerkers geldt die opbrengst, en in welke periode?
- Welk controle-, correctie- en beheerwerk ontstaat?
- Levert de nettowinst samenhangende, planbare ruimte op?
- Waaraan besteden medewerkers die ruimte?
- Welk effect daarvan moet voor studenten zichtbaar worden?
- Hoe wordt voorkomen dat tijdwinst alleen tot taakverzwaring of grotere groepen leidt?
- Wie blijft verantwoordelijk wanneer de feedback onjuist, onvolledig of niet passend is?
- Houdt de opbrengst stand wanneer de toepassing door meerdere teams wordt gebruikt?

Kan een pilot deze vragen nog niet beantwoorden, dan is dat niet meteen een reden om te stoppen. Het betekent wel dat de batenclaim voorlopig moet blijven. De volgende stap is dan niet automatisch opschalen, maar gericht meten en het werk rond de toepassing opnieuw inrichten. Pas wanneer technische snelheid is omgezet in planbare

professionele ruimte én die ruimte aantoonbaar bijdraagt aan beter onderwijs, is er sprake van rendement.

6 Reflectie vraagt om frictie en vertrouwen

Bij reflectie laat rendement zich nog moeilijker vangen in snelheid. Een toepassing kan binnen enkele minuten scherpe vragen stellen en een verzorgd verslag opleveren. Maar wie alleen doorlooptijd, gebruiksgemak of tekstkwaliteit meet, kan precies het verkeerde belonen. Reflectie vraagt niet om een sneller antwoord, maar om de bereidheid het eigen handelen opnieuw te onderzoeken.

Daarvoor is frictie nodig. De student moet aannames ter discussie durven stellen, gemiste signalen onder ogen zien en alternatieven verkennen die minder comfortabel zijn. Tegelijk kan dat alleen in een omgeving waarin hij erop vertrouwt dat persoonlijke ervaringen, twijfels en fouten niet onverwacht elders opduiken of tegen hem worden gebruikt. Zonder frictie blijft reflectie oppervlakkig; zonder vertrouwen wordt zij voorzichtig en strategisch.

Juist daar wordt een ogenschijnlijk didactische toepassing een bestuurlijk vraagstuk. Reflectiegegevens komen immers dicht bij de student en vaak ook bij anderen uit diens opleiding, stage of beroepspraktijk. De vraag is daarom niet alleen of AI het gesprek kan verdiepen, maar ook welke informatie daarvoor werkelijk nodig is, wie daarover beschikt en welke grenzen vooraf vaststaan. Pas binnen die grenzen kan professioneel oordeel groeien zonder dat een keurige tekst daarvoor wordt aangezien.

Stilstaan bij gemiste signalen



6.1 Voorbij het keurige reflectieverslag

Bij reflectie dreigt een vergelijkbare verwarring. Een student voert enkele gegevens in, de toepassing stelt vragen en enkele minuten later ligt er een helder opgebouwd reflectieverslag. De formuleringen zijn zorgvuldig, de leerpunten klinken overtuigend en de vervolgstappen zijn keurig geordend. Toch weten we daarmee nog niet of de student werkelijk heeft onderzocht wat er gebeurde, waarom hij zo handelde en wat hij een volgende keer anders zou doen.

Een verzorgd verslag is een product. Reflectief vermogen blijkt uit het denkproces dat eraan voorafgaat en uit de manier waarop de student zijn handelen vervolgens bijstelt. Wie beide aan elkaar gelijkstelt, beloont vooral het vermogen om een geloofwaardige tekst te produceren – eventueel met hulp van AI.

Reflectie begint waar het ongemakkelijk wordt

Goede reflectie bevestigt niet alleen wat de student al dacht. Zij zorgt voor frictie. De student moet stilstaan bij aannames die misschien niet klopten, informatie die hij over het hoofd zag, gevolgen die hij niet had voorzien en alternatieven die hij te snel verwierp.

Neem een student die tijdens een projectoverleg weinig weerstand ervoer tegen zijn voorstel. In zijn eerste reflectie schrijft hij dat hij de groep kennelijk goed heeft meegenomen. Een reflectietoepassing kan die conclusie bevestigen en helpen om er een fraai verhaal van te maken. Waardevoller is het wanneer de toepassing doorvraagt:

- Waaruit leidde u af dat de anderen overtuigd waren?
- Wie sprak er tijdens het overleg niet?
- Welke andere verklaring is er voor het uitblijven van bezwaren?
- Welk effect had uw positie of voorbereiding op de inbreng van anderen?
- Wat zou u bij een volgend overleg doen om te controleren of er werkelijk draagvlak is?

Door zulke vragen verandert de aard van de activiteit. De student maakt niet langer alleen een verslag van een ervaring, maar onderzoekt zijn interpretatie daarvan. Dat is de mogelijke onderwijswaarde van AI bij reflectie: niet het schrijven overnemen, maar het denken vertragen, verbreden en waar nodig verstoren.

Die verstoring moet overigens wel productief blijven. Een toepassing die eindeloos kritische vragen stelt, kan studenten onzeker maken of tot oppervlakkige antwoorden verleiden. Reflectie vraagt om voldoende veiligheid om fouten en twijfel onder ogen te zien. Frictie zonder vertrouwen wordt al snel een verhoor; vertrouwen zonder frictie levert vooral bevestiging op.

Van verslaggenerator naar kritische gesprekspartner

De keuze voor een reflectietoepassing begint daarom niet bij de vraag of zij goede

teksten kan maken. De eerste vraag is welk professioneel denkproces studenten moeten leren doorlopen. Gaat het om het herkennen van eigen aannames, het wegen van belangen, het verbinden van theorie en praktijk, het verantwoorden van een besluit of het formuleren en beproeven van een alternatief?

Pas daarna kan worden bepaald welke rol AI krijgt. Die rol kan bijvoorbeeld bestaan uit:

- doorvragen op algemene of ontwijkende formuleringen;
- onderscheid laten maken tussen waarneming, interpretatie en oordeel;
- een situatie vanuit het perspectief van een andere betrokkene laten bekijken;
- een keuze laten toetsen aan beroepsnormen of eerder behandelde theorie;
- tegenvoorbeelden of alternatieve verklaringen laten onderzoeken;
- eerdere voornemens terughalen en vergelijken met later handelen;
- de student laten formuleren welk gedrag hij in een volgende situatie wil uitproberen.

Daarbij is terughoudendheid nodig met automatische conclusies. Een AI-toepassing kent de beroepssituatie doorgaans slechts via de beschrijving van de student. Zij heeft het gesprek niet bijgewoond, zag de lichaamstaal niet en weet niet welke organisatorische verhoudingen meespeelden. Wanneer de toepassing toch stellig vaststelt dat de student bijvoorbeeld onvoldoende empathisch, besluitvaardig of professioneel heeft gehandeld, krijgt een onvolledig beeld ten onrechte het gewicht van een beoordeling.

Beter is een toepassing die onzekerheid zichtbaar houdt: *U beschrijft dat uw collega stil bleef. Welke verklaringen zijn daarvoor mogelijk?* Zo'n vraag opent het onderzoek. De uitspraak *Uw collega voelde zich niet gehoord* sluit het juist voortijdig af.

De student moet eigenaar blijven van het oordeel

AI kan alternatieven aandragen, inconsistenties signaleren en een redenering bevragen. Het professionele oordeel moet bij de student en, waar beoordeling aan de orde is, bij de docent blijven. Dat betekent ook dat de student een suggestie van de toepassing moet kunnen afwijzen en die afwijzing moet kunnen onderbouwen.

Juist daarin kan reflectief vermogen zichtbaar worden. Een student die alle suggesties overneemt, toont vooral volgzzaamheid. Een student die uitlegt waarom een voorgesteld alternatief in deze beroepscontext ongeschikt is, laat zien dat hij omstandigheden, normen en gevolgen tegen elkaar kan afwegen.

Docenten krijgen daarmee andere informatie tot hun beschikking. Zij hoeven niet uitsluitend het afgeronde verslag te bekijken, maar kunnen ook aandacht besteden aan beslissende momenten in het proces: welke aanname werd onderzocht, welk alternatief kwam in beeld, welke feedback wees de student af en wat veranderde hij daarna in zijn handelen? Dat hoeft niet te betekenen dat een docent ieder gesprek met een toepassing volledig moet nalezen. Dan ontstaat er nieuwe controlelast en verdwijnt een eventuele tijdswinst onmiddellijk.

Een werkbare inrichting vraagt om selectie. Studenten kunnen bijvoorbeeld zelf twee of drie momenten aanwijzen waarop hun inzicht veranderde, met een korte toelichting en relevante gespreksfragmenten. De docent onderzoekt vervolgens niet of de AI een mooi gesprek heeft gevoerd, maar of de student zijn keuzes kan uitleggen en de opgedane inzichten in een nieuwe situatie kan toepassen.

Beoordeel verandering, niet alleen formulering

Het bewijs voor onderwijswaarde ligt dus niet in de kwaliteit van de gegenereerde tekst. Ook veelvuldig gebruik of een positieve waardering door studenten is nog geen bewijs van sterker reflectief vermogen. Relevantere aanwijzingen zijn bijvoorbeeld dat studenten:

- concreter onderscheid maken tussen gebeurtenis en interpretatie;
- meerdere verklaringen onderzoeken voordat zij conclusies trekken;
- gevolgen voor andere betrokkenen meenemen;
- theorie of beroepsnormen gebruiken om hun handelen te beoordelen;
- eerdere voornemens aantoonbaar toepassen in een volgende situatie;
- hun oordeel kunnen verdedigen wanneer AI een ander alternatief voorstelt.

Daarvoor zijn vergelijkingen in de tijd nodig. Een eenmalig reflectieverslag zegt weinig over ontwikkeling. Interessanter is of de kwaliteit van de afwegingen tijdens een project, stage of studiejaar toeneemt en of de student minder afhankelijk wordt van sturende vragen. Een toepassing rendeert pas werkelijk wanneer zij het zelfstandige professionele oordeel opbouwt, niet wanneer zij blijvend nodig is om ieder inzicht te formuleren.

Voor managers en opleidingsverantwoordelijken volgt daaruit een eenvoudige toets. Vraag bij een voorstel voor een reflectiebot niet als eerste hoe goed de teksten zijn of hoeveel verslagen ermee kunnen worden verwerkt. Vraag welk denk- en handelingsvermogen wordt ontwikkeld, welk gedrag daarvan zichtbaar moet worden, hoe docenten dat vaststellen en wanneer de ondersteuning kan worden afgebouwd.

Daar komt nog een voorwaarde bij. Om gericht door te vragen, krijgt een reflectietoepassing vaak informatie over fouten, onzekerheid, samenwerking, beoordelingen en gebeurtenissen op de werkplek. Juist doordat de toepassing dichtbij het professionele handelen komt, komen de gebruikte gegevens ook dichtbij de persoon. Wie reflectie wil verdiepen, moet daarom vooraf bepalen welke informatie daarvoor werkelijk nodig is, wie die mag inzien en hoelang zij bewaard blijft.

6.2 Gegevens die dichtbij komen

Dat begint met doelbinding: een opleiding moet precies kunnen uitleggen waarvoor zij reflectiegegevens gebruikt. “Om het leren te ondersteunen” is daarvoor te ruim. Gaat het erom dat de student alternatieven leert formuleren, terugkerende handelingspatronen herkent of zich voorbereidt op een gesprek met de docent? Elk doel

vraagt om andere informatie. Zonder die afbakening ontstaat al snel een verzameling verslagen, gespreksfragmenten en analyses die ooit nog van pas zouden kunnen komen.

Juist bij reflectie is dat riskant. Studenten schrijven niet alleen over hun kennis, maar ook over fouten, conflicten, twijfel, werkdruk en relaties met anderen. Een stage-ervaring kan bovendien informatie bevatten over cliënten, patiënten, leerlingen, collega's of opdrachtgevers. Wat didactisch gezien een waardevolle casus is, kan daardoor persoonsgegevens van meerdere betrokkenen bevatten. Soms gaat het zelfs om gegevens over gezondheid, afkomst, overtuigingen of andere persoonlijke omstandigheden die extra bescherming vragen.

De eerste ontwerp-vraag is dan ook niet welke gegevens de toepassing kan verwerken, maar welke gegevens zij minimaal nodig heeft om haar onderwijsfunctie te vervullen. Voor een bot die studenten helpt alternatieven te verkennen, is het bijvoorbeeld meestal niet nodig om de naam van een cliënt, de exacte stageorganisatie of een volledig beoordelingsdossier te kennen. Een korte, geanonimiseerde beschrijving van de situatie kan voldoende zijn. Minder gegevens verwerken beperkt niet alleen het privacyrisico, maar maakt ook het doel van de toepassing helderder.

Dat vraagt om bewuste keuzes in het ontwerp. Een reflectietoepassing kan studenten bijvoorbeeld vóór het invoeren van tekst waarschuwen dat ze geen namen of herkenbare casusgegevens moeten opnemen. Zij kan vragen laten beantwoorden in afgebakende velden in plaats van een compleet logboek te laten uploaden. Ook kan een opleiding besluiten dat de toepassing wel binnen één gesprek context onthoudt, maar geen persoonlijk geheugen over meerdere weken of onderwijsseenheden opbouwt. Technisch gezien is meer geheugen vaak mogelijk. Onderwijskundig is het lang niet altijd nodig.

Vier afspraken vóór de eerste student begint

Voor verantwoord gebruik moeten ten minste vier zaken vooraf helder zijn.

1. Het doel

Studenten en medewerkers moeten weten welke onderwijsfunctie de gegevensverwerking heeft. Wordt de informatie alleen gebruikt om vervolgvragen te stellen, ook om de voortgang te volgen of tevens voor beoordeling en onderzoek? Het later toevoegen van nieuwe doelen vraagt om een nieuwe afweging. Gegevens die voor begeleiding zijn verstrekt, mogen niet ongemerkt een andere functie krijgen. Maak daarom ook duidelijk welk deel van een reflectie privé blijft, wat met een docent wordt gedeeld en welke informatie eventueel meetelt bij een beoordeling.

2. De minimale gegevensset

Leg vast welke informatie nodig is en welke informatie juist buiten de toepassing moet blijven. Dat geldt zowel voor wat studenten zelf invoeren als voor gegevens die automatisch worden toegevoegd, zoals eerdere gesprekken, beoordelingen,

gebruikersgegevens en tijdstippen. “Het systeem kan erbij” is geen reden om gegevens te koppelen. Moeten studenten gevoelige ervaringen aan een externe dienst toevertrouwen, bied dan een gelijkwaardig alternatief, zoals reflectie met een docent, een lokaal opgeslagen formulier of een aanpak waarbij de student alleen zelfgekozen fragmenten invoert.

3. De bewaartermijn

Reflectiegegevens hoeven niet voor onbepaalde tijd beschikbaar te blijven. Soms volstaat opslag gedurende één gesprek. In andere gevallen is een beperkte periode nodig om ontwikkeling zichtbaar te maken. De opleiding moet dan kunnen uitleggen waarom die periode nodig is, wat er daarna gebeurt en of afgeleide analyses eveneens worden verwijderd. Dat geldt ook voor gegevens in reservekopieën, leveranciersomgevingen en geëxporteerde bestanden. Wees terughoudend met een persoonlijk geheugen over langere tijd: wat technisch als handige personalisering geldt, kan voor een student aanvoelen als onverwachte observatie.

4. De toegang

Niet iedereen die technisch toegang kan krijgen, hoeft die ook te hebben. Bepaal welke gegevens studenten, docenten, begeleiders, beheerders, leveranciers en eventuele onderaannemers mogen zien en waarom. Maak ook duidelijk of docenten volledige gesprekken kunnen lezen of alleen een door de student gekozen samenvatting ontvangen. Bij reflectie is dat verschil wezenlijk: begeleiding vraagt niet vanzelf om onbeperkte inzage. De leverancier moet bovendien duidelijk maken waar de gegevens worden verwerkt, welke gegevensstromen worden gebruikt, of reflecties worden benut voor de verdere ontwikkeling van modellen of diensten en welke mogelijkheden er zijn om gegevens werkelijk te verwijderen.

Deze afspraken zijn geen sluitstuk van een technische invoering. Ze bepalen mede of studenten de toepassing als een veilige leeromgeving ervaren. Wie vermoedt dat iedere twijfel later tegen hem kan worden gebruikt, schrijft al snel voorzichtiger, algemener en strategischer. De toepassing krijgt dan misschien keurig ingevulde reflecties, maar verliest precies de openheid die nodig is om aannames en fouten te onderzoeken. Een student kan immers alleen verantwoord kiezen wat hij deelt wanneer hij de gevolgen van die keuze kent.

Voor managers ligt hier een bredere bestuurlijke opdracht. Privacy, informatiebeveiliging, onderwijsontwerp en leveranciersbeheer moeten vanaf het begin samen optrekken. De functionaris gegevensbescherming of privacyadviseur kan beoordelen welke juridische waarborgen nodig zijn. Beveiligingsprofessionals kijken naar toegang, opslag en gegevensstromen. Docenten bepalen welke informatie didactisch noodzakelijk is. Studenten en medezeggenschap kunnen zichtbaar maken waar de formele uitleg niet overeenkomt met de ervaren veiligheid. Geen van deze partijen kan de afweging alleen maken.

Het mogelijke rendement moet vervolgens tegen deze lasten worden afgewogen. Een toepassing die diepere reflectie belooft, maar veel controle, handmatige opschoning en complexe toegangsprocedures vereist, kan alsnog een zwakke zakelijke onderbouwing hebben. Omgekeerd kan een bewust begrensde toepassing minder mogelijkheden bieden, maar juist beter bruikbaar zijn doordat studenten haar vertrouwen en docenten weten op basis waarvan zij mogen handelen.

De bestuurlijke toets is daarom scherp: verzamel niet meer omdat het later misschien nuttig wordt, bewaar niet langer omdat verwijderen lastig is en geef geen toegang omdat het praktisch uitkomt. Vraag steeds welke onderwijswaarde iedere gegevensverwerking dient en of diezelfde waarde met minder gegevens kan worden bereikt. Vooral bij een geheugenfunctie is die toets nodig: zodra een toepassing zonder duidelijke aanleiding een oude onzekerheid, fout of persoonlijke uitspraak terughaalt, verandert een behulpzame coach in een systeem dat kennelijk meer weet dan de student op dat moment wilde delen.

6.3 Uit mijn onderwijspraktijk: de coach die te veel wist

Mijn eerste reactie op dat risico was niet terughoudendheid, maar enthousiasme. Een coachbot die eerdere gesprekken onthoudt, kan immers voortbouwen op doelen, terugkerende vragen en gemaakte afspraken. De student hoeft niet telkens opnieuw uit te leggen waar hij mee bezig is. De bot kan vragen hoe een voorgenomen actie is verlopen en patronen zichtbaar maken die binnen één gesprek verborgen blijven. Juist bij reflectie leek dat waardevol.

In een vroege versie lieten we de coach daarom informatie uit eerdere gesprekken gebruiken. Technisch werkte dat behoorlijk goed. De antwoorden werden persoonlijker en de gesprekken kregen meer samenhang. De bot kon bijvoorbeeld terugkomen op een leerdoel dat een student eerder had geformuleerd of een verband leggen tussen twee situaties die op verschillende momenten waren besproken.

Dat leek aanvankelijk precies wat we wilden bereiken.

Het ongemak ontstond toen een student een nieuwe reflectie begon en de bot verwees naar een onzekerheid die weken eerder in een andere context was genoemd. De verwijzing was inhoudelijk niet onjuist en zelfs relevant voor het gesprek. Toch voelde die voor de student niet als ondersteuning. De student wist niet dat deze uitspraak bewaard was gebleven, had niet verwacht dat de bot haar opnieuw zou gebruiken en kon niet overzien welke andere informatie kennelijk nog beschikbaar was.

Ik zag op dat moment iets wat in een technisch ontwerp gemakkelijk buiten beeld blijft: een geheugenfunctie verandert niet alleen de kwaliteit van het antwoord, maar ook de verhouding tussen student en systeem. De bot stelde nog steeds vragen, maar de student ging zich afvragen wat er achter die vragen zat. Wat wist het systeem al? Welke conclusies had het getrokken? Wie kon die informatie nog meer bekijken? En zou een openhartige uitspraak later terugkomen bij begeleiding of beoordeling?

Daarmee verdween precies de veiligheid die voor goede reflectie nodig is.

Deze ervaring is geen bewijs dat iedere geheugenfunctie studenten wantrouwig maakt. Wel maakte ze duidelijk dat technisch correcte personalisering niet vanzelf als legitiem wordt ervaren. De onderwijswaarde van een geheugenfunctie hangt mede af van voorspelbaarheid en keuzevrijheid. Een student moet niet alleen weten *dát* informatie wordt bewaard, maar ook welke informatie dat is, waarom dat gebeurt, hoelang zij beschikbaar blijft en wanneer zij opnieuw kan worden gebruikt.

We hebben de toepassing daarom niet simpelweg voorzien van een uitgebreidere privacytekst. Zo'n tekst is nodig, maar lost het onderwijsprobleem niet op. Wie midden in een reflectiegesprek zit, baseert zijn vertrouwen niet op een document dat hij weken eerder heeft aangeklikt. De werking moet op het moment zelf begrijpelijk zijn.

Daarom hebben we het ontwerp op drie punten aangepast.

Ten eerste beperkten we het geheugen. Niet alles wat technisch kon worden onthouden, hoefde bewaard te blijven. Tijdelijke gesprekscontext en informatie die over langere tijd wordt gebruikt, werden van elkaar onderscheiden. Persoonlijke details, emotionele uitingen en losse vermoedens mochten niet ongemerkt uitgroeien tot een langdurig studentprofiel. Voor een vervolgspraak kan het nuttig zijn dat een coach een gekozen leerdoel onthoudt; daarvoor hoeft hij niet iedere formulering uit het voorafgaande gesprek te bewaren.

Ten tweede maakten we het gebruik van eerdere informatie zichtbaar. Wanneer de bot op een eerdere uiting voortbouwde, moest voor de student duidelijk zijn waar die verwijzing vandaan kwam. Geen suggestie van alwetendheid, maar een herkenbare koppeling: u hebt dit eerder als aandachtspunt vastgelegd, wilt u daarop doorgaan? Die laatste vraag bleek belangrijk. De student moest een onderwerp kunnen laten rusten zonder zich tegenover het systeem te hoeven verantwoorden.

Ten derde voegden we keuzevrijheid toe. Studenten moesten kunnen zien wat als blijvende context was vastgelegd, onderdelen kunnen verwijderen en een gesprek zonder geheugen kunnen voeren. Ook moest er een duidelijke mogelijkheid zijn om opnieuw te beginnen. Niet omdat studenten daarmee alle gegevensverwerking zelf moeten beheren, maar omdat reflectie alleen werkelijk vrijwillig kan zijn wanneer iemand grenzen kan stellen aan wat wordt meegenomen.

Daarbij kwam een lastigere ontwerp vraag naar voren: wat mag een docent zien? Voor een docent kan inzicht in terugkerende leerproblemen nuttig zijn. Dezelfde inzage kan studenten er echter toe brengen voorzichtiger, strategischer en minder eerlijk te reflecteren. Een coachomgeving wordt dan ongemerkt een observatieomgeving.

We hebben daarom het uitgangspunt aangescherpt dat toegang voortvloeit uit de onderwijsrol, niet uit technische beschikbaarheid. Dat gegevens ergens aanwezig zijn, betekent nog niet dat een docent, beheerder of leverancier ze mag bekijken. Bovendien moet vooraf helder zijn of een gesprek bedoeld is om te oefenen, te reflecteren, begeleiding te ontvangen of prestaties te beoordelen. Die contexten mogen niet

geruisloos in elkaar overlopen.

De belangrijkste les was voor mij dat transparantie geen mededeling achteraf is. Zij moet in de toepassing zelf zitten. De student moet kunnen begrijpen wat er gebeurt zonder eerst voorwaarden, gegevensmodellen of technische documentatie te bestuderen. Gegevensbeperking is evenmin een administratieve maatregel; zij bepaalt mede of studenten vrijuit durven spreken. En keuzevrijheid is geen aardige extra functie, maar een voorwaarde voor autonomie.

Vanaf dat moment werden deze punten vaste onderdelen van ons ontwikkeltraject. Bij iedere nieuwe onderwijsbot stelden we vóór de bouw vragen die we eerder pas na een incident hadden gesteld: welke gegevens heeft deze bot werkelijk nodig, wat onthoudt hij binnen en buiten een gesprek, wie kan dat zien, hoe kan een student corrigeren of verwijderen en wat gebeurt er wanneer iemand niet aan de personalisering wil deelnemen?

Dat vertraagde sommige ontwerpen. Mogelijkheden die op het eerste gezicht aantrekkelijk waren, vielen af of werden bewust begrensd. Toch werd de toepassing daardoor niet zwakker. Een coach die minder weet, kan onderwijsinhoudelijk waardevoller zijn wanneer studenten beter begrijpen wat hij doet en erop vertrouwen dat hun openheid niet onverwacht tegen hen wordt gebruikt.

Zo werd een ongemakkelijk moment meer dan een privacyles. Het liet zien dat toegankelijkheid, autonomie, gelijke behandeling, veiligheid en menselijke waardigheid niet naast het onderwijsontwerp staan. Ze bepalen mede of de beoogde onderwijswaarde überhaupt kan ontstaan.

Grenzen vóór de bouw



6.4 Besluiten vanuit publieke waarden

Daarom horen publieke waarden niet thuis in een laatste controle vlak voor de invoering. Op dat moment zijn de belangrijkste keuzes vaak al gemaakt: de gegevens zijn geselecteerd, het werkproces is ingericht, de leverancier is gekozen en studenten of medewerkers zijn afhankelijk geworden van de toepassing. Een bezwaar leidt dan al snel tot een aanvullende waarschuwing, een toestemmingsscherm of een extra controle achteraf. Het ontwerp zelf blijft buiten schot.

Wie publieke waarden serieus neemt, neemt ze eerder in het proces mee: als voorwaarden waaraan een toepassing moet voldoen voordat de verwachte baten mogen meetellen. Een onderwijsbot die studenten vaker laat oefenen, maar een deel van hen feitelijk uitsluit, levert geen volwaardige onderwijswinst op. Een coach die persoonlijke begeleiding schaalbaar maakt, maar daarvoor onnodig gevoelige gegevens bewaart, heeft geen overtuigende businesscase. En een systeem dat docenten tijd bespaart door ondoorzichtige beslissingen over studenten voor te sorteren, verplaatst het werkelijke probleem naar een plek waar het minder zichtbaar en moeilijker te herstellen is.

Van beginsel naar toetsbare voorwaarde

Begrippen als autonomie, menselijke waardigheid en gelijke behandeling klinken al snel abstract. Toch kunnen bestuurders en ontwerpteams ze vertalen in concrete vragen, waar mogelijk aangevuld met waarneembare criteria.

Publieke waarde	Beslissende vraag	Voorbeeld van een toetsbare voorwaarde
Toegankelijkheid	Kan de beoogde doelgroep de toepassing daadwerkelijk gebruiken?	De kerntaak kan ook worden uitgevoerd door studenten met een functiebeperking, beperkte taalvaardigheid of minder digitale ervaring.
Autonomie	Houdt de student betekenisvolle keuzevrijheid?	Een student kan personalisering weigeren zonder studievertraging, kwaliteitsverlies of een onredelijke extra werklast.
Gelijke behandeling	Werkt de toepassing aantoonbaar niet structureel slechter voor bepaalde groepen?	Uit de evaluatie blijkt niet alleen het gemiddelde resultaat, maar ook of gebruik, uitval en kwaliteit tussen relevante groepen verschillen.
Privacy	Worden alleen gegevens gebruikt die noodzakelijk zijn voor het onderwijsdoel?	Er is vastgelegd welke gegevens worden verwerkt, hoelang ze worden bewaard, wie toegang heeft en hoe correctie en verwijdering plaatsvinden.

Veiligheid	Zijn fouten, misbruik en ongewenste toegang voldoende begrensd?	Risicovolle uitkomsten worden gecontroleerd, incidenten kunnen worden gemeld en er is een werkbare terugvalmogelijkheid wanneer het systeem uitvalt.
Menselijke waardigheid	Blijft de student een persoon die leert en handelt, in plaats van een profiel dat wordt gestuurd?	De toepassing manipuleert niet, gebruikt geen kwetsbaarheid om gedrag af te dwingen en laat ruimte voor context, tegenspraak en menselijk oordeel.

Zo'n tabel is geen afvinklijst waarmee een instelling zich automatisch verantwoord kan verklaren. De betekenis hangt immers af van de onderwijscontext. Bij een vrijwillige oefenbot ligt de risicogrens anders dan bij een toepassing die studieadvies ondersteunt. Een fout antwoord tijdens een vrijblijvende oefening kan vervelend zijn; een fout profiel dat invloed krijgt op begeleiding, toetsing of doorstroom kan langdurige gevolgen hebben.

Naarmate een toepassing meer invloed heeft op kansen, beoordeling of afhankelijkheid, gelden er dus zwaardere eisen. Dan volstaat het niet dat een leverancier verklaart dat het systeem "verantwoord" is. De instelling moet zelf kunnen vaststellen welke fouten kunnen optreden, wie daarvan nadeel ondervindt, hoe die fouten worden ontdekt en wie bevoegd is om ze te herstellen.

Baten kunnen een grensoverschrijding niet compenseren

In een gewone businesscase kunnen baten en lasten tegen elkaar worden afgewogen. Hogere invoeringskosten kunnen bijvoorbeeld aanvaardbaar zijn wanneer de toepassing aantoonbaar veel onderwijswaarde oplevert. Bij publieke waarden geldt die uitruil slechts beperkt.

Sommige tekortkomingen zijn te herstellen. Als een oefenbot slecht bruikbaar blijkt voor studenten die afhankelijk zijn van voorleessoftware, kan het ontwerp mogelijk worden aangepast voordat bredere invoering plaatsvindt. Als instructies onduidelijk zijn, kan extra begeleiding helpen. De toepassing gaat dan niet direct van tafel, maar de batenclaim wordt opgeschort totdat het probleem aantoonbaar is opgelost.

Andere tekortkomingen vormen een harde grens. Denk aan een systeem dat alleen goed werkt wanneer studenten onder druk gevoelige informatie afstaan, dat bepaalde groepen structureel benadeelt zonder herstelmogelijkheid of dat beslissingen met grote gevolgen onttrekt aan menselijk oordeel. De vraag is dan niet hoeveel tijd of geld de toepassing bespaart. De vraag is of de instelling deze werkwijze überhaupt behoort te willen.

Dat onderscheid voorkomt dat publieke waarden worden teruggebracht tot één risicoscore naast kosten, gebruiksgemak en snelheid. Niet elke schade laat zich immers omrekenen in euro's of compenseren met een hoger gemiddeld leerrendement. Een

toepassing gaat niet door wanneer de baten alleen kunnen worden behaald door toegankelijkheid, autonomie, gelijke behandeling, privacy, veiligheid of menselijke waardigheid wezenlijk te verzwakken.

Kijk ook naar de verdeling van baten en lasten

Gemiddelde resultaten kunnen verhullen dat de opbrengsten en nadelen ongelijk zijn verdeeld. Misschien oefenen veel studenten vaker, maar haakt juist de groep af die extra ondersteuning nodig heeft. Misschien bespaart een docent tijd, terwijl studenten meer herstelwerk moeten doen door onjuiste feedback. Of misschien daalt de administratieve belasting binnen een opleiding, maar verschuift die naar functioneel beheer, privacyondersteuning en incidentafhandeling.

Vraag daarom steeds:

- Voor wie zijn de beoogde baten?
- Wie draagt de nieuwe lasten en risico's?
- Wie kan bezwaar maken of een fout laten herstellen?
- Welke groep is in de pilot nauwelijks vertegenwoordigd?
- Wat gebeurt er met studenten die de toepassing niet kunnen of willen gebruiken?
- Blijft er een gelijkwaardig alternatief beschikbaar?

Een alternatief is pas gelijkwaardig wanneer het onderwijsdoel ermee haalbaar blijft zonder verborgen nadeel. Een student die geen toestemming geeft voor uitgebreide gegevensverwerking mag bijvoorbeeld niet worden verwezen naar een trager traject met minder feedback of minder begeleiding. Formele keuzevrijheid zonder reëel alternatief is geen autonomie.

Leg de grens vóór de pilot vast

Bestuurlijke duidelijkheid ontstaat wanneer een instelling vooraf vastlegt welke uitkomsten tot aanpassing of stopzetting leiden. Dat kan heel praktisch. Een pilot wordt bijvoorbeeld niet opgeschaald wanneer een relevante studentgroep de toepassing aantoonbaar minder goed kan gebruiken, wanneer studenten niet begrijpen welke gegevens worden bewaard of wanneer docenten uitkomsten niet zinvol kunnen controleren. Ook onduidelijkheid kan een reden zijn om niet verder te gaan. Wie een wezenlijk risico niet kan onderzoeken, kan evenmin aannemelijk maken dat het beheerst is.

Daarbij moet ook de verantwoordelijkheid helder blijven. Een projectleider kan de toetsing organiseren, een privacyfunctionaris of beveiligingsspecialist kan adviseren en de medezeggenschap kan belangen en zorgen inbrengen. De uiteindelijke beslissing over onderwijsdoel, proportionaliteit en aanvaardbaarheid blijft echter bij de instelling. Die verantwoordelijkheid kan niet worden uitbesteed aan een leverancier, een keurmerk of de techniek zelf.

De kern van het besluit is daarmee eenvoudig, al is de uitvoering dat niet: publieke

waarden zijn geen zachte wensen rond een harde businesscase. Ze vormen de ondergrens ervan. Pas wanneer een toepassing binnen die grens aantoonbare waarde oplevert voor studenten, medewerkers en de instelling, ontstaat er een verantwoord voorstel om door te gaan.

7 Samenwerken met zicht op ieders bijdrage

Bij samenwerking ligt de waarde van AI niet alleen in wat een groep oplevert, maar vooral in wat er tussen studenten gebeurt. Een verzorgd advies, een samenhangend verslag of een overtuigende presentatie kan de indruk wekken dat de samenwerking geslaagd is. Toch blijft dan een wezenlijke vraag open: heeft ieder groepslid geleerd kennis in te brengen, verschillen te bespreken, keuzes te onderbouwen en verantwoordelijkheid te dragen?

AI maakt die vraag urgenter. Taken kunnen sneller worden verdeeld, teksten moeiteloos worden samengevoegd en argumenten binnen enkele seconden worden geordend. Dat kan studenten helpen, maar het kan ook precies die momenten wegnemen waarop samenwerking leerzaam wordt: wanneer perspectieven botsen, iemand bezwaar maakt, een keuze moet worden herzien of de groep samen tot een besluit komt.

Voor managers en bestuurders is een beter groepsproduct daarom onvoldoende grond voor een batenclaim. Er moet zicht blijven op zowel de gezamenlijke prestatie als de bijdrage en ontwikkeling van iedere student. Zonder dat zicht dreigt AI vooral de productie te verbeteren, terwijl het groepsleren verschaalt. De bestuurlijke opgave is dan ook toepassingen zo te kiezen en in te richten dat ze het gesprek ondersteunen zonder het over te nemen.

Het gesprek achter het groepswerk



7.1 Waar groepswork zijn leerwaarde verliest

Vier signalen van geringe leerwaarde

De kwaliteit van groepswork wordt beter zichtbaar wanneer niet alleen naar het eindproduct wordt gekeken, maar ook naar vier onderdelen van het proces.

1. Taakverdeling

Een efficiënte taakverdeling houdt rekening met beschikbare tijd en verschillende deskundigheden. Voor het leren is bovendien van belang wie welke taak krijgt en waarom. Wanneer studenten steeds kiezen voor wat zij al goed kunnen, levert de groep misschien snel resultaat, maar oefenen zij weinig nieuws. Ook kan een vroege verdeling ertoe leiden dat niemand het volledige vraagstuk meer overziet.

De relevante vraag is dus niet alleen of de taken eerlijk zijn verdeeld. Minstens zo belangrijk is of de verdeling bijdraagt aan de leerdoelen en of studenten elkaars werk nog kunnen begrijpen, beoordelen en zo nodig overnemen.

2. Deelname

Aanwezigheid is geen deelname. Een student kan ieder overleg bijwonen en toch nauwelijks inhoudelijk bijdragen. Omgekeerd kan iemand buiten de bijeenkomsten belangrijk analysewerk doen dat tijdens de beoordeling onzichtbaar blijft. Wie alleen uren, berichten of geproduceerde tekst telt, krijgt daarom een onvolledig beeld.

Betekenisvolle deelname blijkt uit handelingen zoals vragen stellen, een redenering toetsen, bronnen beoordelen, feedback verwerken, verbanden leggen en een collega helpen diens bijdrage te verbeteren. AI kan zulke handelingen ondersteunen, maar ook vervangen. Een automatisch gemaakte samenvatting bewijst bijvoorbeeld niet dat studenten naar elkaar hebben geluisterd of gezamenlijk betekenis hebben gegeven aan wat er is besproken.

3. Besluitvorming

In professioneel groepswork moeten studenten keuzes maken onder onzekerheid. Zij vergelijken mogelijkheden, wegen argumenten en risico's af en leggen uit waarom zij voor een bepaalde aanpak kiezen. Wanneer één groepslid de richting bepaalt of wanneer de groep zonder bespreking het eerste aannemelijke AI-voorstel overneemt, ontbreekt een wezenlijk deel van de samenwerking.

Niet ieder besluit hoeft uitvoerig te worden vastgelegd. Wel moet de groep bij belangrijke keuzes kunnen reconstrueren welke alternatieven zijn overwogen, welke argumenten de doorslag gaven en wie bij de afweging betrokken was. Daarmee wordt zichtbaar of AI als denksteun is gebruikt of als onzichtbare beslisser.

4. Gezamenlijke verantwoordelijkheid

Groepsleden dragen niet alleen verantwoordelijkheid voor hun eigen stukje, maar ook

voor het geheel. Dat betekent dat zij fouten in andere onderdelen helpen herkennen, elkaar aanspreken op afspraken en het gezamenlijke resultaat kunnen verdedigen. Die verantwoordelijkheid verdwijnt wanneer studenten zich bij problemen terugtrekken achter de taakverdeling: “Dat onderdeel was niet van mij.”

AI maakt zo’n terugtrekkende beweging nog gemakkelijker. Een student kan niet alleen naar een medestudent wijzen, maar ook naar het hulpmiddel: de samenvatting, formulering of analyse kwam immers daarvandaan. Voor professioneel handelen is dat ontoereikend. De groep blijft verantwoordelijk voor wat zij gebruikt, aanpast en inlevert, ongeacht wie of wat de eerste versie heeft gemaakt.

Een groep kan presteren zonder samen te leren

Veel groepsopdrachten nodigen uit tot een eenvoudige verdeling: ieder groepslid maakt een onderdeel, waarna één student de stukken samenvoegt. Dat is organisatorisch efficiënt, maar didactisch vaak arm. Studenten werken dan naast elkaar in plaats van met elkaar. Ze kennen hun eigen onderdeel, maar kunnen niet altijd uitleggen hoe de andere onderdelen tot stand kwamen, welke afwegingen zijn gemaakt en waarom het eindproduct als geheel verdedigbaar is.

AI kan deze werkwijze aantrekkelijker maken. Studenten kunnen sneller deeltaken uitwerken, teksten in dezelfde stijl laten herschrijven en ontbrekende verbindingen laten aanvullen. Het product wint daardoor aan samenhang, terwijl de feitelijke samenwerking verder afneemt. De groep lijkt beter te functioneren doordat de zichtbare oneffenheden zijn weggewerkt. Precies daarin schuilt het risico: AI kan een zwak proces verhullen met een overtuigend resultaat.

Voor een opleiding is dat geen detail. Wanneer samenwerking een leerdoel is, moet de opdracht studenten dwingen om informatie en oordelen uit te wisselen. Er moet sprake zijn van inhoudelijke onderlinge afhankelijkheid: groepsleden hebben elkaar nodig om tot een verantwoord resultaat te komen. Als ieder onderdeel zonder overleg kan worden gemaakt en een hulpmiddel vervolgens de montage verzorgt, wordt vooral taakverdeling geoefend. Dat kan nuttig zijn, maar het is iets anders dan professioneel samenwerken.

Het eerste vraagstuk is daarom niet welke AI-functie een projectgroep kan gebruiken. Eerst moet duidelijk zijn wat studenten door de samenwerking behoren te leren en welk gedrag daarvoor nodig is. Gaat het om taken verdelen, perspectieven combineren, vakinhoudelijke keuzes bespreken, omgaan met verschillen van inzicht of gezamenlijk verantwoordelijkheid dragen? Pas als die leerwaarde scherp is omschreven, kan worden beoordeeld of AI het groepsproces ondersteunt of er juist een deel van overneemt.

Diagnose vóór toepassing

Een manager of opleiding hoeft groepsprocessen niet volledig meetbaar te maken voordat er met AI mag worden geëxperimenteerd. Wel is een geloofwaardige uitgangsdiagnose nodig. Zonder zo’n diagnose valt immers niet vast te stellen of een

toepassing de samenwerking verbetert.

Die diagnose kan compact blijven. Bekijk een aantal bestaande groepen en onderzoek waar het leerproces werkelijk hapert. Worden taken te vroeg opgesplitst? Blijven sommige studenten structureel buiten de inhoudelijke discussie? Worden keuzes onvoldoende besproken? Kunnen groepsleden het eindresultaat niet gezamenlijk verantwoorden? Docentwaarnemingen, korte studentgesprekken, tussentijdse producten en een analyse van belangrijke beslismomenten geven doorgaans meer inzicht dan een algemene tevredenheidsenquête.

Daarbij moet de opleiding onderscheid maken tussen drie soorten opbrengsten. Een beter eindproduct is een productopbrengst. Minder tijd kwijt zijn aan plannen of vastleggen is een procesopbrengst. Beter leren overleggen, beoordelen en verantwoordelijkheid nemen is een onderwijsopbrengst. Ze kunnen samengaan, maar dat gebeurt niet vanzelf. Een toepassing die alleen het productietempo verhoogt, verdient dan ook geen batenclaim op het gebied van samenwerken.

Pas na deze analyse komt AI in beeld. De kansrijkste rol ligt niet in het overnemen van het gezamenlijke denkwerk, maar in het zichtbaar en hanteerbaar maken van het proces eromheen. AI kan bijvoorbeeld helpen om rollen te verduidelijken, openstaande vragen te signaleren en afspraken of voortgang samen te vatten. De wezenlijke handelingen moeten echter bij de studenten blijven: zij brengen argumenten in, reageren op elkaar, onderhandelen over verschillen, nemen besluiten en verantwoorden gezamenlijk de uitkomst.

7.2 AI als procesondersteuner

De vraag is dus niet óf AI aan het groepswork deelneemt, maar welk deel van het proces ermee wordt ondersteund. Een bruikbare toepassing helpt studenten overzicht te houden, verschillen zichtbaar te maken en het gesprek voor te bereiden. Zij vervangt dat gesprek niet.

Een systeem dat in enkele seconden een taakverdeling, analyse en eindtekst produceert, kan het groepsproduct versnellen terwijl het nauwelijks bijdraagt aan samenwerken. Onderwijskundig is een beperktere rol vaak waardevoller, omdat studenten dan zelf moeten blijven handelen.

Ondersteunen rond het beslismoment

Een compact ontwerpmodel maakt duidelijk waar AI wel en niet wordt ingezet:

- **Voor het beslismoment:** AI helpt werkzaamheden, rollen en openstaande vragen te ordenen. Studenten bespreken zelf de taakverdeling en verbinden zich aan de gemaakte afspraken.
- **Tijdens het beslismoment:** AI rubricert argumenten, maakt verschillen zichtbaar en signaleert ontbrekende criteria of informatie. Studenten beoordelen de

argumenten, wegen belangen af en nemen het besluit.

- **Na het beslismoment:** AI maakt een eerste samenvatting van de afweging, afspraken en vervolgacties. Studenten controleren die samenvatting en verantwoorden gezamenlijk de uitkomst.

De toepassing ondersteunt daarmee steeds het proces rond een besluit, niet het besluit zelf. Een automatisch gemaakte planning is immers nog geen gezamenlijke afspraak. Ook kan een samenvatting de betekenis van een afspraak verschuiven. “Student A onderzoekt de kosten” is iets anders dan “Student A besluit welke oplossing financieel het beste is”. Studenten moeten zulke verschillen herkennen en corrigeren.

Maak bijdragen zichtbaar zonder alles te meten

Procesondersteuning kan individuele bijdragen beter zichtbaar maken, maar vraagt om terughoudendheid. Een grote hoeveelheid tekst, spreektijd of bewerkingen zegt weinig over de kwaliteit van iemands bijdrage. De student die één scherpe vraag stelt, kan meer aan het gezamenlijke denken toevoegen dan degene die voortdurend aan het woord is. Ook mag procesinformatie niet ongemerkt veranderen in permanente gedragsbewaking.

In veel gevallen is een compact groepslogboek voldoende. Daarin legt de groep per belangrijk beslismoment vast:

1. welke keuze voorlag;
2. welke alternatieven zijn besproken;
3. welke criteria zijn gebruikt;
4. welk besluit is genomen;
5. wie welke bijdrage aan de afweging heeft geleverd;
6. welke onzekerheid of minderheidsopvatting is blijven bestaan.

AI kan zo’n logboek structureren of een eerste samenvatting maken. Studenten controleren en verbeteren die samenvatting en bevestigen gezamenlijk dat zij de afweging herkennen. Zo ondersteunt de technologie de transparantie van het proces, zonder te pretenderen dat zij samenwerking objectief kan beoordelen.

De grens ligt bij het gezamenlijke denkwerk

Bij groepsleren worden luisteren, reageren, belangen afwegen, meningsverschillen hanteren, keuzes maken en verantwoordelijkheid dragen niet aan AI overgedragen. Juist daar ontstaan professionele bekwaamheden die studenten later in hun beroep nodig hebben.

Een discussie vervalt bijvoorbeeld wanneer studenten hun afzonderlijke standpunten invoeren en het systeem vragen om “de beste gezamenlijke conclusie”. Het resultaat kan evenwichtig klinken, maar de studenten hebben dan niet onderzocht waarom zij van mening verschillen. Een betere werkwijze is dat AI de verschillen ordent zonder ze op te lossen. Het systeem kan benoemen welke argumenten tegenover elkaar staan,

welke aannames nog niet zijn getoetst en welke informatie ontbreekt. Daarna bespreekt de groep de kwestie zelf.

Hetzelfde geldt voor het eindproduct. AI kan inconsistenties signaleren, onderdelen samenvoegen of de leesbaarheid verbeteren. De groep moet echter kunnen uitleggen waarom het product zo is opgebouwd, welke keuzes eraan ten grondslag liggen en welke bijdrage ieder lid heeft geleverd. Kan slechts één student dat, dan is er wel een groepsproduct, maar nog geen overtuigend bewijs van groepsleren.

Rendement vraagt om taakherontwerp

De mogelijke baten liggen op drie niveaus. Studenten krijgen meer houvast bij complexe samenwerking, docenten hoeven minder tijd te besteden aan het reconstrueren van belangrijke afwegingen en de opleiding krijgt zicht op terugkerende knelpunten. Daar staan nieuwe werkzaamheden tegenover: studenten moeten AI-samenvattingen controleren, docenten moeten procesinformatie interpreteren en de opleiding moet afspraken maken over gegevensgebruik, bewaartermijnen en toegang.

De businesscase draait daarom niet om de hoeveelheid notuleerwerk die AI overneemt. Belangrijker is of het groepsproces leerzamer wordt. Dat kan blijken uit eerder zichtbare meningsverschillen, beter onderbouwde besluiten en meer studenten die het gezamenlijke resultaat zelfstandig kunnen toelichten. Docenttijd is alleen een opbrengst wanneer die daadwerkelijk vrijkomt en vervolgens wordt gebruikt voor inhoudelijke begeleiding.

Voor een manager of opleidingsverantwoordelijke volgen daar drie beslisvragen uit. Welke samenwerkingshandeling moet aantoonbaar verbeteren? Welke procesondersteuning is daarvoor minimaal nodig? En welke handelingen moeten studenten juist zonder tussenkomst van AI blijven uitvoeren?

Wie die grens niet vooraf trekt, loopt het risico dat een groep haar gezamenlijke opdracht efficiënt uitbesteedt. Eén goede opdracht aan een AI-systeem kan dan een verzorgd product opleveren, terwijl kritiek en gezamenlijke besluitvorming vrijwel geheel verdwijnen.

7.3 Uit mijn onderwijspraktijk: één prompt, één groepsproduct

In een van mijn eerste groepsopdrachten leek het aanvankelijk alsof de inzet van AI uitstekend werkte. Studenten moesten gezamenlijk een analyse maken en daarover een advies formuleren. Een groep voerde de opdracht vrijwel letterlijk in bij een taalmodel, stelde nog een aanvullende vraag en verwerkte het antwoord in het eindproduct. Binnen korte tijd lag er een samenhangende tekst met een heldere structuur, bruikbare argumenten en een keurige conclusie.

Het resultaat zag er beter uit dan ik gezien de bestede tijd had verwacht. Van het

leerproces zag ik nauwelijks iets.

Toen ik de studenten vroeg hoe zij tot hun afweging waren gekomen, bleek dat er weinig te reconstrueren viel. Ze hadden niet eerst afzonderlijk naar het vraagstuk gekeken. Er waren geen verschillende analyses naast elkaar gelegd en er was niet besproken welke criteria voor het advies het zwaarst moesten wegen. Eén student had de prompt ingevoerd, een tweede had het antwoord vluchtig gecontroleerd en een derde had enkele formuleringen aangepast. De overige groepsleden hadden vooral ingestemd met het resultaat.

Daarmee was de opdracht wel af, maar de samenwerking grotendeels verdwenen. De studenten hadden een gezamenlijk product ingeleverd zonder gezamenlijk kennis op te bouwen, afwegingen te maken en keuzes te maken. De snelheid waarmee dat gebeurde, maakte de tekortkoming aanvankelijk minder zichtbaar: het verzorgde eindresultaat suggereerde immers dat de groep goed had gefunctioneerd.

De volgende keer veranderde ik daarom niet in de eerste plaats de toegestane technologie, maar de opbouw van de opdracht. Iedere student moest eerst afzonderlijk een korte analyse maken. Daarbij mocht AI worden gebruikt om vragen te verkennen of mogelijke argumenten te vinden, maar de student moest zelf vastleggen welke redenering hij of zij overtuigend vond, welke aannames daarin zaten en waarover nog twijfel bestond.

Pas daarna deelde de groep de afzonderlijke bijdragen. De studenten kregen de opdracht om niet meteen één tekst samen te stellen, maar eerst de verschillen zichtbaar te maken. Welke argumenten kwamen in meerdere analyses terug? Waar spraken de groepsleden elkaar tegen? Welke conclusie steunde op een onbewezen aanname? En welk perspectief dreigde buiten beeld te blijven?

AI kreeg vervolgens een beperkte rol. Studenten konden het systeem bijvoorbeeld vragen om tegenargumenten te formuleren, ontbrekende belanghebbenden te benoemen of twee oplossingsrichtingen systematisch te vergelijken. Het systeem mocht het gesprek scherper maken, maar niet namens de groep bepalen wat het beste advies was. Die keuze moesten de studenten zelf maken en verantwoorden.

Door die volgorde veranderde de groepsopdracht. De afzonderlijke voorbereiding gaf ieder groepslid iets om in te brengen. De vergelijking van bijdragen maakte verschillen bespreekbaar. Door om kritiek te vragen, voorkwam de groep dat de eerste aannemelijke oplossing meteen de gezamenlijke oplossing werd. Uiteindelijk moest de groep expliciet aangeven welke argumenten zij had overgenomen, welke zij had verworpen en waarom.

Ik zag daardoor meer inhoudelijke gesprekken ontstaan. Sommige studenten ontdekten dat hun eerste analyse vooral een herhaling was van wat het AI-systeem had voorgesteld. Anderen konden juist uitleggen waarom zij een automatisch gegenereerd argument niet bruikbaar vonden in de beroepscontext van de opdracht. De kwaliteit van het groepsproduct bleef belangrijk, maar werd niet langer als enige aanwijzing voor

leren gebruikt.

Niet alles werkte meteen. Wanneer de individuele voorbereiding te uitgebreid was, verdeelden studenten het werk alsnog: ieder werkte één onderdeel uit, waarna de stukken zonder veel discussie aan elkaar werden geplakt. Wanneer de verplichte verantwoording te gedetailleerd werd, ontstond er bovendien een administratief verslag van het proces. Studenten noteerden dan keurig wie wat had gedaan, zonder dat daaruit bleek of zij werkelijk naar elkaar hadden geluisterd of hun standpunt hadden herzien.

De bruikbare vorm lag ertussenin. De individuele bijdrage moest klein genoeg zijn om vergelijking mogelijk te maken, maar inhoudelijk genoeg om een eigen standpunt te vergen. De procesverantwoording moest zich niet richten op ieder stapje, maar op enkele betekenisvolle momenten: een belangrijk verschil van inzicht, ontvangen kritiek, een bijgestelde keuze en de uiteindelijke onderbouwing van het groepsbesluit.

Ook de mondelinge toelichting bleek waardevol. Ik vroeg de groep niet alleen om het resultaat te presenteren, maar stelde ook vragen aan afzonderlijke studenten. Waarom heeft de groep deze optie gekozen? Welk bezwaar woog het zwaarst? Waarover waren jullie het aanvankelijk oneens? Welke bijdrage van een ander heeft uw eigen oordeel veranderd? Daardoor werd snel zichtbaar of het groepsproduct werkelijk gezamenlijk eigendom was.

Deze aanpak leverde geen hard bewijs dat iedere student meer leerde. Mijn waarneming was wel dat verschillen in deelname eerder zichtbaar werden en dat studenten vaker inhoudelijk op elkaar reageerden. Ook kon ik gericht begeleiden, omdat ik niet pas bij het eindproduct zag dat een redenering tekortschoot. De tussenstappen boden aanknopingspunten voor vragen en feedback.

Daar stond nieuw werk tegenover. De opdracht moest zorgvuldiger worden ontworpen, studenten hadden uitleg nodig over de rol van AI en ik moest bepalen welke procesinformatie werkelijk relevant was. Wie alle prompts, wijzigingen en bijdragen wil controleren, creëert al snel een toezichtstaak die niet uitvoerbaar is. Bovendien kan permanente registratie het vertrouwen binnen een groep aantasten. Studenten gaan dan produceren voor het logboek in plaats van samenwerken aan het vraagstuk.

De belangrijkste les was voor mij dan ook niet dat iedere student altijd een volledig individueel tussenproduct moet maken. Het gaat erom dat de onderwijsopzet deelname, onderlinge afhankelijkheid en gezamenlijke besluitvorming noodzakelijk maakt. AI kan daarbij helpen door alternatieven, kritiek en structuur aan te bieden. Zodra het systeem echter de analyse, de afweging én de formulering in één keer overneemt, blijft er voor de groep weinig meer over dan instemming.

Een goede groepsopdracht levert daarom twee soorten waarde op. Er ontstaat een resultaat dat geen van de studenten alleen op dezelfde manier had kunnen maken, én ieder groepslid ontwikkelt het vermogen om bij te dragen, kritiek te verwerken en het gezamenlijke besluit te verdedigen. Pas wanneer beide zichtbaar zijn, is een verzorgd

groepsproduct meer dan het antwoord op één goede prompt.

Samen beslissen met AI



7.4 Besluiten over groepswaarde

Een bestuurlijke afweging

Besluit	Criteria
Stoppen	Vooraf het groepsproduct verbetert, terwijl deelname, onderlinge afhankelijkheid en individueel leren uit beeld verdwijnen.
Aanpassen	De onderwijswaarde is aannemelijk, maar rollen, controlemomenten, procesinformatie of verantwoordelijkheden zijn nog niet goed ingericht.
Voortzetten als pilot	Studenten dragen aantoonbaar actiever bij, groepen maken betere afwegingen en docenten kunnen hun aandacht gericht inzetten. De extra begeleidings- en beheerlast blijft uitvoerbaar.
Opschalen	De individuele én gezamenlijke leerwaarde blijft zichtbaar bij meerdere groepen en verschillende docenten. Er is zicht op afwijkende uitkomsten, menselijke verantwoordelijkheid is helder belegd en de structurele beheerlast is realistisch geraamd.

Een bestuurder hoeft niet te verlangen dat ieder effect vooraf exact in geld of uren wordt uitgedrukt. Wel moet duidelijk zijn welke verandering wordt verwacht, hoe die wordt vastgesteld en welke extra lasten daarvoor worden aanvaard. Bij een pilot kan

voorlopig bewijs volstaan. Voor opschaling is meer nodig: resultaten over meerdere groepen, ervaringen van verschillende docenten, zicht op afwijkende uitkomsten en een realistische raming van structureel beheer.

De beslissende toets blijft eenvoudig: maakt deze toepassing studenten beter in samenwerken, of maakt zij het alleen gemakkelijker om gezamenlijk iets in te leveren?

8 Toetsen als produceren goedkoop wordt

Die vraag wordt nog dringender zodra het ingeleverde werk meetelt voor een formele beoordeling. Generatieve AI kan in korte tijd een verzorgd rapport, een overtuigend advies of werkende code produceren. Daarmee wordt het product niet waardeloos, maar wel minder vanzelfsprekend als bewijs dat de student de vereiste professionele bekwaamheid bezit.

Voor opleidingen raakt dit aan de kern van hun opdracht. Een diploma verklaart immers niet dat iemand aantrekkelijke producten kan inleveren, maar dat diegene zelfstandig, verantwoord en op het vereiste niveau kan handelen. Als niet meer duidelijk is wat een student zelf begrijpt, afweegt en uitvoert, komt niet alleen de beoordeling onder druk te staan. Ook studenten, het beroepenveld en de samenleving moeten erop kunnen vertrouwen dat het diploma zijn betekenis behoudt.

De bestuurlijke reflex kan zijn om AI-gebruik strenger te controleren of uit te sluiten. Maar voordat er nieuwe detectiemiddelen, registraties of toezichtslagen worden ingevoerd, moet een fundamentele vraag worden gesteld: welk bewijs heeft een opleiding eigenlijk nodig om een betrouwbaar oordeel over professioneel handelen te geven? Die vraag dwingt tot een heroverweging van toetsing – niet vanuit wantrouwen tegenover studenten, maar vanuit de waarde die een hbo-diploma moet blijven vertegenwoordigen.

Een verzorgd rapport als toetsbewijs



8.1 Wat de student zelf moet kunnen aantonen

Zodra produceren goedkoop en overtuigend wordt, verschuift de kernvraag. Niet langer staat centraal of de student een goed product kan inleveren, maar wat dit product bewijst over wat de student zelf begrijpt, kan en verantwoordt. Een verzorgd adviesrapport, werkend programma of overtuigend ontwerp kan nog steeds waardevol zijn, alleen vormt het op zichzelf steeds minder overtuigend bewijs van professionele bekwaamheid.

Dat vraagt niet om een algemeen verbod op AI. Het vraagt om scherpere keuzes per toets: **welke bekwaamheid moet de student aantonen, welk AI-gebruik past daarbij en welk bewijs van eigen handelen blijft noodzakelijk?**

Begin bij de bekwaamheid

Veel toetsproblemen ontstaan doordat het eindproduct in de loop der jaren gelijk is gaan staan aan de bekwaamheid. De student levert een beroepsproduct in en krijgt daarop een beoordeling. Zolang het maken van dat product veel eigen denk- en schrijfwerk vergde, kon het product redelijk veel zeggen over het onderliggende leerproces. Met generatieve AI wordt die aanname minder houdbaar.

De opleiding moet daarom terug naar de bedoeling van de toets. Wat moet de student na afloop zelfstandig kunnen doen? Gaat het om:

- een vraagstuk analyseren;
- relevante kennis selecteren en toepassen;
- een professioneel gesprek voeren;
- een ontwerp of interventie maken;
- risico's en belangen afwegen;
- een besluit nemen en onderbouwen;
- handelen in een onvoorspelbare situatie;
- de kwaliteit van eigen werk en dat van anderen beoordelen?

Dat onderscheid is wezenlijk. Wie wil toetsen of een student helder kan formuleren, stelt andere grenzen aan AI dan wie wil toetsen of die student de inhoud van een beleidsadvies kan beoordelen. Bij de eerste toets kan automatische herschrijving de te beoordelen prestatie grotendeels overnemen. Bij de tweede kan dezelfde functie juist vergelijkbaar zijn met professioneel gereedschap, mits de student de voorstellen kritisch onderzoekt en verantwoordelijkheid neemt voor het resultaat.

Een bruikbare toetsdefinitie benoemt daarom niet alleen het product, maar ook het vereiste handelen. Niet: *de student levert een projectplan in*. Wel: *de student analyseert een praktijkvraagstuk, maakt onderbouwde keuzes, vertaalt die naar een uitvoerbaar projectplan en verdedigt de gekozen aanpak tegenover betrokkenen*. Het projectplan blijft bestaan, maar is niet langer het volledige bewijs.

Leg het passende AI-gebruik vooraf vast

Nadat de professionele bekwaamheid is bepaald, kan de opleiding besluiten welke rol AI binnen de toets krijgt. Daarbij zijn grofweg vier mogelijkheden.

Rol van AI	Betekenis voor de toets
Niet toegestaan	De student voert de te beoordelen handeling zonder AI uit, omdat juist die zelfstandige beheersing moet worden vastgesteld.
Beperkt toegestaan	AI mag bijvoorbeeld helpen bij taalcorrectie, ordening of het zoeken naar alternatieven, maar niet bij de kernhandeling.
Toegestaan als professioneel hulpmiddel	De student mag AI inzetten zoals dat ook in de beroepspraktijk aannemelijk is en moet het resultaat controleren en verantwoorden.
AI-gebruik is onderdeel van de bekwaamheid	De toets beoordeelt mede of de student AI doelgericht, kritisch, veilig en verantwoord kan inzetten.

Niet iedere toets binnen een opleiding hoeft dezelfde keuze te maken. Integendeel: een toetsprogramma wordt sterker wanneer het verschillende situaties omvat. Studenten moeten soms zonder ondersteuning laten zien dat zij de basis beheersen. In andere situaties moeten zij juist aantonen dat zij met hulpmiddelen tot een professioneel resultaat kunnen komen. Wie alleen zonder AI toetst, kan de aansluiting met de beroepspraktijk missen. Wie AI altijd toestaat, loopt het risico dat wezenlijke basiskennis en zelfstandige handelingsvaardigheid niet meer zichtbaar worden.

De gekozen grens moet bovendien begrijpelijk zijn. Formuleringen als *AI mag worden gebruikt zolang het ondersteunend is* bieden te weinig houvast. Wat voor de student ondersteuning is, kan voor de beoordelaar overname van de kernhandeling zijn. Benoem daarom concreet welke handelingen wel en niet mogen worden uitbesteed, welke gegevens niet mogen worden ingevoerd, wat de student moet vastleggen en waarop hij of zij aanspreekbaar blijft.

Maak eigen begrip en oordeel zichtbaar

Zelfstandig aantonen betekent niet dat een student ieder woord zonder hulpmiddelen moet hebben geschreven. Het betekent dat de opleiding met voldoende betrouwbaarheid kan vaststellen dat de student beschikt over de kennis, vaardigheden en het oordeel die bij de leeruitkomst horen.

Daarvoor zijn ten minste drie soorten bewijs nodig.

Eigen begrip blijkt wanneer de student begrippen kan uitleggen, verbanden kan leggen, fouten kan herkennen en kennis in een andere situatie kan toepassen. Een ingestudeerde samenvatting of overtuigend gegenereerde tekst is daarvoor niet genoeg. De student moet kunnen laten zien waarom een redenering klopt, waar de grenzen

ervan liggen en wat er verandert wanneer de omstandigheden wijzigen.

Eigen handelen wordt zichtbaar wanneer de student de relevante beroepshandeling uitvoert. Dat kan bijvoorbeeld een gesprek, analyse, demonstratie, presentatie, ontwerpbeslissing of interventie zijn. Bij complexe opdrachten hoeft niet iedere stap zonder ondersteuning plaats te vinden, maar de opleiding moet wel kunnen vaststellen welke wezenlijke handelingen de student beheerst.

Eigen professioneel oordeel blijkt uit de keuzes die de student maakt en de gevolgen waarvoor die verantwoordelijkheid neemt. Welke bron is betrouwbaar genoeg? Welke uitkomst van een AI-systeem wordt verworpen? Welk belang weegt zwaarder? Wanneer is aanvullende menselijke controle nodig? Juist hier ligt een belangrijk verschil tussen een aannemelijk product en aantoonbare bekwaamheid: de student moet niet alleen een antwoord kunnen presenteren, maar ook kunnen beoordelen of dat antwoord in deze beroepssituatie verantwoord is.

Stem het bewijs af op het beroepsrisico

Niet iedere bekwaamheid vereist dezelfde bewijslast. Bij een oefening met beperkte gevolgen kan een digitaal product, aangevuld met enkele controlevragen, voldoende zijn. Bij handelingen die gevolgen hebben voor gezondheid, veiligheid, rechtspositie, financiële belangen of toegang tot voorzieningen, is steviger bewijs nodig. Dan liggen directe observatie, mondelinge verantwoording of handelen onder realistische omstandigheden meer voor de hand.

De bewijslast moet dus toenemen naarmate de gevolgen van onbekwaam handelen groter zijn. Dat voorkomt twee uitersten. Enerzijds hoeft niet iedere student bij iedere opdracht uitgebreid te worden ondervraagd. Dat zou de toetslast onnodig vergroten. Anderzijds mag doelmatigheid geen reden zijn om bij risicovolle bekwaamheden genoeg te nemen met een product waarvan het auteurschap en de totstandkoming nauwelijks iets zeggen over zelfstandig handelen.

Voor bestuurders is daarbij van belang dat scherpere toetsing niet vanzelf duurder hoeft te zijn, maar wel om herontwerp vraagt. Minder tijd besteden aan algemene vermoedens van ongeoorloofd AI-gebruik kan ruimte geven voor gerichte waarnemingen, korte mondelinge controles en steekproeven. Die verschuiving lukt alleen wanneer examencommissies, toetsdeskundigen, opleidingen en informatiemanagement gezamenlijk bepalen welk bewijs noodzakelijk en uitvoerbaar is.

Van AI-regel naar toetsbesluit

Voor iedere toets moeten uiteindelijk vijf vragen beantwoord zijn:

1. Welke professionele bekwaamheid staat centraal?
2. Welke onderdelen moet de student zelfstandig beheersen?
3. Welk AI-gebruik is verboden, beperkt, toegestaan of juist vereist?

4. Welk bewijs maakt begrip, handelen en professioneel oordeel zichtbaar?
5. Hoe zwaar moet dat bewijs zijn gezien de gevolgen van mogelijk onbekwaam handelen?

Pas wanneer deze vragen helder zijn beantwoord, kan een opleiding zinvol spreken over toegestane hulpmiddelen, fraude of auteurschap. Algemene AI-regels bieden daarvoor te weinig precisie. De toets zelf moet duidelijk maken wat van de student wordt verlangd en waarom.

Het eindproduct verdwijnt daarmee niet uit beeld. Het krijgt alleen een bescheidener en zuiverder plaats: als één bron van informatie naast ander bewijs. De betrouwbaarheid ontstaat vervolgens uit de combinatie van het product met wat tijdens de totstandkoming, uitvoering en verantwoording zichtbaar wordt.

8.2 Procesbewijs naast het eindproduct

Geen enkele bewijsbron hoeft op zichzelf doorslaggevend te zijn. Een overtuigend eindproduct kan immers door de student zijn gemaakt, met passende ondersteuning van AI, maar het kan ook grotendeels buiten het zicht van de opleiding tot stand zijn gekomen. Omgekeerd bewijst een aarzelende mondelinge toelichting niet dat de student de stof onvoldoende beheerst. Betrouwbare beoordeling ontstaat daarom door verschillende vormen van bewijs te combineren.

De kern is **triangulatie**: dezelfde professionele bekwaamheid wordt vanuit meerdere invalshoeken zichtbaar gemaakt. Het eindproduct laat zien wat de student heeft opgeleverd. Tussenproducten tonen hoe keuzes en redeneringen zich hebben ontwikkeld. Een observatie maakt zichtbaar hoe de student handelt. Een korte mondelinge verantwoording geeft inzicht in begrip, afwegingen en eigenaarschap. Als deze bronnen een samenhangend beeld opleveren, kan de beoordelaar met meer vertrouwen een uitspraak doen.

Begin met een authentieke taak

Procesbewijs krijgt pas betekenis wanneer de taak zelf representatief is voor de beroepspraktijk. Een authentieke taak vraagt niet alleen om kennis te reproduceren, maar om die kennis toe te passen in een situatie waarin keuzes, onzekerheden en gevolgen een rol spelen.

Denk aan een verpleegkundestudent die niet alleen een zorgplan inlevert, maar ook toelicht hoe risico's zijn afgewogen. Een student bedrijfskunde kan een advies schrijven en vervolgens reageren op kritische vragen van een opdrachtgever. Een informaticastudent kan programmatuur opleveren, wijzigingen doorvoeren en uitleggen waarom voor een bepaalde architectuur is gekozen.

AI mag bij zulke taken soms worden gebruikt, juist omdat dat in de beroepspraktijk eveneens gebeurt. De toets moet dan zichtbaar maken of de student:

- het vraagstuk zelf kan analyseren;
- de invoer voor AI doelgericht kan formuleren;
- uitkomsten kan controleren en verbeteren;
- relevante bronnen en beroepsnormen kan toepassen;
- alternatieven kan afwegen;
- verantwoordelijkheid kan nemen voor het uiteindelijke besluit.

Daarmee verschuift het toetsobject. Niet het vermogen om zonder hulpmiddelen een verzorgd document te produceren staat automatisch centraal, maar het vermogen om met beschikbare hulpmiddelen verantwoord professioneel te handelen. Voor bepaalde basishandelingen of risicovolle situaties kan zelfstandig handelen zonder AI overigens nog steeds noodzakelijk zijn. Dat moet de opleiding dan expliciet toetsen.

Tussenproducten maken ontwikkeling zichtbaar

Een tussenproduct hoeft geen uitgebreid dossier te zijn. Juist een beperkt aantal goed gekozen momenten kan voldoende inzicht geven. Een eerste probleemanalyse, een conceptadvies, een ontwerpbesluit of een beargumenteerde herziening laat vaak meer zien dan een volledig logboek waarin iedere stap wordt vastgelegd.

Het doel is niet dat studenten voortdurend bewijzen dat zij niets ongeoorloofds hebben gedaan. Het doel is dat de beoordelaar kan volgen hoe het professionele oordeel zich heeft ontwikkeld. Daarom zijn vooral tussenproducten waardevol waarin de student keuzes maakt of feedback verwerkt.

Een bruikbare selectie kan bestaan uit:

Bewijsbron	Wat hiermee zichtbaar wordt
Eerste analyse of aanpak	Begrip van het vraagstuk en gekozen werkwijze
Concept of voorlopig besluit	Redenering, toepassing van kennis en kwaliteit van keuzes
Ontvangen feedback	Welke aanwijzingen de student kreeg
Herzien product	Vermogen om feedback kritisch te verwerken
Korte toelichting op AI-gebruik	Doel, controle en invloed van het hulpmiddel
Eindproduct	Bereikt professioneel niveau en kwaliteit van het resultaat

Meer vastleggen is niet automatisch beter. Wanneer studenten voor iedere kleine handeling schermafbeeldingen, gesprekken en versies moeten bewaren, groeit vooral de administratieve last. Bovendien kan zo'n dossier een schijn van zekerheid geven: ook uitgebreide registraties kunnen onvolledig, geselecteerd of met hulp van anderen samengesteld zijn. De opleiding moet dus alleen bewijs vragen dat daadwerkelijk iets zegt over de bekwaamheid die wordt beoordeeld.

Mondelinge verantwoording als gerichte controle

Een korte mondelinge verantwoording is vooral sterk wanneer de vragen voortkomen uit het ingeleverde werk. De beoordelaar kan de student vragen een opvallende keuze toe te lichten, een alternatief te bespreken of een onderdeel toe te passen op een gewijzigde situatie. Daarmee wordt niet alleen feitelijke kennis gecontroleerd, maar ook de samenhang tussen product, begrip en professioneel oordeel.

Zo'n gesprek hoeft geen tweede volledige toets te worden. Enkele gerichte vragen kunnen al veel informatie opleveren:

- Waarom koos u voor deze aanpak?
- Welke informatie was onzeker of onvolledig?
- Welke uitvoer van AI hebt u verworpen en waarom?
- Wat zou er veranderen als deze voorwaarde wegvalt?
- Welk risico ontstaat wanneer uw advies onjuist blijkt?
- Welk onderdeel zou u zonder hulpmiddel opnieuw kunnen uitvoeren?

De kwaliteit van de mondelinge controle hangt wel af van de uitvoering. Studenten moeten weten waarop zij worden bevraagd, beoordelaars hebben houvast nodig voor vergelijkbare vraagstelling en de antwoorden moeten proportioneel meewegen. Anders dreigen verschillen in taalvaardigheid, spanning of beoordelaarsgedrag de uitkomst te veel te beïnvloeden.

Observeren waar handelen centraal staat

Bij sommige bekwaamheden is een product eenvoudigweg onvoldoende bewijs. Een gespreksverslag bewijst niet dat een student een moeilijk gesprek kan voeren. Een projectplan toont niet dat iemand in een team adequaat handelt. Een zorgvuldig geschreven veiligheidsanalyse laat nog niet zien hoe de student reageert wanneer zich onverwacht een risico voordoet.

Observatie is dan noodzakelijk, bijvoorbeeld tijdens een simulatie, praktijkopdracht, demonstratie of gesprek met een cliënt. De beoordelaar kijkt niet naar ieder detail, maar naar vooraf vastgestelde gedragingen en beslismomenten. AI kan daarbij onderdeel van de situatie zijn: de student gebruikt bijvoorbeeld een digitaal advies, controleert dat advies en besluit vervolgens of handelen verantwoord is.

Ook hier geldt proportionaliteit. Intensieve observatie is vooral gerechtvaardigd wanneer onbekwaam handelen grote gevolgen kan hebben of wanneer de bekwaamheid anders nauwelijks aantoonbaar is. Voor minder risicovolle onderdelen kunnen een opname, demonstratie of gerichte steekproef volstaan.

Steekproeven beperken de uitvoeringslast

Niet ieder onderdeel van ieder product hoeft volledig te worden gecontroleerd. Een opleiding kan vooraf bepalen welke onderdelen altijd worden verantwoord en welke

onderdelen steekproefsgewijs worden uitgediept. De student weet dan dat iedere belangrijke keuze onderwerp van gesprek kan zijn, zonder dat de docent alle procesinformatie uitvoerig hoeft te beoordelen.

Een goede steekproef is niet willekeurig in de alledaagse betekenis. De beoordelaar selecteert onderdelen die veel zeggen over de centrale bekwaamheid, opvallend afwijken of een verhoogd beroepsrisico kennen. Bij twijfel kan de steekproef worden uitgebreid. Zo ontstaat een getrapte werkwijze: beperkt controleren waar het beeld consistent is en verdiepen waar bronnen elkaar tegenspreken.

Voor bestuurders is dit een belangrijk ontwerpprincipe. Procesbewijs mag niet uitlopen op permanente digitale bewaking of een onuitvoerbare beoordelingslast. De vraag is steeds welke combinatie van bewijs voldoende zekerheid biedt, tegen aanvaardbare kosten en met respect voor privacy en gelijke behandeling.

Wanneer product, proces, verantwoording en waargenomen handelen bij elkaar passen, is algemene controle op vermoedelijk AI-gebruik meestal weinig informatief. En wanneer die bronnen elkaar tegenspreken, is aanvullend onderzoek nodig – niet een automatisch oordeel op basis van een technische waarschijnlijkheidsscore.

8.3 Uit mijn onderwijspraktijk: detectie zonder zekerheid

Mijn eerste reactie op de opkomst van generatieve AI was herkenbaar: als studenten zulke hulpmiddelen gebruiken, moeten we dat gebruik toch kunnen vaststellen? Ik experimenteerde daarom met automatische detectie. Teksten werden ingevoerd in hulpmiddelen die een waarschijnlijkheidsscore gaven voor door AI gegenereerde inhoud. Zo'n percentage oogde precies en leek houvast te bieden. Bij nader onderzoek bleek die zekerheid schijn.

Dezelfde tekst kon bij verschillende detectoren een andere uitslag krijgen. Ook veranderde de score soms nadat enkele zinnen waren herschreven, terwijl de inhoud en de manier waarop de tekst tot stand was gekomen nauwelijks verschilden. Menselijk geschreven passages werden geregeld als verdacht aangemerkt. Omgekeerd kon een met AI gemaakte tekst na beperkte bewerking zonder waarschuwing door de controle komen. De detector mat dus niet wat ik als beoordelaar werkelijk moest weten: of de student de vereiste kennis en professionele bekwaamheid zelfstandig kon aantonen.

Dat onderscheid werd extra zichtbaar bij studenten die bondig en voorspelbaar formuleerden. Een gelijkmatige schrijfstijl, korte zinnen of veelgebruikte formuleringen konden een hoge waarschijnlijkheidsscore opleveren. Daarmee ontstond het risico dat taalvaardigheid, schrijfstijl of taalachtergrond onbedoeld meespeelde in de verdenking. Een technisch signaal dreigde zo gevolgen te krijgen voor studenten zonder dat voldoende duidelijk was waarop het precies berustte.

Ik zag bovendien wat zo'n score met het gesprek deed. Zodra een detector een tekst als waarschijnlijk door AI geschreven aanmerkte, verschoof de aandacht van de kwaliteit

van het bewijs naar de vraag of de student iets had overtreden. De student moest vervolgens aantonen dat het systeem zich vergiste. Dat is een problematische omkering. Niet de opleiding onderbouwt dan haar oordeel, maar de student moet zich verdedigen tegen een berekening waarvan de werking, foutmarge en gegevensbasis meestal niet controleerbaar zijn.

Daarom ben ik gestopt met het beschouwen van detectiescores als bewijs. Hooguit konden ze aanleiding geven om nog eens goed naar een werkstuk te kijken, zoals een opvallende stijlbreuk of een onverwachte sprong in kwaliteit die dat ook kan zijn. Maar ook dan mocht het signaal nooit rechtstreeks leiden tot een oordeel over fraude, auteurschap of beheersing. Daarvoor was aanvullend, inhoudelijk onderzoek nodig.

Die conclusie dwong mij terug naar het toetsontwerp. Als het eindproduct niet meer vanzelf laat zien wat een student zelfstandig kan, heeft de opleiding een rijker bewijsarrangement nodig. Ik ging daarom meer werken met tussenproducten, ontwerpkeuzes, bronverantwoording, mondelinge toelichting en korte demonstraties. Niet om iedere handeling van studenten vast te leggen, maar om op betekenisvolle momenten te kunnen zien hoe zij redeneerden, welke afwegingen zij maakten en of zij hun resultaat konden verdedigen.

Binnen de ontwikkeling naar de opleiding Toegepaste AI werd dat geen tijdelijke noodmaatregel, maar een uitgangspunt voor valide beoordeling. Studenten moeten in hun toekomstige beroep immers met AI kunnen werken. Een algemeen verbod zou dan slecht aansluiten bij de beroepspraktijk. Tegelijk mag het gebruik van AI niet verhullen dat een student essentiële handelingen nog niet zelfstandig beheerst. We moesten dus per bekwaamheid bepalen wat studenten zonder hulpmiddel moeten kunnen, waar AI als professioneel gereedschap is toegestaan en welke verantwoordelijkheid altijd bij de student blijft.

Bij een programmeeropdracht kon een student bijvoorbeeld AI gebruiken om een eerste oplossing te verkennen. Voor de beoordeling was vervolgens niet alleen de ingeleverde code van belang. De student moest kunnen uitleggen waarom de gekozen aanpak passend was, fouten kunnen aanwijzen, alternatieven kunnen vergelijken en de oplossing onder gewijzigde voorwaarden kunnen aanpassen. Wie alleen een overtuigend product had gegenereerd, liep daar vast. Wie de onderliggende bekwaamheid beheerste, kon het werk reconstrueren, beargumenteren en verbeteren.

Ook deze aanpak is niet waterdicht. Een mondeling gesprek kan worden ingestudeerd, processtukken kunnen achteraf worden samengesteld en intensievere beoordeling kost docenttijd. Ik heb dan ook niet waargenomen dat één aanvullende toetsvorm het probleem oploste. De winst zat in de combinatie van verschillende bronnen die elkaar konden bevestigen of juist tegenspreken. Een eindproduct, enkele tussenstappen, een gerichte toelichting en een kleine nieuwe opdracht leverden samen een geloofwaardiger beeld op dan één technisch oordeel over de herkomst van een tekst.

Daarmee veranderde ook de rol van de docent. Die hoefde niet langer als digitale

rechercheur te bewijzen waarmee een product was gemaakt. De professionele taak werd weer inhoudelijk: vaststellen of de student begrijpt wat hij heeft gedaan, relevante keuzes kan verantwoorden en in een nieuwe situatie passend kan handelen. Dat vraagt beoordelingsdeskundigheid en goede vragen, maar het sluit veel beter aan bij wat toetsing in het hbo hoort te doen.

De bestuurlijke les is dat detectietechnologie geen zwak toetsontwerp kan repareren. Een instelling die zulke technologie toch inzet, moet vooraf duidelijk maken wat een score betekent, welke foutmarges bekend zijn, wie het signaal beoordeelt en hoe de student het oordeel kan betwisten. Zonder die waarborgen ontstaat schijnzekerheid met mogelijk grote gevolgen voor rechtsbescherming, gelijke behandeling en vertrouwen.

De verschuiving van detectie naar controleerbaar procesbewijs betekende overigens niet dat iedere student voortdurend moest worden gevolgd. Juist doordat we scherper bepaalden welk bewijs doorslaggevend was, konden we gericht controleren. Bij taken met beperkte gevolgen volstond soms een korte toelichting of steekproef. Waar zelfstandig handelen of beroepsveiligheid zwaar woog, was meer directe observatie nodig. De passende hoeveelheid toezicht bleek daarmee geen technische keuze, maar een afweging tussen het risico van een onjuist oordeel, de benodigde zekerheid en de uitvoerbaarheid voor student en docent.

Beoordelen door doorvragen



8.4 Besluiten over proportioneel toezicht

Proportioneel toezicht begint bij een eenvoudige vraag: **wat zijn de gevolgen als de beoordeling onjuist is?** Naarmate die gevolgen groter worden, moet de opleiding sterker bewijs verlangen dat de student zelfstandig en professioneel kan handelen. Niet omdat iedere student bij voorbaat wordt gewantrouwd, maar omdat de waarde van een diploma mede berust op de betrouwbaarheid van de beoordeling.

Bij een formatieve oefening kan een onjuist oordeel meestal worden hersteld in de volgende feedbackronde. Bij een afsluitende toets, een bindend studieadvies of een beoordeling die toegang geeft tot risicovol beroepshandelen ligt dat anders. Dan kan een fout grote gevolgen hebben voor de student, toekomstige cliënten, patiënten, opdrachtgevers of de samenleving. Een automatisch gegenereerd product, een ondoorzichtige detectiescore of een vluchtige mondelinge controle biedt in zulke situaties onvoldoende zekerheid.

Dat betekent overigens niet dat meer toezicht altijd beter is. Permanente observatie, uitgebreide registratie en meerdere controlelagen kunnen de werkdruk sterk verhogen, studenten ongelijk treffen en een sfeer van wantrouwen creëren. De bestuurlijke opdracht is daarom niet om toezicht te maximaliseren, maar om **de lichtste vorm te kiezen die nog voldoende betrouwbaar bewijs oplevert**.

Drie niveaus van toezicht

Een bruikbaar onderscheid bestaat uit drie niveaus. De precieze invulling verschilt per opleiding, maar de onderliggende redenering blijft gelijk.

Gevolgen van een onjuist oordeel	Passende vorm van toezicht	Mogelijk bewijs
Beperkt en goed herstelbaar	Licht toezicht en steekproeven	Korte toelichting, versiegeschiedenis, gerichte controlevraag of vergelijking met eerder werk
Betekenisvol voor studievoortgang of beheersing van een kerntaak	Gericht docenttoezicht	Mondelinge verdediging, aanvullende uitvoering, observatie van cruciale stappen of beoordeling van proces én product
Groot of moeilijk herstelbaar, bijvoorbeeld bij beroepsveiligheid of diplomering	Direct en zorgvuldig ingericht toezicht	Uitvoering onder gecontroleerde omstandigheden, praktijkobservatie, gesprek met beoordelaars en meerdere samenhangende bewijzen

Dit model koppelt toezicht dus niet rechtstreeks aan het gebruik van AI. Een student die AI gebruikt, hoeft niet automatisch zwaarder te worden gecontroleerd dan een student die dat niet doet. Doorslaggevend zijn het doel van de toets, de beroepsmatige betekenis van de bekwaamheid en de gevolgen van een onjuist oordeel.

Ook zonder AI kan een fraai verslag verhullen dat een student een taak niet zelfstandig beheerst. Omgekeerd kan verantwoord AI-gebruik juist onderdeel zijn van professioneel

handelen, mits de student de uitkomsten kan controleren, keuzes kan uitleggen en verantwoordelijkheid neemt voor het resultaat. Het toezicht moet aantonen wat de opleiding werkelijk wil beoordelen.

Vijf afwegingen voor goedkeuring

Voordat een manager, examencommissie of opleidingsbestuur een toetsvorm goedkeurt, zijn ten minste vijf afwegingen nodig.

1. Betrouwbaarheid

Levert de toets voldoende stabiel bewijs op? Een enkele prestatie is kwetsbaar voor toeval, spanning, technische problemen of een uitzonderlijk gunstige opdracht. Zeker bij belangrijke beslissingen verdient een combinatie van bewijzen de voorkeur: een product, een toelichting, een waargenomen handeling en eventueel eerder werk. Daarbij moet duidelijk zijn welk bewijs doorslaggevend is en hoe beoordelaars tot een oordeel komen.

2. Werkdruk en uitvoerbaarheid

Extra toezicht kost tijd. Een mondelinge verdediging van tien minuten lijkt beperkt, maar wordt bij honderden studenten al snel een omvangrijke taak. Daar komen planning, verslaglegging, herkansingen en scholing van beoordelaars nog bij. De volledige werklast moet daarom vóór invoering zichtbaar worden gemaakt.

Dat vraagt ook om taakherontwerp. Misschien kan een uitgebreide controle worden beperkt tot kerntaken, grensgevallen en steekproeven. Mogelijk kan een deel van de bestaande productbeoordeling vervallen wanneer directe observatie sterker bewijs oplevert. Nieuw toezicht simpelweg boven op de bestaande toetsing stapelen, levert zelden een duurzame werkwijze op.

3. Rechtsbescherming

Wanneer toezicht of technologie bijdraagt aan een beslissing over fraude, studievoortgang of diplomering, moet de student kunnen begrijpen hoe dat oordeel tot stand kwam. Een signaal mag nooit ongemerkt veranderen in een beschuldiging. Er moet ruimte zijn voor menselijke beoordeling, hoor en wederhoor en bezwaar tegen feitelijke of procedurele fouten.

Vooraf bij automatische analyses is terughoudendheid nodig. Als niet duidelijk is waarop een score berust, hoe vaak het systeem fouten maakt of welke groepen daardoor mogelijk vaker worden geraakt, mag die score niet als zelfstandig bewijs dienen. De instelling blijft verantwoordelijk voor het besluit, ook wanneer een leverancier het hulpmiddel heeft ontwikkeld.

4. Toegankelijkheid en gelijke behandeling

Een aanvullende mondelinge toets, tijdsdruk of verplichte observatie kan voor sommige studenten een grotere hindernis vormen dan voor andere. Dat hoeft de vorm niet

ongeschikt te maken, maar de opleiding moet wel onderzoeken of studenten met een beperking, taalachterstand of andere ondersteuningsbehoefte dezelfde bekwaamheid op een passende manier kunnen aantonen.

Gelijke beoordeling betekent niet dat iedere student exact dezelfde route moet volgen. Wel moeten de eisen gelijkwaardig zijn en aanpassingen vooraf zijn geregeld. Anders dreigt toezicht vooral te meten hoe goed een student met de toetsomstandigheden omgaat, in plaats van de beoogde professionele bekwaamheid.

5. Uitlegbaarheid

Studenten en docenten moeten weten wat wordt gecontroleerd, waarom dat nodig is en hoe AI wel of niet mag worden gebruikt. Vage regels als ‘AI-gebruik is beperkt toegestaan’ zijn daarvoor onvoldoende. Beter is om per toets te benoemen welke hulpmiddelen zijn toegestaan, welke handelingen zelfstandig moeten worden uitgevoerd, welke verantwoording wordt verwacht en waarop de beoordelaar doorvraagt.

Uitlegbaarheid helpt trouwens niet alleen de student. Ook beoordelaars hebben gemeenschappelijke criteria en herkenbare voorbeeldsituaties nodig. Zonder die basis ontstaan verschillen tussen docenten, opleidingen of locaties, met onzekerheid en bezwaarprocedures als voorspelbaar gevolg.

De bestuurlijke ondergrens

Een toetsvorm komt pas voor goedkeuring in aanmerking wanneer drie zaken tegelijk op orde zijn. Ten eerste moet zij valide bewijs opleveren van de beoogde kennis, vaardigheid of professionele bekwaamheid. Ten tweede moet het toezicht uitvoerbaar zijn zonder een onbeheersbare controlelast te veroorzaken. Ten derde moeten rechtsbescherming, toegankelijkheid en menselijke verantwoordelijkheid aantoonbaar zijn geborgd.

De kernbeslissing luidt daarmee niet: *hoe voorkomen we dat studenten AI gebruiken?* De betere vraag is: *welk bewijs hebben we nodig om verantwoord te kunnen verklaren dat deze student zelfstandig kan handelen?* Vanuit die vraag kan de opleiding bepalen welke combinatie van product, proces, toelichting en observatie nodig is.

Wanneer de betrouwbaarheid, uitvoerbaarheid en waarborgen niet overtuigend kunnen worden aangetoond, is de toetsvorm nog niet rijp voor invoering. Wanneer dat wel kan, hoeft AI de betrouwbaarheid van toetsing niet te ondermijnen. De opkomst ervan kan opleidingen juist dwingen scherper te bepalen wat een diploma werkelijk moet bewijzen: niet dat een student zonder hulpmiddelen een overtuigend product kan inleveren, maar dat die student kennis en hulpmiddelen verantwoord kan inzetten, professioneel kan handelen en daarvoor zelf rekenschap kan afleggen.

9 Kunnen handelen buiten de oefenomgeving

Zelfstandig handelen blijkt niet alleen in een toets. De beslissende proef volgt wanneer de vertrouwde opdracht, omgeving en ondersteuning verdwijnen. Dan moet een student in een nieuwe beroepssituatie zelf herkennen wat er speelt, relevante kennis kiezen, hulpmiddelen verantwoord inzetten en de kwaliteit van zijn handelen beoordelen.

Juist AI kan verhullen of dat vermogen werkelijk is opgebouwd. Een student die met gerichte aanwijzingen, voorbeelden en onmiddellijke terugkoppeling uitstekend presteert, kan veel hebben geleerd. Maar het sterke resultaat kan ook deels rusten op ondersteuning die straks ontbreekt. Het verschil wordt pas zichtbaar wanneer de context verandert en de student minder houvast krijgt.

Voor opleidingen en bestuurders staat daarmee meer op het spel dan de werking van één toepassing. Als studenten vooral leren handelen binnen een vaste voorziening, kan opschaling hun afhankelijkheid vergroten in plaats van hun professionele bekwaamheid. De relevante vraag is daarom niet alleen of AI tijdens het oefenen betere prestaties oplevert, maar wat een student buiten die oefenomgeving nog zelfstandig kan waarnemen, afwegen en doen.

Zelfstandig in een nieuwe beroepssituatie



9.1 Toepassing in een andere context volgt niet vanzelf

Een student kan in een vertrouwde onderwijsomgeving overtuigend presteren en toch vastlopen zodra de omstandigheden veranderen. Dat is geen uitzondering, maar een bekend probleem bij leren: kennis en werkwijzen die sterk verbonden zijn met één opdracht, hulpmiddel of vaste volgorde worden niet vanzelf herkend als bruikbaar in een andere situatie.

Juist bij AI kan dat verschil lang onzichtbaar blijven. Een oefenbot stelt op het juiste moment een vraag, wijst op ontbrekende informatie en geeft direct terugkoppeling. Daardoor kan de student een complexe casus stap voor stap oplossen. In de beroepspraktijk verschijnt diezelfde casus echter niet met een duidelijke opdracht, een compleet dossier en een knop voor feedback. De student moet daar eerst zelf herkennen wat er aan de hand is, welke kennis relevant is, welke informatie nog ontbreekt en wanneer overleg nodig is.

Dat herkennen en opnieuw toepassen noemen we transfer. Het gaat niet alleen om onthouden wat eerder is geleerd, maar om kennis en werkwijzen bruikbaar te maken in een situatie die op belangrijke punten anders is. Pas dan blijkt of een student werkelijk handelingsvermogen heeft opgebouwd.

Een goed product kan misleiden

Stel dat studenten met een AI-toepassing leren adviesgesprekken voor te bereiden. De toepassing helpt hen de klantvraag te ordenen, belanghebbenden te onderscheiden, risico's te benoemen en passende vervolgvragen te formuleren. Na enkele oefenronden leveren studenten aantoonbaar betere voorbereidingen in.

Dat is waardevol, maar het bewijst nog niet dat zij in een echt gesprek professioneel kunnen handelen. Misschien herkennen zij de onderliggende structuur alleen doordat de toepassing hun steeds dezelfde vragen stelt. Misschien nemen zij suggesties over zonder te begrijpen waarom die in deze situatie passend zijn. Of zij kunnen een keurige analyse maken wanneer alle gegevens beschikbaar zijn, maar raken de draad kwijt als een cliënt aarzelend antwoordt, informatie achterhoudt of onverwacht een nieuw probleem inbrengt.

Het risico is dan dat de opleiding het resultaat van de samenwerking tussen student en toepassing aanziet voor een bekwaamheid van de student. Dat verschil doet ertoe. Een AI-hulpmiddel mag het professionele handelen ondersteunen, maar de student moet wel kunnen beoordelen wanneer het hulpmiddel bruikbaar is, wat er aan de uitkomst ontbreekt en wanneer ervan moet worden afgeweken.

Een overtuigend product binnen één toepassing levert daarom hoogstens bewijs voor presteren **met die toepassing, onder die omstandigheden**. Voor een uitspraak over professionele bekwaamheid is aanvullend bewijs nodig.

Transfer vraagt om herkenning, selectie en aanpassing

In een nieuwe situatie moet een student minstens vier dingen kunnen doen.

Ten eerste moet de student relevante overeenkomsten herkennen. Een praktijkcasus ziet er zelden precies zo uit als een oefencasus. Namen, belangen, tijdsdruk en beschikbare informatie verschillen. Onder die zichtbare verschillen moet de student de beroepsmatige kern kunnen herkennen: welk soort vraagstuk speelt hier eigenlijk?

Ten tweede moet de student bruikbare kennis en werkwijzen selecteren. Niet alles wat eerder is geleerd, is in iedere situatie even relevant. Professioneel handelen vraagt nu eenmaal om keuzes. Welke analyse is nodig, welke stap kan worden overgeslagen en welke bron moet juist opnieuw worden geraadpleegd?

Ten derde moet de student de gekozen aanpak aanpassen. Een gespreksstructuur die werkt bij een coöperatieve cliënt kan tekortschieten bij weerstand of wantrouwen. Een standaardanalyse die geschikt is voor een overzichtelijk project kan onverantwoord zijn wanneer veiligheidsrisico's, conflicterende belangen of onvolledige gegevens een rol spelen.

Ten vierde moet de student de gevolgen beoordelen en bijsturen. Transfer is dus niet voltooid zodra een bekende werkwijze in een nieuwe context wordt ingezet. De student moet ook kunnen vaststellen of die werkwijze daar het gewenste effect heeft.

AI kan elk van deze stappen ondersteunen. De toepassing kan verschillen tussen casussen zichtbaar maken, alternatieven aandragen, tegenargumenten formuleren en de student vragen zijn keuze te onderbouwen. Maar als AI steeds bepaalt wat relevant is en welke stap volgt, oefent de student mogelijk vooral het opvolgen van aanwijzingen. Dan groeit de prestatie binnen de oefenomgeving, terwijl de zelfstandige inzetbaarheid achterblijft.

Breng bewust variatie aan

Wie transfer wil bevorderen, moet variatie niet aan het toeval overlaten. Tien vergelijkbare opdrachten leveren wel routine op, maar routine binnen een smalle reeks omstandigheden is nog geen brede professionele inzetbaarheid. Studenten moeten oefenen met situaties die voldoende overeenkomen om een onderliggende werkwijze te herkennen en voldoende verschillen om bewuste aanpassing af te dwingen.

Dat kan bijvoorbeeld door te variëren in:

- de volledigheid en betrouwbaarheid van de informatie;
- de belangen en reacties van betrokkenen;
- de beschikbare tijd en middelen;
- het beroepsrisico van een verkeerde beslissing;
- de vorm waarin het vraagstuk wordt aangeboden;
- de hulpmiddelen die wel of niet beschikbaar zijn;
- de noodzaak om zelfstandig te handelen of anderen te raadplegen.

Variatie alleen is overigens niet genoeg. Studenten moeten ook leren benoemen waarom

een aanpak in de ene situatie wel werkte en in de andere niet. Zonder die vergelijking kunnen verschillende casussen als losse ervaringen naast elkaar blijven staan. Gerichtte vragen helpen om de verbinding te leggen: wat bleef hetzelfde, wat veranderde, welke aanname moest worden herzien en welk signaal leidde tot een andere keuze?

De opbrengst van AI ligt hier niet in het onbeperkt voortbrengen van nieuwe casussen. Meer varianten zijn pas nuttig als de verschillen didactisch betekenisvol zijn en aansluiten bij de bekwaamheid die de opleiding wil opbouwen. Een grote hoeveelheid willekeurige oefensituaties kan studenten juist laten zoeken naar oppervlakkige patronen. De docent of het opleidingsteam moet daarom bepalen welke contextwisselingen ertoe doen in het beroep.

Meet buiten de vertrouwde toepassing

Ook de evaluatie moet de overstap naar een andere context zichtbaar maken. Wanneer studenten oefenen én worden beoordeeld met dezelfde interface, dezelfde vraagvolgorde en dezelfde vorm van terugkoppeling, meet de opleiding vooral hoe goed zij binnen die inrichting functioneren.

Sterker bewijs ontstaat wanneer studenten een verwante maar nieuwe situatie krijgen en zelf moeten bepalen wat nodig is. Dat kan een praktijksimulatie zijn, een gesprek met een onbekende gesprekspartner, een onvolledig dossier, een project voor een externe opdrachtgever of een beroepsopdracht waarin verschillende oplossingen verdedigbaar zijn. Belangrijk is dat niet alleen het eindproduct wordt bekeken. Ook de waarneming, afweging, keuze van hulpmiddelen, uitvoering en bijsturing moeten zichtbaar worden.

Voor managers heeft dit een directe consequentie voor de businesscase. Veel gebruik, hoge tevredenheid en betere resultaten binnen de toepassing zijn bemoedigende signalen, maar nog geen bewijs van transfer. De relevante vraag is of studenten in een andere, beroepsmatig betekenisvolle context beter handelen dan voorheen. Blijkt dat niet, dan heeft de toepassing mogelijk vooral de oefentaak gemakkelijker gemaakt.

Daarbij hoeft het doel niet te zijn dat studenten altijd zonder AI werken. In veel beroepen zullen digitale hulpmiddelen een normaal onderdeel van het werk zijn. Zelfstandigheid betekent dan niet: alles uit het hoofd en zonder ondersteuning doen. Het betekent wel dat de student het vraagstuk kan doorgronden, een hulpmiddel bewust kiest, de uitkomst kritisch controleert en verantwoordelijk blijft voor het handelen dat erop volgt.

Om dat zichtbaar te maken, moet de ondersteuning tijdens het leerproces niet voortdurend gelijk blijven. Aanwijzingen, voorbeelden en automatische terugkoppeling die in het begin nuttig zijn, kunnen later het zicht op zelfstandigheid ontnemen. De volgende stap is daarom niet om alle steun plotseling weg te halen, maar om haar doelgericht en geleidelijk af te bouwen.

9.2 Steun geleidelijk afbouwen

Geleidelijk afbouwen voorkomt twee fouten. In het eerste geval blijft de ondersteuning zo ruim dat studenten vooral leren de toepassing te volgen. In het tweede verdwijnt alle hulp in één keer, waardoor een terugval moeilijk te duiden is. Heeft de student de bekwaamheid nog onvoldoende opgebouwd, begrijpt hij de nieuwe context niet of raakt hij alleen de vertrouwde bediening kwijt? Door de steun in stappen te verminderen, wordt zichtbaar wat de student inmiddels zelf kan en waar nog gerichte oefening nodig is.

Van voordoen naar zelfstandig bijsturen

In het begin kan veel ondersteuning didactisch verstandig zijn. Een student krijgt bijvoorbeeld een uitgewerkt voorbeeld, een vaste reeks vragen, aanwijzingen bij iedere stap en onmiddellijk commentaar op een keuze. Daarmee wordt de aandacht gericht op kenmerken die een beginnend beroepsbeoefenaar nog niet vanzelf ziet.

Die ondersteuning moet echter steeds een tijdelijk doel hebben. Bij iedere vorm van hulp hoort daarom de vraag: **welke handeling moet de student later zelf uitvoeren?** Een aanwijzing om ontbrekende informatie te verzamelen is nuttig als de student daardoor leert een onvolledige probleembeschrijving te herkennen. Blijft de aanwijzing altijd verschijnen, dan kan de student afhankelijk worden van het signaal en leert hij mogelijk niet zelf vaststellen dat er informatie ontbreekt.

Het afbouwen kan langs vier stappen verlopen:

1. Voordoen en toelichten.

De toepassing maakt een aanpak zichtbaar, benoemt aandachtspunten en laat zien hoe een deskundige relevante signalen waarneemt, gegevens afweegt en tot een keuze komt. De student volgt de redenering en verklaart waarom de stappen nodig zijn. In deze fase moet duidelijk zijn welke aanwijzingen, voorbeelden, vragen en terugkoppeling beschikbaar zijn en welke professionele handelingen de student uiteindelijk zelf moet uitvoeren.

2. Handelen met aanwijzingen.

De student voert de taak zelf uit, maar krijgt nog vragen, tussenstappen of waarschuwingen. De ondersteuning stuurt het proces zonder het antwoord volledig over te nemen. De opleiding vermindert de hulp zodra uit de prestaties blijkt dat de student zelf relevante informatie herkent, een aanpak kiest en die uitvoert. Daarbij wordt expliciet gemaakt welke hulpmiddelen ook in de beroepspraktijk normaal en toegestaan zijn.

3. Handelen met terugkoppeling achteraf.

De aanwijzingen tijdens de uitvoering verdwijnen. Pas nadat de student een keuze heeft gemaakt, volgt terugkoppeling. Daardoor wordt zichtbaar wat hij zonder directe sturing waarneemt en afweegt, hoe hij de kwaliteit van zijn uitkomst controleert en of hij zijn aanpak na feedback kan bijsturen. Een nieuwe

taak of context moet voorkomen dat alleen vertrouwde reacties worden herhaald.

4. Zelfstandig handelen en verantwoorden.

De student bepaalt de aanpak, kiest eventueel een passend hulpmiddel, controleert de uitkomst en onderbouwt waarom zijn keuze in deze context passend is. Feedback komt van een docent, begeleider, opdrachtgever of medestudent en richt zich op het geheel van de prestatie. De opleiding legt vast wie de uitkomst beoordeelt en waar menselijke observatie en deskundig oordeel nodig blijven.

Deze stappen hoeven niet voor iedere student in hetzelfde tempo te verlopen. De ene student heeft langer ondersteuning nodig bij het analyseren van de situatie, terwijl een ander juist moeite houdt met het beoordelen van alternatieven of het bijsturen na feedback. Het doel is niet dat iedereen op hetzelfde moment even weinig hulp ontvangt, maar dat de opleiding kan uitleggen waarom bepaalde ondersteuning nog nodig is en wanneer die wordt verminderd.

Bouw niet alleen de hoeveelheid hulp af

Ondersteuning kent verschillende vormen. Juist daarom is “de bot uitzetten” een te grove ingreep. Een toepassing kan immers helpen bij het herkennen van relevante gegevens, het ordenen van een aanpak, het formuleren van alternatieven, het controleren van een product of het reflecteren op het resultaat. Binnen elk van de vier stappen moet de opleiding daarom bepalen welke vorm van hulp wordt verminderd en welk onderdeel van het handelen daardoor zelfstandig zichtbaar moet worden.

Een student kan bijvoorbeeld al zelfstandig een cliëntgesprek voeren, maar nog steun nodig hebben om achteraf patronen in het gesprek te herkennen. Andersom kan iemand een analyse goed verwoorden, terwijl tijdens het gesprek belangrijke signalen worden gemist. Alleen kijken naar het eindproduct verhuult dat verschil.

Bij het doorlopen van de vier stappen moet daarom zichtbaar worden of de student relevante signalen en ontbrekende informatie herkent, gegevens verbindt aan kennis, criteria en beroepsnormen, een verdedigbare aanpak uitvoert, de uitkomst controleert, zo nodig bijstuurt en zijn keuzes verantwoordt. Zo ontstaat een nauwkeuriger beeld dan met de algemene vraag of een student “het zonder AI kan”.

In de beroepspraktijk zal de student vaak wél hulpmiddelen gebruiken. Waar het om gaat, is dat het professionele oordeel niet ongemerkt naar het hulpmiddel verschuift.

Varieer de context voordat u zelfstandigheid concludeert

Afbouwen binnen één vertrouwde omgeving is nog niet genoeg. Studenten kunnen namelijk leren welke reactie een toepassing verwacht, zonder de achterliggende werkwijze werkelijk te beheersen. Zij herkennen dan de opbouw van de oefening, vaste formuleringen of terugkerende soorten feedback. Zodra de vorm verandert, valt die houvast weg.

Daarom moet de steun niet alleen worden verminderd, maar moet het handelen ook in uiteenlopende situaties worden beproefd. Denk aan een casus met onvolledige informatie, een afwijkende opdrachtgever, tegenstrijdige belangen of tijdsdruk. Ook kan dezelfde bekwaamheid in een andere vorm worden gevraagd: eerst schriftelijk, daarna mondeling of in een praktijksimulatie.

De student laat transfer zien wanneer hij onder veranderde omstandigheden zelf kan bepalen wat uit eerdere oefeningen bruikbaar is. Een foutloze herhaling van een bekende aanpak is daarvoor onvoldoende. Juist de onderbouwing bij een afwijkende situatie maakt zichtbaar of de student regels volgt of professioneel afweegt.

Maak de afbouw onderdeel van het ontwerp

Voor een opleiding is dit geen detail dat aan individuele docenten kan worden overgelaten. Zodra AI structureel ondersteuning geeft, moet ook structureel zijn vastgelegd hoe die ondersteuning in de vier stappen wordt verminderd en hoe zelfstandigheid wordt aangetoond. Anders kan een toepassing uitstekende gebruikscijfers opleveren, terwijl onduidelijk blijft wat studenten zonder de vaste hulp beheersen.

Een bruikbaar ontwerp beschrijft daarom per stap welke steun de student krijgt, welke professionele handeling daarmee wordt opgebouwd en op basis van welke voortgang de steun wordt verminderd. Ook moet vaststaan welke hulpmiddelen beschikbaar blijven, in welke nieuwe taak of context zelfstandigheid zichtbaar wordt en wie de prestatie beoordeelt. Zo vormen de vier stappen niet alleen een didactische opbouw, maar ook een toetsbaar ontwerp voor invoering en evaluatie.

Daarmee verandert de managementvraag. Niet alleen het aantal gebruikers, de tevredenheid of de snelheid van afronden telt, maar ook het aantal studenten dat bij minder ondersteuning een gelijkwaardige of betere prestatie levert. Relevant is bovendien waar studenten terugvallen, welke begeleiding dan nodig blijkt en hoeveel werk het onderhouden van verschillende ondersteuningsniveaus vraagt.

Terugval is daarbij niet meteen een mislukking. Zij kan juist waardevolle informatie opleveren over een te grote stap, een onvoldoende gevarieerde oefenreeks of een bekwaamheid die nog niet stevig genoeg is opgebouwd. Problematisch wordt het pas wanneer de opleiding die terugval niet onderzoekt en toch concludeert dat de toepassing rendement heeft opgeleverd.

De beslissende proef komt daarom wanneer de vertrouwde steun niet meer automatisch verschijnt. Dan blijkt of de student zelf een aanpak kan kiezen, onzekerheid kan herkennen en hulp kan inroepen zonder het eigenaarschap over de taak uit handen te geven. Soms wordt pas op dat moment zichtbaar hoe sterk de opgebouwde afhankelijkheid werkelijk is.

9.3 Uit mijn onderwijspraktijk: zonder bot opnieuw

beginnen

Ik zag die afhankelijkheid vooral bij studenten die binnen een vertrouwde onderwijsbot overtuigend konden werken. Zolang de opdracht, de kennisbron en de manier van terugkoppelen herkenbaar waren, stelden zij gerichte vragen, verbeterden zij hun uitwerking en kwamen zij tot goed onderbouwde keuzes. Hun prestaties wekten de indruk dat zij de aanpak beheersten.

Dat beeld veranderde zodra ik dezelfde soort beroepsopdracht in een andere omgeving aanbood. De casus was nieuw geformuleerd, de informatie stond op een andere plek of de vaste bot was niet beschikbaar. Sommige studenten begonnen dan niet met het analyseren van het vraagstuk, maar met het zoeken naar vervangende houvast. Welke vraag moest ik stellen? Waar stond het voorbeeld? Welke bot hoorde hierbij? Een student kon dus goed hebben geleerd hoe hij met één voorziening moest werken, zonder de onderliggende werkwijze zelfstandig te beheersen.

Dat onderscheid is belangrijk. Vaardig gebruik van een bot kan immers waardevol zijn, zeker wanneer vergelijkbare hulpmiddelen ook in de beroepspraktijk beschikbaar zijn. Maar het is iets anders dan zelfstandig een situatie kunnen doorgronden, een aanpak kunnen kiezen en de kwaliteit van een uitkomst kunnen beoordelen. De bot had in sommige gevallen niet alleen het leren ondersteund, maar ook een deel van de regie overgenomen.

Mijn eerste reactie was de bot verder te verbeteren. Ik maakte de aanwijzingen duidelijker, voegde voorbeelden toe en probeerde de terugkoppeling beter op het niveau van de student af te stemmen. Binnen de oefenomgeving hielp dat. Studenten konden sneller verder en liepen minder vaak vast. Buiten die omgeving veranderde er echter weinig. Achteraf was dat logisch: ik had de ondersteuning verfijnd, terwijl het probleem juist was dat studenten zonder die ondersteuning niet goed wisten hoe zij moesten beginnen.

Daarom verschoof ik de aandacht van het antwoord naar de aanpak. Studenten moesten niet alleen een product opleveren, maar ook benoemen welke stappen zij zetten. Wat is hier precies het vraagstuk? Welke informatie ontbreekt nog? Welke criteria gebruikt u om mogelijkheden af te wegen? Op welk punt vertrouwt u de uitkomst niet zonder aanvullende controle? En wanneer is menselijke of vakinhoudelijke hulp nodig?

Die vragen vormden geen vast invulrijtje dat opnieuw afhankelijkheid kon veroorzaken. We oefenden ermee in wisselende vormen. Soms werkte een student met een onderwijsbot, soms met een algemeen AI-hulpmiddel en soms zonder digitaal hulpmiddel. De ene keer bood ik een uitgewerkte casus aan, de andere keer slechts een onvolledige praktijksituatie. Ook veranderde de rol van AI: van uitlegger naar tegenspreker, van bron van voorbeelden naar gesprekspartner die juist onjuiste of onvolledige suggesties kon geven.

Die variatie maakte het leren aanvankelijk minder soepel. Studenten hadden meer tijd nodig en vroegen vaker of zij “op de goede weg” zaten. Toch was juist dat ongemak

leerzaam. Het dwong hen om niet op de vorm van de oefening te vertrouwen, maar om kenmerken van de situatie te herkennen. Langzamerhand gingen zij dezelfde denkstappen in verschillende omgevingen gebruiken: eerst afbakenen, dan informatie verzamelen, vervolgens mogelijkheden vergelijken, een keuze onderbouwen en ten slotte de uitkomst controleren.

Hulpmiddelvrije momenten bleken daarbij onmisbaar. Ik gebruikte die niet om AI uit het onderwijs te weren of om studenten op snelheid te betrappen. Het doel was zichtbaar te maken wat zij zelf konden waarnemen, beredeneren en besluiten. Soms duurde zo'n moment maar kort. Een student moest bijvoorbeeld vóór het openen van een bot zijn eerste analyse opschrijven, een controleplan maken of mondeling toelichten welke risico's hij in de casus zag. Pas daarna mocht AI worden ingezet.

Daardoor werd ook beter zichtbaar waarvoor het hulpmiddel werkelijk waarde toevoegde. De ene student had AI nodig om alternatieven te verkennen, maar kon de uiteindelijke afweging zelfstandig maken. Een ander produceerde met AI snel een nette uitwerking, maar kon niet uitleggen waarom een advies in deze beroepssituatie verantwoord was. Weer een ander bleek zonder bot prima te kunnen starten en gebruikte het hulpmiddel vooral om de eigen redenering kritisch te laten bevragen. Aan de buitenkant konden hun producten vergelijkbaar zijn; hun handelingsvermogen verschilde sterk.

Niet alles werkte meteen. Wanneer ik de vertrouwde ondersteuning te abrupt wegnam, ontstond er vooral onzekerheid. Studenten die nog onvoldoende vakkennis hadden opgebouwd, konden dan ook met een goede strategie niet verder. Omgekeerd leverde langdurige toegang tot uitgebreide aanwijzingen soms keurige resultaten op zonder dat duidelijk werd welk deel van de prestatie aan de student kon worden toegeschreven. Ik moest dus beter onderscheiden of een student vastliep door ontbrekende kennis, een gebrekkige aanpak, onvoldoende oefening of een te grote afhankelijkheid van het hulpmiddel.

Die ervaring heeft het onderwijsontwerp van de opleiding Toegepaste AI mede gevormd. AI is daarin niet alleen een onderwerp dat studenten moeten begrijpen en evenmin alleen een handig middel om opdrachten sneller uit te voeren. Studenten moeten leren handelen in situaties waarin hulpmiddelen veranderen, gegevens onvolledig zijn, uitkomsten onzeker blijven en professionele verantwoordelijkheid niet kan worden uitbesteed. Dat vraagt om oplopende complexiteit én om momenten waarop de steun bewust van vorm verandert of tijdelijk verdwijnt.

Voor mij werd daarmee ook de opbrengst anders zichtbaar. Ik keek niet meer alleen naar de kwaliteit van het product dat binnen de botomgeving tot stand kwam. Minstens zo relevant werd of studenten in een nieuwe casus zelf konden beginnen, een passende strategie konden kiezen, AI doelgericht konden inschakelen en een uitkomst konden afwijzen wanneer die niet deugde. De vraag was niet of zij zonder AI alles alleen konden, maar of zij ook mét wisselende hulpmiddelen eigenaar van hun handelen bleven.

Dat is een bescheidener opbrengstclaim dan “studenten presteren beter met een bot”, maar bestuurlijk gezien wel een bruikbaarere. Een vaste bot kan uitstekende oefencondities scheppen. Pas wanneer studenten de opgebouwde werkwijze meenemen naar een andere taak, een ander hulpmiddel of een hulpmiddelvrij moment, ontstaat er grond om te spreken van leeroverdracht. En juist die overdracht bepaalt of een toepassing alleen binnen één onderwijsontwerp werkt of werkelijk bijdraagt aan duurzame professionele bekwaamheid.

Een ondeugdelijke AI-uitkomst afwijzen



9.4 Besluiten op basis van leeroverdracht

Bestuurders doen er goed aan drie niveaus van bewijs uit elkaar te houden en daar direct een besluit aan te verbinden.

Bewijsniveau	Wat wordt zichtbaar?	Bestuurlijke betekenis
Prestatie binnen de toepassing	Studenten oefenen meer, werken sneller of leveren een beter product op	Stoppen als de opbrengst beperkt blijft en verbetering weinig perspectief biedt; aanpassen als de toepassing bruikbare oefencondities biedt, maar leeroverdracht uitblijft

Prestatie bij een nieuwe taak	Studenten passen de aanpak toe op een onbekende casus	Voortzetten terwijl de opleiding bij meerdere taken, groepen en momenten verder bewijs verzamelt
Prestatie in een andere omgeving	Studenten handelen adequaat met andere of minder ondersteuning, zo mogelijk in de beroepspraktijk	Opschalen als de resultaten standhouden en de benodigde begeleiding uitvoerbaar blijft

Deze vier besluiten vormen één vaste ladder: stoppen, aanpassen, voortzetten of opschalen. Verder beproeven is daarbij geen afzonderlijk vijfde besluit. Het hoort bij aanpassen wanneer eerst het onderwijsontwerp moet veranderen en bij voortzetten wanneer de eerste resultaten positief zijn, maar nog niet sterk genoeg voor opschaling.

Een toepassing die alleen binnen de oorspronkelijke oefenomgeving resultaat oplevert, kan didactisch nuttig zijn, maar rechtvaardigt nog geen brede invoering. Voor opschaling is sterker bewijs nodig: studenten moeten de opgebouwde bekwaamheid ook laten zien bij een nieuwe taak, in een andere context en met minder of andere ondersteuning. Anders schaalt de instelling mogelijk vooral de afhankelijkheid van één toepassing op, en niet het leren.

Leeroverdracht hoort daarom thuis in de businesscase. Niet als aanvullende evaluatie achteraf, maar als een van de criteria waarop een manager deze beslis ladder doorloopt.

Van productkwaliteit naar zelfstandig handelen

Bij veel experimenten ligt de eerste aandacht op het directe resultaat. Studenten leveren met AI bijvoorbeeld een beter advies, een vollediger ontwerp of een professioneler verslag op. Dat is relevant, maar het blijft onduidelijk waardoor de verbetering is ontstaan. Heeft de student beter leren analyseren en beslissen, of heeft het hulpmiddel vooral een sterker product voortgebracht?

Om dat onderscheid te maken, moet een opleiding niet alleen het eindproduct beoordelen. Ook het handelen van de student telt mee:

- herkent de student wat voor soort vraagstuk hij voor zich heeft?
- kiest hij zelf een passende aanpak?
- bepaalt hij bewust wanneer AI wel en niet bruikbaar is?
- controleert hij de uitkomsten op inhoud, bronnen en toepasbaarheid?
- kan hij een gebrekkig voorstel afwijzen en verbeteren?
- kan hij zijn keuzes professioneel verantwoorden?
- lukt dit ook bij een onbekende casus, een ander hulpmiddel of zonder directe AI-ondersteuning?

Deze vragen maken zichtbaar of de student een werkwijze heeft opgebouwd die overdraagbaar is. Juist daarin zit het verschil tussen een toepassing die het onderwijs tijdelijk ondersteunt en een toepassing die duurzame professionele bekwaamheid helpt

ontwikkelen.

Maak de afstand tot de oefensituatie groter

Leeroverdracht kan niet overtuigend worden vastgesteld door studenten nogmaals vrijwel dezelfde opdracht te geven. Hoe sterker de nieuwe situatie op de oefening lijkt, hoe minder zeker het is dat zij hun kennis en aanpak werkelijk zelfstandig inzetten.

Een opleiding kan de afstand stapsgewijs vergroten. Eerst verandert alleen de casus, daarna ook de gegevens, het vakgebied, het beschikbare hulpmiddel of de omgeving waarin de student moet handelen. Vervolgens kan de begeleiding worden verminderd. Bij iedere stap wordt duidelijker wat de student zelf beheerst en welk deel van de prestatie afhankelijk blijft van de oorspronkelijke inrichting.

Dat hoeft niet te betekenen dat studenten altijd zonder AI moeten werken. In veel beroepen zal juist het verantwoord inzetten van AI tot de professionele praktijk behoren. Een hulpmiddelvrij moment kan wel nodig zijn om basiskennis, redeneringen of kritische controle afzonderlijk zichtbaar te maken. Daarnaast moet de student kunnen aantonen dat hij niet vastzit aan één bot, leverancier of vooraf ontworpen gesprekspad.

De toets is dus niet: kan de student dezelfde prestatie zonder enig hulpmiddel herhalen? De betere vraag luidt: blijft de student regie voeren wanneer taak, context, ondersteuning en hulpmiddel veranderen?

Het eerste bewijsniveau kan voldoende zijn om de toepassing aan te passen en opnieuw te beproeven. Voor voortzetting is doorgaans ten minste een positieve aanwijzing op het tweede niveau nodig. Naarmate de toepassing duurder wordt, meer opleidingen raakt of wordt ingezet bij risicovol professioneel handelen, wordt ook het derde niveau belangrijker voordat opschaling verantwoord is.

Daarbij past enige terughoudendheid. Een geslaagde nieuwe casus bewijst niet meteen dat de bekwaamheid in iedere beroepssituatie standhoudt. Omgekeerd hoeft een tegenvallende prestatie niet te betekenen dat de hele toepassing waardeloos is. Misschien was de oefening te smal, werd de ondersteuning te abrupt afgebouwd of sloot de nieuwe taak onvoldoende aan op wat studenten hadden kunnen leren. Leeroverdracht is daarmee niet alleen een uitkomstmaat, maar ook informatie waarmee het onderwijsontwerp kan worden verbeterd.

Kijk naar de volledige opbrengst

Wanneer studenten hun bekwaamheid meenemen naar andere situaties, ontstaat er waarde op drie niveaus. Studenten worden minder afhankelijk van vaste aanwijzingen en kunnen bewuster omgaan met wisselende hulpmiddelen. Docenten krijgen beter zicht op wat studenten zelf beheersen en kunnen hun begeleiding richten op hardnekkige tekortkomingen. De instelling kan aannemelijker maken dat een investering bijdraagt aan de opleidingskwaliteit in plaats van alleen aan de productie

binnen één digitale omgeving.

Daar staan wel lasten tegenover. Nieuwe transferopdrachten moeten worden ontworpen, docenten moeten het handelen en niet alleen het product beoordelen en gegevens uit verschillende onderwijs- en praktijksituaties moeten zorgvuldig worden geïnterpreteerd. Ook kunnen hulpmiddelvrije momenten weerstand oproepen wanneer studenten niet begrijpen waarom die nodig zijn. Het meten van leeroverdracht kost dus tijd. Op die last valt echter moeilijk te bezuinigen, want zonder deze stap blijft de belangrijkste opbrengstclaim onbewezen.

Stoppen, aanpassen, voortzetten of opschalen

De uitkomst van de evaluatie leidt niet automatisch tot opschaling. Vier besluiten zijn mogelijk.

Stoppen ligt voor de hand wanneer de betere prestatie volledig wegvalt zodra studenten buiten de toepassing werken, terwijl aanpassing weinig perspectief biedt of de lasten niet opwegen tegen de beperkte opbrengst.

Aanpassen past wanneer de toepassing wel goede oefencondities biedt, maar studenten onvoldoende worden voorbereid op variatie. De opleiding kan dan meer uiteenlopende casussen aanbieden, strategieën explicieter laten verwoorden, ondersteuning geleidelijk verminderen of verschillende hulpmiddelen laten gebruiken. Na die aanpassing wordt de toepassing opnieuw beproefd.

Voortzetten is verstandig wanneer de eerste aanwijzingen voor leeroverdracht positief zijn, maar het bewijs nog beperkt is. De toepassing blijft dan afgebakend terwijl de opleiding over een langere periode en bij meerdere groepen gegevens verzamelt. Verder beproeven betekent hier dus dat de proef wordt voortgezet zonder de toepassing al breder in te voeren.

Opschalen komt pas in beeld wanneer studenten ook bij nieuwe taken en in andere omgevingen adequaat blijven handelen, docenten de uitkomsten herkennen en de benodigde begeleiding uitvoerbaar is. Bovendien moeten de baten standhouden wanneer niet langer alleen de meest gemotiveerde studenten en docenten deelnemen.

De kern van het besluit is eenvoudig: schaal niet op omdat een toepassing binnen haar eigen omgeving goed werkt. Schaal op wanneer aantoonbaar is dat studenten meenemen wat zij daar hebben geleerd. Pas dan investeert de instelling niet alleen in een hulpmiddel, maar in handelingsvermogen dat ook buiten dat hulpmiddel waarde houdt.

10 Eerst opnieuw diagnosticeren, dan besluiten

Een pilot eindigt niet wanneer de laatste student de toepassing heeft gebruikt. Dan begint juist de bestuurlijke beoordeling. De vraag is niet hoeveel activiteit de toepassing heeft opgeleverd, maar of het oorspronkelijke probleem aantoonbaar kleiner is geworden – en tegen welke prijs.

Dat onderscheid is belangrijk. Veel gebruik, positieve reacties en een technisch goed werkende toepassing kunnen vertrouwen wekken, terwijl de onderwijswaarde beperkt blijft. Omgekeerd kan een bescheiden proef wel degelijk betekenisvol zijn wanneer studenten beter leren handelen, medewerkers bruikbare ruimte winnen en de benodigde begeleiding uitvoerbaar blijft.

Daarom moet de instelling terug naar het vertrekpunt. Welke verandering werd gezocht, voor wie en waaraan zou die zichtbaar worden? Alleen vanuit die oorspronkelijke diagnose kunnen resultaten, kosten, risico's en nieuw beheerwerk eerlijk worden gewogen. Ook stoppen moet daarbij een volwaardige uitkomst zijn. Niet iedere geslaagde proef verdient voortzetting en niet iedere tegenvaller betekent dat de gekozen richting onjuist was. Een verantwoord besluit vraagt vooral om de discipline om opnieuw te kijken – met betere gegevens, scherpere vragen en zonder de investering tot nu toe als argument voor de volgende stap te gebruiken.

De pilot opnieuw beoordeeld



10.1 De Ausgangssituatie tegenover het behaalde resultaat

Daarvoor moet de opleiding terug naar de vraag waarmee het experiment begon. Niet: *werd de toepassing gebruikt?* Maar: *is het oorspronkelijke onderwijsprobleem aantoonbaar kleiner geworden?* Alleen door de nieuwe situatie naast de Ausgangssituatie te leggen, ontstaat een bruikbare basis voor een besluit.

Dat klinkt vanzelfsprekend, maar in de praktijk verschuift de aandacht gemakkelijk. Een opleiding begint bijvoorbeeld met het probleem dat studenten te weinig kunnen oefenen met professionele gesprekken. Na de proefperiode rapporteert het projectteam hoeveel gesprekken met de bot zijn gevoerd en hoe studenten de toepassing waarderen. Dat zijn relevante gegevens, maar ze beantwoorden nog niet de oorspronkelijke vraag. Daarvoor moet ook worden vastgesteld of studenten vaker hebben geoefend, gericht konden bijsturen en uiteindelijk beter handelden in een gesprek met een echte gesprekspartner.

De vergelijking begint daarom bij vier elementen:

- het oorspronkelijke vraagstuk;
- de doelgroep en Ausgangssituatie;
- de beoogde verandering;
- de maatstaf waarmee die verandering zichtbaar zou worden.

Als deze elementen bij de start niet scherp zijn vastgelegd, wordt evalueren achteraf lastig. Toch hoeft dat niet te betekenen dat een pilot waardeloos is. De opleiding kan alsnog reconstrueren wat er vóór de invoering gebeurde, bijvoorbeeld aan de hand van eerdere studentproducten, toetsresultaten, roostergegevens, docentregistraties en gesprekken met betrokkenen. Zo'n reconstructie is minder sterk dan een vooraf uitgevoerde nulmeting, maar wel beter dan een besluit dat uitsluitend op enthousiasme berust.

Vergelijk wat werkelijk vergelijkbaar is

Een stijging of daling krijgt pas betekenis als duidelijk is waarmee wordt vergeleken. Een groep gemotiveerde vrijwilligers vergelijken met een volledige studentengroep levert bijvoorbeeld al snel een vertekend beeld op. Hetzelfde geldt wanneer de ene groep een nieuwe opdracht kreeg, meer begeleiding ontving of door andere docenten werd beoordeeld.

Bij de vergelijking moet daarom ten minste worden nagegaan:

- of de studentgroepen voldoende overeenkomen;
- of opdracht, beoordelingscriteria en beschikbare tijd gelijkwaardig waren;
- of docenten en beoordelaars op dezelfde manier te werk gingen;
- of studenten naast de AI-toepassing nog andere ondersteuning kregen;
- of verschillen in studiejaar, voorkennis of deelnamebereidheid de uitkomst kunnen verklaren.

Volledige gelijkheid is in het onderwijs zelden haalbaar. Dat hoeft ook niet. Wel moet zichtbaar zijn welke andere verklaringen mogelijk zijn. Een evaluatie wordt geloofwaardiger wanneer zij onzekerheden benoemt dan wanneer zij een exact effect suggereert dat met de beschikbare gegevens niet kan worden aangetoond.

Ook de gekozen periode telt mee. Een gunstige uitkomst in de eerste weken kan voortkomen uit nieuwsgierigheid, extra aandacht van het projectteam of intensieve ondersteuning door de ontwikkelaars. Pas wanneer het gebruik onderdeel wordt van het gewone onderwijs, blijkt of studenten blijven oefenen, docenten de werkwijze kunnen volhouden en de kwaliteit niet terugloopt.

Verschillende gegevens beantwoorden verschillende vragen

Niet ieder gegeven levert hetzelfde soort bewijs. De volgende bronnen vullen elkaar aan, maar mogen niet als onderling uitwisselbaar worden behandeld.

Informatiebron	Wat deze zichtbaar maakt	Wat daarmee nog niet is aangetoond
Persoonlijke waarnemingen	Wat een docent, ontwikkelaar of manager zag gebeuren	Dat het waargenomen patroon algemeen of blijvend is
Ervaringen van studenten en medewerkers	Bruikbaarheid, ervaren steun, vertrouwen en knelpunten	Dat studenten daadwerkelijk beter leren of medewerkers structureel tijd winnen
Gebruiksgegevens	Hoe vaak, hoe lang en op welke momenten de toepassing is gebruikt	Dat intensief gebruik ook doelmatig of leerzaam gebruik is
Leeruitkomsten	Verandering in kennis, producten, prestaties of professioneel handelen	Dat de verandering door de AI-toepassing is veroorzaakt
Vergelijkend onderzoek	Verschillen tussen situaties, groepen of perioden	Dat de uitkomst zonder meer naar andere opleidingen kan worden overgedragen

Een docent kan bijvoorbeeld waarnemen dat studenten beter voorbereid naar een werkcollege komen. Studenten kunnen aangeven dat zij dankzij directe terugkoppeling eerder fouten herkennen. De gebruiksgegevens kunnen vervolgens laten zien dat zij gemiddeld vier oefenronden doorlopen in plaats van één. Dat geheel vormt een aannemelijk patroon, maar de beslissende vraag blijft of de kwaliteit van hun handelen ook is verbeterd.

Daarvoor zijn leeruitkomsten nodig die aansluiten bij het oorspronkelijke onderwijsdoel. Bij schrijfonderwijs kan het gaan om de kwaliteit van argumentatie en revisie, niet alleen om de snelheid waarmee een tekst tot stand komt. Bij gespreksvoering telt hoe de student luistert, doorvraagt en professioneel reageert, niet

hoeveel gesprekken zijn afgerond. Bij programmeren gaat het niet alleen om werkende code, maar ook om het vermogen fouten te analyseren, ontwerpkeuzes uit te leggen en oplossingen zelfstandig aan te passen.

Sneller produceren is dus hooguit een tussenuitkomst. Soms is die snelheid waardevol, bijvoorbeeld wanneer studenten daardoor meer oefenronden kunnen doorlopen. Maar dan moet zichtbaar worden dat die extra rondes leiden tot beter handelen. Anders heeft de toepassing vooral de productie versneld.

Zoek naar samenhang, niet naar één overtuigend cijfer

Een goede evaluatie steunt meestal op meerdere gegevensbronnen. Dat voorkomt dat één opvallend percentage het besluit gaat beheersen. Hoge tevredenheid kan samengaan met beperkte leerwinst. Meer gebruik kan wijzen op betrokkenheid, maar ook op afhankelijkheid of een onduidelijke taak. Een hogere toetsscore kan positief zijn, terwijl uit een mondelinge toelichting blijkt dat studenten hun keuzes niet zelfstandig kunnen verantwoorden.

Juist de samenhang is informatief. Als studenten vaker oefenen, hun producten aantoonbaar verbeteren en docenten waarnemen dat zij bij een zelfstandige beroepsopdracht sterker handelen, groeit het vertrouwen dat de toepassing onderwijswaarde heeft. Lopen de gegevens uiteen, dan is nader onderzoek nodig. Dat verschil is geen hinderlijke ruis die moet worden weggepoetst, maar een aanwijzing dat de werking nog niet goed wordt begrepen.

Leg daarom vooraf vast welke uitkomst voldoende is om door te gaan. Zonder zo'n grens wordt iedere verbetering achteraf al snel als succes uitgelegd. Een opleiding kan bijvoorbeeld bepalen dat de toepassing pas wordt voortgezet wanneer een substantieel deel van de studenten extra oefent, de kwaliteit op vooraf gekozen criteria verbetert en de docentbelasting binnen een aanvaardbare bandbreedte blijft. De precieze grens verschilt per toepassing en beroepsrisico, maar moet wel vóór het definitieve besluit bespreekbaar zijn.

De uitkomst van deze vergelijking is geen eenvoudig rapportcijfer. Zij laat zien voor welke studenten de toepassing werkt, bij welke leerhandelingen, onder welke begeleiding en met welk bewijsniveau. Ook wordt zichtbaar waar nog onzekerheid bestaat. Daarmee kan de opleiding het besluit nauwkeuriger formuleren: niet “de bot werkt”, maar bijvoorbeeld “de toepassing helpt tweedejaarsstudenten om vaker en gericht te oefenen met gespreksstructuur, terwijl nog niet is aangetoond dat zij in onverwachte praktijksituaties zelfstandiger handelen”.

Pas daarna kan de bestuurlijke afweging worden voltooid. Want zelfs een aantoonbaar beter leerresultaat rechtvaardigt niet vanzelf voortzetting of opschaling. Tegenover de opbrengst staat de volledige rekening van ontwikkeling, scholing, toezicht, beheer, toegankelijkheid en kwaliteitsbewaking.

10.2 De volledige rekening

De volledige rekening bestaat uit meer dan licentiekosten en ontwikkeluren. Juist bij AI-toepassingen ontstaat een aanzienlijk deel van de lasten ná de eerste ingebruikname: medewerkers moeten leren hoe zij de toepassing verantwoord inzetten, uitkomsten moeten worden gecontroleerd, bronnen en instructies moeten actueel blijven en incidenten en vragen moeten worden afgehandeld. Wie alleen de kosten van de pilot telt, beoordeelt dus niet de toepassing die de opleiding straks werkelijk moet beheren.

Een bruikbare raming kijkt daarom naar de hele levensduur: van ontwerp en invoering tot dagelijks gebruik, onderhoud en uiteindelijk vervanging of beëindiging. Daarbij gaat het niet alleen om geld. Ook schaarse uren van docenten, onderwijsontwikkelaars, informatiemanagers, functioneel beheerders en privacy- en beveiligingsspecialisten horen op de rekening. Een toepassing kan financieel betaalbaar lijken en toch onhaalbaar zijn omdat zij structureel een beroep doet op mensen die daarvoor geen ruimte hebben.

Kostencategorie	Wat moet worden meegerekend?	Veelvoorkomende onderschatting
Ontwikkeling en inrichting	ontwerp, testen, koppelingen, bronselectie, instructies en aanpassingen	alleen de bouwuren van de eerste versie tellen
Licenties en gebruik	abonnementen, verbruikskosten, opslag en prijsstijgingen	rekenen met pilotprijzen of beperkt gebruik
Scholing en invoering	training, oefentijd, handleidingen en begeleiding van teams	alleen een eenmalige introductiebijeenkomst opnemen
Toezicht en kwaliteitscontrole	steekproeven, beoordeling van uitkomsten, incidentafhandeling en herziening	controlewerk als tijdelijke opstartlast behandelen
Gegevensbeheer	toestemming, bewaartermijnen, toegangsbeheer, verwijdering en controle van gegevensstromen	aannemen dat de leverancier alles afdoende regelt
Toegankelijkheid en ondersteuning	alternatieve werkwijzen, gebruikersondersteuning en voorzieningen voor studenten met beperkingen	uitgaan van één standaardgebruiker

Onderhoud en beheer	actualiseren van bronnen, testen na wijzigingen, accountbeheer en rapportage	verwachten dat een werkende toepassing vanzelf blijft werken
Leveranciersafhankelijkheid	migratie, uitvoer van gegevens, contractbeheer en vervangingskosten	alleen naar de huidige prijs en functionaliteit kijken
Taakherontwerp	nieuwe rollen, werkafspraken, afstemming en overdracht tussen mens en systeem	tijdwinst tellen zonder het nieuwe werk zichtbaar te maken

Vooraf dat laatste punt verdient aandacht. AI verwijdert zelden een complete taak. Meestal verandert zij de verdeling van het werk. Een bot kan bijvoorbeeld eerste feedback op een conceptproduct geven, waardoor studenten sneller een nieuwe versie maken. Dat lijkt docentenwerk te besparen. Tegelijk moeten docenten de feedbackcriteria ontwerpen, controleren of de bot geen onjuiste aanwijzingen geeft, studenten leren hoe zij de feedback beoordelen en ingrijpen bij twijfelgevallen. Bovendien kunnen studenten vaker werk inleveren, juist doordat de feedbackcyclus korter wordt. Het oude nakijkwerk neemt dan af, maar er ontstaat meer ontwerp-, controle- en begeleidingswerk.

Dat hoeft geen bezwaar te zijn. Het kan zelfs precies de bedoeling zijn wanneer routinematige correctie plaatsmaakt voor gesprekken over lastige afwegingen en professionele groei. Maar noem het dan geen eenvoudige besparing. De relevante vraag is welke tijd vrijkomt, welk nieuw werk ervoor in de plaats komt en of die nieuwe taakverdeling aantoonbaar meer waarde oplevert. Tijdwinst rendeert pas wanneer de vrijgekomen ruimte ook werkelijk wordt gebruikt voor begeleiding, onderwijsontwerp of deskundig oordeel.

Hetzelfde geldt voor scholing. Een korte demonstratie is zelden voldoende. Medewerkers moeten niet alleen weten op welke knop zij drukken, maar ook wanneer zij een uitkomst niet mogen vertrouwen, welke gegevens zij wel en niet kunnen invoeren en hoe zij studenten begeleiden bij verantwoord gebruik. Nieuwe medewerkers moeten later dezelfde scholing krijgen. Verandert de toepassing of het onderliggende model, dan moet een deel van de instructie opnieuw worden gegeven. Scholing is daarmee geen eenmalige invoeringspost, maar voor een deel een terugkerende beheerlast.

Ook kwaliteitscontrole blijft nodig. AI-uitkomsten kunnen veranderen doordat bronnen worden bijgewerkt, instellingen worden aangepast of een leverancier een nieuw model invoert. Een toepassing die bij de pilot goed presteerde, doet dat niet vanzelf een jaar later nog steeds. De opleiding moet daarom bepalen wie periodiek test, met welke voorbeelden dat gebeurt, welke afwijkingen aanvaardbaar zijn en wie de toepassing kan stilzetten. Bij een oefenbot met beperkte gevolgen kan een steekproef volstaan. Bij gebruik in studieadvies, beoordeling of risicovolle beroepssituaties ligt de vereiste controle veel hoger.

Toegankelijkheid hoort eveneens vanaf het begin in de raming. Een spraaktoepassing kan voor sommige studenten een drempel verlagen, maar voor anderen juist een nieuwe barrière vormen. Denk aan studenten met een auditieve beperking, spraakproblemen, beperkte digitale vaardigheden of behoefte aan een prikkelarme leeromgeving. Als een alternatief nodig is, moet dat alternatief worden ontworpen, getest en ondersteund. Een toepassing is niet doelmatig wanneer de besparing bij de ene groep leidt tot extra handwerk of uitsluiting bij een andere.

Daar komen leverancierskeuzes bij. Een lage instapprijs kan aantrekkelijk zijn, terwijl later kosten ontstaan door stijgend gebruik, beperkte mogelijkheden om gegevens uit te voeren, verplichte koppelingen of een moeizame overstap. Vraag daarom niet alleen wat de leverancier nu rekent, maar ook wat er gebeurt wanneer het aantal gebruikers vertienvoudigt, de instelling haar gegevens wil meenemen, functionaliteit verdwijnt of het contract wordt beëindigd. Afhankelijkheid is niet per definitie verkeerd, maar moet wel zichtbaar en bestuurbaar zijn.

Een volledige rekening hoeft overigens niet tot schijnprecisie te leiden. Niet iedere post kan vooraf exact in euro's worden uitgedrukt. Werk daarom met bandbreedten en benoem onzekerheden. Maak in elk geval onderscheid tussen eenmalige invoeringskosten, terugkerende vaste lasten en kosten die met het gebruik toenemen. Voeg daaraan de benodigde uren per betrokken functie toe. Zo wordt zichtbaar of opschaling werkelijk schaalvoordeel oplevert of vooral meer controle, ondersteuning en beheer veroorzaakt.

De kernvraag luidt dan niet: *Is de AI-toepassing goedkoper dan de huidige werkwijze?* Een betere vraag is: *Wegen de aantoonbare baten voor studenten, medewerkers en de instelling op tegen alle financiële lasten, het nieuwe werk, de risico's en de afhankelijkheden?* Soms is een toepassing duurder, maar toch verdedigbaar omdat studenten aantoonbaar beter leren of omdat de toegankelijkheid verbetert. Soms levert zij uren op, maar is voortzetting onverstandig omdat het leerresultaat niet verbetert of de controlelast te groot wordt.

Pas wanneer zowel de opbrengsten als deze volledige rekening zichtbaar zijn, kan er bestuurlijk een eerlijke keuze worden gemaakt. Die keuze hoeft niet uit te komen op een eenvoudig ja of nee. Het bewijs en de verhouding tussen baten, lasten en risico's kunnen immers verschillende besluiten rechtvaardigen: stoppen, aanpassen, voortzetten of opschalen.

10.3 Vier besluiten, niet één succesverhaal

Daarmee ontstaat geen rangorde van slecht naar goed. **Stoppen, aanpassen, voortzetten en opschalen zijn vier volwaardige besluiten**, elk passend bij een andere combinatie van bewijs, baten, lasten en risico's. Opschalen is dus niet vanzelf het gewenste eindpunt en stoppen is geen bestuurlijke nederlaag. De kwaliteit van de besluitvorming blijkt juist uit de mate waarin het gekozen besluit aansluit bij wat

werkelijk is vastgesteld.

De beslismatrix

De matrix steunt op twee vragen. Ten eerste: hoe overtuigend is het bewijs dat de beoogde waarde ontstaat? Ten tweede: hoe gunstig is de verhouding tussen die waarde en de volledige lasten en risico's?

Bij het bewijs gaat het niet alleen om tevredenheid of gebruik. Een toepassing kan populair zijn zonder dat studenten er beter door leren. Andersom kan een toepassing aantoonbaar bijdragen aan professionele bekwaamheid, terwijl studenten haar aanvankelijk lastig vinden. Daarom moet vooraf duidelijk zijn welke uitkomst telt: meer gerichte oefening, betere prestaties, kortere feedbackcycli, meer toegankelijkheid, bruikbare professionele ruimte of een andere concreet omschreven opbrengst.

Ook de tweede vraag vraagt om meer dan een financiële rekensom. Een gunstige verhouding betekent dat de baten opwegen tegen licenties, ontwikkeling, ondersteuning, toezicht, beheer en risico's. Daarbij telt bovendien of de opbrengsten terechtkomen bij degene voor wie ze bedoeld waren. Tijdwinst voor een docent heeft bijvoorbeeld weinig betekenis als die tijd volledig opgaat aan foutcontrole of technische ondersteuning.

Besluit	Stand van het bewijs	Verhouding tussen baten, lasten en risico's	Bestuurlijke betekenis
Stoppen	De beoogde waarde treedt niet of onvoldoende op, of kan niet geloofwaardig worden aangetoond	De lasten of risico's zijn te groot in verhouding tot de opbrengst	Middelen, aandacht en capaciteit vrijmaken voor een kansrijker vraagstuk
Aanpassen	Er zijn aanwijzingen voor waarde, maar het bewijs is nog beperkt of de toepassing werkt slechts voor een deel van de doelgroep	De verhouding kan gunstiger worden door herontwerp, betere afbakening of ander toezicht	Gericht veranderen en opnieuw meten
Voortzetten	De waarde is binnen de bestaande context voldoende aannemelijk of aangetoond	De balans is voor het huidige bereik verantwoord, maar schaalbaarheid is nog niet bewezen	Gecontroleerd blijven gebruiken binnen duidelijke grenzen

Opschalen	De baten zijn overtuigend aangetoond en blijven onder verschillende omstandigheden voldoende overeind	De volledige balans is gunstig en de risico's zijn beheersbaar	Het bereik doelbewust vergroten en structurele verantwoordelijkheid organiseren
------------------	---	--	---

Stoppen voorkomt dat een zwakke toepassing blijft hangen

Stoppen ligt voor de hand wanneer de toepassing het oorspronkelijke vraagstuk niet oplost, de onderwijswaarde gering blijkt of de bijkomende lasten niet te rechtvaardigen zijn. Dat geldt ook wanneer een toepassing technisch goed functioneert. Technische werking is immers geen bewijs van onderwijsrendement.

Een stopbesluit kan eveneens nodig zijn als de risico's niet voldoende beheersbaar zijn. Denk aan onduidelijk auteurschap, onvoldoende betrouwbare feedback, problematisch gegevensgebruik, uitsluiting van bepaalde studenten of een afhankelijkheid van een leverancier die niet past bij de publieke verantwoordelijkheid van de instelling. Een toepassing hoeft dan niet eerst schade te veroorzaken voordat beëindiging verdedigbaar is.

Leg bij stoppen vast waarom het experiment eindigt en wat ervan is geleerd. Daarmee voorkomt u dat dezelfde toepassing enkele maanden later onder een andere naam opnieuw begint. Behoud alleen wat bruikbaar is, zoals een aangescherpte probleemdefinitie, een meetinstrument of inzicht in de benodigde begeleiding.

Aanpassen is een nieuw ontwerpbesluit

Aanpassen is passend wanneer er wel waarde zichtbaar is, maar nog niet in de gewenste vorm, omvang of verhouding. Misschien oefenen studenten vaker, maar te oppervlakkig. Misschien krijgen zij sneller feedback, maar verwerken zij die nauwelijks. Of docenten winnen tijd bij het opstellen van oefenmateriaal, terwijl de controle ervan meer werk vraagt dan verwacht.

De aanpassing moet dan gericht zijn op de oorzaak van het probleem. Dat kan betekenen dat de doelgroep smaller wordt gekozen, de kennisbron beter wordt afgebakend, het gebruiksmoment verandert of de docent meer zicht krijgt op de interacties en resultaten. Soms moet niet de toepassing, maar het onderwijsproces eromheen opnieuw worden ontworpen.

Maak van aanpassen geen automatische verlenging van de pilot. Benoem vooraf welke verandering wordt aangebracht, welk effect daarvan wordt verwacht en wanneer opnieuw wordt besloten. Zonder zo'n afbakening kan een experiment jarenlang in een tussentoestand blijven, met terugkerende kosten maar zonder duidelijke opbrengst.

Voortzetten is begrensd vertrouwen

Voortzetten past bij een toepassing die binnen de huidige context voldoende waarde oplevert, terwijl er nog geen grond is om het bereik te vergroten. De toepassing kan bijvoorbeeld goed werken binnen één cursus, met een ervaren docent en een overzichtelijke studentengroep. Dat rechtvaardigt gebruik in die omgeving, maar nog niet invoering bij andere opleidingen.

Dit besluit vraagt daarom om grenzen. Voor welke doelgroep blijft de toepassing beschikbaar? Wie bewaakt de kwaliteit? Welke gebruiksgegevens en leerresultaten worden gevolgd? Wanneer volgt een nieuwe beoordeling? Voortzetten betekent niet dat de toepassing voortaan buiten beeld mag raken. Het is een besluit om gecontroleerd door te gaan zolang de vastgestelde voorwaarden gelden.

Juist deze uitkomst voorkomt dat een geslaagde lokale toepassing te snel als instellingsbrede oplossing wordt behandeld. Wat onder begeleiding van een betrokken ontwikkelteam werkt, hoeft immers niet stand te houden zodra het gebruik onderdeel wordt van de gewone bedrijfsvoering.

Opschalen vraagt ander bewijs dan een pilot

Opschalen is pas verdedigbaar wanneer niet alleen de toepassing, maar ook de werkwijze eromheen bestand is tegen een groter bereik. De opbrengsten moeten voldoende standhouden bij andere docenten, studentgroepen en onderwijscontexten. Bovendien moeten ondersteuning, gegevensbeheer, toegankelijkheid, kwaliteitscontrole, scholing en verantwoordelijkheden structureel kunnen worden georganiseerd.

Daarmee verandert ook de aard van het besluit. Een pilot kan draaien op betrokkenheid, improvisatie en tijdelijk projectgeld. Bij opschaling zijn eigenaarschap, budget, beheerafspraken en medezeggenschap nodig. Ook moet duidelijk zijn wat er gebeurt wanneer de leverancier voorwaarden wijzigt, de prestaties teruglopen of nieuwe risico's ontstaan.

Opschalen betekent trouwens niet altijd instellingsbreed invoeren. Het kan ook gaan om uitbreiding naar enkele vergelijkbare cursussen, opleidingen of studentgroepen. Gefaseerde opschaling maakt het mogelijk om opnieuw te meten voordat de instelling zich verder vastlegt.

Leg ook de grens voor heroverweging vast

Welk besluit ook wordt genomen, noteer ten minste de onderbouwing, de belangrijkste onzekerheden, de verantwoordelijke eigenaar en het moment van herbeoordeling. Voeg daar concrete stop- of bijstelcriteria aan toe. Bijvoorbeeld: de toepassing wordt herzien als de controlelast boven een afgesproken niveau komt, als een bepaalde groep studenten structureel minder profiteert of als de kwaliteit van de feedback aantoonbaar terugloopt.

Zo wordt de matrix geen afvinklijst, maar een manier om onzekerheid bestuurbaar te maken. Niet elk experiment hoeft een succesverhaal op te leveren. Wel moet elk

experiment eindigen in een verdedigbaar besluit – ook wanneer dat besluit luidt dat de toepassing niet langer de moeite waard is.

Besluit met heroverwegingsgrens



10.4 Uit mijn onderwijspraktijk: het experiment dat moest stoppen

Dat besluit hebben wij zelf ook een keer moeten nemen. Juist bij een toepassing die op het eerste gezicht geslaagd leek.

We hadden een onderwijsbot ontwikkeld waarmee studenten snel uitleg, voorbeelden en suggesties voor opdrachten konden opvragen. De bot was gebaseerd op het cursusmateriaal en sloot aan bij de begrippen die we in de lessen gebruikten. Studenten hoefden daardoor niet in verschillende documenten te zoeken en konden ook buiten de contacturen vragen stellen. De eerste reacties waren positief. De toepassing werd veel gebruikt en studenten gaven aan dat zij de lage drempel prettig vonden.

Aanvankelijk zagen we dat als een sterk signaal. Wie een toepassing vrijwillig en herhaaldelijk gebruikt, zal er toch wel iets aan hebben. Maar gebruik is nog geen leeropbrengst. Toen we beter keken naar wat studenten met de bot deden, ontstond een ander beeld.

Veel vragen waren gericht op een direct bruikbaar antwoord. Studenten vroegen om een samenvatting, een voorbeeld of een mogelijke uitwerking en namen het resultaat

vervolgens over in hun opdracht. Ze hoefden nauwelijks te analyseren, af te wegen of zelf een redenering op te bouwen. De bot hielp hen dus wel verder met het product, maar daagde hen onvoldoende uit om de onderliggende professionele bekwaamheid te ontwikkelen.

Dat verschil werd vooral zichtbaar tijdens gesprekken met studenten. Wanneer een docent doorvroeg waarom een bepaalde keuze was gemaakt, konden studenten de formulering uit hun uitwerking soms wel herhalen, maar niet altijd onderbouwen. De bot had een hindernis weggenomen, terwijl die hindernis juist een nuttig onderdeel van het leren had moeten zijn.

Dat betekende niet dat iedere student er verkeerd mee werkte. Sommigen gebruikten de uitleg als opstapje en controleerden de uitkomsten zorgvuldig. Anderen stelden vervolgvragen en vergeleken verschillende mogelijkheden. Maar de toepassing zelf lokte dat gedrag niet uit. Ze maakte de kortste route naar een antwoord wel erg aantrekkelijk.

Ook voor de medewerkers viel de rekening hoger uit dan verwacht. Het cursusmateriaal moest worden bijgewerkt, antwoorden moesten steekproefsgewijs worden gecontroleerd en onjuiste of onduidelijke reacties vroegen om aanpassingen. Wanneer de cursus veranderde, veranderde de bot niet vanzelf mee. Bovendien kwamen er vragen van studenten over verschillen tussen wat de docent had gezegd en wat de bot antwoordde.

Het beheer bestond daardoor niet uit één technische handeling, maar uit een reeks kleine onderwijskundige en organisatorische taken. Iemand moest bepalen welke bronnen geldig waren, wijzigingen verwerken, meldingen onderzoeken en nagaan of correcties het gewenste effect hadden. Docenten wonnen bij het beantwoorden van eenvoudige vragen enige tijd, maar leverden een deel daarvan weer in bij controle, onderhoud en herstel.

Tijdens de pilot kon dat nog worden opgevangen. Een klein team kende de toepassing goed en loste problemen snel op. Bij uitbreiding naar meer cursussen zou die werkwijze niet houdbaar zijn geweest. Iedere nieuwe cursus zou immers eigen bronnen, controles, verantwoordelijken en onderhoudsmomenten vragen. Wat in de pilot een paar extra handelingen waren, dreigde bij opschaling een structurele beheerfunctie te worden.

We hebben daarom de oorspronkelijke vraag er opnieuw bij gepakt. Het doel was niet dat studenten vaker met een bot zouden werken. Het doel was dat zij zelfstandig begrippen leerden toepassen, keuzes konden onderbouwen en betere beroepsproducten maakten. Voor die verandering zagen we te weinig aanwijzingen. Het gebruik was hoog, de waardering was overwegend positief, maar de bijdrage aan het beoogde leren bleef onzeker. Daar stonden aantoonbare beheerlast en het risico op oppervlakkige verwerking tegenover.

De conclusie was ongemakkelijk maar helder: deze toepassing moest in deze vorm stoppen.

Dat besluit riep aanvankelijk weerstand op. Een populaire voorziening beëindigen voelt al snel als het terugdraaien van innovatie. Ook was er al tijd in de ontwikkeling gestoken. Juist die investering kan een valkuil worden. Wie vooral kijkt naar wat er al is gedaan, zoekt gemakkelijk argumenten om door te gaan. Bestuurlijk telt echter niet hoeveel moeite een toepassing heeft gekost, maar of verdere inzet nog te rechtvaardigen is.

We hebben de bot daarom niet stilzwijgend laten uitdoven. We legden vast waarom we stopten: te weinig zichtbare leerwaarde, een te passieve leerhandeling en een beheerlast die bij uitbreiding sterk zou toenemen. Ook benoemden we wat wél bruikbaar was gebleken. Studenten waardeerden de beschikbaarheid buiten de les, de aansluiting op het cursusmateriaal en de mogelijkheid om zonder gêne een eenvoudige vraag te stellen.

Die onderdelen namen we mee naar volgende toepassingen, maar met een ander ontwerp. Nieuwe bots gaven minder snel een volledig antwoord. Ze stelden eerst een tegenvraag, lieten studenten een keuze toelichten of boden een oefensituatie waarin de student zelf moest handelen. De aandacht verschoof daarmee van informatie krijgen naar actief oefenen, feedback verwerken en verantwoorden.

Achteraf was het stoppen minstens zo belangrijk als de experimenten die we voortzetten. Het dwong ons om onderwijsbots niet langer als afzonderlijke hulpmiddelen te beoordelen, maar als onderdelen van een samenhangend onderwijsconcept. Iedere toepassing moest een herkenbare leerhandeling ondersteunen en bijdragen aan de opbouw van professionele bekwaamheid. Gebruiksgemak en enthousiasme bleven relevant, maar waren niet langer doorslaggevend.

Deze tiende praktijkstap leverde dan ook geen nieuw product op, maar een strengere ontwerpregel: een toepassing die studenten vooral helpt sneller tot een antwoord te komen, verdient geen vaste plaats alleen omdat zij populair is. Pas wanneer duidelijk wordt wat studenten daardoor beter leren doen, welke menselijke begeleiding nodig blijft en welke volledige rekening daarbij hoort, ontstaat een verdedigbare basis voor voortzetting.

Stoppen was daarmee geen mislukking van het ontwikkeltraject. Het was het gevolg van serieus opnieuw diagnosticeren. De oorspronkelijke bedoeling, het feitelijke gebruik, de waargenomen leerwaarde en de beheerlast wezen niet dezelfde kant op. Precies dan moet een experiment eindigen – niet met een succesverhaal, maar met een besluit.

Het besluit om te stoppen



10.5 Van proefresultaat naar besluit

Geen nieuwe diagnose zonder nieuwe situatie

Na het besluit begint niet automatisch weer een volledig diagnosetraject. Wanneer dezelfde toepassing voor dezelfde doelgroep, in dezelfde onderwijscontext en met hetzelfde doel wordt voortgezet, volstaat gerichte monitoring. De instelling houdt dan bij of de verwachte baten standhouden, of risico's veranderen en of de beheerlast binnen de afgesproken grenzen blijft.

Een nieuwe diagnose is wél nodig wanneer een van drie zaken wezenlijk verandert:

- **de doelgroep**, bijvoorbeeld wanneer een toepassing van derdejaarsstudenten naar eerstejaarsstudenten gaat;
- **de context**, bijvoorbeeld wanneer een oefenbot ook bij formele toetsing wordt ingezet;
- **de doelstelling**, bijvoorbeeld wanneer een hulpmiddel voor extra oefening voortaan docenttijd moet besparen.

In zulke gevallen is eerder succes geen voldoende onderbouwing. Andere studenten kunnen meer uitleg nodig hebben, een toetscontext stelt zwaardere eisen aan betrouwbaarheid en auteurschap en een besparingsdoel vraagt om andere metingen dan een leerdoel. Opschaling is dan feitelijk een nieuwe toepassing, ook als de techniek gelijk blijft.

11 De kennisbron bepaalt de grens

Een besluit tot voortzetting of opschaling maakt een AI-toepassing niet vanzelf betrouwbaar. Vanaf dat moment wordt het juist belangrijker te weten waarop haar antwoorden zijn gebaseerd. Een overtuigend geformuleerd antwoord kan immers onjuist, verouderd of ongeschikt voor de opleiding zijn – zonder dat een student dat aan de formulering merkt.

Daarmee komt een andere bestuurlijke vraag op tafel: welke kennis mag binnen de instelling als gezaghebbend gelden, en wie houdt die kennis actueel? Dat is geen eenmalige technische inrichting. Onderwijsregelingen veranderen, cursusmateriaal wordt herzien en beroepsstandaarden verschuiven. Wat vandaag klopt, kan morgen ongemerkt achterhaald zijn.

De grens van een toepassing wordt daarom mede bepaald door haar kennisbron en door de afspraken rond het beheer daarvan. Die grens raakt ook aan leveranciers: waar worden gegevens verwerkt, wie kan wijzigingen doorvoeren en behoudt de instelling voldoende zeggenschap? Wie AI structureel inzet, verbindt zich niet alleen aan een hulpmiddel, maar ook aan een kennisvoorziening die voortdurend onderhoud, controle en eigenaarschap vraagt.

Actualiseren van de kennisbron



11.1 Van leercyclus naar ontwerpkeuze

Bij een toepassing die de pilotfase voorbij is, dringt zich onvermijdelijk een nieuwe vraag op: waarop mogen studenten en medewerkers eigenlijk vertrouwen? Zolang een onderwijsbot vooral wordt gebruikt om ideeën te verkennen of extra oefenvragen te maken, lijkt die vraag misschien minder dringend. Zodra dezelfde toepassing uitleg geeft over een vak, procedures toelicht of studenten voorbereidt op professioneel handelen, bepaalt de kennisbron echter rechtstreeks wat de toepassing verantwoord kan doen.

Daarmee verschuift de aandacht van de leercyclus naar het ontwerp. Niet alleen de vorm van de interactie telt, maar vooral de grond waarop een antwoord rust.

Een goed geformuleerd antwoord kan toch onjuist zijn

Generatieve AI formuleert antwoorden op basis van patronen in de gegevens waarmee het model is ontwikkeld. Het model reconstrueert daarbij niet eenvoudigweg een gecontroleerd feitenbestand. Het berekent welk vervolg waarschijnlijk past bij de vraag en de voorgaande tekst. Daardoor kunnen antwoorden helder, volledig en overtuigend klinken, terwijl de inhoud onjuist, verouderd of niet van toepassing is op de opleiding.

Dat verschil is bestuurlijk belangrijk. Een student kan aan een vloeiend antwoord immers niet zien hoe sterk de onderbouwing ervan is. De toon blijft vaak even stellig wanneer het model een algemeen bekend gegeven weergeeft, een redelijke veronderstelling doet of een feit verzint.

Ook de vraag van de gebruiker beïnvloedt het antwoord. Een kleine verandering in formulering kan tot een andere uitleg, andere voorbeelden of zelfs een andere conclusie leiden. Vrij formuleren biedt daardoor veel ruimte voor verkenning, maar weinig zekerheid over herhaalbaarheid en herkomst.

Dat maakt AI met algemene modelkennis niet waardeloos. Voor brainstormen, het verkennen van perspectieven, taalondersteuning en het bedenken van oefensituaties kan een brede kennisbasis juist nuttig zijn. Alleen past daarbij een bescheiden claim: de toepassing helpt de gebruiker bij het denken en produceren, maar geldt niet vanzelf als gezaghebbende bron.

Geselecteerde bronnen verkleinen de ruimte

Een opleiding kan de toepassing ook laten werken met een afgebakende verzameling bronnen, zoals cursusmateriaal, beleidsdocumenten, boeken, beroepsstandaarden of opleidingscompetenties. Bij een vraag zoekt het systeem dan eerst relevante passages in die verzameling en gebruikt het die om een antwoord op te bouwen.

Die werkwijze verandert de aard van de toepassing. De bot wordt niet alwetend, maar juist doelbewust beperkt. Hij hoeft niet te antwoorden vanuit alles wat het taalmodel mogelijk heeft geleerd, maar kan zich richten op materiaal dat de opleiding heeft geselecteerd.

Dat heeft drie mogelijke voordelen.

Ten eerste sluiten antwoorden beter aan op de gekozen terminologie, leerdoelen en context. Een eerstejaarsstudent krijgt dan geen algemeen betoog over projectmanagement, maar uitleg die aansluit bij de methode en begrippen die binnen de opleiding worden gebruikt.

Ten tweede kan de herkomst van uitspraken zichtbaarder worden. Wanneer een antwoord naar een concrete passage verwijst, kan de student controleren waarop het is gebaseerd. Zo'n bronverwijzing is overigens pas waardevol als die werkelijk naar de gebruikte informatie leidt. Een lijstje bronnen onder een antwoord is geen bewijs wanneer onduidelijk blijft welke bewering uit welke bron komt.

Ten derde wordt correctie beter uitvoerbaar. Als de uitleg niet klopt omdat een opleidingsdocument verouderd is, kan de instelling dat document aanpassen. Dat is beheersbaarder dan hopen dat een leverancier ooit het onderliggende taalmodel bijwerkt.

Afbakening is geen garantie

Werken met geselecteerde bronnen lost het betrouwbaarheidsvraagstuk niet volledig op. De toepassing kan een relevante passage missen, verschillende passages verkeerd combineren of meer zekerheid suggereren dan de bron toelaat. Ook kan het systeem terugvallen op algemene modelkennis wanneer de geselecteerde documenten geen antwoord bevatten.

Bovendien kan de bronverzameling zelf gebrekkig zijn. Een cursusomgeving bevat bijvoorbeeld oude presentaties, voorlopige opdrachten, tegenstrijdige instructies en documenten waarvan niemand meer weet of ze nog gelden. Door zulke informatie aan een bot te koppelen, wordt de bron niet vanzelf beter. De toepassing ontsluit dan vooral bestaande onduidelijkheid – alleen sneller en overtuigender.

De juiste ontwerpvrage is daarom niet: *Kunnen we onze documenten aan AI koppelen?* De vraag luidt: *Welke kennis mag deze toepassing gebruiken, voor welk doel en met welke mate van zekerheid?*

Daarbij helpt een eenvoudige indeling:

Gebruik	Passende kennisbasis	Verantwoorde bewering
Ideeën verkennen of voorbeelden bedenken	Brede modelkennis	Mogelijke perspectieven en voorstellen
Oefenen met vakinhoud	Geselecteerd lesmateriaal, aangevuld met duidelijke onzekerheidsmeldingen	Uitleg en oefening binnen de gekozen afbakening

Informatie geven over een opleiding of instelling	Actuele, beheerde instellingsbronnen	Controleerbare informatie met bronverwijzing
Ondersteunen van risicovol professioneel handelen	Gevalideerde vakbronnen en deskundige controle	Vorbereidende ondersteuning, geen zelfstandig eindoordeel
Formele beoordeling of besluitvorming	Vastgestelde criteria, controleerbare bewijsstukken en menselijke verantwoordelijkheid	Hoogstens advies of signalering binnen een gereguleerd proces

De laatste kolom is doorslaggevend. De gekozen kennisbron bepaalt niet alleen welke antwoorden waarschijnlijk beter worden, maar ook welke belofte de instelling aan gebruikers mag doen.

De bronkeuze bepaalt de businesscase

Een toepassing met algemene modelkennis kan snel en tegen beperkte aanvangskosten worden beproefd. Zodra de instelling betrouwbare, contextgebonden antwoorden verlangt, ontstaat echter nieuw werk. Documenten moeten worden geselecteerd, opgeschoond, ontsloten en bijgewerkt. Bronverwijzingen moeten worden getest. Er moet worden bepaald wat het systeem doet wanneer bronnen elkaar tegenspreken of wanneer de bronnen geen antwoord bieden.

Daarmee verschuift ook het mogelijke rendement. Studenten hoeven wellicht minder lang naar de juiste uitleg te zoeken en docenten krijgen minder herhaalde basisvragen. Daartegenover staan uren voor inhoudelijk beheer, technische inrichting, kwaliteitscontrole en foutafhandeling. Alleen wanneer die volledige lasten worden meegeteld, ontstaat een geloofwaardige businesscase.

Voor managers zijn daarom vier beslisvragen van belang:

1. **Welke uitspraken moet de toepassing betrouwbaar kunnen doen?**
2. **Welke bronnen zijn daarvoor gezaghebbend, actueel en toegestaan?**
3. **Hoe kan de gebruiker de herkomst en onzekerheid van een antwoord controleren?**
4. **Welke menselijke controle is nodig als een fout gevolgen heeft voor leren, beoordeling of professioneel handelen?**

Wie deze vragen niet kan beantwoorden, heeft nog geen beheerde kennisvoorziening, maar een taalmodel met documenten eromheen. Dat kan een bruikbaar experiment zijn, alleen rechtvaardigt het nog geen structurele afhankelijkheid.

Zodra geselecteerde bronnen de basis vormen voor onderwijs, begeleiding of bedrijfsvoering, moet bovendien duidelijk zijn wie bepaalt wat als geldig en actueel geldt. Die verantwoordelijkheid kan niet aan het systeem worden overgelaten. De techniek kan informatie vinden en formuleren; de instelling moet vaststellen welke

kennis zij ermee ontsluit, wie veranderingen controleert en wat er gebeurt wanneer de bron tekortschiet.

11.2 Wie beheert wat als waar geldt

Dat vraagt om eigenaarschap dat verder gaat dan het eenmalig goedkeuren van een verzameling documenten. Een kennisbron verandert immers voortdurend.

Onderwijsregelingen worden aangepast, procedures krijgen een nieuwe versie, beroepsstandaarden verschuiven en docenten verbeteren hun materiaal. Zonder ingericht bronbeheer kan een toepassing daardoor gisteren nog het juiste antwoord geven en vandaag ongemerkt verouderde informatie gebruiken.

De kernvraag is dan ook niet alleen wie de techniek beheert, maar vooral wie bevoegd is om vast te stellen welke kennis geldig is. Die bevoegdheid hoort bij de instelling en moet duidelijk bij mensen en functies worden belegd.

Drie verantwoordelijkheden

Voor een beheerde kennisvoorziening zijn ten minste drie verantwoordelijkheden nodig: vakinhoudelijk eigenaarschap, informatiebeheer en technisch of functioneel beheer. Bij kleine toepassingen kunnen deze rollen door enkele personen worden gecombineerd. De verantwoordelijkheden zelf mogen echter niet door elkaar gaan lopen.

De vakinhoudelijke eigenaar bepaalt welke bronnen inhoudelijk gezaghebbend zijn. Dat kan een opleidingscoördinator, curriculumcommissie, examencommissie, diensthoofd of andere proceseigenaar zijn, afhankelijk van het onderwerp. Deze eigenaar beoordeelt niet elk afzonderlijk antwoord, maar stelt de inhoudelijke grenzen vast: welke documenten tellen mee, welke versie geldt en over welke onderwerpen mag de toepassing geen stellige uitspraak doen?

Bij onderwijsinhoud kan bijvoorbeeld worden vastgelegd dat het vastgestelde cursusmateriaal leidend is, aangevuld met een geselecteerde vakpublicatie. Voor informatie over toetsing kan juist de vastgestelde onderwijs- en examenregeling gelden, terwijl informele instructies van een docent alleen als toelichting worden gebruikt. Zo'n rangorde voorkomt dat twee tegenstrijdige bronnen binnen het systeem stilzwijgend dezelfde status krijgen.

De informatiemanager of informatiebeheerder bewaakt de samenhang rond de bron. Daarbij gaat het onder meer om versiebeheer, toegangsrechten en de verbinding met andere informatiesystemen. Deze rol helpt ook vast te stellen waar de authentieke bron wordt beheerd. Een kopie in een onderwijsbot mag bijvoorbeeld niet ongemerkt een eigen leven gaan leiden naast het oorspronkelijke document in Canvas, SharePoint of een ander instellingssysteem.

De functionele of technische beheerder zorgt ervoor dat de toepassing de

geselecteerde bronnen daadwerkelijk kan vinden en gebruiken. Deze beheerder bewaakt de werking, verwerkt wijzigingen en onderzoekt gemelde fouten. Daar hoort ook bij dat oude informatie tijdig uit de toepassing verdwijnt.

Die verdeling maakt meteen duidelijk waarom bronbeheer geen zuiver technisch vraagstuk is. Een technisch beheerder kan vaststellen dát twee documenten elkaar tegenspreken, maar niet altijd welk document inhoudelijk voorrang krijgt. Andersom kan een docent bepalen wat vakinhoudelijk juist is, maar niet vanzelf beoordelen of een wijziging overal in de technische keten is verwerkt.

Leg de bronketen vast

Een eenvoudige bronregistratie voorkomt veel onduidelijkheid. Daarin staat per bron in ieder geval wie de inhoudelijk eigenaar is, waar de authentieke versie staat, voor welk doel de bron mag worden gebruikt en wanneer die opnieuw moet worden beoordeeld of verwijderd. Zo'n registratie hoeft bij een eerste proef niet omvangrijk te zijn. Een overzicht van één pagina kan volstaan, zolang daarmee maar zichtbaar wordt wie in actie moet komen. De uitgewerkte beheerketen voor wijzigingen, controles en foutafhandeling staat in het addendum.

Vooraf het wijzigingsproces verdient aandacht. Als een toetsprocedure op maandag verandert, is het niet vanzelfsprekend dat een toepassing die wijziging op dinsdag correct verwerkt. Er kan tijd zitten tussen vaststelling, publicatie, technische verwerking en controle. In die periode moet duidelijk zijn welke informatie studenten en medewerkers krijgen. Soms is het verstandiger de toepassing tijdelijk geen antwoord te laten geven dan haar op basis van een mogelijk verouderde bron te laten doorpraten.

Onzekerheid moet zichtbaar worden

Ook met zorgvuldig geselecteerde bronnen blijven er grenzen. Een document kan onduidelijk zijn, een situatie kan niet zijn voorzien of twee regels kunnen op gespannen voet staan. De toepassing moet zulke onzekerheid niet gladstrijken met een overtuigend geformuleerd antwoord.

Daarom hoort bij bronbeheer ook een afspraak over doorverwijzen. Bij vragen met beperkte gevolgen kan de toepassing bijvoorbeeld aangeven welke bron zij heeft gebruikt en de gebruiker uitnodigen die te controleren. Bij vragen over formele beoordeling, studievoortgang, rechten of veiligheidskritisch professioneel handelen kan directe menselijke tussenkomst nodig zijn.

De gebruiker moet bovendien fouten eenvoudig kunnen melden. Zo'n melding mag niet verdwijnen in een algemene postbus, maar moet terechtkomen bij de verantwoordelijke voor beoordeling en herstel. Wanneer een fout gevolgen kan hebben gehad, moet de instelling ook vergelijkbare antwoorden onderzoeken en betrokken gebruikers informeren.

Daarmee ontstaat waardevolle managementinformatie. Niet alleen het aantal gebruikers

telt, maar eveneens het aantal foutmeldingen, de aard van de fouten, de hersteltijd en de hoeveelheid beheerwerk. Een toepassing die veel vragen afvangt maar wekelijks omvangrijke inhoudelijke correcties vraagt, kan minder renderen dan de gebruikscijfers op het eerste gezicht suggereren.

Eigenaarschap vraagt capaciteit

Een naam in een verantwoordelijkhedenmatrix is nog geen werkend beheer. De aangewezen eigenaar moet tijd, deskundigheid en bevoegdheid hebben om bronnen te beoordelen en wijzigingen af te dwingen. Dat werk komt vaak boven op bestaande taken. Juist daardoor wordt de beheerlast in een businesscase gemakkelijk onderschat.

Maak daarom vooraf zichtbaar hoeveel bronnen worden ontsloten, hoe vaak ze veranderen en welke gevolgen een fout kan hebben. Een beperkte bot voor één stabiele cursus vraagt iets anders dan een instellingsbrede voorziening die studenten informeert over toetsregels, inschrijving en studievoortgang. Naarmate het bereik en de mogelijke gevolgen groeien, moeten ook de controle, vervangingsregeling, rapportage en foutafhandeling zwaarder worden ingericht.

Voor de manager volgen daaruit vier concrete besluiten:

- **Benoem de inhoudelijk eigenaar** die bepaalt wat als geldig geldt.
- **Beleg het dagelijkse beheer** van versies, toegang, verwerking en incidenten.
- **Reserveer capaciteit** voor controle, correctie en communicatie met gebruikers.
- **Stel grenzen vast** waarbinnen de toepassing zelfstandig antwoord mag geven.

Pas wanneer deze verantwoordelijkheden zijn geregeld, kan een instelling spreken van een beheerde kennisvoorziening. Een leverancier kan daarbij techniek, opslag of ondersteuning leveren, maar niet het institutionele oordeel overnemen over wat waar, geldig en verantwoord is. Wie dat onderscheid niet contractueel en organisatorisch bewaakt, besteedt niet alleen techniek uit, maar ongemerkt ook een deel van de eigen regie.

De inhoudelijk eigenaar beslist



11.3 Leveranciers kiezen zonder de regie weg te geven

Die regie blijft alleen behouden wanneer zij niet afhankelijk is van goede bedoelingen of mondelinge afspraken. De voorwaarden waaronder een leverancier gegevens verwerkt, modellen wijzigt, kosten verhoogt en ondersteuning beëindigt, moeten vooraf inzichtelijk en contractueel afdwingbaar zijn. Dat geldt zowel voor een gespecialiseerde onderwijsbot als voor een algemene AI-voorziening die via een bestaand softwarepakket de instelling binnenkomt.

De kernvraag is dan ook niet alleen welke leverancier de beste demonstratie geeft. Belangrijker is of de instelling de toepassing gedurende de hele levensduur kan begrijpen, controleren, aanpassen en zo nodig buiten gebruik kan stellen. Een aantrekkelijke proefomgeving kan binnen enkele maanden beschikbaar zijn, terwijl de gevolgen van een verkeerde leverancierskeuze pas jaren later zichtbaar worden.

Maak gegevensgebruik concreet

Algemene toezeggingen over zorgvuldig gegevensgebruik zijn onvoldoende. De instelling moet weten welke gegevens de toepassing verwerkt, waar die worden opgeslagen, hoelang ze worden bewaard en waarvoor de leverancier ze nog meer mag gebruiken. Bij onderwijsbots kunnen ingevoerde gegevens immers bestaan uit vragen van studenten, conceptteksten, beoordelingen, gesprekken, leeractiviteiten en informatie over studievoortgang. Ook wanneer studenten geen naam invullen, kan de combinatie van gegevens soms alsnog naar een persoon herleidbaar zijn.

Minstens zo belangrijk is wat er met ingevoerde gegevens gebeurt nadat het antwoord is gegenereerd. Worden de gegevens gebruikt om een model of dienst verder te ontwikkelen? Kunnen medewerkers van de leverancier ze inzien voor kwaliteitscontrole? Worden andere partijen ingeschakeld voor opslag, analyse of ondersteuning? En kan de instelling aantoonbaar laten verwijderen wat niet meer nodig is?

Daarbij hoort een heldere beperking van het doel. Gegevens die voor begeleiding zijn verzameld, mogen niet ongemerkt worden gebruikt voor profilering, commerciële ontwikkeling of andere analyses. Ook intern moet de instelling terughoudend zijn. Dat informatie technisch beschikbaar is, betekent nog niet dat een opleiding die mag gebruiken om studenten te volgen of te beoordelen.

Beveiliging is meer dan een certificaat

Certificeringen en beveiligingsverklaringen kunnen nuttig zijn, maar zij vervangen geen beoordeling van de feitelijke toepassing. Het maakt bijvoorbeeld verschil of studenten alleen openbaar cursusmateriaal raadplegen of ook vertrouwelijke informatie uit een stage, cliëntgesprek of bedrijfsopdracht invoeren. De beveiliging moet passen bij de gevoeligheid van de gegevens én bij de mogelijke gevolgen van verlies, misbruik of ongewenste openbaarmaking.

Vraag daarom niet alleen of de leverancier veilig werkt, maar ook hoe toegang wordt geregeld, incidenten worden gemeld en gegevens worden gescheiden. Wie kan beheerdersrechten krijgen? Welke handelingen worden vastgelegd? Hoe snel wordt een kwetsbaarheid hersteld? Binnen welke termijn hoort de instelling dat er iets is misgegaan? En kan zij zelf controleren of de afgesproken maatregelen daadwerkelijk worden toegepast?

Voor toepassingen met grotere gevolgen is ook een onafhankelijke beoordeling nodig. Een leveranciersverklaring is nu eenmaal geen bewijs dat alle risico's voor de specifieke onderwijscontext voldoende zijn afgedekt. Privacy-, beveiligings- en informatiemanagementspecialisten moeten daarom vroeg aansluiten, niet pas wanneer de pilot al is gestart en docenten hun onderwijs ervan afhankelijk hebben gemaakt.

Houd bronnen, instellingen en resultaten overdraagbaar

Afhankelijkheid ontstaat niet alleen doordat gegevens bij een leverancier staan. Ook instructies, broncollecties, configuraties, gebruiksgegevens en opgebouwde werkprocessen vertegenwoordigen waarde. Wanneer deze onderdelen niet kunnen worden geëxporteerd, moet de instelling bij een overstap feitelijk opnieuw beginnen.

Overdraagbaarheid vraagt meer dan een knop waarmee losse documenten kunnen worden gedownload. De export moet bruikbaar zijn in een gangbaar formaat en voldoende context bevatten om de voorziening elders te reconstrueren. Denk aan versies van bronnen, ingestelde gedragsregels, autorisaties, correcties, evaluatiegegevens en de koppelingen met andere onderwijssystemen.

Leg bovendien vast wat er bij beëindiging gebeurt. Hoeveel tijd krijgt de instelling om gegevens over te zetten? Welke ondersteuning levert de leverancier daarbij? Wanneer worden kopieën verwijderd? Blijft de dienst tijdens de overgang beschikbaar? Zonder zulke afspraken bestaat een uitstapmogelijkheid alleen op papier en is die in de praktijk onuitvoerbaar.

Een eenvoudige toets helpt: kan de instelling binnen een redelijke termijn overstappen zonder onaanvaardbaar verlies van gegevens, onderwijscontinuïteit of opgebouwde kennis? Is het antwoord onzeker, dan is er sprake van leveranciersafhankelijkheid die als expliciete last in de businesscase thuishoort.

Leg zeggenschap over wijzigingen vast

AI-diensten veranderen sneller dan veel traditionele onderwijssystemen. Een leverancier kan het onderliggende model vervangen, functies toevoegen, antwoordregels aanpassen of beperkingen invoeren. Zo'n wijziging kan de kwaliteit verbeteren, maar ook leiden tot ander gedrag, hogere kosten of nieuwe risico's.

De instelling moet daarom weten welke veranderingen de leverancier zelfstandig mag doorvoeren en wanneer zij vooraf moet worden geïnformeerd of moet instemmen. Bij wezenlijke wijzigingen moet er voldoende tijd zijn om opnieuw te testen. Dat geldt

zeker wanneer een toepassing wordt gebruikt bij studieadvies, feedback, begeleiding of processen die de rechtspositie van studenten raken.

Ook moet duidelijk zijn welke versie in gebruik is en hoe wijzigingen worden vastgelegd. Zonder versiebeheer kan een opleiding nauwelijks verklaren waarom dezelfde vraag op verschillende momenten een ander antwoord opleverde. Evenmin kan zij betrouwbaar onderzoeken of een verandering de kwaliteit heeft verbeterd of verslechterd.

Controle betekent hier niet dat de instelling het model technisch volledig moet kunnen doorgronden. Wel moet zij kunnen vaststellen welke bronnen en instellingen worden gebruikt, welke belangrijke beperkingen bekend zijn, hoe prestaties worden gevolgd en hoe fouten kunnen worden gemeld en hersteld.

Beoordeel de kosten over de hele levensduur

Een lage instapprijs zegt weinig over de uiteindelijke kosten. Tijdens een pilot zijn gebruik, ondersteuning en koppelingen vaak beperkt. Bij opschaling kunnen kosten ontstaan voor licenties, verbruik, opslag, beveiliging, integratie, opleiding van medewerkers, bronbeheer, kwaliteitscontrole en ondersteuning van gebruikers.

Let vooral op prijsmodellen die meebewegen met het aantal vragen, de hoeveelheid verwerkte tekst, het gebruik van rekenkracht of aanvullende functies. Naarmate studenten een toepassing vaker gebruiken – wat onderwijskundig juist wenselijk kan zijn – kunnen de kosten sterk oplopen. Vraag daarom om scenario's voor beperkt, gemiddeld en intensief gebruik en leg vast hoe prijswijzigingen worden aangekondigd.

Neem daarnaast het nieuwe werk mee. Wie controleert de antwoorden, actualiseert bronnen, behandelt incidenten en ondersteunt docenten? Als deze taken buiten de raming blijven, lijkt de voorziening goedkoper dan zij werkelijk is. De juiste vergelijking gaat niet tussen een licentiebedrag en nul euro, maar tussen de volledige lasten en de aantoonbare waarde voor studenten, medewerkers en de instelling.

Aanbesteden met behoud van regie

Bij een aanbesteding verdienen bestuurlijke voorwaarden daarom een volwaardige plaats naast functionaliteit en prijs. Een overtuigende demonstratie mag nooit zwaarder wegen dan aantoonbare afspraken over gegevensgebruik, beveiliging, overdraagbaarheid, wijzigingen en beëindiging.

Laat leveranciers bij voorkeur dezelfde praktijksituaties uitwerken. Vraag hoe zij omgaan met een foutieve bron, een beveiligingsincident, een plotselinge prijsstijging, een modelwijziging en een overstap naar een andere aanbieder. Juist uit zulke scenario's blijkt hoeveel regie de instelling werkelijk houdt.

Een leverancier is pas passend wanneer de instelling niet alleen kan instappen, maar ook kan bijsturen en uitstappen. Voor goedkeuring moeten daarom minstens vier zaken vaststaan:

- de instelling houdt zeggenschap over bronnen, doelen en toegestane vormen van gebruik;
- gegevensgebruik, beveiliging en controle zijn concreet en toetsbaar geregeld;
- kosten en wijzigingen blijven gedurende de looptijd bestuurbaar;
- een overstap of beëindiging is technisch, organisatorisch en financieel uitvoerbaar.

Pas dan ondersteunt de leverancier de onderwijsvoorziening zonder er de feitelijke eigenaar van te worden.

11.4 Besluiten over bron en leverancier

Daarmee is de leverancierskeuze nog niet afgerond. Ook bij een passende leverancier kan de toepassing immers onbetrouwbaar worden wanneer de onderliggende kennisbron veroudert. De kwaliteit van een AI-toepassing wordt niet alleen bepaald bij de aanschaf, maar vooral in de maanden en jaren daarna.

Een goede startcollectie is niet genoeg

Stel dat een opleiding een onderwijsbot ontwikkelt op basis van het eigen cursusmateriaal. Bij de start controleert een groep docenten alle documenten. Verouderde presentaties worden verwijderd, begrippen worden gelijkgetrokken en de bot verwijst in zijn antwoorden naar de gebruikte bronnen. De eerste ervaringen zijn positief: studenten vinden informatie sneller en docenten hoeven veel terugkerende vragen niet meer afzonderlijk te beantwoorden.

Een jaar later is de situatie anders. De opleiding heeft de beoordelingscriteria aangepast, enkele beroepsstandaarden zijn vernieuwd en in de digitale leeromgeving staat inmiddels een nieuwe versie van de studiewijzer. In de broncollectie van de bot staan echter nog oude documenten. Niemand heeft formeel de opdracht gekregen die collectie bij te werken.

De bot blijft ondertussen overtuigende antwoorden geven. Juist daarin schuilt het risico. Een ontbrekend document valt vaak nog wel op, maar verouderde informatie kan volledig en geloofwaardig klinken. Studenten merken het verschil niet altijd en docenten gaan er mogelijk van uit dat een collega het beheer uitvoert. Zo ontstaat een onderwijsvoorziening waarvan iedereen gebruikmaakt, maar waarvoor niemand aantoonbaar verantwoordelijk is.

Dit is geen technisch incident. Het is een bestuurlijk tekort in het bronbeheer.

Bronhouderschap vraagt om een naam en een opdracht

Voor iedere structurele toepassing moet daarom duidelijk zijn wie bronhouder is. Dat hoeft niet één persoon te zijn die alle documenten zelf controleert. Wel moet één functie of organisatieonderdeel verantwoordelijk zijn voor de samenhang, actualiteit en

kwaliteit van de gebruikte collectie.

De bronhouder bepaalt onder meer:

- welke documenten als gezaghebbend gelden;
- wie inhoudelijke wijzigingen mag goedkeuren;
- hoe nieuwe versies worden toegevoegd;
- wanneer oude versies vervallen;
- hoe fouten worden gemeld en hersteld;
- welke bronnen tijdelijk of definitief worden uitgesloten;
- hoe gebruikers kunnen zien waarop een antwoord is gebaseerd.

Daarbij moet de taak werkelijk zijn belegd. Een naam in een projectplan is onvoldoende wanneer er geen tijd, bevoegdheid en ondersteuning beschikbaar zijn. Bronhouderschap brengt terugkerend werk met zich mee, zoals wijzigingen beoordelen, dubbele documenten verwijderen, verwijzingen controleren en gebruikersmeldingen afhandelen. Dat werk hoort in de businesscase thuis.

Voor opleidingsspecifieke bronnen ligt het houderschap meestal bij de opleiding. Bij instellingsbrede bronnen, bijvoorbeeld over regelingen, voorzieningen of personeelsbeleid, kan een centrale dienst verantwoordelijk zijn. Worden beide soorten bronnen in één toepassing gecombineerd, dan is ook een procedure nodig voor tegenstrijdige informatie. De gebruiker mag niet zelf hoeven uitzoeken welke versie bestuurlijk geldig is.

Controle volgt het veranderingsrisico

Niet iedere bron hoeft even vaak te worden gecontroleerd. Een vastgesteld begrippenkader verandert minder snel dan een jaarplanning, toetsregeling of externe beroepsstandaard. Een vaste jaarlijkse controle kan voor de ene collectie ruim voldoende zijn en voor de andere veel te laat komen.

De controlefrequentie moet daarom aansluiten bij drie kenmerken:

1. **De veranderlijkheid van de bron.** Hoe vaak worden inhoud, regels of procedures aangepast?
2. **De gevolgen van een fout.** Wat gebeurt er wanneer een student of medewerker verouderde informatie volgt?
3. **De zichtbaarheid van de fout.** Wordt een afwijking snel herkend of kan die lange tijd geloofwaardig blijven?

Voor bronnen met grote gevolgen is alleen periodieke controle niet genoeg. Daar moet een wijziging in het bronsysteem zo snel mogelijk leiden tot beoordeling van de AI-toepassing. Denk aan toetsregels, veiligheidsvoorschriften, stage-eisen en informatie over rechten en plichten. De verantwoordelijke hoeft niet na iedere wijziging het hele systeem opnieuw te onderzoeken, maar moet wel kunnen vaststellen welke antwoorden en processen daardoor worden geraakt.

Leg bovendien vast wat er gebeurt wanneer de actualiteit niet kan worden gegarandeerd. Soms is tijdelijk uitschakelen verstandiger dan doorgaan met een waarschuwing die nauwelijks wordt gelezen. Een onderwijsvoorziening die op een kritiek punt niet betrouwbaar is, wordt niet verantwoord alleen doordat onderaan het scherm staat dat antwoorden fouten kunnen bevatten.

Ook leverancierswijzigingen raken de bron

De instelling beheert mogelijk de documenten, terwijl de leverancier bepaalt hoe die documenten worden verwerkt, doorzocht en weergegeven. Een wijziging in het model, de zoekmethode of de rangschikking van bronnen kan daardoor de antwoorden veranderen zonder dat de collectie zelf is aangepast.

Daarom moet de leverancier zichtbaar maken welke relevante wijzigingen er plaatsvinden en welke gevolgen die kunnen hebben. De instelling moet kunnen nagaan:

- welke versie van een bron wordt gebruikt;
- wanneer die bron is toegevoegd of gewijzigd;
- hoe snel een correctie doorwerkt;
- of verwijderde informatie daadwerkelijk niet meer wordt gebruikt;
- welke systeemwijzigingen de selectie of weergave van bronnen beïnvloeden;
- hoe de werking na een wijziging kan worden getest;
- of eerdere versies en besluiten reconstrueerbaar blijven.

Ook moet duidelijk zijn wie een fout mag herstellen en hoe lang dat duurt. Wanneer alleen de leverancier correcties kan uitvoeren, ontstaat afhankelijkheid op een punt dat direct de onderwijskwaliteit raakt. Een afgesproken reactietijd helpt, maar eigen controlemogelijkheden blijven nodig. De instelling moet immers kunnen vaststellen dat een correctie niet alleen administratief is verwerkt, maar ook werkelijk in de antwoorden doorwerkt.

Goedkeuren onder voorwaarden

Een pilot kan beginnen met een beperkte, handmatig gecontroleerde broncollectie. Voor structureel gebruik ligt de lat hoger. Dan moet het beheer blijven functioneren wanneer het oorspronkelijke projectteam uiteenvalt, een betrokken docent vertrekt of de leverancier zijn product wijzigt.

Gebruik bij de goedkeuring de volgende controle:

Onderwerp	Controle vraag	Aantoonbaar bewijs
Bronhouderschap	Wie is verantwoordelijk en beschikt over tijd en bevoegdheid?	Vastgelegde opdracht en beschikbare capaciteit
Versies en wijzigingen	Wie bepaalt welke versie geldig is en merkt wijzigingen op?	Versiehistorie en wijzigingsprocedure

Controle en correctie	Hoe snel wordt een fout gemeld, beoordeeld en zichtbaar hersteld?	Controleplanning, meldproces en testresultaat
Leverancier	Kunnen wij wijzigingen, verwerking en correcties zelf controleren?	Contractafspraken en eigen controle mogelijkheden
Continuïteit	Wat gebeurt er wanneer de bronhouder of leverancier uitvalt?	Overdrachts-, beëindigings- en vervangingsplan

Neem daarbij ook de jaarlijkse beheerlast op in de businesscase. Zonder overtuigend bewijs is er nog geen structurele voorziening, maar een tijdelijk werkende pilot. Een betrouwbare kennisbron is namelijk geen bezit dat eenmaal wordt opgeleverd. Die blijft alleen betrouwbaar zolang iemand haar onderhoudt, controleert en zo nodig begrenst.

12 Interactievormen doelgericht vergelijken

Ook een betrouwbare toepassing kan onderwijsinhoudelijk tekortschieten. De bron bepaalt wat een toepassing kan weten, maar de interactie bepaalt wat de student zelf moet doen. Krijgt die ruimte om te analyseren, af te wegen en te formuleren, of neemt de toepassing juist het denkwerk over? Een correct antwoord is immers nog geen bewijs van leren.

Daarmee verschuift de ontwerp vraag van inhoud naar leerhandeling. Tekst, spraak en gestructureerde stappen bieden elk andere mogelijkheden. Een mondeling gesprek kan een beroepssituatie dichter benaderen, terwijl tekst meer tijd geeft om terug te lezen en een redenering te verbeteren. Een vaste reeks vragen kan houvast bieden, maar ook te veel sturen. Vrije interactie komt natuurlijk over, maar maakt niet vanzelf zichtbaar wie de analyse uitvoert.

Voor een opleiding is de aantrekkelijkste vorm daarom niet automatisch de waardevolste. Van belang is welke activiteit de student aantoonbaar verricht, hoe die aansluit bij de beroepspraktijk en welke menselijke begeleiding nodig blijft. Pas daarna komen gebruiksgemak, technische mogelijkheden en kosten in beeld. De kernvraag luidt niet of de toepassing een goed gesprek kan voeren, maar of zij een leerproces organiseert waarin de student zelf leert handelen.

Redenering herzien na teruglezen



12.1 Niet elk gesprek is begeleiding

Een betrouwbare bron is dus noodzakelijk, maar zij zegt nog weinig over de kwaliteit van de interactie. Een toepassing kan inhoudelijk correcte antwoorden geven en toch nauwelijks bijdragen aan het leren. Dat gebeurt vooral wanneer de student een vraag stelt, onmiddellijk een afgerond antwoord ontvangt en vervolgens verdergaat zonder zelf te analyseren, kiezen, formuleren of controleren.

Juist de vrije gespreksvorm maakt dit risico gemakkelijk onzichtbaar. Zo'n gesprek oogt persoonlijk en toegankelijk. De toepassing reageert vriendelijk, past haar formulering aan en stelt soms een vervolgvraag. Daardoor ontstaat al snel de indruk dat de student wordt begeleid. Maar een opeenvolging van vragen en antwoorden is nog geen begeleiding. Daarvoor moet de interactie de student aanzetten tot leerhandelingen die passen bij het onderwijsdoel.

De schijn van een goed gesprek

Stel dat een student een advies moet opstellen voor een organisatievraagstuk. In een vrije vraag-en-antwoordvorm kan hij de situatie beschrijven en vragen: 'Wat is hier de beste aanpak?' De toepassing levert vervolgens een analyse, enkele alternatieven en een keurige aanbeveling. Het resultaat kan bruikbaar zijn, maar het belangrijkste denkwerk is dan door de toepassing verricht.

De student heeft mogelijk wel informatie ingevoerd en het antwoord gelezen, maar nog niet aantoonbaar:

- het vraagstuk afgebakend;
- relevante informatie van bijzaken onderscheiden;
- alternatieven ontwikkeld;
- gevolgen en risico's afgewogen;
- een keuze onderbouwd;
- de kwaliteit van het antwoord gecontroleerd.

Als juist deze handelingen centraal staan in de leeruitkomst, werkt een toegankelijk gesprek averechts. De toepassing verkort dan niet alleen de route naar een antwoord, maar omzeilt een deel van de route waarlangs professionele bekwaamheid moet ontstaan.

Kijk naar wat de student doet

Bij de beoordeling van een interactievorm moet het waarneembare handelen van de student centraal staan. Niet de natuurlijkheid van het gesprek, de snelheid van de reactie of de indruk van persoonlijke aandacht, maar de bijdrage aan de beoogde leerhandeling geeft de doorslag.

Een bruikbare analyse begint met vier vragen:

- **Wat moet de student zelf doen?** Denk aan analyseren, formuleren, beslissen, oefenen, reageren, verbeteren of verantwoorden.
- **Wat neemt de toepassing over?** Maakt zij vooral iets mogelijk, of produceert zij het inhoudelijke resultaat?
- **Wat wordt er zichtbaar vastgelegd?** Is alleen het eindantwoord beschikbaar, of ook de redenering, de tussenstappen en de verwerking van feedback?
- **Wie beoordeelt de kwaliteit?** Kan de student dat zelf, is controle door een docent nodig of moet de interactie worden begrensd?

Deze vragen maken ook het verschil zichtbaar tussen gemak voor de student en waarde voor het onderwijs. Een toepassing die direct een sterke tekst genereert, kan de taak gemakkelijker maken. Een toepassing die eerst argumenten laat ordenen, zwakke plekken laat aanwijzen en daarna gerichte feedback geeft, kan het leren versterken. Het tweede proces vraagt doorgaans meer inspanning van de student – en juist daarin kan de onderwijswaarde liggen.

Dat betekent niet dat vrije interactie geen waarde heeft. Zij kan studenten helpen wanneer zij voorkennis willen activeren, een begrip willen verhelderen, verschillende formuleringen willen onderzoeken of een eerste richting zoeken. Ook kan de lage drempel belangrijk zijn voor studenten die niet snel een vraag aan een docent stellen. De waarde zit dan in de snelle beschikbaarheid en de mogelijkheid om door te vragen.

Begeleiding vraagt om didactische terughoudendheid

Goede begeleiding geeft niet steeds zo snel mogelijk het beste antwoord. Soms moet een begeleider juist informatie achterhouden, een tegenvraag stellen, een onvolledige redenering zichtbaar maken of de student laten kiezen. Dat geldt ook voor een digitale interactie.

Een toepassing die leren ondersteunt, kan bijvoorbeeld eerst vragen:

1. Hoe omschrijft u zelf het probleem?
2. Welke gegevens gebruikt u voor die diagnose?
3. Welke twee verklaringen zijn nog meer mogelijk?
4. Op grond van welk criterium kiest u tussen de alternatieven?
5. Wat zou een vakgenoot tegen uw redenering kunnen inbrengen?

Pas daarna kan de toepassing inhoudelijke feedback, een voorbeeld of aanvullende uitleg geven. De student blijft zo eigenaar van de redenering en moet tussen de reacties door aantoonbaar werk verrichten.

Dat vraagt om meer structuur dan een leeg invoerveld met de uitnodiging ‘Waarmee kan ik u helpen?’. Mogelijke maatregelen zijn een vaste volgorde van stappen, verplichte tussenproducten, keuzevragen met onderbouwing, een begrensd aantal aanwijzingen of een opdracht om het ontvangen antwoord te controleren aan de hand van bronnen en vakcriteria. Bij complexere of risicovollere leerdoelen kan bovendien worden vastgelegd wanneer een docent meekijkt.

Meer structuur is overigens niet automatisch beter. Een te strak gesprek kan studenten door een reeks voorspelbare stappen leiden zonder dat zij nog werkelijk hoeven na te denken. Zij leren dan vooral het patroon van de toepassing te volgen. De relevante vraag is daarom niet hoeveel structuur er is, maar welk gedrag die structuur uitlokt.

Voor managers heeft dit een direct gevolg. Een demonstratie van een soepel gesprek is geen overtuigende onderbouwing voor invoering. Laat bij een pilot niet alleen zien wat de toepassing zegt, maar ook wat de student vóór, tijdens en na iedere reactie moet doen. Meet vervolgens niet uitsluitend gebruik en waardering, maar onderzoek of studenten betere analyses maken, feedback daadwerkelijk verwerken, hun keuzes sterker onderbouwen en uiteindelijk zelfstandiger handelen.

De passende interactievorm volgt daarmee uit het onderwijsdoel. Soms is een open gesprek voldoende. In andere situaties is een stapsgewijze oefening, een rollenspel, een simulatie of een gestructureerde feedbackcyclus nodig. Die vormen kunnen dezelfde onderliggende technologie gebruiken, maar zij vragen iets wezenlijk anders van de student – en leveren daarom ook een andere onderwijswaarde op.

12.2 Interactievormen naast elkaar

De keuze tussen uitleg, oefening, rollenspel, simulatie, feedback en coaching is dus geen kwestie van voorkeur. Iedere vorm organiseert een andere leerhandeling. Bij uitleg probeert de student iets te begrijpen, bij oefening past hij kennis actief toe en bij een simulatie moet hij onder wisselende omstandigheden beslissen en handelen. Wie deze vormen onder de noemer ‘gesprek met een bot’ schaaft, mist precies het verschil waarop de onderwijswaarde berust.

Interactievorm	Wat de student vooral doet	Benodigde structuur	Aandachtspunt voor toegankelijkheid
Uitleg	Vragen stellen, verbanden leggen en begrip controleren	Laag tot gemiddeld; begripsvragen en bronverwijzingen zijn nodig	Tekst geeft leestijd, spraak kan drempels verlagen; taalniveau moet aanpasbaar zijn
Oefening	Ophalen, toepassen, redeneren en fouten herstellen	Hoog; niveau, volgorde, hints en herkansing moeten zijn ontworpen	Korte stappen en directe terugkoppeling helpen, maar mogen niet tot oppervlakkig klikken leiden
Rollenspel	Gesprekken voeren, reageren en professioneel gedrag uitproberen	Gemiddeld tot hoog; rollen, situatie en gespreksdoel moeten vaststaan	Spraak benadert de beroepshandeling, tekst biedt meer denktijd en rust

Simulatie	Beslissen, gevolgen afwegen en handelen in een veranderende situatie	Hoog; scenario, spelregels, vertakkingen en beoordelingscriteria zijn noodzakelijk	De hoeveelheid informatie en tijdsdruk moeten instelbaar zijn
Feedback	Eigen werk vergelijken, tekortkomingen beoordelen en verbeteren	Hoog; criteria, feedbackvolgorde en revisiestappen moeten expliciet zijn	Feedback moet begrijpelijk, doseerbaar en terugvindbaar zijn
Coaching	Doelen formuleren, keuzes onderzoeken, voortgang bewaken en reflecteren	Gemiddeld tot hoog; grenzen en escalatiemomenten zijn onmisbaar	Een voorspelbare werkwijze helpt, maar menselijke begeleiding moet bereikbaar blijven

Uitleg en oefening lijken op elkaar, maar vragen iets anders

Bij uitleg staat betekenisvorming centraal, bij oefening actief ophalen en toepassen. De tabel laat het belangrijkste ontwerpverschil zien: een uitlegvorm moet de student zijn begrip laten controleren, terwijl een oefenvorm hem eerst zelf moet laten handelen. Pas daarna volgt een hint, toelichting of nieuwe poging.

Tekst biedt bij uitleg ruimte om te herlezen en vragen in eigen tempo te formuleren; spraak kan de drempel verlagen, maar is vluchtiger. Bij oefening is vooral de volgorde bepalend. Als de student meteen om de uitwerking kan vragen, verdwijnt het denkwerk waarop de onderwijswaarde berust. Het niveau moet kunnen oplopen en een fout antwoord moet aanleiding geven tot onderzoek en herstel, niet alleen tot het tonen van het juiste antwoord.

Rollenspel en simulatie brengen handelen dichterbij

Bij een rollenspel oefent de student een afgebakende sociale of professionele rol. Spraak ligt voor de hand wanneer tempo, formulering en reageren onder druk deel uitmaken van de beroepshandeling. Tekst kan juist passend zijn in een eerdere leerfase, wanneer de student eerst een gespreksstrategie en de opbouw van argumenten moet leren doorzien.

Een simulatie gaat verder: daarin veranderen omstandigheden door beslissingen van de student. Daarom zijn duidelijke spelregels, geloofwaardige scenario's en beoordelingscriteria nodig. Voor managers is vooral het verschil in ontwerp- en beheerlast relevant. Een rollenspel kan rond één gesprekssituatie worden ingericht, terwijl een betrouwbare simulatie meer inhoudelijke validatie, testen en onderhoud vergt. Die extra investering is alleen gerechtvaardigd als het onderwijsdoel vraagt om handelen in een veranderende situatie.

Feedback vraagt criteria, coaching vraagt grenzen

Bij feedback bepaalt een vaste verbetercyclus de waarde: de student vergelijkt zijn werk met bekende criteria, beoordeelt de terugkoppeling, past zijn werk aan en verantwoordt zijn keuzes. Zonder die stappen blijft algemene feedback gemakkelijk overtuigend klinken zonder dat zichtbaar wordt of de student beter leert.

Coaching is opener van aard, maar mag niet grenzeloos worden. Een coachende interactie kan helpen bij leerdoelen, planning en reflectie, maar kan de professionele verantwoordelijkheid van een docent, studieloopbaanbegeleider of hulpverlener niet overnemen. Daarom moeten onderwerp, duur en doorverwijzing vooraf zijn geregeld. Juist een vriendelijke en natuurlijk klinkende interactie kan verwachtingen oproepen die de toepassing niet kan waarmaken.

Kies ook binnen één toepassing bewust

Eén onderwijsbot kan verschillende interactievormen combineren, maar dat is alleen verstandig wanneer de overgangen herkenbaar zijn. Een toepassing kan bijvoorbeeld eerst uitleg geven, daarna een oefenvraag stellen en ten slotte feedback geven op de redenering. De student moet dan wel weten in welke fase hij zit, wat er van hem wordt verwacht en waarop de reactie is gebaseerd.

Toegankelijkheid moet daarbij niet achteraf als extra voorziening worden toegevoegd. Tekst biedt rust, terugzoekbaarheid en controle over het tempo, maar stelt eisen aan leesvaardigheid. Spraak kan deelname vergemakkelijken, maar brengt ook omgevingsgeluid, problemen met accentherkenning, privacy en vluchtigheid mee. Een sterk ontworpen toepassing biedt daarom waar mogelijk meerdere manieren om informatie te ontvangen en te geven, zonder dat het onderwijsdoel verwatert.

De bestuurlijke toets blijft eenvoudig: maakt de gekozen vorm de beoogde leerhandeling waarschijnlijker en beter zichtbaar? Als het antwoord daarop niet duidelijk is, voegt een vrijer of natuurlijker gesprek weinig toe. Vaak blijkt juist het omgekeerde: door fasen, rollen, criteria en keuzemomenten scherper af te bakenen, ontstaat niet minder maar meer betekenisvolle interactie.

12.3 Ontwerples: minder vrijheid, beter gesprek

Die afbakening bleek in de praktijk belangrijker dan de schijnbare natuurlijkheid van het gesprek. Een onderwijsbot die studenten alle ruimte geeft, kan vriendelijk en soepel reageren, maar levert daardoor nog niet vanzelf een goed leerproces op. Juist wanneer de student zelf mag bepalen waar het gesprek begint, welke informatie hij geeft en wanneer hij naar een conclusie springt, blijft vaak onzichtbaar hoe die conclusie tot stand is gekomen.

Voor leren is dat een wezenlijk probleem. Een docent wil niet alleen het antwoord zien, maar ook de redenering erachter: welke gegevens heeft de student geselecteerd, welke

aannames maakt hij, welke alternatieven heeft hij overwogen en waarop baseert hij zijn uiteindelijke keuze? Een open gesprek kan die stappen overslaan zonder dat dit onmiddellijk opvalt. De bot houdt het gesprek immers wel gaande en vult ontbrekende verbanden soms ongemerkt zelf aan. Het resultaat klinkt dan overtuigender dan de prestatie van de student rechtvaardigt.

Uit mijn onderwijspraktijk

Bij een van mijn onderwijsbots werkte ik met twee interactievormen. In de open vorm konden studenten een praktijksituatie inbrengen, vragen stellen en zelf bepalen welke richting het gesprek opging. In de gestructureerde vorm doorliepen zij vaste fasen: eerst de situatie beschrijven, daarna relevante informatie selecteren, vervolgens mogelijke verklaringen of handelingsopties vergelijken en pas daarna een keuze onderbouwen.

Het verschil zat niet vooral in de inhoud die de bot kon geven, maar in het denkwerk dat de interactievorm zichtbaar maakte. In de open gesprekken zag ik grote verschillen. Sommige studenten gaven veel context, onderzochten meerdere mogelijkheden en vroegen gericht om kritiek. Zij kregen een inhoudelijk bruikbaar gesprek. Andere studenten begonnen met een korte vraag, accepteerden het eerste voorstel en werkten vrijwel direct toe naar een oplossing. De bot reageerde in beide gevallen behulpzaam.

Daardoor leek dezelfde interactievorm voor beide groepen te werken, terwijl de feitelijke leerhandeling sterk verschilde. De ene student onderzocht een vraagstuk uitvoerig, de andere verzamelde vooral formuleringen voor een antwoord. Uit een afgerond gesprek viel niet altijd af te leiden of de student beter had leren analyseren of alleen sneller een aannemelijk resultaat had bereikt. De vrijheid in het gesprek maakte de leeropbrengst dus minder zichtbaar.

In de gestructureerde vorm mocht de bot niet meteen een oplossing geven. In plaats daarvan stelde hij per fase gerichte wedervragen, zoals:

- Welke informatie uit de situatie is volgens u doorslaggevend?
- Welke aanname maakt u bij deze conclusie?
- Welke andere verklaring is mogelijk?
- Wat kan er gebeuren als uw keuze onjuist blijkt?
- Op grond van welk criterium kiest u tussen deze twee opties?

Deze interactievorm leverde minder vrije gesprekken op, maar maakte de tussenstappen beter waarneembaar. Studenten konden niet meer zo gemakkelijk van probleem naar antwoord springen. Zichtbaar werd welke informatie zij selecteerden, waar zij aannames deden en hoe zij alternatieven afwogen. Daarmee ontstond ook meer houvast voor feedback van een docent of medestudent, omdat de reactie niet alleen over het eindproduct ging, maar ook over de gevolgde aanpak.

Dit was geen vergelijkend onderzoek waaruit een algemeen effect kan worden afgeleid. Het was wel een duidelijke ontwerpwaarneming: de open vorm bood ruimte om een eigen route te kiezen, terwijl de gefaseerde vorm beter liet zien wat de student zelf

analyseerde, afwoog en besloot. Daardoor kon de tweede vorm gericht worden geëvalueerd en verbeterd. Welke vorm passend is, hangt dus niet af van hoe natuurlijk het gesprek aanvoelt, maar van het denkwerk dat voor het onderwijsdoel zichtbaar moet worden.

Structuur is een didactische keuze

Vaste fasen hebben alleen waarde wanneer ze voortkomen uit de leerhandeling. Het gaat er niet om elk gesprek in zoveel mogelijk stappen op te delen. Een te strak gesprek kan studenten juist dwingen een kunstmatige route te volgen, terwijl professioneel handelen soms vraagt om terugkeren, heroverwegen of omgaan met onvolledige informatie.

De relevante vraag is daarom niet hoeveel vrijheid een toepassing biedt, maar welke vrijheid de student nodig heeft om de beoogde bekwaamheid te ontwikkelen. Voor een eerste oefening in klinisch redeneren, adviseren, ontwerpen of besluitvorming kan een duidelijke volgorde noodzakelijk zijn. Naarmate studenten meer ervaring opdoen, kan de structuur geleidelijk worden losgelaten. Zij krijgen dan meer ruimte om zelf te bepalen welke informatie nodig is, welke stappen ze zetten en wanneer ze hun keuze herzien.

Zo kan dezelfde toepassing verschillende begeleidingsniveaus ondersteunen:

1. **Sterk gestructureerd:** de bot bepaalt de fasen en stelt per fase gerichte vragen.
2. **Gedeeltelijk gestructureerd:** de student kiest de route, maar moet vaste aandachtspunten behandelen.
3. **Open:** de student voert zelfstandig het gesprek en de bot grijpt alleen in bij ontbrekende onderbouwing of een duidelijke denkfout.

Dat onderscheid voorkomt dat openheid wordt verward met een hoog niveau. Vrij kunnen praten met een bot zegt weinig over professionele zelfstandigheid. Die blijkt pas wanneer een student zelf relevante vragen stelt, onzekerheid herkent, alternatieven afweegt en verantwoordelijkheid neemt voor een besluit.

Minder vrijheid vraagt meer ontwerpwerk

Een gestructureerde interactie ontstaat niet door enkele standaardvragen aan een taalmodel toe te voegen. Er moeten keuzes worden gemaakt over de gewenste denkstappen, de volgorde daarvan en de momenten waarop de bot wel of niet mag helpen. Ook moet duidelijk zijn hoe de toepassing reageert wanneer een student een fase overslaat, irrelevante informatie geeft of vraagt direct het antwoord te geven.

Daarmee verschuift een deel van het werk. Studenten hebben wellicht minder herstelhulp nodig tijdens het oefenen, maar docenten en onderwijsontwikkelaars moeten vooraf meer aandacht besteden aan het gespreksontwerp. Vervolgens moet worden gecontroleerd of de bot zich aan de afgesproken fasen houdt en of de vragen passen bij het niveau van de student. Na wijzigingen in cursusinhoud,

beoordelingscriteria of beroepsstandaarden moet ook de interactie worden bijgewerkt.

De baten van meer structuur moeten daarom worden afgewogen tegen dit extra ontwerp- en beheerwerk. Die investering is vooral te rechtvaardigen wanneer de toepassing vaak wordt gebruikt, studenten er herhaaldelijk mee oefenen en de vastgelegde tussenstappen bruikbare informatie opleveren voor feedback en begeleiding. Voor een eenmalig gesprek kan een uitvoerig gespreksontwerp meer kosten dan het oplevert.

Ook mag structuur geen schijnzekerheid geven. Een student kan alle fasen netjes doorlopen en toch zwak redeneren. De bot kan bovendien een oppervlakkig antwoord ten onrechte accepteren of een relevante, afwijkende aanpak afremmen. Menselijke beoordeling blijft nodig om te bepalen of de gekozen structuur werkelijk overeenkomt met deskundig handelen in het vakgebied.

De ontwerples is dan ook eenvoudig: meer gespreksvrijheid is niet vanzelf studentgericht en minder vrijheid is niet vanzelf beperkend. Een goede interactievorm geeft precies genoeg ruimte om zelfstandig te handelen en precies genoeg structuur om het leren zichtbaar te maken. Pas vanuit die verhouding kan worden bepaald welke vorm de meeste leerwaarde heeft en of het benodigde ontwerp, toezicht en beheer bestuurlijk te verantwoorden zijn.

12.4 Kiezen op grond van leerwaarde

De beste interactievorm is niet de vorm die technisch het meest kan, maar de vorm die studenten aanzet tot de leerhandeling die de opleiding wil versterken. Soms is dat een vrij tekstgesprek, soms een gesproken simulatie en soms juist een strak opgebouwd formulier met keuzes, tussenstappen en gerichte vragen. De beroepscontext, toegankelijkheid en gewenste menselijke inbreng bepalen mede welke vorm passend is. Daarna volgt de bestuurlijke toets: wegen de aantoonbare baten op tegen het ontwerpwerk, de controle en het beheer?

Die volgorde is belangrijk. Zodra een opleiding begint met de vraag of een toepassing tekst, spraak of beeld moet gebruiken, dreigt de vorm leidend te worden. Gebruiksgemak, technische mogelijkheden en aantrekkelijkheid voor studenten zijn relevante eigenschappen, maar ze zeggen nog weinig over de onderwijswaarde. De eerste vraag blijft wat de student door de interactie moet leren doen.

Van passende vorm naar houdbare keuze

Voor de uiteindelijke beslissing volstaan vijf scherpe vragen:

- 1. Welke waarneembare leerhandeling moet deze interactievorm uitlokken?**
- 2. Welke kenmerken van tekst, spraak, beeld of structuur zijn daarvoor noodzakelijk?**
- 3. Welke studenten worden ermee geholpen en voor wie is een gelijkwaardig**

alternatief nodig?

- 4. Welke menselijke beoordeling, correctie en begeleiding blijven onmisbaar?**
- 5. Welke aantoonbare opbrengst rechtvaardigt het terugkerende ontwerp- en beheerwerk?**

De antwoorden moeten niet alleen de gekozen vorm onderbouwen, maar ook duidelijk maken hoe de opleiding de leerwaarde vaststelt. Een positieve studentenwaardering is daarbij nuttig, maar onvoldoende. Studenten kunnen een toepassing prettig vinden omdat zij snel antwoord geeft, weinig weerstand biedt of een opdracht gemakkelijker maakt. Dat bewijst nog niet dat zij beter leren handelen. Waardering moet daarom steeds naast leerresultaten, gebruiksgegevens en professionele beoordeling worden gelegd.

Ook moet zichtbaar zijn welk werk tussen student, docent en toepassing verschuift. Tijdwinst op één plek kan immers nieuw werk elders veroorzaken, zoals het controleren van gespreksroutes, het onderzoeken van fouten en het actualiseren van inhoud. Professionele ruimte ontstaat pas wanneer vrijgekomen tijd werkelijk kan worden ingezet voor begeleiding, ontwerp of beoordeling – niet wanneer zij onmiddellijk wordt gevuld met controle, foutafhandeling en technisch beheer.

Zonder concreet antwoord op de eerste vraag is de keuze nog niet rijp. Kan de noodzakelijke menselijke inbreng niet worden georganiseerd, dan is de toepassing onverantwoord. Bestaat de opbrengst alleen uit enthousiasme, snelheid of technisch bereik, dan ontbreekt voorsnóg een houdbare batenclaim. De passende interactievorm is uiteindelijk de eenvoudigste vorm die de bedoelde leerhandeling geloofwaardig ondersteunt, toegankelijk kan worden aangeboden en onder deskundige menselijke regie beheersbaar blijft.

Vijf vragen vóór de keuze



13 De context bepaalt of een toepassing past

Diezelfde nuchterheid is nodig bij de vraag waar een toepassing wordt ingezet. Een AI-toepassing heeft immers op zichzelf geen waarde. Wat passend is, hangt af van de studenten, de beroepspraktijk, het onderwijsontwerp, de schaal en de gevolgen van fouten. Eenzelfde voorziening kan daardoor in de ene opleiding een wezenlijk knelpunt oplossen en in de andere vooral extra werk veroorzaken.

Dat maakt context tot een bestuurlijke kwestie. Instellingsbrede beschikbaarheid mag niet worden verward met algemene toepasbaarheid. Een aantrekkelijke demonstratie, een enthousiaste kartrekker of een gunstige prijs zegt nog weinig over de bijdrage aan het leren. Daarvoor moet eerst duidelijk zijn welk vraagstuk speelt, voor wie verbetering nodig is en hoe die verbetering zichtbaar kan worden.

Bovendien stopt de afweging niet bij de onderwijsinhoud. Iedere toepassing raakt ook processen, ondersteuning, gegevensbeheer, leveranciersafspraken en continuïteit. Soms ligt daar zelfs de voornaamste waarde, bijvoorbeeld wanneer medewerkers sneller betrouwbare informatie kunnen vinden of administratieve handelingen zorgvuldiger verlopen. Maar ook dan geldt: lokale tijdwinst is nog geen beter proces. De beslissende vraag is waar ruimte voor maatwerk nodig blijft en wat de instelling juist gezamenlijk moet organiseren om AI verantwoord en doelmatig te kunnen gebruiken.

Een overtuigende AI-demonstratie



13.1 Verschillen tussen opleidingen doen ertoe

Diezelfde terughoudendheid is nodig wanneer instellingen toepassingen tussen opleidingen vergelijken. Wat bij de ene opleiding een logische investering is, kan bij een andere opleiding weinig onderwijswaarde hebben of zelfs onverantwoord zijn. Niet omdat de techniek daar slechter werkt, maar omdat studenten, beroepshandelingen, onderwijsontwerp en risico's verschillen.

Een instellingsbrede voorziening kan technisch beschikbaar zijn voor alle opleidingen. Daaruit volgt nog niet dat zij overal op dezelfde manier moet worden gebruikt. De relevante eenheid voor de eerste inhoudelijke afweging is daarom meestal de opleiding, soms zelfs een specifieke fase, leerlijn of onderwijseenheid daarbinnen.

Dezelfde toepassing, een andere waarde

Neem een gespreksbot waarmee studenten een beroepssituatie oefenen. In een zorgopleiding kan zo'n toepassing studenten helpen om slecht nieuws te bespreken, door te vragen of begrijpelijke uitleg te geven. In een commerciële opleiding kan dezelfde techniek worden ingezet voor een advies- of onderhandelingsgesprek. In een technische opleiding ligt de waarde wellicht niet bij het gesprek zelf, maar bij het mondeling toelichten en verdedigen van een ontwerpbeslissing.

Technisch lijken deze toepassingen op elkaar. Onderwijskundig zijn het drie verschillende ontwerpen. De rol van de bot, de beoordelingscriteria, het vereiste realisme, de gevolgen van onjuiste feedback en de benodigde docentbegeleiding lopen uiteen. Ook de batenclaim verschilt. De ene opleiding zoekt vooral meer oefenmogelijkheden, de andere een kortere feedbackcyclus en een derde een betere voorbereiding op een complexe praktijksituatie.

Daarmee veranderen ook de eisen aan bewijs. Een hoog gebruikscijfer kan bij alle drie zichtbaar zijn, maar zegt nog weinig over de bereikte professionele bekwaamheid. Daarvoor moet de opleiding vaststellen of studenten aantoonbaar beter leren luisteren, analyseren, beslissen, uitleggen of verantwoorden. Welk gedrag relevant is, kan alleen vanuit het beroep en het opleidingsprofiel worden bepaald.

Zes verschillen die de keuze bepalen

Bij de beoordeling van een AI-toepassing doen ten minste zes kenmerken van een opleiding ertoe.

Kenmerk	Betekenis voor de AI-keuze
Studentpopulatie	Bepaalt welke ondersteuning, taal, toegankelijkheid en mate van zelfstandigheid nodig zijn.
Beroepspraktijk	Bepaalt welke handelingen realistisch moeten worden geoefend en welke fouten ernstige gevolgen kunnen hebben.

Didactisch concept	Bepaalt of AI aansluit op bijvoorbeeld projectonderwijs, praktijkleren, zelfstudie of intensieve begeleiding.
Toetsing	Bepaalt waar oefenen eindigt en formele beoordeling begint en welk bewijs van eigen bekwaamheid nodig blijft.
Schaal en gebruiksfrequentie	Bepalen of ontwerp-, licentie- en beheerlast over voldoende gebruik kan worden verdeeld.
Docentcapaciteit en risicoprofiel	Bepalen hoeveel controle, correctie, begeleiding en onderhoud werkelijk uitvoerbaar zijn.

De studentpopulatie verdient daarbij meer aandacht dan een gemiddelde instellingsbeschrijving kan bieden. Eerstejaarsstudenten die nog weinig vaktaal beheersen, hebben iets anders nodig dan afstudeerders die zelfstandig beroepskeuzes moeten kunnen verantwoorden. Studenten met uiteenlopende taalvaardigheid kunnen baat hebben bij extra uitleg en oefening, maar lopen ook het risico dat een systeem verschillen maskeert in plaats van verkleint. Een toepassing die veel zelfstandigheid veronderstelt, past bovendien niet vanzelf bij studenten die juist nog moeten leren plannen, bronnen beoordelen en hulp vragen.

Ook de beroepspraktijk stelt grenzen. In sommige opleidingen kunnen studenten relatief veilig variëren, uitproberen en opnieuw beginnen. In andere domeinen moet al tijdens het oefenen scherp worden gelet op veiligheid, vertrouwelijkheid, wettelijke normen of de gevolgen voor cliënten en burgers. Naarmate een beroepshandeling gevoeliger is, moet de opleiding hogere eisen stellen aan de inhoud, de afbakening van het systeem en het menselijke toezicht.

Aansluiten op het didactisch concept

Een toepassing rendeert alleen wanneer zij past in de manier waarop de opleiding het leren organiseert. Een bot voor individuele uitleg kan nuttig zijn binnen een programma met veel zelfstudie. In projectonderwijs kan dezelfde bot juist het gezamenlijke denkproces verarmen wanneer studenten hun vragen afzonderlijk aan het systeem stellen en minder met elkaar overleggen. Daar kan een toepassing die teams helpt om aannames zichtbaar te maken of verschillende oplossingsrichtingen te vergelijken beter passen.

De vraag is dus niet alleen wat de toepassing kan, maar ook welk bestaand leerproces zij verandert. Wordt een leerhandeling versterkt, vervangen of onbedoeld overgeslagen? Ontstaat er meer gerichte oefening, of vooral een snellere route naar een product? Krijgt de docent betere aanknopingspunten voor begeleiding, of verdwijnt juist het zicht op de redenering van de student?

Ook toetsing beïnvloedt de keuze. Een hulpmiddel dat tijdens het leren wenselijk is, is tijdens een formele beoordeling mogelijk niet toegestaan. Omgekeerd kan verantwoord AI-gebruik zelf onderdeel zijn van de vereiste beroepsbekwaamheid. De opleiding moet daarom duidelijk maken wanneer studenten vrij mogen experimenteren, wanneer zij hun gebruik moeten verantwoorden en wanneer zij zelfstandig zonder ondersteuning

moeten handelen. Zonder die afbakening ontstaan verschillen tussen docenten, onduidelijkheid voor studenten en twijfel over de waarde van behaalde resultaten.

Schaal is meer dan studentenaantallen

Een grote opleiding kan ontwikkel- en beheerkosten over veel gebruikers verdelen. Toch levert schaal niet automatisch een betere businesscase op. Wanneer verschillende docententeams eigen instructies, bronnen en beoordelingscriteria gebruiken, kan juist veel afstemming en onderhoud nodig zijn. Een kleine opleiding kan daarentegen een heldere en stabiele toepassing hebben die meerdere jaren intensief wordt gebruikt.

Daarom moeten behalve studentenaantallen ook de gebruiksfrequentie, levensduur en herbruikbaarheid worden meegewogen. Een toepassing die wekelijks terugkomt in een doorlopende leerlijn biedt een ander perspectief dan een hulpmiddel voor één opdracht. Hetzelfde geldt voor een ontwerp dat door meerdere verwante opleidingen kan worden gedeeld, mits de beroepscontext en beoordelingscriteria niet worden platgeslagen tot een algemeen model.

Instellingsbrede schaal ontstaat dan ook niet door één toepassing overal op dezelfde manier uit te rollen. Zij ontstaat wanneer opleidingen gebruik kunnen maken van gedeelde technische voorzieningen, afspraken en deskundigheid, terwijl zij inhoudelijk blijven bepalen welke leerhandeling wordt ondersteund. De instelling kan bijvoorbeeld zorgen voor contractbeheer, privacytoetsing, beveiliging, toegankelijkheid, gebruiksondersteuning en controle op leveranciers. Het opleidingsteam blijft verantwoordelijk voor de vakinhoud, didactische inbedding, begeleiding en beoordeling.

Beschikbare docentcapaciteit is een harde voorwaarde

De kwaliteit van een toepassing hangt uiteindelijk niet alleen af van het ontwerp, maar ook van de mensen die haar moeten gebruiken en onderhouden. Docenten moeten voorbeelden selecteren, instructies aanscherpen, uitkomsten controleren, studenten begeleiden en fouten herstellen. Bij veranderingen in het curriculum of de beroepspraktijk moeten bronnen en opdrachten worden bijgewerkt.

Dat werk verdwijnt niet doordat AI een deel van de interactie overneemt. Soms verschuift het van herhaalde uitleg naar ontwerp en kwaliteitscontrole. Dat kan waardevol zijn, vooral wanneer docenten daardoor meer ruimte krijgen voor complexe begeleiding. Maar die ruimte ontstaat pas als de nieuwe taken expliciet zijn belegd en de totale werkbelasting werkelijk daalt of het werk inhoudelijk verbetert.

Een opleiding die nauwelijks capaciteit heeft voor inhoudelijke controle, kan daarom beter kiezen voor een eenvoudiger en duidelijk afgebakende toepassing dan voor een geavanceerd systeem dat voortdurend toezicht vraagt. Bestuurlijke ambitie kan ontbrekende docenttijd niet compenseren.

De context bepaalt zo niet alleen welke toepassing aantrekkelijk lijkt, maar vooral

welke toepassing uitvoerbaar en verantwoord is. Pas wanneer deze verschillen zichtbaar zijn gemaakt, kunnen opleidingsteams mogelijke toepassingen zinvol naast elkaar zetten. Dan kan de afweging verschuiven van algemene belangstelling naar een gerichte rangschikking: welk vraagstuk is het meest urgent, waar is de onderwijswaarde het grootst, welke opbrengst kan geloofwaardig worden aangetoond en wat kan het team daadwerkelijk dragen?

13.2 Een besliskader per opleiding

Een bruikbaar besliskader maakt die rangschikking expliciet. Het voorkomt dat een opleiding kiest voor de toepassing met de aantrekkelijkste demonstratie, de meest enthousiaste kartrekker of de luidste leverancier. Niet de technische mogelijkheden vormen het vertrekpunt, maar een beperkt aantal concrete vraagstukken waarvoor het team aantoonbare verbetering zoekt.

De eerste stap is daarom niet het verzamelen van AI-ideeën, maar het benoemen van knelpunten. Denk aan studenten die te weinig kunnen oefenen met lastige beroepsgesprekken, feedback die te laat komt om nog te worden verwerkt of docenten die telkens dezelfde basisvragen beantwoorden. Pas daarna wordt bekeken welke toepassing bij het vraagstuk kan passen.

Rangschikken op vier hoofdcriteria

Een opleiding kan mogelijke toepassingen eerst beoordelen op vier hoofdcriteria: urgentie, onderwijswaarde, bewijsbaarheid en uitvoerbaarheid.

Criterium	Kernvraag	Waarop het team let
Urgentie	Hoe belangrijk is het vraagstuk?	Omvang, ernst, betrokken studenten, gevolgen als er niets verandert
Onderwijswaarde	Welke verbetering verwachten we in het leren of handelen?	Meer gerichte oefening, betere feedback, meer zelfstandigheid, sterkere professionele bekwaamheid
Bewijsbaarheid	Kunnen we vaststellen of die verbetering optreedt?	Uitgangsmeting, gebruiksgegevens, kwaliteit van studentwerk, doorlooptijd, ervaringen van studenten en docenten
Uitvoerbaarheid	Kan het opleidingsteam dit verantwoord invoeren en onderhouden?	Docentcapaciteit, beschikbare bronnen, technische ondersteuning, kosten, beheer en verandervermogen

Deze criteria hoeven niet allemaal hetzelfde gewicht te krijgen. Bij een opleiding waarin studenten risicovolle professionele handelingen leren, kunnen betrouwbaarheid en menselijke controle zwaarder wegen. Bij een grote opleiding met veel eerstejaars kan schaal juist belangrijker zijn, mits de toepassing voldoende toegankelijk en beheersbaar

blijft.

Urgentie vraagt om enige discipline. Een vaak gehoorde klacht is nog niet vanzelf het belangrijkste probleem. Het team moet nagaan hoeveel studenten ermee te maken hebben, hoe vaak het voorkomt en wat de gevolgen zijn. Een moeizame administratieve handeling kan veel irritatie veroorzaken, maar weinig effect hebben op het leren. Omgekeerd kan een probleem dat minder zichtbaar is – bijvoorbeeld te weinig oefening met klinisch redeneren of professioneel adviseren – grote gevolgen hebben voor de bekwaamheid waarmee studenten afstuderen.

Ook onderwijswaarde moet concreet worden omschreven. ‘Persoonlijker onderwijs’ of ‘efficiënter leren’ is te algemeen om een besluit op te baseren. Benoem liever welke leerhandeling verandert. Krijgen studenten vaker gelegenheid om een redenering uit te werken? Kunnen zij een gesprek herhalen nadat zij feedback hebben ontvangen? Leren zij betere keuzes te onderbouwen? Of krijgen docenten eerder zicht op hardnekkige misvattingen?

Bewijsbaarheid vooraf ontwerpen

Een toepassing is bestuurlijk sterker wanneer de verwachte opbrengst vóór de pilot toetsbaar wordt gemaakt. Dat betekent niet dat ieder experiment meteen vergelijkend onderzoek vereist. Wel moet vooraf duidelijk zijn welke aanwijzingen voldoende zijn om een volgend besluit te nemen.

Bij een oefenbot kan het team bijvoorbeeld vastleggen hoeveel studenten ermee oefenen, hoe vaak zij een poging herhalen en of hun prestaties bij een daaropvolgende menselijke beoordeling verbeteren. Bij een toepassing voor conceptfeedback kan worden gekeken naar de duur van de feedbackcyclus, de kwaliteit van opeenvolgende versies en de tijd die docenten besteden aan controle en aanvullende begeleiding.

Daarbij horen verschillende soorten bewijs naast elkaar. Gebruiksgegevens laten zien wat studenten doen, maar niet zonder meer wat zij leren. Ervaringen van studenten zeggen iets over bruikbaarheid, maar vormen nog geen bewijs van betere prestaties. Docentwaarnemingen kunnen belangrijke patronen zichtbaar maken, maar blijven mede afhankelijk van verwachtingen en omstandigheden. Juist de combinatie maakt een afweging geloofwaardiger.

Het team moet bovendien vooraf een grens bepalen. Wanneer is de opbrengst groot genoeg om door te gaan? Welke uitkomst leidt tot aanpassing? En bij welke uitkomst stopt de opleiding? Zonder zulke afspraken verandert een pilot gemakkelijk in een toepassing die blijft bestaan omdat er al tijd en geld in is gestoken.

Risico's en procesaansluiting als correctie

Een hoge verwachte onderwijswaarde is niet voldoende. Risico's en de aansluiting op bestaande processen kunnen de rangschikking veranderen of een toepassing uitsluiten. Een eenvoudig scoresysteem helpt om toepassingen te vergelijken, maar mag geen

schijnzekerheid geven. Een hoog totaalcijfer compenseert geen onaanvaardbaar risico.

Daarom werkt het besliskader met twee lagen. Eerst worden de vier hoofdcriteria beoordeeld, bijvoorbeeld op een schaal van één tot vijf. Daarna volgt een afzonderlijke controle op uitsluitingsgronden en randvoorwaarden:

- worden persoonsgegevens of vertrouwelijke praktijkgegevens verwerkt;
- kan onjuiste uitvoer ernstige gevolgen hebben voor studenten of anderen;
- blijft duidelijk wanneer een docent moet ingrijpen;
- past de toepassing binnen toetsbeleid en afspraken over auteurschap;
- zijn bronmateriaal, eigenaarschap en onderhoud geregeld;
- sluit de toepassing aan op de digitale leeromgeving en bestaande ondersteuning;
- is de toepassing toegankelijk voor de beoogde studentengroep;
- kan de opleiding stoppen of overstappen zonder gegevens, leermateriaal of werkwijzen kwijt te raken?

Een toepassing die slecht aansluit op bestaande onderwijsprocessen vraagt vaak meer veranderwerk dan aanvankelijk zichtbaar is. Als studenten voor één leerhandeling naar een afzonderlijk systeem moeten, docenten gegevens handmatig moeten overzetten en ondersteuning bij iedere wijziging opnieuw instructies moet maken, ontstaat er al snel een kwetsbare voorziening. De technische werking kan dan prima zijn, terwijl de toepassing als geheel toch niet rendeert.

Van lijst naar portfolio

Het resultaat van de afweging is geen ranglijst die automatisch bepaalt wat er wordt ingevoerd. Het is een portfolio waarin het opleidingsteam zijn keuzes gericht kan vastleggen. Daarin zijn doorgaans vier categorieën te onderscheiden:

1. **Nu beproeven:** urgent vraagstuk, duidelijke onderwijswaarde, meetbare opbrengst en haalbare invoering.
2. **Eerst voorbereiden:** kansrijke toepassing waarvoor bronnen, processen, expertise of toezicht nog onvoldoende zijn geregeld.
3. **Kleinschalig onderzoeken:** mogelijke waarde, maar nog te weinig zekerheid over gebruik, effect of risico.
4. **Niet doen:** geringe toegevoegde waarde, disproportionele lasten of risico's die niet verantwoord kunnen worden beheerst.

Zo'n portfolio dwingt ook tot beperking. Een opleiding die tien toepassingen tegelijk beproeft, verdeelt haar aandacht, ondersteuning en evaluatie over te veel initiatieven. Twee zorgvuldig gekozen pilots leveren doorgaans meer bestuurlijke kennis op dan tien losse experimenten zonder uitgangsmeting of vervolgcriteria.

De uiteindelijke keuze moet daarom niet alleen antwoord geven op de vraag welke toepassing aantrekkelijk is. Zij moet duidelijk maken welk probleem voorrang krijgt, voor wie waarde moet ontstaan, welk bewijs nodig is en welke capaciteit het team

beschikbaar stelt. Vervolgens komt de organisatorische werkelijkheid in beeld: wie beheert de toepassing, wie controleert de kwaliteit, welke processen veranderen en waar komt het nieuwe werk terecht? Juist daar raakt het onderwijsbesluit aan de bedrijfsvoering.

Twee pilots krijgen voorrang



13.3 Bedrijfsvoering als randvoorwaarde en toepassingsgebied

Bedrijfsvoering is daarbij niet alleen een verzameling randvoorwaarden. Ze kan zelf ook een toepassingsgebied zijn. AI kan bijvoorbeeld helpen bij roostervragen, het ontsluiten van regelingen, het voorbereiden van standaardcommunicatie, het controleren van administratieve gegevens en het ordenen van managementinformatie. Dat kan medewerkers tijd besparen en processen toegankelijker maken. Maar zodra een instelling zo'n toepassing invoert, ontstaan er ook nieuwe taken: bronnen onderhouden, uitkomsten controleren, fouten herstellen, gebruikers ondersteunen en wijzigingen in processen begeleiden.

De zakelijke onderbouwing moet daarom verder gaan dan de vraag of een handeling sneller kan. Van belang is of het hele proces beter wordt, waar werk verdwijnt, waar nieuw werk ontstaat en wat de gevolgen zijn voor studenten, medewerkers en de continuïteit van het onderwijs.

Kijk naar de hele procesketen

Een toepassing kan lokaal efficiënt lijken en toch verderop in het proces extra werk veroorzaken. Denk aan een AI-hulpmiddel dat binnenkomende studentvragen automatisch indeelt en een conceptantwoord opstelt. De behandelaar hoeft dan minder tijd te besteden aan zoeken en formuleren. Als antwoorden echter regelmatig moeten worden gecorrigeerd, studenten vaker terugkomen met vervolgvragen of uitzonderingen niet goed worden herkend, verschuift het werk slechts.

Daarom moet niet alleen de verwerkingssnelheid worden gemeten, maar ook de kwaliteit van het resultaat. Relevante vragen zijn bijvoorbeeld:

- Wordt de totale doorlooptijd korter?
- Neemt het aantal fouten, herstelhandelingen of vervolgvragen af?
- Krijgen studenten tijdig een begrijpelijk en bruikbaar antwoord?
- Kunnen medewerkers uitzonderingen sneller herkennen en afhandelen?
- Blijft duidelijk wie verantwoordelijk is voor de uiteindelijke beslissing?
- Houdt de verbetering stand wanneer het gebruik toeneemt?

Juist de uitzonderingen bepalen vaak de werkelijke beheerlast. Een proef met een kleine, overzichtelijke groep kan goed verlopen doordat ervaren medewerkers fouten snel opmerken. Bij opschaling komen er meer gebruikers, meer verschillende vragen en meer afwijkende situaties. Wat tijdens de pilot informeel werd opgelost, moet dan onderdeel worden van een beheerd proces.

Tijdwinst is pas waarde als zij bruikbaar wordt

Een claim als ‘dit bespaart twintig minuten per medewerker per week’ zegt op zichzelf weinig. Kleine tijdwinsten zijn niet altijd om te zetten in minder personele inzet of meer beschikbare begeleiding. De bespaarde minuten kunnen verspreid over de werkweek vallen, terwijl controle, overleg en foutafhandeling juist geconcentreerde aandacht vragen.

Een geloofwaardige onderbouwing maakt daarom duidelijk:

Vraag	Wat moet zichtbaar worden?
Wie wint tijd?	Een concrete functie of gebruikersgroep
Bij welke taak?	Een afgebakende processtap
Hoe vaak gebeurt dat?	Volume, frequentie en piekbelasting
Welk nieuw werk ontstaat?	Controle, correctie, beheer, scholing en ondersteuning
Waarvoor wordt de ruimte gebruikt?	Bijvoorbeeld studentbegeleiding, analyse of procesverbetering
Wat verandert voor het onderwijs?	Betere bereikbaarheid, minder uitval, snellere ondersteuning of juist nieuwe risico's

Personele gevolgen verdienen daarbij een bredere blik dan formatie alleen. Medewerkers kunnen routinematige taken kwijtraken, maar krijgen er complexer beoordelingswerk voor terug. Dat kan de functie aantrekkelijker maken, mits zij daarvoor voldoende deskundigheid, bevoegdheid en tijd krijgen. Gebeurt dat niet, dan wordt AI een extra controlelaag boven op het bestaande werk.

Het gaat dus om taakherontwerp. Welke handelingen kunnen veilig worden ondersteund, welke vragen moeten direct naar een medewerker en welke beslissingen mogen nooit zonder menselijke beoordeling tot stand komen? Pas wanneer die verdeling expliciet is gemaakt, kan worden beoordeeld of er werkelijk professionele ruimte ontstaat.

Continuïteit vraagt om een terugvalmogelijkheid

Bij toepassingen die roosters, studentinformatie, inschrijving, voortgang of communicatie ondersteunen, telt niet alleen de gemiddelde prestatie. Ook beschikbaarheid is van belang. Een storing vlak voor een inschrijftermijn of tijdens een toetsperiode heeft immers andere gevolgen dan een storing in een vrijblijvende oefenomgeving.

Voor iedere toepassing moet daarom duidelijk zijn wat er gebeurt wanneer het systeem niet beschikbaar is, verkeerde informatie geeft of onverwacht verandert. Dat vraagt om meer dan een technische reservevoorziening. Er moet ook een werkbare terugvalprocedure zijn: medewerkers weten welke bron leidend is, studenten kunnen ergens terecht en noodzakelijke processen kunnen tijdelijk zonder de AI-toepassing doorgaan.

Ook afhankelijkheid van een leverancier hoort in deze afweging. Een toepassing kan goedkoop beginnen, maar kostbaar worden wanneer gebruikstarieven stijgen, koppelingen veranderen of gegevens moeilijk naar een andere omgeving zijn over te brengen. Continuïteit omvat daarom ten minste technisch beheer, functioneel beheer, brononderhoud, contractbeheer, gebruikersondersteuning en periodieke controle van de uitkomsten.

Bedrijfsvoering staat ten dienste van het onderwijs

Efficiëntie in de ondersteuning is geen zelfstandig einddoel. De uiteindelijke vraag is wat studenten en medewerkers ervan merken. Een beter planningsproces kan bijvoorbeeld zorgen voor minder roosterwijzigingen, tijdiger inzetbare begeleiders en een gelijkmatiger studielast. Snellere informatievoorziening kan voorkomen dat studenten deadlines missen of onnodig van loket naar loket worden verwezen. In zulke gevallen ondersteunt een organisatorische verbetering rechtstreeks de toegankelijkheid en kwaliteit van het onderwijs.

Het omgekeerde is ook mogelijk. Een centraal informatiesysteem kan goedkoper zijn, maar studenten opzadelen met algemene antwoorden die onvoldoende aansluiten op

hun opleiding. Automatische planning kan de beschikbare capaciteit efficiënt verdelen, maar didactisch ongewenste roosters opleveren. Een administratieve toepassing kan medewerkers tijd besparen, terwijl studenten minder gemakkelijk een mens bereiken wanneer hun situatie afwijkt.

De zakelijke onderbouwing moet die wisselwerking zichtbaar maken. Neem daarom naast kosten en productiviteit ook proceskwaliteit, toegankelijkheid, foutkans, werkdruk, continuïteit en onderwijs effecten op. Benoem bovendien wie eventuele nadelen draagt. Een besparing bij een centrale dienst is geen instellingswinst wanneer opleidingsteams of studenten het herstelwerk overnemen.

Voor bestuurders volgen daar vier beslisvragen uit:

- 1. Verbetert de volledige procesketen, of slechts één handeling?**
- 2. Leidt tijdswinst aantoonbaar tot bruikbare ruimte of betere dienstverlening?**
- 3. Zijn controle, beheer, ondersteuning en terugvalprocedures structureel belegd?**
- 4. Draagt de verandering bij aan het onderwijs, zonder lasten ongemerkt naar studenten of opleidingsteams te verschuiven?**

Veel van deze voorwaarden hoeven niet door iedere opleiding afzonderlijk te worden georganiseerd. Gegevensbeheer, beveiliging, contracten, technische koppelingen en gebruikersondersteuning lenen zich vaak voor een gemeenschappelijke basis. De keuze welke toepassing didactisch past en welke afweging verantwoord is, blijft echter sterk afhankelijk van de opleiding en haar context.

13.4 Instellingsbrede basis, lokale keuzes

Daarmee ontstaat een tweedeling die voor verantwoorde opschaling essentieel is: organiseer centraal wat iedere opleiding nodig heeft, maar laat de onderwijsinhoudelijke keuze zo dicht mogelijk bij het opleidingsteam. Zonder gemeenschappelijke basis ontstaan versnippering, dubbele kosten en onduidelijke verantwoordelijkheden. Zonder lokale ruimte ontstaat het omgekeerde probleem: één centraal gekozen toepassing wordt gebruikt voor onderwijsvragen waarvoor zij niet geschikt is.

De bestuurlijke opgave is dus niet om alles te centraliseren of alles vrij te laten. Het gaat erom per besluit vast te stellen op welk niveau het thuishoort.

Wat de instelling gemeenschappelijk organiseert

Een opleiding hoeft niet zelf uit te zoeken hoe een leverancier met persoonsgegevens omgaat, welke beveiligingseisen gelden of hoe gegevens na afloop van een contract worden verwijderd. Ook technische koppelingen, toegangsbeheer, contractvoorwaarden, continuïteit en ondersteuning kunnen doorgaans beter instellingsbreed worden geregeld. Dat voorkomt dat ieder team dezelfde expertise moet

opbouwen en afzonderlijk met leveranciers onderhandelt.

Tot de gemeenschappelijke basis behoren in ieder geval:

- eisen aan privacy, informatiebeveiliging en gegevensgebruik;
- afspraken over eigenaarschap, bewaartermijnen en verwijdering van gegevens;
- kaders voor intellectueel eigendom en het gebruik van onderwijs- en studentproducten;
- leveranciersbeoordeling, contractbeheer en afspraken over continuïteit;
- toegangsbeheer en koppelingen met bestaande onderwijs- en informatiesystemen;
- minimale eisen aan toegankelijkheid, uitlegbaarheid en menselijke controle;
- technische ondersteuning, incidentafhandeling en terugvalprocedures;
- registratie van toepassingen, kosten, gebruik en verantwoordelijke eigenaren;
- een heldere procedure voor experimenteren, evalueren en beëindigen.

Door deze voorzieningen gezamenlijk te organiseren, ontstaat niet alleen doelmatigheid. De instelling beschermt er ook studenten en medewerkers mee. Zij hoeven dan niet per cursus opnieuw te achterhalen welke gegevens een toepassing verzamelt, waar die worden verwerkt en bij wie zij terecht kunnen wanneer er iets misgaat.

Een gemeenschappelijke basis maakt bovendien vergelijking mogelijk. Wanneer toepassingen volgens dezelfde definities worden gevolgd, kan de instelling beter beoordelen welke kosten werkelijk worden gemaakt, hoeveel ondersteuning nodig is en welke opbrengsten standhouden. Zonder zo'n basis blijft ieder experiment een afzonderlijk verhaal, met eigen maatstaven en vaak optimistische conclusies.

Wat de opleiding zelf moet kunnen bepalen

De didactische keuze kan niet op dezelfde manier worden gestandaardiseerd. Een opleiding Verpleegkunde stelt andere eisen aan simulaties en professioneel handelen dan een economische, technische of kunstopleiding. Ook binnen één instelling verschillen beroepsrisico's, toetsvormen, studentgroepen en opvattingen over goed onderwijs.

Daarom moet het opleidingsteam kunnen bepalen:

- welk onderwijsprobleem prioriteit heeft;
- welke leerhandeling door AI mag worden ondersteund;
- welke professionele bekwaamheid studenten daarbij ontwikkelen;
- op welk moment studenten zelfstandig oefenen en wanneer begeleiding nodig is;
- welke bronnen inhoudelijk gezaghebbend zijn;
- hoe de docent de uitkomsten bespreekt, controleert of corrigeert;
- welke aanwijzingen nodig zijn om vast te stellen of studenten beter leren;
- wanneer een toepassing niet past, ook al is zij technisch beschikbaar.

Juist dit laatste vraagt bestuurlijke terughoudendheid. Een centraal ingekochte

voorziening wordt al snel een standaardantwoord. Omdat de licentie er toch is, ontstaat druk om de toepassing breed te gebruiken. Maar beschikbaarheid is geen onderwijsargument. De relevante vraag blijft of de voorziening een aantoonbaar vraagstuk oplost voor deze studenten, in deze leerfase en binnen deze beroepscontext.

Lokale ruimte betekent overigens niet dat ieder team naar eigen inzicht ieder hulpmiddel mag invoeren. De instelling bepaalt de ondergrens voor verantwoord gebruik; de opleiding onderbouwt daarbinnen de didactische keuze. Teams krijgen dus geen vrijstelling van gemeenschappelijke eisen, maar wel ruimte om te beslissen of, waar en hoe zij een goedgekeurde voorziening inzetten.

Vier niveaus van besluitvorming

Een bruikbare verdeling bestaat uit vier niveaus:

Niveau	Hoofdverantwoordelijkheid	Voorbeelden van besluiten
Instelling	Gemeenschappelijke kaders en voorzieningen	Beveiliging, privacy, contracten, architectuur, toegankelijkheid en continuïteit
Academie of domein	Afstemming tussen verwante opleidingen	Gezamenlijke kennisbronnen, ondersteuning, deskundigheidsontwikkeling en evaluatie
Opleiding	Didactische en beroepsgerichte inpassing	Leerdoelen, leerhandelingen, begeleidingsniveau, toetsrelatie en bewijs van opbrengsten
Cursus of onderwijseenheid	Uitvoering en bijsturing	Instructies, oefenvormen, docentcontrole, studentvoorlichting en dagelijkse feedback

Deze verdeling voorkomt twee bekende fouten. Bij volledige centralisatie komt het besluit te ver van het onderwijs af te staan. Bij volledige decentralisatie worden risico's, kosten en beheer onderschat en ontstaat een verzameling toepassingen die onderling nauwelijks te ondersteunen is.

Niet ieder besluit past vanzelf in één vakje. Een spraaktoepassing kan bijvoorbeeld centraal worden beoordeeld op beveiliging en toegankelijkheid, terwijl de opleiding vaststelt of gesproken interactie nodig is voor het leerdoel. De docent bepaalt vervolgens hoe studenten ermee oefenen en wanneer hij ingrijpt. Eén toepassing kent zo verschillende beslis-eigenaren, elk met een eigen verantwoordelijkheid.

Standaardiseren zonder didactiek vast te zetten

De gemeenschappelijke basis werkt het best als zij uit normen en herbruikbare voorzieningen bestaat, niet uit een dichtgetimmerde werkwijze. Een instelling kan

bijvoorbeeld voorschrijven dat de gebruikte kennisbron controleerbaar en actueel moet zijn, zonder voor alle opleidingen dezelfde kennisbron te kiezen. Zij kan menselijke controle verplicht stellen bij beslissingen met grote gevolgen, zonder iedere cursus dezelfde controleprocedure op te leggen.

Ook een beperkte verzameling goedgekeurde voorzieningen kan doelmatig zijn. Daarmee vermindert de instelling contractuele versnippering en kan zij ondersteuning concentreren. Zo'n verzameling mag echter geen gesloten catalogus worden die vernieuwing verhindert. Er blijft een eenvoudige, gecontroleerde route nodig om andere toepassingen te beproeven wanneer een opleiding daarvoor een overtuigende onderwijsreden heeft.

Bij die uitzonderingsroute horen wel voorwaarden. Het team beschrijft het vraagstuk, de doelgroep, het beoogde voordeel, het gegevensgebruik, het benodigde toezicht en de wijze van evalueren. Ook wordt vooraf afgesproken wie de proef kan beëindigen en wat er daarna met gegevens, accounts en ontwikkeld materiaal gebeurt. Zo blijft experimenteren mogelijk zonder dat tijdelijke initiatieven ongemerkt structurele voorzieningen worden.

Besturen op samenhang

De kwaliteit van deze inrichting blijkt uiteindelijk uit de samenhang tussen centrale ondersteuning en lokale verantwoordelijkheid. Een instelling die alleen kaders vaststelt, maar geen bruikbare voorzieningen, ondersteuning of deskundigheid aanbiedt, verschuift het werk alsnog naar opleidingen. Een opleiding die alleen gebruikmaakt van centrale voorzieningen, maar geen onderwijsdoel of effectmeting formuleert, besteedt haar didactische verantwoordelijkheid feitelijk uit.

Voor bestuurders zijn daarom vier vragen richtinggevend:

- 1. Welke voorwaarden en voorzieningen zijn voor alle opleidingen gelijk en worden aantoonbaar doelmatiger wanneer de instelling ze gezamenlijk organiseert?**
- 2. Welke keuzes hangen zo sterk samen met leerdoelen, beroepscontext en studentgroep dat het opleidingsteam ze moet maken?**
- 3. Zijn bevoegdheden, kosten en verantwoordelijkheden op ieder niveau expliciet belegd, ook wanneer een toepassing faalt of moet worden beëindigd?**
- 4. Biedt de gemeenschappelijke basis opleidingen daadwerkelijk ruimte om betere keuzes te maken, of dwingt zij vooral uniform gebruik af?**

Een sterke instellingsbrede basis verkleint de risico's en de beheerlast, maar neemt de onderwijsinhoudelijke afweging niet over. Precies daarin ligt de balans: centraal organiseren wat niet telkens opnieuw hoeft, lokaal beslissen wat alleen vanuit het onderwijs goed kan worden beoordeeld.

14 Menselijke begeleiding doelbewust inrichten

Na die lokale afweging volgt een minder zichtbare, maar bepalende ontwerpkeuze: welke rol blijft de mens vervullen? AI kan studenten direct laten oefenen, conceptwerk van terugkoppeling voorzien en medewerkers ondersteunen bij routinetaken. Toch verdwijnt het professionele werk daarmee niet. Het verschuift naar het stellen van kaders, het controleren van uitkomsten, het ondersteunen van gebruikers en het herstellen van fouten.

Juist die verschuiving blijft in plannen en businesscases gemakkelijk buiten beeld. Kortere productietijd wordt dan al snel als tijdwinst geboekt, terwijl docenten nieuwe controle- en beheertaken krijgen. Zelfstandig gebruik wordt als efficiënt beschouwd, zonder eerst te bepalen wat een fout voor de student kan betekenen. En wanneer verantwoordelijkheden breed worden gedeeld, is soms onduidelijk wie daadwerkelijk mag ingrijpen.

Menselijke begeleiding is daarom geen vangnet dat pas nodig is als de techniek tekortschiet. Zij maakt deel uit van het onderwijsontwerp en van de bedrijfsvoering. De kernvraag is niet hoeveel menselijk werk AI kan overnemen, maar welk menselijk oordeel nodig blijft, waar dat in het proces wordt ingezet en wie daarop aanspreekbaar is. Pas wanneer dat expliciet is ingericht, kan vermeende tijdwinst uitgroeien tot werkelijke professionele ruimte.

Docent controleert AI-feedback



14.1 Vier niveaus van menselijke betrokkenheid

Die lokale beslissing betreft niet alleen *of* een opleiding AI inzet, maar vooral hoeveel menselijke betrokkenheid daarbij nodig blijft. Dat begeleidingsniveau kan per toepassing sterk verschillen. Een bot waarmee studenten vrijwillig begrippen oefenen, vraagt immers iets anders dan een toepassing die feedback geeft op beroepshandelingen of informatie verwerkt over studievertraging en persoonlijk welzijn.

Daarom zijn er vier niveaus van menselijke betrokkenheid. Deze vormen geen groeipad waarbij volledig zelfstandig gebruik vanzelf het einddoel is. Het passende niveau hangt af van het onderwijsdoel, de mogelijke gevolgen van fouten, de gevoeligheid van de gegevens en het vermogen van studenten en medewerkers om de uitkomsten kritisch te beoordelen.

Niveau	Gebruik	Rol van de medewerker	Passend wanneer
1. Zelfstandig gebruik	De student of medewerker gebruikt de toepassing zonder elk resultaat te laten controleren	Kaders stellen, instructie geven en beschikbaar zijn bij vragen of incidenten	Fouten weinig gevolgen hebben, gegevens niet gevoelig zijn en de gebruiker uitkomsten zelf kan beoordelen
2. Periodieke controle	De toepassing werkt grotendeels zelfstandig, maar het gebruik en de resultaten worden steekproefsgewijs bekeken	Patronen controleren, afwijkingen onderzoeken en instructies of toepassing bijstellen	Fouten herstelbaar zijn en controle achteraf voldoende bescherming biedt
3. Directe docentcontrole	Een medewerker beoordeelt de relevante uitkomst voordat deze onderwijsinhoudelijke of formele gevolgen krijgt	Goedkeuren, corrigeren, aanvullen of afwijzen	Onjuiste uitkomsten het leren, de begeleiding of een belangrijke beslissing kunnen beïnvloeden
4. Uitsluitend onder begeleiding	De toepassing wordt alleen gebruikt terwijl een deskundige medewerker het proces begeleidt	De interactie sturen, grenzen bewaken en onmiddellijk ingrijpen	De situatie complex of gevoelig is en fouten ernstige of moeilijk herstelbare gevolgen kunnen hebben

Zelfstandig betekent niet zonder verantwoordelijkheid

Op het eerste niveau gebruikt een student de toepassing zelfstandig. Denk aan een oefenbot die extra vragen aanbiedt over behandelde leerstof. Een onjuist antwoord is vervelend, maar meestal snel te herstellen. De student kan de uitleg vergelijken met het cursusmateriaal, een andere bron raadplegen of de docent om opheldering vragen.

Ook dan blijft menselijke verantwoordelijkheid nodig. Het opleidingsteam bepaalt waarvoor de toepassing geschikt is, welke bronnen leidend zijn en hoe studenten onbetrouwbare antwoorden kunnen herkennen en melden. Verder moet duidelijk zijn wie ingrijpt wanneer dezelfde fout vaak terugkomt. Zelfstandig gebruik betekent dus dat niet iedere afzonderlijke interactie wordt gecontroleerd. Het betekent niet dat niemand verantwoordelijk is voor de kwaliteit van de voorziening.

Dit niveau is alleen verantwoord als studenten voldoende voorkennis en kritisch vermogen hebben. Juist beginnende studenten kunnen een overtuigend geformuleerd antwoord moeilijk op waarde schatten. Dezelfde oefenbot kan daardoor bij vierdejaarsstudenten zelfstandig worden gebruikt, terwijl eerstejaarsstudenten meer begeleiding nodig hebben.

Periodiek controleren op patronen

Bij periodieke controle beoordeelt een medewerker niet ieder resultaat, maar bekijkt hij op vaste momenten steekproeven, gebruiksgegevens, meldingen en terugkerende afwijkingen. Dit past bijvoorbeeld bij een toepassing die studenten formatieve feedback geeft op conceptteksten. De student kan daarmee verder, terwijl de docent controleert of de feedback in grote lijnen aansluit bij de beoordelingscriteria en geen stelselmatige missers bevat.

De kracht van dit niveau is dat het zelfstandig oefenen mogelijk maakt zonder de toepassing aan haar lot over te laten. Daar staat een nieuwe beheertaak tegenover. Het team moet afspreken hoe vaak wordt gecontroleerd, hoeveel resultaten in een steekproef vallen, welke afwijkingen aanleiding geven om in te grijpen en wie verbeteringen doorvoert.

Een periodieke controle die alleen op papier bestaat, biedt weinig bescherming. “De docent kijkt af en toe mee” is geen werkbare afspraak. Leg daarom minimaal vast:

- wie de controle uitvoert;
- welke resultaten, meldingen en gebruikspatronen worden bekeken;
- hoe vaak dat gebeurt;
- welke grenswaarden of signalen tot actie leiden;
- hoe fouten worden gecorrigeerd en aan gebruikers worden gemeld;
- wanneer het gebruik tijdelijk wordt beperkt of stilgelegd.

Controleren voordat er gevolgen zijn

Op het derde niveau beoordeelt een docent of andere deskundige de uitkomst voordat

deze gevolgen krijgt voor de student. Dat is nodig wanneer AI bijvoorbeeld een analyse voorbereidt van een beroepsproduct, mogelijke leerproblemen signaleert of een voorstel doet voor persoonlijke begeleiding.

De toepassing kan dan informatie ordenen, aandachtspunten benoemen of een eerste terugkoppeling opstellen. De medewerker blijft degene die de betekenis ervan beoordeelt en beslist wat ermee gebeurt. Een automatisch gegenereerde aanwijzing wordt dus niet rechtstreeks een studieadvies, beoordelingsbesluit of interventie.

Directe controle is meer dan een snelle blik op het scherm. De medewerker moet voldoende tijd, vakkennis en beslissingsruimte hebben om de uitkomst werkelijk te kunnen afwijzen. Als de werkdruk zo hoog is dat medewerkers voorstellen bijna automatisch overnemen, is er formeel wel menselijk toezicht, maar werkt dat in de praktijk niet. De mens wordt dan een doorgeefluik in plaats van een deskundige beslisser.

Alleen inzetten terwijl een deskundige begeleidt

Het vierde niveau is bedoeld voor situaties waarin de interactie zelf begeleiding vraagt. Denk aan een simulatie van een gevoelig cliëntgesprek, het bespreken van psychische of sociale problematiek of het oefenen met beslissingen die in de beroepspraktijk grote gevolgen kunnen hebben. De medewerker bewaakt dan niet alleen de einduitkomst, maar ook het verloop van de interactie.

De begeleider kan ingrijpen wanneer de toepassing een onveilige richting kiest, wanneer de student onvoldoende afstand houdt tot het advies of wanneer de oefensituatie onverwacht emotioneel belastend wordt. Ook kan hij zichtbaar maken waar de AI vereenvoudigt, aannames doet of professioneel relevante informatie mist.

Deze vorm is intensief en levert daarom zelden een eenvoudige besparingsclaim op. De mogelijke opbrengst ligt eerder in rijkere oefensituaties, gerichtere observatie en een beter gesprek over professioneel oordeel. Het rendement moet dan ook op die punten worden vastgesteld.

Gevolgen en gegevens bepalen de ondergrens

Voor de keuze tussen de vier niveaus zijn twee vragen doorslaggevend:

- 1. Wat kan er gebeuren als de toepassing een fout maakt of verkeerd wordt gebruikt?**
- 2. Welke gegevens zijn nodig en hoe schadelijk zijn onjuist gebruik, verlies of ongewenste verspreiding daarvan?**

Hoe groter en moeilijker herstelbaar de gevolgen, hoe intensiever de menselijke betrokkenheid moet zijn. Hetzelfde geldt wanneer de toepassing persoonsgegevens, gezondheidsinformatie, beoordelingen, vertrouwelijke praktijkgegevens of informatie over kwetsbare studenten verwerkt. Bij zulke gegevens is begeleiding overigens niet

voldoende: ook doelbinding, toegangsbeheer, bewaartermijnen, beveiliging en afspraken met leveranciers moeten op orde zijn.

Het gekozen niveau moet bovendien opnieuw worden beoordeeld wanneer de toepassing, doelgroep of onderwijscontext verandert. Een succesvolle proef met gemotiveerde studenten en een betrokken docent bewijst niet dat zelfstandig gebruik op grote schaal verantwoord is. Opschaling kan juist leiden tot minder deskundige gebruikers, meer uitzonderingssituaties en een grotere afstand tussen fouten en degene die ze moet herstellen.

De bestuurlijke vraag is daarom niet hoeveel menselijk toezicht in algemene zin wenselijk is, maar welk niveau per toepassing aantoonbaar nodig en uitvoerbaar is. Daarbij hoort ook een eerlijke raming van het werk: instructie, steekproeven, beoordeling, incidentafhandeling en onderhoud kosten tijd. Pas wanneer die nieuwe taken naast de vervallen taken worden gelegd, wordt zichtbaar of AI werkelijk ruimte vrijmaakt – en waarvoor die ruimte vervolgens kan worden gebruikt.

14.2 Tijdwinst is nog geen vrije ruimte

Een geloofwaardige tijdsraming onderscheidt ten minste vier categorieën:

Taakcategorie	Centrale vraag	Voorbeeld
Vervallen taken	Welk werk hoeft niet meer te worden gedaan?	Iedere week handmatig vergelijkbare oefenvragen opstellen
Verkorte taken	Welk werk blijft bestaan, maar kost minder tijd?	Een eerste opzet voor een casus maken
Veranderde taken	Welk werk krijgt een andere vorm of vereist andere deskundigheid?	Van zelf schrijven naar beoordelen, selecteren en verbeteren
Nieuwe taken	Welk werk ontstaat door het gebruik van AI?	Uitvoer controleren, incidenten afhandelen en instructies onderhouden

Een team moet dus niet alleen vragen hoeveel tijd AI bespaart, maar vooral **welke taken vervallen, welke taken veranderen en welke nieuwe taken ontstaan**. Pas daarna kan het bepalen of er werkelijk ruimte vrijkomt en waarvoor die ruimte wordt gebruikt.

Bruto tijdwinst en netto ruimte

Bruto tijdwinst ontstaat wanneer een bestaande taak sneller wordt uitgevoerd. Een docent die met AI in twintig minuten een eerste reeks oefencasuïstiek maakt waarvoor eerder twee uur nodig was, boekt op die taak een duidelijke winst. Die constatering is nuttig, maar nog onvolledig.

De casussen moeten immers inhoudelijk kloppen, aansluiten bij het niveau van de studenten en vrij zijn van ongewenste vertekeningen. Misschien moeten ze ook worden ingevoerd in de digitale leeromgeving, van toelichting worden voorzien en tijdens het gebruik worden gevolgd. Blijkt een vraag dubbelzinnig, dan moet iemand haar aanpassen en nagaan welke studenten daardoor verkeerd zijn gestuurd.

Netto ruimte is wat overblijft nadat al dat nieuwe werk is meegerekend. Daarbij gaat het niet alleen om de uren, maar ook om het moment waarop ze nodig zijn. Een uur controlewerk in een toch al drukke toetsweek kan zwaarder wegen dan twee uur productiewerk in een rustige voorbereidingsperiode.

Door de vier taakcategorieën expliciet te maken, voorkomt een team dat alleen de opvallende versnelling wordt meegeteld. Juist nieuwe, verspreide taken blijven anders gemakkelijk buiten beeld. Ze worden dan door enthousiaste docenten tussendoor uitgevoerd en verschijnen niet in de begroting, terwijl de toepassing er wel van afhankelijk is.

Vrijgekomen tijd krijgt vooraf een bestemming

Zelfs aantoonbare netto tijdwinst leidt niet vanzelf tot beter onderwijs. Als niet is afgesproken waarvoor de ruimte wordt gebruikt, verdwijnt zij meestal in de bestaande werkdruk. Docenten beantwoorden meer berichten, lossen achterstanden op of krijgen er andere taken bij. Dat kan begrijpelijk zijn, maar het is iets anders dan de beoogde onderwijsopbrengst.

De bestemming van de tijd hoort daarom onderdeel te zijn van het ontwerp en van de businesscase. Een opleiding kan bijvoorbeeld besluiten dat vrijgekomen uren worden ingezet voor:

- extra gesprekken met studenten die vastlopen;
- gerichte bespreking van complexe beroepssituaties;
- verbetering van opdrachten en beoordelingscriteria;
- ontwikkeling van nieuwe oefenvormen;
- afstemming tussen docenten over leeruitkomsten en feedback;
- deskundigheidsontwikkeling rond AI, toetsing of vakinhoud;
- analyse van patronen in studentvragen en leerproblemen.

Daarmee verandert de redenering. De claim luidt dan niet langer: *AI bespaart docenten tijd*. De claim wordt concreter: *AI verkort het maken van eerste oefenmaterialen, waarna docenten de vrijgekomen ruimte gebruiken voor begeleiding van studenten die moeite hebben met het onderbouwen van hun keuzes*.

Zo'n claim is bestuurlijk beter te beoordelen. Er is immers zichtbaar wie voordeel moet hebben, bij welke taak de ruimte ontstaat en welke onderwijsactiviteit ervoor in de plaats komt.

Herontwerp het werk op teamniveau

Taakherontwerp is geen individuele efficiëntieoefening. Wanneer iedere docent zelf bepaalt hoe AI wordt gebruikt, ontstaan al snel dubbele controles, verschillende kwaliteitsnormen en afhankelijkheid van enkele voorlopers. De ene docent bouwt een uitgebreide verzameling instructies op, terwijl een collega hetzelfde werk opnieuw doet. Een derde gebruikt de toepassing zonder controle, omdat onduidelijk is wie daarvoor tijd heeft gekregen.

Het team moet daarom vastleggen welke taken lokaal bij de docent blijven en welke gezamenlijk of centraal worden georganiseerd. Niet iedere docent hoeft bijvoorbeeld dezelfde kennisbron te beheren, dezelfde standaardinstructies te testen of dezelfde veelvoorkomende fouten te onderzoeken. Tegelijk kan inhoudelijke eindcontrole niet zonder meer worden overgedragen aan een centrale dienst die de beroepscontext onvoldoende kent.

Een werkbare verdeling verbindt de taak steeds aan de benodigde deskundigheid. Vakinhoudelijke beoordeling hoort bij vakdeskundigen. Technisch beheer kan elders worden belegd. Onderwijskundigen kunnen helpen bij het ontwerpen van leeractiviteiten en informatiemanagers bij gegevensstromen en toegangsrechten. De manager blijft verantwoordelijk voor voldoende capaciteit, duidelijk eigenaarschap en een realistische verdeling van structureel werk.

Daarbij verdient ook de overgangsfase aandacht. Invoering kost aanvankelijk vaak meer tijd dan de oude werkwijze. Docenten moeten leren werken met de toepassing, materiaal aanpassen en afspraken ontwikkelen. Die tijdelijke investering mag niet worden verward met structurele beheerlast, maar mag evenmin worden genegeerd. Een pilot die alleen slaagt dankzij avonduren van enkele betrokken medewerkers heeft nog geen houdbare businesscase.

Maak ruimte zichtbaar in de planning

Tijdwinst wordt pas professionele ruimte wanneer zij herkenbaar terugkomt in taakafspraken, roosters en prioriteiten. Dat vraagt om drie concrete afspraken.

Ten eerste wordt vooraf een uitgangssituatie vastgesteld: hoeveel tijd kost de bestaande taak, hoe vaak wordt zij uitgevoerd en door wie? Een grove schatting kan voldoende zijn voor een eerste proef, maar bij opschaling is een onderbouwde raming nodig.

Ten tweede worden tijdens de proef zowel vervallen als nieuwe werkzaamheden bijgehouden. Ook kleine taken tellen mee wanneer ze vaak terugkomen, zoals het beantwoorden van gebruikersvragen of het herstellen van ongeschikte uitvoer.

Ten derde besluit het management waaraan de netto ruimte wordt besteed. Zonder zo'n besluit blijft tijdwinst een theoretisch saldo. Met een duidelijke bestemming kan worden vastgesteld of studenten daadwerkelijk meer begeleiding krijgen, het onderwijsontwerp verbetert of medewerkers hun deskundigheid verder ontwikkelen.

De relevante bestuurlijke vraag is dan ook niet of AI een taak sneller uitvoert. De vraag

is of het volledige werkproces zo is heringericht dat medewerkers aantoonbaar meer tijd kunnen besteden aan activiteiten waarin hun professionele oordeel het verschil maakt. Dat wordt vooral zichtbaar wanneer alle stappen rond terugkoppeling, controle en herstel in de berekening worden meegenomen.

14.3 Verborgen werk in de terugkoppelingsketen

Automatische terugkoppeling lijkt eenvoudig: een systeem reageert binnen enkele seconden op een uitwerking waarvoor een docent anders meerdere minuten nodig heeft. Maar rond iedere automatische reactie ontstaat een specifieke keten van inrichting, controle, uitleg, herstel en verbetering.

Juist in die keten gaat veel werk schuil.

De terugkoppeling begint vóór de student iets indient

Goede automatische terugkoppeling vraagt eerst om een heldere inhoudelijke basis. Het systeem moet weten wat de leerdoelen zijn, aan welke criteria een uitwerking moet voldoen en welk niveau van redeneren of handelen van de student wordt verwacht. Die basis moet worden onderhouden wanneer onderwijsmateriaal, opdrachten of beroepsnormen veranderen.

Ook moet de terugkoppeling passen bij het leerdoel, het niveau en het moment in het leerproces. Soms is één gerichte aanwijzing waardevoller dan een volledig overzicht. Dat didactische ontwerp blijft mensenwerk.

Na de automatische reactie begint nieuw werk

Tijdens het gebruik ontstaan nieuwe taken, zoals vragen van studenten beantwoorden, ongeschikte reacties signaleren, fouten herstellen en de inrichting bijstellen. Zulke afwijkingen worden niet altijd spontaan gemeld. Studenten kunnen een reactie immers aannemelijk vinden zonder dat zij de kwaliteit ervan kunnen beoordelen.

Daarom is controle nodig, maar niet iedere automatische reactie hoeft vooraf door een docent te worden goedgekeurd. De intensiteit van het toezicht moet aansluiten bij het doel en het risico. Bij vrijblijvende oefeningen kunnen steekproeven volstaan. Naarmate de terugkoppeling dichter bij formele beoordeling, studievoortgang of risicovol professioneel handelen komt, neemt de vereiste controle toe.

Richt het proces opnieuw in

De opleiding moet de terugkoppelingsketen als geheel ontwerpen in plaats van alleen een bestaande docentaak te automatiseren. Daarvoor moet duidelijk zijn wie verantwoordelijk is voor de inhoudelijke inrichting, de instructie aan studenten, de kwaliteitscontrole, het herstel van fouten en het periodieke onderhoud. Zonder die taakverdeling komt het nieuwe werk al snel terecht bij één enthousiaste docent en verdwijnt een deel van de veronderstelde tijdwinst.

Bereken de nettobaat van de hele keten

Een toepassing rendert wanneer studenten vaker bruikbare terugkoppeling krijgen en de controle-, ondersteunings- en hersteltaken beheersbaar blijven. De bestuurlijke afweging moet daarom niet alleen de bespaarde nakijktijd omvatten, maar ook de structurele inzet voor inrichting, toezicht, gebruikersondersteuning en verbetering.

Daarvoor hoeft de automatische reactie niet volmaakt te zijn. De keten eromheen moet wel fouten kunnen signaleren, gevolgen kunnen beperken en de kwaliteit stelselmatig kunnen verbeteren.

Zodra die keten wordt vastgelegd, komt onvermijdelijk de volgende bestuurlijke vraag op tafel: wie is eigenaar van iedere stap en wie grijpt in wanneer de terugkoppeling tekortschiet?

Steekproef op automatische terugkoppeling

Achter snelle terugkoppeling zit controlewerk



14.4 Besluiten over verantwoordelijkheid

Het antwoord moet per toepassing vooraf vastliggen. Niet in algemene formuleringen als ‘de opleiding en de ondersteunende diensten zijn gezamenlijk verantwoordelijk’, maar in concrete taken, beslisrechten en escalatieroutes. Gedeelde verantwoordelijkheid klinkt zorgvuldig, maar betekent in de praktijk nogal eens dat niemand zich bevoegd voelt om het gebruik stil te leggen, fouten te herstellen of extra middelen beschikbaar te stellen.

Daarom heeft iedere structurele AI-toepassing ten minste één herkenbare eigenaar nodig. Die eigenaar is aanspreekbaar op de werking van de toepassing als geheel en zorgt dat didactiek, techniek, gegevensgebruik en beveiliging met elkaar verbonden blijven. Dat betekent niet dat één persoon alle werkzaamheden uitvoert. Het betekent wel dat iemand bewaakt dat ze worden uitgevoerd en dat knelpunten niet tussen organisatieonderdelen blijven liggen.

Verdeel vier soorten verantwoordelijkheid

Bij onderwijsgerichte AI-toepassingen zijn doorgaans vier verantwoordelijkheidsgebieden te onderscheiden.

Verantwoordelijkheidsgebied	Centrale vraag	Logische eigenaar
Didactiek en onderwijskwaliteit	Draagt de toepassing bij aan de beoogde leerhandeling en blijft de docent in staat professioneel te oordelen?	Opleiding, onder verantwoordelijkheid van de opleidingsmanager
Functioneel beheer en ondersteuning	Werkt de toepassing voor studenten en medewerkers, worden vragen afgehandeld en worden wijzigingen zorgvuldig ingevoerd?	Onderwijs- of ondersteunende dienst
Techniek, gegevens en koppelingen	Zijn gegevensstromen bekend, toegangsrechten passend en koppelingen beheersbaar?	Informatiemanagement en functioneel of technisch beheer
Beveiliging, continuïteit en institutionele kaders	Passen de risico's binnen de grenzen van de instelling en is ingrijpen mogelijk bij ernstige incidenten?	Bestuur of gemandateerde portefeuillehouder, ondersteund door deskundige functies

Deze verdeling is een vertrekpunt, geen universeel organogram. Instellingen verschillen immers in omvang en inrichting. Bij een kleine opleiding kan één medewerker verschillende rollen vervullen; bij een instellingsbrede toepassing zijn vaak meerdere diensten betrokken. In beide gevallen moet duidelijk zijn wie beslist, wie uitvoert, wie adviseert en wie wordt geïnformeerd.

Vooraf het onderscheid tussen inhoudelijk eigenaarschap en technisch beheer is belangrijk. Een beheerder kan controleren of een toepassing beschikbaar is, maar kan niet bepalen of de terugkoppeling didactisch deugt. Omgekeerd kan een docent de

onderwijswaarde beoordelen, maar niet zelfstandig vaststellen of gegevensverwerking, toegangsbeheer en leveranciersafspraken verantwoord zijn. De toepassing is pas bestuurbaar wanneer beide lijnen samenkomen.

Leg ook het recht om in te grijpen vast

Verantwoordelijkheid zonder bevoegdheid werkt niet. Wie fouten moet signaleren, moet daarom ook weten wat er daarna kan gebeuren. Bij onjuiste terugkoppeling kan een docent bijvoorbeeld besluiten de uitkomst niet te gebruiken en studenten daarover informeren. Bij een patroon van ernstige fouten moet de opleiding het gebruik tijdelijk kunnen beperken. Gaat het om een gegevenslek, ongeoorloofde toegang of een storing met grote gevolgen, dan treedt een institutionele incidentprocedure in werking.

Leg voor de invoering ten minste vast:

- wie meldingen van studenten en medewerkers ontvangt;
- wie de ernst van een fout of incident beoordeelt;
- wie onjuiste inhoud corrigeert en betrokken gebruikers informeert;
- wie het gebruik tijdelijk kan beperken of stilleggen;
- wie contact onderhoudt met de leverancier;
- wie bepaalt wanneer de toepassing weer veilig en verantwoord kan worden gebruikt;
- wie registreert wat er is gebeurd en welke aanpassing nodig is.

Daarbij hoeft niet iedere fout langs het bestuur. Een verkeerde uitleg bij een oefenvraag vraagt een andere reactie dan het uitlekken van persoonsgegevens of het op grote schaal geven van misleidende studieadviezen. Juist daarom is een escalatiemodel nodig: kleine afwijkingen worden in het onderwijsproces hersteld, terugkerende tekortkomingen leiden tot herontwerp en ernstige incidenten worden institutioneel behandeld.

Maak scholing onderdeel van het eigenaarschap

Ook scholing krijgt vaak pas aandacht nadat de toepassing beschikbaar is gesteld. Medewerkers ontvangen dan een korte instructie over de bediening, terwijl zij vooral moeten leren wanneer zij de uitkomst wel of niet kunnen gebruiken. Dat vraagt kennis van de onderwijsfunctie, herkenning van onbetrouwbare antwoorden, passende controle en een handelwijze bij twijfel.

Wijs daarom expliciet aan wie verantwoordelijk is voor:

- de instructie van nieuwe gebruikers;
- periodieke bijscholing bij gewijzigde functies of risico's;
- ondersteuning van docenten bij inhoudelijke controle;
- informatie aan studenten over gebruik, grenzen en meldmogelijkheden;
- het vaststellen of medewerkers voldoende toegerust zijn voor hun rol.

Scholing is geen eenmalige voorwaarde bij de ingebruikname. De toepassing verandert, de leverancier past het onderliggende systeem aan en gebruikers vinden nieuwe werkwijzen die niet allemaal waren voorzien. Zonder terugkerende scholing gaat het feitelijke gebruik geleidelijk steeds verder afwijken van het oorspronkelijke ontwerp.

Plan herbeoordeling voordat die nodig lijkt

Een positieve proef vormt geen blijvende rechtvaardiging voor structureel gebruik. Na invoering kunnen de kosten toenemen, kan de kwaliteit teruglopen of kan blijken dat docenten meer controlewerk uitvoeren dan verwacht. Ook kunnen onderwijsdoelen, wetgeving, leveranciersvoorwaarden en institutionele kaders veranderen.

Neem daarom bij het besluit tot invoering meteen een herbeoordelingsmoment op. Leg vast wie dan informatie verzamelt, welke criteria worden gebruikt en wie over voortzetting beslist. Zo'n herbeoordeling kijkt niet alleen naar aantallen gebruikers, maar in ieder geval naar:

- de bereikte onderwijsopbrengst;
- de netto gewonnen of verloren medewerkerstijd;
- fouten, klachten en incidenten;
- de belasting van beheer en ondersteuning;
- de kwaliteit en noodzakelijkheid van menselijke controle;
- wijzigingen in gegevensgebruik en leveranciersvoorwaarden;
- de beschikbaarheid van een werkbaar alternatief of een uitstaproute.

De uitkomst kan zijn dat de toepassing doorgaat, wordt aangepast, beperkter wordt ingezet of stopt. Ook stoppen vraagt eigenaarschap: gegevens moeten zo nodig worden verwijderd of overgedragen, koppelingen afgesloten, gebruikers geïnformeerd en onderwijsprocessen opgevangen.

Structurele invoering begint dus niet met de vraag wie de rekening betaalt, maar met de vraag wie aanspreekbaar blijft nadat het enthousiasme over de proef is verdwenen. Als er geen eigenaar is voor scholing, incidenten, onderhoud en herbeoordeling, is er nog geen structurele toepassing. Er is dan slechts een experiment dat is blijven voortbestaan.

15 Periodiek herbeoordelen en samen leren

Een toepassing die langer in gebruik blijft, wordt nog niet vanzelf een duurzame toepassing. De onderwijscontext verandert, studenten gebruiken het hulpmiddel anders dan verwacht en leveranciers passen modellen, prijzen of voorwaarden aan. Ook binnen de instelling verschuiven verantwoordelijkheden en neemt het aanvankelijke enthousiasme af. Een besluit dat bij de start verdedigbaar was, kan daardoor na verloop van tijd minder goed uitpakken.

Periodieke herbeoordeling is dan ook geen teken van twijfel, maar van behoorlijk bestuur. De kernvraag is niet alleen of de toepassing nog werkt, maar of zij nog steeds voldoende waarde oplevert voor studenten, medewerkers en de instelling – rekening houdend met alle kosten, risico's en beheertaken die inmiddels zichtbaar zijn geworden.

Daarbij ontstaat al snel een tweede vraag. Als een proef bij één opleiding goede resultaten laat zien, wat mag een andere opleiding daar dan uit afleiden? Instellingen willen begrijpelijkerwijs voortbouwen op opgedane ervaring. Toch is een lokaal succes geen algemeen bewijs. Wat overdraagbaar lijkt, kan sterk afhangen van de doelgroep, het onderwijsontwerp, de beroepscontext en de beschikbare begeleiding. Samen leren vraagt daarom om meer dan het delen van succesverhalen: er moet scherp worden vastgelegd wat er precies is geprobeerd, onder welke voorwaarden en met welk bewijs.

De periodieke herbeoordeling



15.1 Een proef levert beperkt bewijs

Een toepassing kan in die ene omgeving goed werken en toch elders weinig opleveren. Een proef levert immers geen bewijs dat ‘AI werkt’, maar kennis over een bepaalde toepassing, voor een bepaalde groep, binnen een bepaalde onderwijs- en organisatiecontext. Zodra een van die elementen verandert, kunnen ook de opbrengsten, lasten en risico’s veranderen.

Dat onderscheid is bestuurlijk belangrijk. Positieve resultaten bij één opleiding worden al snel gebruikt als argument om dezelfde toepassing breder beschikbaar te stellen. Dat lijkt doelmatig: er is al ontwikkelwerk verricht, ervaring opgedaan en mogelijk zelfs geëvalueerd. Maar wat overdraagbaar is, moet nog blijken. Een resultaat uit de ene opleiding is een aanwijzing voor de andere, geen garantie.

Wat een proef wél aannemelijk kan maken

Een goed uitgevoerde proef kan waardevolle kennis opleveren. Zij kan laten zien of studenten de toepassing gebruiken, of de beoogde leerhandeling daadwerkelijk plaatsvindt en welke fouten of knelpunten zich voordoen. Ook kan zichtbaar worden hoeveel docenttijd nodig is voor voorbereiding, controle en begeleiding en welke beheerlast bij structureel gebruik te verwachten valt.

De kracht van die kennis hangt af van de manier waarop de proef is ingericht. Een proef met twintig gemotiveerde studenten, begeleid door de ontwikkelaar zelf, zegt iets anders dan een toepassing die een volledig studiejaar door verschillende docenten wordt gebruikt. Ook het bewijsniveau verschilt. Positieve reacties van studenten tonen aan dat zij de toepassing waarderen, maar nog niet dat zij daardoor beter leren. Veel gebruik laat zien dat de toepassing kennelijk toegankelijk of aantrekkelijk is, maar niet dat studenten de beoogde professionele bekwaamheid ontwikkelen.

Om de uitkomst goed te kunnen duiden, moet daarom duidelijk zijn:

- welke studenten deelnamen en hoeveel ervaring zij al hadden;
- welke leerhandeling werd ondersteund;
- welk resultaat vooraf werd verwacht;
- met welk alternatief de toepassing is vergeleken;
- hoeveel begeleiding studenten en docenten kregen;
- welke uitzonderingen, fouten en uitvallers zijn waargenomen;
- welke nieuwe controle-, beheer- en ondersteuningstaken ontstonden;
- hoe lang de proef duurde en of het gebruik in die periode veranderde.

Zonder die context blijft vooral een succesverhaal over. Dat kan inspireren, maar het biedt een bestuurder weinig houvast voor een besluit over invoering elders.

Overnemen begint met het toetsen van aannames

Bij overdracht naar een andere opleiding hoeft niet alles opnieuw vanaf nul te worden

onderzocht. De instelling kan technische voorzieningen, contractuele afspraken, handleidingen, scholingsmateriaal en eerder vastgestelde risico's hergebruiken. Ook kan een eerdere proef helpen om een volgende toepassing gericht en sneller te beoordelen. Hergebruik van kennis is juist een belangrijke reden om instellingsbreed samen te leren.

Wel moeten de aannames onder de eerdere batenclaim opnieuw worden getoetst. Als de oorspronkelijke proef uitwees dat studenten vaker oefenden, moet worden nagegaan of de nieuwe doelgroep dezelfde oefenbehoefte heeft. Als docenten tijd wonnen doordat de bot veelvoorkomende vragen afhandelde, moet duidelijk worden of die vragen ook in de nieuwe opleiding voorkomen. En als fouten weinig gevolgen hadden omdat het om formatief oefenen ging, mag dat risicoprofiel niet zonder meer worden overgenomen wanneer de toepassing dicht tegen toetsing, studieadvies of beroepshandelen aan komt te liggen.

Een compacte overdrachtstoets kan vier onderdelen onderscheiden.

Onderdeel	Vraag bij toepassing in een andere opleiding
Doelgroep	Hebben deze studenten vergelijkbare voorkennis, taalvaardigheid, ondersteuningsbehoeften en digitale vaardigheden?
Leerhandeling	Moeten zij hetzelfde leren doen, of lijkt alleen het onderwerp op elkaar?
Uitvoering	Beschikt het team over voldoende tijd, deskundigheid en eigenaarschap voor begeleiding en controle?
Risico en bewijs	Zijn de gevolgen van fouten vergelijkbaar en is het eerdere bewijs sterk genoeg voor deze nieuwe context?

Vooraf het verschil tussen onderwerp en leerhandeling verdient aandacht. Twee opleidingen kunnen allebei een bot gebruiken voor 'feedback', terwijl de onderwijsfunctie sterk uiteenloopt. In de ene opleiding controleert de bot of een student vakbegrippen correct toepast. In de andere reageert hij op een complex beroepsproduct waarin ook ethische, juridische en contextgebonden afwegingen een rol spelen. De technische functie lijkt dezelfde, maar de vereiste kwaliteit, menselijke controle en bewijsvoering zijn dat niet.

Toets de toepassing in de lokale context

Wanneer de eerdere resultaten veelbelovend zijn en de belangrijkste aannames overeenkomen, volstaat vaak een beperktere toetsing in de lokale context. De nieuwe opleiding hoeft dan niet opnieuw de hele ontwikkelgeschiedenis te doorlopen, maar onderzoekt gericht waar de context afwijkt en of de verwachte waarde daar standhoudt.

Zo'n lokale toetsing kan beginnen met één onderwijseenheid, een beperkte studentengroep of een duidelijk afgebakende leerhandeling. Daarbij worden dezelfde kernindicatoren gebruikt als in de eerdere proef, zodat vergelijking mogelijk blijft. Voeg alleen indicatoren toe voor de verschillen die er werkelijk toe doen. Bij een grotere of

meer diverse studentengroep kan toegankelijkheid bijvoorbeeld zwaarder wegen. Bij een opleiding met gevoelige praktijkgegevens zijn privacy en gegevensbeheer belangrijker. En wanneer meer docenten de toepassing gebruiken, moet worden vastgesteld of de werkwijze ook zonder intensieve ondersteuning van de oorspronkelijke ontwikkelaar uitvoerbaar blijft.

De uitkomst hoeft trouwens niet te zijn dat de toepassing overal hetzelfde wordt ingericht. Juist bij onderwijs kan verantwoord opschalen betekenen dat een gemeenschappelijke voorziening lokaal anders wordt gebruikt. De techniek en het beheer kunnen centraal zijn geregeld, terwijl opleidingen zelf bepalen voor welke leerhandelingen de toepassing geschikt is, welke bronnen worden gebruikt en wanneer een docent moet meekijken.

Instellingsbreed leren vraagt daarmee om twee bewegingen tegelijk. Ervaringen moeten actief worden gedeeld, zodat opleidingen niet telkens dezelfde fouten maken. Tegelijk moeten lokale verschillen serieus worden genomen, zodat één geslaagde proef niet wordt verheven tot algemeen bewijs. Standaardiseer wat aantoonbaar gemeenschappelijk is en toets opnieuw wat voor de onderwijscontext bepalend blijft.

Ook na die toetsing in de lokale context is het oordeel niet definitief. Studentgroepen veranderen, teams doen ervaring op, leveranciers passen hun systemen aan en kosten kunnen verschuiven. Het bewijs waarmee een toepassing ooit is ingevoerd, kan daardoor gaandeweg minder goed passen bij de werkelijkheid waarin zij wordt gebruikt.

15.2 Veranderingen blijven volgen

De verzamelde informatie krijgt pas betekenis wanneer zij op vaste momenten opnieuw wordt gewogen. Werk daarom na invoering met een compact proces: volg een beperkte set indicatoren, bespreek afwijkingen en wijzigingen op vooraf bepaalde momenten, weeg baten en lasten opnieuw en besluit vervolgens of de toepassing ongewijzigd doorgaat, moet worden aangepast of opnieuw ter discussie komt. Leg daarnaast vast welke signalen aanleiding geven om eerder in te grijpen en wie daarop handelt.

Daarom moet de onderbouwing van een AI-toepassing niet verdwijnen zodra het eerste invoeringsbesluit is genomen. De vraag verschuift alleen. Tijdens een proef staat centraal of de toepassing voldoende waarde *kan* opleveren. Na invoering gaat het erom of die waarde in het gewone onderwijs daadwerkelijk behouden blijft.

Dat is niet vanzelfsprekend. Studenten kunnen de toepassing anders gaan gebruiken dan voorzien, docenten kunnen nieuwe controlewerkzaamheden krijgen en een leverancier kan het model, de prijs of de gebruiksvoorwaarden aanpassen. Ook een aanvankelijk positief resultaat kan daardoor omslaan. Wie alleen bij de start meet, ziet zulke veranderingen pas wanneer klachten, kosten of kwaliteitsproblemen al zijn opgelopen.

Volg dezelfde drie soorten waarde

De kernindicatoren blijven aansluiten bij de drie niveaus waarop AI moet renderen: student, medewerker en instelling. Daarmee ontstaat geen volledig nieuw meetstelsel, maar een voortzetting van de oorspronkelijke batenafweging.

Niveau	Wat blijft u volgen?	Mogelijke signalen
Student	Leren, oefenen en professionele bekwaamheid	gebruiksfrequentie, aantal oefenpogingen, kwaliteit van producten of prestaties, verwerking van feedback, zelfstandig handelen en verschillen tussen studentgroepen
Medewerker	Bruikbare professionele ruimte	tijd voor begeleiding en ontwerp, hoeveelheid controlewerk, herstelwerk, ervaren werkdruk en duidelijkheid over verantwoordelijkheden
Instelling	Kwaliteit, toegankelijkheid en doelmatigheid	kosten per gebruiker, bereik, uitval, incidenten, beheerlast, afhankelijkheid van leveranciers en aansluiting op onderwijs- en informatiebeleid

Niet iedere toepassing vraagt om alle indicatoren. Een oefenbot voor een grote eerstejaarscursus vraagt bijvoorbeeld om andere gegevens dan een gesprekssimulatie in een kleine zorgopleiding. De eerste kan vooral worden beoordeeld op oefenintensiteit, kwaliteit van antwoorden en benodigde docentondersteuning. Bij de tweede wegen ook professioneel gedrag, emotionele veiligheid en de kwaliteit van de nabespreking zwaar.

De gekozen indicatoren moeten wel dezelfde bestuurlijke vraag blijven beantwoorden: levert de toepassing nog steeds de waarde op waarvoor zij is ingevoerd?

Gebruik is nog geen opbrengst

Na invoering ontstaat gemakkelijk de verleiding om vooral te volgen hoeveel studenten de toepassing gebruiken. Die gegevens zijn meestal eenvoudig beschikbaar en ogen overtuigend in een voortgangsrapportage. Veel gebruik bewijst echter niet dat studenten beter leren. Het kan evengoed betekenen dat de toepassing aantrekkelijk, gemakkelijk of zelfs noodzakelijk is geworden.

Gebruiksgegevens zijn daarom vooral een aanwijzing om verder te kijken. Als studenten vaker oefenen, is de vervolgvraag of zij ook gericht oefenen en aantoonbaar vooruitgaan. Als zij sneller feedback krijgen, moet worden onderzocht of zij die feedback begrijpen en verwerken. Als docenten minder tijd besteden aan een eerste reactie, moet zichtbaar worden of die tijd werkelijk beschikbaar komt voor persoonlijke begeleiding of dat zij opgaat aan controle, correctie en het oplossen van technische problemen.

Ook een dalend gebruik vraagt om uitleg. Misschien hebben studenten de ondersteuning minder nodig doordat zij zelfstandiger zijn geworden. Dat kan juist een positief resultaat zijn. Het kan ook betekenen dat de antwoorden niet bruikbaar zijn, dat het systeem lastig toegankelijk is of dat studenten zijn uitgeweken naar een algemeen beschikbare toepassing buiten het zicht van de instelling. Eén cijfer vertelt

zelden het hele verhaal.

Houd baten en lasten bij elkaar

Een eerdere businesscase kan ongemerkt verouderen. Licentiekosten kunnen stijgen, terwijl het aantal actieve gebruikers achterblijft. Een koppeling met de digitale leeromgeving kan extra beheer vragen. Nieuwe modelversies kunnen betere antwoorden geven, maar ook ander gedrag vertonen. Bovendien ontstaan na verloop van tijd vaak taken die tijdens de proef nog nauwelijks zichtbaar waren: bronnen actualiseren, foutieve antwoorden onderzoeken, instructies aanpassen, gebruikers ondersteunen en afspraken over gegevensgebruik controleren.

Daarom moeten opbrengsten en lasten steeds in samenhang worden gevolgd. Een besparing op nakijkwerk is bijvoorbeeld minder overtuigend wanneer docenten vervolgens veel tijd kwijt zijn aan het controleren van door AI gegenereerde feedback. Andersom kan een toepassing die financieel weinig bespaart toch waardevol zijn wanneer studenten hierdoor substantieel meer kunnen oefenen in een beroepssituatie die anders moeilijk, duur of risicovol te organiseren is.

De relevante vraag is dus niet alleen of de kosten hoger of lager worden. Het gaat erom welke waarde tegenover de volledige kosten en beheerlast staat en of die verhouding nog aanvaardbaar is.

Afwijkingen zijn informatie

Incidenten en uitvoeringsproblemen horen niet thuis in een apart dossier dat pas bij een ernstige verstoring wordt geopend. Ze vormen informatie over de werking van de toepassing. Denk aan onjuiste bronverwijzingen, ongepaste feedback, uitval van een koppeling, ontoegankelijkheid voor een groep studenten of onduidelijkheid over wie een fout moet herstellen.

Losse meldingen zeggen nog niet dat een toepassing moet worden beëindigd. Terugkerende patronen kunnen de oorspronkelijke afweging echter wel veranderen. Drie vergelijkbare incidenten in verschillende opleidingen kunnen wijzen op een gemeenschappelijk technisch probleem. Hetzelfde incident binnen één opleiding kan juist samenhangen met de gekozen opdracht, kennisbron of begeleidingsvorm.

Leg bij een incident daarom compact vast wat nodig is om het patroon, de gevolgen en de vereiste actie te beoordelen:

Wat legt u vast?	Waarover moet dit duidelijkheid geven?
Betrokken studenten en medewerkers	Wie ondervond gevolgen en waren bepaalde groepen extra kwetsbaar?
Leerhandeling of werктаak	In welke onderwijs- of werkcontext ontstond het probleem?

Ernst en herstelbaarheid	Hoe groot waren de gevolgen en konden die worden hersteld?
Menselijke controle	Werd het probleem tijdig opgemerkt en zo niet, waarom niet?
Herstelinspanning	Hoeveel tijd en ander werk waren nodig om het probleem op te lossen?
Oorzaak	Was die lokaal, technisch of organisatorisch?
Vervolgactie	Welke aanpassing is nodig om herhaling te voorkomen?

Zo wordt incidentregistratie een bron voor verbetering in plaats van alleen een administratieve verplichting.

Let op verschuivingen tussen groepen

Gemiddelde resultaten kunnen belangrijke verschillen verhullen. Een toepassing kan voor taalvaardige studenten goed werken, terwijl studenten die meer structuur nodig hebben juist vastlopen. Spraakbediening kan de toegankelijkheid voor de ene groep vergroten en voor een andere groep problemen opleveren. Ook kunnen studenten verschillen in hun vermogen om antwoorden te controleren, goede vervolgvragen te stellen en onjuiste suggesties te herkennen.

Bij periodieke monitoring hoort daarom aandacht voor de verdeling van baten en risico's. Niet alleen: gaan de gemiddelde resultaten vooruit? Maar ook: voor wie werkt dit goed, wie profiteert nauwelijks en wie loopt extra risico? Juist zulke verschillen bepalen of aanvullende instructie, een andere interactievorm of intensievere docentbegeleiding nodig is.

Maak veranderingen bestuurlijk zichtbaar

Monitoring heeft weinig waarde wanneer gegevens wel worden verzameld, maar niet leiden tot een gesprek over de toepassing. Een bruikbare rapportage hoeft niet omvangrijk te zijn. Vaak volstaat een compact overzicht met de oorspronkelijke doelen, de actuele resultaten, opvallende verschillen tussen groepen, volledige kosten, beheerlast, incidenten en relevante wijzigingen aan het systeem.

Daarbij moet helder zijn wie een verandering beoordeelt. Een technisch beheerder kan vaststellen dat een leverancier een nieuw model heeft ingevoerd, maar niet alleen bepalen wat dit betekent voor de didactische kwaliteit. Een docent kan veranderingen in studentgedrag signaleren, maar heeft niet altijd zicht op contractuele of privacygevolgen. Voor een volledig oordeel moeten onderwijs, beheer, informatiemanagement en waar nodig privacy, beveiliging en medezeggenschap met elkaar worden verbonden.

Continu volgen betekent overigens niet dat iedere toepassing permanent door een brede commissie moet worden bewaakt. Het toezicht moet in verhouding staan tot de schaal,

het beroepsrisico, de gebruikte gegevens en de gevolgen van fouten. Zonder een vast ritme blijven dashboards niet meer dan rapportages en worden afwijkingen gemakkelijk als losse voorvallen behandeld. De volgende stap is daarom om te bepalen *wanneer* een toepassing opnieuw ter discussie komt en hoe de betrokken deskundigen wijzigingen gezamenlijk beoordelen.

Gezamenlijke beoordeling van een modelwijziging



15.3 Een vast ritme van herbeoordeling

Een vast ritme voorkomt dat een toepassing ongemerkt structureel wordt omdat niemand het besluit nog ter discussie stelt. De instelling legt daarom vooraf vast wanneer de herbeoordeling plaatsvindt, wie eraan deelneemt, welke informatie beschikbaar moet zijn en welke besluiten mogelijk zijn. Zo wordt herbeoordeling een normaal onderdeel van beheer en kwaliteitszorg, niet een reactie die pas op gang komt nadat er iets is misgegaan.

Herbeoordeling op vaste momenten

Niet iedere AI-toepassing vraagt om hetzelfde ritme. Bij een kleinschalige toepassing voor vrijwillige extra oefening kan een jaarlijkse beoordeling voldoende zijn. Een toepassing die feedback geeft binnen een groot aantal onderwijseenheden, persoonsgegevens verwerkt of invloed heeft op beslissingen over studievoortgang vraagt om een kortere cyclus.

In de beginfase ligt een hogere frequentie voor de hand. Een opleiding kan bijvoorbeeld na drie maanden beoordelen of de toepassing technisch en organisatorisch werkt, na een semester of de beoogde leerhandelingen daadwerkelijk plaatsvinden en na een studiejaar of de baten opwegen tegen de volledige lasten. Zodra het gebruik stabiel is en de risico's beheersbaar blijken, kan de frequentie omlaag. Dat ontslaat de instelling overigens niet van de plicht om tussentijds in te grijpen wanneer de omstandigheden veranderen.

Het ritme kan zoveel mogelijk aansluiten bij bestaande processen, zoals de periodieke evaluatie van een onderwijseenheid, de jaarlijkse kwaliteitscyclus, de begrotingsvoorbereiding, contractbesprekingen of de actualisering van een onderwijs- en examenregeling. Daarmee voorkomt de instelling dat voor iedere toepassing een afzonderlijke overlegstructuur ontstaat. Wel moet de AI-toepassing binnen zo'n bestaand proces herkenbaar op de agenda staan. Als zij slechts opgaat in een algemene bespreking van digitalisering, blijven onderwijsresultaten, nieuwe beheertaken en veranderende risico's al snel buiten beeld.

Een werkbaar ritme bevat ten minste drie soorten momenten:

1. **Een vroege gebruiksbeoordeling**, gericht op bereik, gebruik, technische werking en eerste knelpunten.
2. **Een inhoudelijke effectbeoordeling**, gericht op leren, professionele ruimte, kwaliteit en ongewenste gevolgen.
3. **Een bestuurlijke herafweging**, waarin baten, kosten, risico's, beheerlast en strategische betekenis samenkomen.

Bij die momenten hoeft niet steeds dezelfde groep betrokken te zijn. Het team kan het gebruik en de didactische werking beoordelen, terwijl een manager of portefeuillehouder beslist over voortzetting en middelen. Privacy-, beveiligings- en inkoopdeskundigen worden betrokken waar hun oordeel nodig is. Belangrijker dan de precieze overlegvorm is dat de verschillende inzichten uiteindelijk in één afweging bijeenkomen.

Niet alleen de opbrengst opnieuw bekijken

Bij herbeoordeling gaat het niet om de vraag of de toepassing nog steeds populair is. Veel gebruik kan wijzen op waarde, maar ook op afhankelijkheid, gemak of een gebrek aan alternatieven. Evenmin is een hoge tevredenheid voldoende bewijs dat studenten beter leren of dat medewerkers werkelijk professionele ruimte winnen.

De oorspronkelijke doelen vormen daarom het vertrekpunt. Wilde de opleiding studenten vaker laten oefenen, dan moet zichtbaar zijn of zij dat inderdaad doen en of de oefeningen bijdragen aan de beoogde professionele bekwaamheid. Was het doel om feedbackcycli te verkorten, dan telt niet alleen de snelheid van de terugkoppeling. Ook moet worden vastgesteld of studenten de feedback begrijpen, verwerken en daardoor beter gaan handelen. Was tijdwinst voor docenten een belangrijk argument, dan moet

duidelijk zijn hoeveel tijd er vrijkomt, welke nieuwe controle- en beheertaken zijn ontstaan en waaraan de gewonnen tijd vervolgens wordt besteed.

Daarnaast worden de uitgangspunten uit het eerdere besluit opnieuw getoetst. Misschien waren de kosten gebaseerd op een tijdelijk contract, bleek menselijke controle intensiever dan verwacht of is het aantal gebruikers sterk gestegen. Ook kan een leverancier de werking, voorwaarden of gegevensverwerking hebben gewijzigd. In al deze gevallen is de oorspronkelijke businesscase niet langer vanzelfsprekend geldig.

Een compacte herbeoordeling behandelt daarom steeds dezelfde vragen:

- Is het onderwijs- of organisatievraagstuk nog hetzelfde?
- Wordt de beoogde doelgroep bereikt?
- Zijn de verwachte baten zichtbaar en voldoende onderbouwd?
- Kunnen medewerkers op een werkbare manier verantwoordelijk toezicht blijven houden?
- Welke nieuwe kosten, correcties en beheertaken zijn ontstaan?
- Zijn de risico's, incidenten en afhankelijkheden veranderd?
- Past de toepassing nog bij de publieke waarden en afspraken van de instelling?
- Is voortzetting nog steeds beter dan aanpassen, vervangen of stoppen?

Door deze vragen steeds opnieuw te gebruiken, worden beoordelingen in de loop van de tijd beter vergelijkbaar. De instelling ziet dan niet alleen de toestand van dat moment, maar ook de richting waarin de toepassing zich ontwikkelt.

Aanleidingen voor tussentijdse herbeoordeling

Naast geplande beoordelingen zijn er aanleidingen nodig voor een tussentijdse heroverweging. Een ernstig incident is de duidelijkste aanleiding, maar niet de enige. Ook een reeks kleine fouten kan erop wijzen dat de toepassing structureel tekortschiet. Andere mogelijke aanleidingen zijn een opvallende verandering in studentgedrag, een sterke stijging van kosten, nieuwe wet- of regelgeving, gewijzigde leveranciersvoorwaarden, een beveiligingsmelding, klachten over toegankelijkheid of signalen dat docenten het noodzakelijke toezicht niet meer kunnen uitvoeren.

Vooraf afgesproken drempels maken sneller handelen mogelijk. Denk aan een maximumaantal ernstige inhoudelijke fouten, een grens voor extra docenttijd, een minimale gebruiksgraad of een afspraak dat iedere wezenlijke wijziging in model, gegevensbron of opslaglocatie opnieuw wordt beoordeeld. Zulke drempels zijn geen automatische beslisregels. Ze zorgen er wel voor dat een ongunstige ontwikkeling niet eindeloos wordt afgedaan als een tijdelijk opstartprobleem.

Bij toepassingen met grote gevolgen moet bovendien duidelijk zijn wie het gebruik tijdelijk kan beperken of stilleggen. Dat besluit hoeft niet te wachten op de eerstvolgende vergadering. De procedure voor snel ingrijpen hoort daarom net zo goed bij de inrichting als de reguliere evaluatiecyclus.

Bevestigen, aanpassen of terugdraaien

Een herbeoordeling is alleen geloofwaardig wanneer alle uitkomsten werkelijk mogelijk zijn. In de praktijk ontstaat gemakkelijk een voorkeur voor voortzetting: er is al geïnvesteerd, gebruikers zijn eraan gewend en stoppen kan worden uitgelegd als mislukking. Juist daardoor blijven toepassingen soms bestaan terwijl hun oorspronkelijke doel niet meer wordt bereikt.

De uitkomst kan zijn dat de eerdere keuze wordt bevestigd. De baten zijn dan nog steeds zichtbaar, de lasten blijven aanvaardbaar en de voorwaarden zijn op orde. Aanpassen ligt voor de hand wanneer de waarde aannemelijk is, maar het ontwerp, de doelgroep, het toezicht of de technische inrichting moet veranderen. Soms is tijdelijk terugschalen verstandig, bijvoorbeeld tot één opleiding of één type leerhandeling. Stoppen is passend wanneer de opbrengst onvoldoende is, de beheerslast structureel te hoog wordt of wezenlijke risico's niet redelijk kunnen worden beperkt.

Leg bij ieder besluit niet alleen de uitkomst vast, maar ook de onderbouwing, de resterende onzekerheden en het volgende beoordelingsmoment. Zo ontstaat een spoor van bestuurlijk leren. Dat spoor is later ook nodig om te bepalen welke ervaringen bruikbaar zijn voor andere opleidingen. Zonder vastgelegde doelen, ontwerpkeuzes, resultaten en voorwaarden is immers niet te onderscheiden wat werkelijk overdraagbaar is en wat alleen in de oorspronkelijke context werkte.

15.4 Besluiten over overdraagbaarheid

Drie mogelijke besluiten

Bij iedere voorgenomen overdracht zijn grofweg drie besluiten mogelijk.

Besluit	Wanneer passend	Wat de ontvangende opleiding doet
Overnemen	Doelgroep, leerhandeling, context, risico's en randvoorwaarden komen sterk overeen	De toepassing invoeren met een beperkte lokale controle
Aanpassen en beproeven	Het werkzame principe lijkt bruikbaar, maar inhoud, begeleiding of organisatie verschilt	Het ontwerp aanpassen en de belangrijkste baten en risico's opnieuw meten
Opnieuw experimenteren	De context of het beroepsrisico wijkt wezenlijk af, of het bewijs is nog zwak	Een nieuwe, afgebakende proef uitvoeren zonder de eerdere resultaten als lokale opbrengst te beschouwen

Volledig overnemen zal minder vaak passend zijn dan op het eerste gezicht lijkt. Zelfs

bij verwante opleidingen kunnen verschillen in curriculum, studentpopulatie en begeleiding groot zijn. Een toepassing die goed werkt bij derdejaarsstudenten met veel praktijkervaring kan eerstejaarsstudenten juist afhankelijk maken. Een bot die bruikbare formatieve feedback geeft op oefenproducten kan bovendien ongeschikt zijn voor een context waarin dezelfde producten meetellen voor een formele beoordeling.

Aanpassen en beproeven is daarom vaak de verstandigste route. De ontvangende opleiding gebruikt dan wat al bekend is, maar behandelt de lokale werking nog als een te toetsen aanname. Dat vraagt geen omvangrijke pilot wanneer de risico's beperkt zijn. Wel moet vooraf duidelijk zijn welke uitkomst wordt verwacht, wat wordt gemeten en bij welke signalen de toepassing wordt aangepast of gestopt.

Opnieuw experimenteren is nodig wanneer wezenlijke kenmerken veranderen. Denk aan toepassing in een ander beroep, gebruik van gevoelige gegevens, een overgang van tekst naar spraak, inzet bij formele toetsing of veel zelfstandiger gebruik door studenten. Eerder succes verlaagt dan hooguit de ontwikkelkosten; het bewijst niet dat de nieuwe toepassing verantwoord of effectief is.

Om tussen deze besluiten te kiezen, moet duidelijk zijn waarom de toepassing in de oorspronkelijke omgeving werkte. Dezelfde bot, dezelfde leverancier en zelfs hetzelfde onderwijsdoel kunnen in een andere opleiding tot een andere uitkomst leiden. Wat op één plek rendeert, is dus niet automatisch een bewezen oplossing voor de hele instelling.

Van succesverhaal naar overdraagbare ervaring

Een ervaring wordt pas bruikbaar voor anderen wanneer ten minste vijf onderdelen bekend zijn:

1. **De doelgroep en uitgangssituatie.** Voor welke studenten, medewerkers of processen was de toepassing bedoeld? Welke voorkennis, problemen en bestaande werkwijzen waren er?
2. **De ontwerpkeuzes.** Welke leerhandeling ondersteunde de toepassing, welke kennisbronnen werden gebruikt, hoe verliep de interactie en welk menselijk toezicht was ingericht?
3. **Het bewijsniveau.** Gaat het om positieve reacties, zichtbaar gebruik, gemeten tijdswinst, betere leerprestaties of een vergelijking met de eerdere werkwijze?
4. **De beperkingen en mislukkingen.** Waar haakten studenten af, welke antwoorden bleken onbetrouwbaar, welk extra werk ontstond en welke groepen profiteerden niet?
5. **De voorwaarden.** Welke deskundigheid, gegevens, technische koppelingen, licenties, begeleiding en beheerinspanning waren nodig om de resultaten te bereiken?

Vooraf het vierde onderdeel wordt gemakkelijk overgeslagen. Teams delen liever wat goed ging dan wat zij moesten herstellen. Toch zijn juist de mislukkingen vaak het

meest overdraagbaar. Een opleiding die ontdekt dat studenten feedback van een bot kritiekloos overnemen, levert niet alleen een waarschuwing op, maar ook een ontwerpvoorwaarde: studenten moeten leren de feedback te controleren en docenten moeten kunnen zien hoe zij die verwerken. Een andere opleiding kan die voorwaarde meteen meenemen, ook als haar onderwijsinhoud verschilt.

Maak daarbij onderscheid tussen het **product** en het **werkzame principe**. Een specifieke oefenbot voor verpleegkundige gespreksvoering is wellicht niet bruikbaar binnen bedrijfskunde. Het principe erachter kan dat wel zijn: studenten oefenen herhaaldelijk een professioneel gesprek, krijgen onmiddellijk terugkoppeling op vooraf bepaalde criteria en bespreken complexe of twijfelachtige situaties met een docent. Het product blijft lokaal, terwijl het didactische ontwerp overdraagbaar kan zijn.

Dat betekent niet dat iedere opleiding opnieuw vanaf nul moet beginnen. Juist door ervaringen zorgvuldig vast te leggen, kan een instelling voorkomen dat teams dezelfde fouten maken, dezelfde voorwaarden opnieuw ontdekken en afzonderlijk vergelijkbare voorzieningen inkopen. De bestuurlijke opgave is om te bepalen welk deel van een ervaring als gemeenschappelijke kennis kan worden gebruikt, welk deel lokaal moet worden aangepast en welk deel opnieuw onderzocht moet worden.

Gemeenschappelijk leren zonder schijnzekerheid

Instellingsbreed leren vraagt om een gemeenschappelijke manier van vastleggen. Dat hoeft geen omvangrijk verantwoordingsdossier te worden. Een compact overdrachtsdocument kan volstaan, mits daarin het vraagstuk, de doelgroep, het ontwerp, de waargenomen baten, de volledige lasten, de risico's, de mislukkingen en de noodzakelijke voorwaarden herkenbaar zijn opgenomen. Voeg ook toe wat nog niet bekend is. Onzekerheid die expliciet wordt gemaakt, is bestuurlijk beter hanteerbaar dan zekerheid die alleen wordt verondersteld.

Voor managers ontstaat daarmee een dubbele verantwoordelijkheid. Enerzijds moeten zij kennisdeling organiseren, zodat opleidingen niet als losse eilanden blijven experimenteren. Anderzijds moeten zij voorkomen dat een lokaal resultaat door herhaling in presentaties de status van algemeen bewijs krijgt. Drie enthousiaste opleidingen vormen nog geen instellingsbrede effectmeting wanneer alle drie dezelfde beperkte waarnemingen rapporteren.

Ook centrale voorzieningen verdienen deze toets. Een gedeeld platform, een goedgekeurde leverancier of een gezamenlijke kennisbank kan technische en juridische lasten verminderen. Daarmee is de onderwijswaarde nog niet aangetoond. Centrale beschikbaarheid en lokale geschiktheid zijn twee verschillende besluiten. De instelling kan de veilige en beheersbare basis gezamenlijk regelen, terwijl opleidingen verantwoordelijk blijven voor de didactische inpassing, begeleiding en beoordeling van de opbrengsten.

De kernvraag bij overdraagbaarheid luidt dan ook niet: *Kunnen andere opleidingen dit*

gebruiken? Vrijwel altijd kan dat technisch wel. De betere vragen zijn: Welk probleem wordt elders hiermee opgelost, welke onderdelen van het eerdere resultaat zijn daar werkelijk relevant en wat moet opnieuw worden vastgesteld?

Zo groeit een instelling niet door ieder experiment zo snel mogelijk uit te rollen, maar door ervaringen steeds preciezer te maken. Wat aantoonbaar gemeenschappelijk is, wordt gedeeld. Wat van de context afhangt, wordt aangepast. Wat nog onzeker is, wordt opnieuw onderzocht. Daarmee wordt periodieke herbeoordeling meer dan controle achteraf: zij vormt de basis voor gezamenlijk leren zonder lokale ervaring te verwarren met algemeen bewijs.

16 Bekwaamheid groeit geleidelijk

Wat voor instellingen geldt, geldt ook voor studenten: een resultaat in de ene situatie bewijst nog niet dat het geleerde in een volgende situatie standhoudt. Wie een methode correct toepast in een afgebakende opdracht, kan alsnog vastlopen wanneer informatie onvolledig is, belangen botsen of tijdsdruk ontstaat. Juist daar wordt zichtbaar of kennis en vaardigheden zijn uitgegroeid tot professionele bekwaamheid.

Voor die groei is meer nodig dan een reeks steeds grotere opdrachten. De opleiding moet bewust bepalen welke complexiteit zij toevoegt, welke ondersteuning de student daarbij krijgt en wanneer die ondersteuning juist moet verdwijnen. AI kan veel houvast bieden, maar daardoor ook verhullen wat een student nog niet zelfstandig kan. Een goed product bewijst immers niet automatisch dat iemand professioneel heeft gehandeld.

Voor bestuurders verschuift de aandacht zo van afzonderlijke toepassingen naar de leerlijn als geheel. Niet: waar kunnen we nog een hulpmiddel inzetten? Wel: welk handelen moet de student hier leren, onder welke omstandigheden en met hoeveel steun? Pas vanuit die vragen wordt duidelijk of AI werkelijk bijdraagt aan de ontwikkeling – of vooral helpt de volgende opdracht af te ronden.

Vastlopen onder tijdsdruk



16.1 Van deelvaardigheid naar beroepsbekwaamheid

Ook voor de ontwikkeling van studenten geldt dat een geslaagde oefening niet zonder meer overdraagbaar is naar een volgende fase. Een student kan een begrip correct uitleggen, een gesprekstechniek beheersen of een analyse volgens een vast stappenplan uitvoeren, maar daarmee nog niet adequaat handelen in een beroepssituatie waarin al die elementen tegelijk nodig zijn. Beroepsbekwaamheid ontstaat pas wanneer kennis, vaardigheden en professioneel gedrag steeds meer met elkaar verbonden raken.

Dat vraagt om een doelbewuste opbouw. Losse toepassingen voor uitleg, oefening, feedback en simulatie krijgen pas didactische betekenis wanneer duidelijk is welke ontwikkeling zij ondersteunen, wat de student in iedere fase zelfstandig moet kunnen en hoe de opleiding vaststelt dat de volgende stap verantwoord is. Niet het aantal AI-toepassingen bepaalt dan de kwaliteit van het curriculum, maar de samenhang tussen de leerhandelingen.

Eerst afzonderlijk, daarna in samenhang

Aan het begin van een opleiding of leerlijn oefenen studenten vaak met afgebakende deelvaardigheden. Zij herkennen begrippen, passen een methode toe, stellen vragen of beoordelen een eenvoudige casus. Dat is geen kunstmatige versimpeling, maar een noodzakelijke tussenstap. Wie nog alle aandacht nodig heeft voor de structuur van een adviesgesprek, kan moeilijk tegelijkertijd luisteren naar onderliggende belangen, omgaan met weerstand en professioneel afwegen wat wel en niet verantwoord is.

AI kan in deze fase zorgen voor veel korte oefenmomenten. Een student kan bijvoorbeeld:

- begrippen toepassen op verschillende voorbeelden;
- een redenering stap voor stap opbouwen;
- hoofd- en bijzaken in een casus onderscheiden;
- vragen formuleren voor een intake;
- een eenvoudig beroepsproduct aan criteria toetsen;
- gerichte feedback verwerken in een tweede versie.

De waarde zit hier vooral in herhaling en variatie. Studenten hoeven niet te wachten tot een docent tijd heeft om een nieuwe oefening te geven of een eerste poging van commentaar te voorzien. Wel blijft de oefening nauw afgebakend: de student weet wat er wordt geoefend en waarop de feedback betrekking heeft.

In een volgende fase worden deelvaardigheden gecombineerd. De student analyseert dan niet alleen een probleem, maar vertaalt de analyse ook naar een voorstel, licht keuzes toe en reageert op vragen. Een simulatie kan meerdere handelingen bijeenbrengen: informatie verzamelen, doorvragen, een afweging maken, communiceren en achteraf verantwoorden waarom een bepaalde aanpak is gekozen.

Juist daar wordt zichtbaar of eerder geoefende kennis beschikbaar blijft zodra de

situatie meer aandacht vraagt. Een student kan de regels immers kennen en ze toch vergeten tijdens een lastig gesprek. Andersom kan iemand overtuigend communiceren zonder de inhoudelijke afweging goed te onderbouwen. Beide prestaties lijken op het eerste gezicht geslaagd, maar wijzen op een andere ontwikkelbehoefte.

Ontwikkeling zichtbaar maken

Een opleiding moet daarom niet alleen beschrijven *wat* studenten oefenen, maar ook *hoe* hun handelingsvermogen zich ontwikkelt. Een bruikbare opbouw onderscheidt ten minste vier fasen:

Fase	Kenmerkende prestatie van de student	Passende rol van AI	Wat de opleiding zichtbaar maakt
Verkennen	Herkent begrippen, stappen en relevante kenmerken	Uitleg geven, voorbeelden aanbieden en voorkennis activeren	Welke basiskennis aanwezig is en waar misvattingen zitten
Gericht oefenen	Voert een afgebakende handeling uit met steun	Oefeningen aanpassen, hints geven en directe feedback bieden	Of de student de deelvaardigheid steeds consistentere beheerst
Combineren	Verbindt meerdere kennis- en vaardigheidselementen in een taak	Varianten en simulaties aanbieden, feedback op samenhang geven	Of de student keuzes kan onderbouwen en feedback kan verwerken
Zelfstandig handelen	Past het geleerde toe in een nieuwe beroepssituatie en verantwoordt de gevolgen	Een uitdagende oefenomgeving bieden of als kritische gesprekspartner optreden	Of de student zelfstandig, zorgvuldig en contextgevoelig handelt

Het doorlopen van deze fasen leidt niet automatisch tot beroepsbekwaamheid door voldoende gebruik van een oefenbot. De overgang moet blijken uit het handelen van de student. Dat vraagt om observeerbare aanwijzingen: een betere probleemanalyse, minder behoefte aan aanwijzingen, een onderbouwde keuze, adequaat gedrag in een simulatie of succesvolle toepassing in een nieuwe context.

Daarbij is het verstandig verschillende soorten informatie te combineren. Gebruiksgegevens kunnen laten zien hoe vaak een student oefent, maar niet zonder meer wat hij leert. Een door AI gegenereerde voortgangsscore kan een patroon signaleren, maar vormt nog geen zelfstandig oordeel over bekwaamheid. Producten, mondelinge toelichtingen, observaties, reflecties en prestaties in realistische opdrachten bieden samen een rijker beeld.

Steun moet geleidelijk verdwijnen

Bij de opbouw van bekwaamheid hoort ook dat ondersteuning afneemt. In de eerste fase mag een toepassing uitleg geven, tussentapen voorstellen en veel gerichte aanwijzingen bieden. Later moet de student vaker zelf bepalen wat het probleem is, welke kennis relevant is en welke aanpak passend lijkt.

Dat onderscheid is belangrijk. Wanneer AI bij iedere opdracht de analyse ordent, alternatieven formuleert en de redenering aanscherpt, kan het eindproduct verbeteren terwijl de zelfstandigheid van de student nauwelijks groeit. De toepassing neemt dan precies die denkhandelingen over die de opleiding wil ontwikkelen.

Een leerlijn moet daarom expliciet vastleggen welke steun in iedere fase is toegestaan. Zo kan een student aanvankelijk werken met voorbeeldvragen en directe aanwijzingen, vervolgens alleen nog feedback achteraf ontvangen en ten slotte zonder AI handelen. Daarna kan de toepassing opnieuw worden ingezet om de gemaakte keuzes kritisch te bevragen. AI verschuift dan van instructeur naar oefenpartner en uiteindelijk naar tegenspreker, terwijl de student steeds meer verantwoordelijkheid draagt.

Beoordeel de samenhang, niet alleen de onderdelen

Voor opleidingsmanagers en curriculumteams ligt hier een concrete ontwerp vraag. Bekijk niet alleen waar een onderwijsbot in een vak past, maar welke plaats de ondersteunde leerhandeling inneemt in de gehele ontwikkel lijn. Een aantrekkelijke toepassing kan binnen een vak goed werken en toch weinig bijdragen aan het eindniveau, bijvoorbeeld omdat dezelfde eenvoudige handeling telkens terugkomt of omdat studenten nooit zonder ondersteuning hoeven te presteren.

De relevante vragen zijn dan:

- Welke professionele bekwaamheid wordt hiermee opgebouwd?
- Welke deelvaardigheid staat in deze fase centraal?
- Wat moet de student al beheersen en wat volgt hierna?
- Welke ondersteuning biedt AI en wanneer wordt die verminderd?
- Aan welk gedrag of product zien we dat de student vooruitgaat?
- Kan de student de handeling ook zonder AI en in een nieuwe situatie uitvoeren?
- Wie beoordeelt of de samenhang tussen kennis, vaardigheden en professioneel gedrag voldoende is?

Zo ontstaat geen verzameling digitale oefenmomenten, maar een herkenbare ontwikkel lijn. Iedere fase bouwt voort op de vorige, vraagt meer zelfstandigheid en levert bewijs voor een volgende stap. Pas daarna kan de moeilijkheid werkelijk worden opgevoerd – niet alleen door opdrachten omvangrijker te maken, maar door studenten te laten handelen in situaties met meer onzekerheid, tegenstrijdige belangen en verantwoordelijkheid.

16.2 Complexiteit bewust opvoeren

Moeilijker onderwijs is niet hetzelfde als méér onderwijs. Een omvangrijker project, een langer verslag of een extra gegevensbestand verhoogt vooral de werklast. Professionele complexiteit ontstaat pas wanneer studenten verschillende soorten kennis en vaardigheden moeten combineren, onvolledige informatie moeten beoordelen en verantwoordelijkheid moeten nemen voor de gevolgen van hun keuzes.

Daarom moet een ontwikkellijn niet alleen beschrijven *wat* studenten uitvoeren, maar ook *onder welke omstandigheden* zij dat doen. In een vroege fase kan een opdracht nog afgebakend zijn: het probleem is gegeven, de beschikbare informatie is betrouwbaar en de kwaliteitscriteria zijn duidelijk. In een volgende fase ontbreken gegevens, spreken bronnen elkaar tegen of blijken de belangen van betrokkenen uiteen te lopen. Nog later moet de student onder tijdsdruk samenwerken, keuzes uitleggen en bijsturen wanneer de situatie verandert.

De taak blijft misschien herkenbaar, maar het gevraagde professionele handelen verandert wezenlijk.

Vijf dimensies van toenemende complexiteit

Complexiteit kan langs verschillende dimensies worden opgebouwd. Die hoeven niet allemaal tegelijk te worden verzwaard. Juist door ze bewust te doseren, blijft duidelijk welke ontwikkeling van de student wordt gevraagd.

Dimensie	Eenvoudige situatie	Complexere situatie
Onzekerheid	Het probleem en de benodigde informatie zijn gegeven	De student moet zelf bepalen wat het probleem is en welke informatie nog ontbreekt
Belangen	Betrokkenen streven hetzelfde doel na	Belangen botsen en een keuze heeft voor iedere partij andere gevolgen
Tijdsdruk	Er is voldoende tijd voor analyse en herstel	De student moet prioriteren en besluiten met beperkte voorbereiding
Samenwerking	Taken zijn vooraf verdeeld	Rollen zijn afhankelijk van elkaar en moeten tijdens het werk worden afgestemd
Verantwoordelijkheid	De gevolgen blijven binnen de oefensituatie	De student moet risico's afwegen, keuzes verantwoorden en rekening houden met beroepsnormen

Deze dimensies maken zichtbaar waarom een opdracht complex is. Dat helpt studenten, docenten en curriculumteams om verder te kijken dan de hoeveelheid werk. Een opdracht met veel onderdelen kan didactisch eenvoudig blijven wanneer alle stappen zijn voorgeschreven. Een korte praktijksituatie kan daarentegen veeleisend zijn wanneer de student zelf moet vaststellen wat verantwoord handelen inhoudt.

Van toepassing naar afweging

In een eerste fase kan een student bijvoorbeeld een AI-toepassing beoordelen aan de hand van vaste criteria. De gegevens zijn compleet, de gebruiker is duidelijk omschreven en de mogelijke risico's zijn al benoemd. De student leert dan vooral een methode correct toe te passen.

In een volgende fase ontvangt de student een minder geordende casus. De opdrachtgever vraagt om snellere dienstverlening, medewerkers vrezen extra controlewerk en gebruikers verwachten persoonlijke ondersteuning. Niet alle beschikbare gegevens zijn even betrouwbaar. De student moet nu niet alleen criteria toepassen, maar ook bepalen welke informatie ertoe doet, welke aannames worden gemaakt en wiens belang zwaar moet wegen.

Daarna kan de opleiding een extra laag toevoegen. De student werkt in een team, krijgt te maken met een veranderende opdracht en moet binnen beperkte tijd een advies uitbrengen. Een technisch aantrekkelijke oplossing blijkt bijvoorbeeld moeilijk te beheren of gebruikt gegevens waarvoor geen duidelijke grondslag bestaat. Dan is er niet één juist antwoord. De kwaliteit blijkt uit de manier waarop de student onzekerheden onderzoekt, alternatieven vergelijkt, grenzen stelt en het uiteindelijke besluit onderbouwt.

In de meest gevorderde fase draagt de student verantwoordelijkheid voor een resultaat dat voor anderen betekenis heeft. Dat hoeft niet meteen een volledig zelfstandig beroepsproduct te zijn. Het kan ook gaan om een advies aan een echte opdrachtgever, een ontwerp dat door gebruikers wordt beproefd of een simulatie waarin beslissingen zichtbare gevolgen krijgen. Doorslaggevend is dat de student niet meer uitsluitend laat zien *dat* een methode wordt beheerst, maar ook *wanneer* die methode passend is, *wanneer niet* en wat er moet gebeuren als de uitkomst anders uitpakt dan verwacht.

Niet alle onzekerheid is leerzaam

Meer onzekerheid toevoegen is overigens niet vanzelfsprekend beter. Een onduidelijke opdracht kan studenten uitdagen om het vraagstuk te verkennen, maar kan ook eenvoudigweg slecht ontworpen zijn. Het verschil zit in de bedoeling en de begeleiding. Productieve onzekerheid vraagt professioneel oordeel. Onbedoelde onduidelijkheid dwingt studenten vooral te raden wat de docent bedoelt.

Curriculumteams moeten daarom vooraf bepalen welke onzekerheid onderdeel is van de beroepssituatie. Gaat het om onvolledige informatie, veranderende omstandigheden, botsende waarden of een opdrachtgever die zijn eigen vraag nog niet scherp heeft? Dan hoort ook duidelijk te zijn welk gedrag van de student wordt verwacht. Denk aan doorvragen, aannames expliciet maken, ontbrekende gegevens verzamelen, risico's benoemen of een besluit voorlopig uitstellen.

Hetzelfde geldt voor tijdsdruk en samenwerking. Tijdsdruk is pas didactisch zinvol wanneer studenten leren prioriteren. Zij moeten dan kunnen uitleggen welke analyses zij wel uitvoerden, welke zij achterwege lieten en welk risico daarmee ontstond.

Samenwerking krijgt pas professionele betekenis wanneer studenten van elkaar afhankelijk zijn, informatie moeten overdragen en elkaar op de kwaliteit van besluiten mogen aanspreken. Alleen een groepsopdracht geven, voert de complexiteit dus nog niet op.

Beoordeel de kwaliteit van het handelen

Naarmate situaties complexer worden, volstaat de beoordeling van het eindproduct steeds minder. Twee studenten kunnen tot hetzelfde advies komen, terwijl de een zorgvuldig heeft afgewogen en de ander toevallig goed heeft gegokt. Omgekeerd kan een verdedigbare aanpak door onverwachte omstandigheden tot een minder gunstige uitkomst leiden.

De opleiding moet daarom ook kijken naar het proces van professioneel handelen:

- Welke informatie verzamelt de student en wat laat hij bewust buiten beschouwing?
- Welke belangen en risico's worden herkend?
- Hoe worden aannames en onzekerheden zichtbaar gemaakt?
- Wanneer vraagt de student hulp of deskundige tegenspraak?
- Hoe worden keuzes binnen het team afgestemd?
- Kan de student uitleggen welke beroepsnormen of publieke waarden de doorslag geven?
- Wordt een besluit herzien wanneer nieuwe informatie daartoe aanleiding geeft?

Daarmee verschuift de beoordeling van het reproduceren van een oplossing naar het onderbouwen van een handelwijze. Dat is wezenlijk voor het hbo, want ook in de beroepspraktijk zijn problemen zelden volledig afgebakend en blijven gevolgen niet beperkt tot één opdracht.

Voer de complexiteit beheerst op

Voor een opleidingsmanager draait de bestuurlijke vraag dan ook niet alleen om het eindniveau. Minstens zo belangrijk is of het curriculum een begaanbare route daarnaartoe biedt. Als onzekerheid, tijdsdruk, samenwerking en verantwoordelijkheid allemaal tegelijk worden toegevoegd, is achteraf nauwelijks vast te stellen waardoor een student vastloopt. De complexiteit moet daarom doelgericht en gefaseerd toenemen.

Per fase hoort duidelijk te zijn welke nieuwe moeilijkheid erbij komt, welk professioneel gedrag daarbij past en welke ondersteuning nog beschikbaar is. Vervolgens kan worden vastgesteld of studenten die steun gebruiken om beter te handelen of er afhankelijk van worden. Dat laatste is bij AI-toepassingen een reëel risico: een systeem dat ieder probleem structureert, alle alternatieven aandraagt en voortdurend verbeterpunten noemt, kan een ingewikkelde opdracht hanteerbaar maken zonder dat de student zelf met die complexiteit leert omgaan.

De volgende ontwerpstep is daarom onlosmakelijk verbonden met het opvoeren van complexiteit. Naarmate studenten meer verantwoordelijkheid dragen, moet de ondersteuning niet automatisch meegroeien. Soms is aanvankelijk juist veel structuur nodig. Later moet die structuur worden verminderd, veranderd in kritische tegenspraak of bewust worden achtergehouden. Alleen dan wordt zichtbaar of de student werkelijk zelfstandiger leert handelen wanneer er geen eenvoudig of eenduidig antwoord voorhanden is.

16.3 Ondersteuning op- en afbouwen

Ondersteuning afbouwen betekent niet dat AI steeds minder wordt gebruikt. De functie verandert. In een vroege leerfase helpt AI de student om een taak te overzien, begrippen toe te passen en fouten snel te herkennen. In een latere fase kan hetzelfde systeem juist twijfel zaaien, tegenargumenten geven of weigeren een oplossing aan te dragen. Niet de hoeveelheid interactie, maar de aard van de ondersteuning bepaalt dan of de student zelfstandiger wordt.

Eerst houvast bieden

Beginnende studenten hebben vaak nog onvoldoende kennis en ervaring om een complexe beroepsopdracht zelfstandig te structureren. AI kan in die fase houvast bieden door bijvoorbeeld:

- een opdracht op te delen in herkenbare stappen;
- vakbegrippen in de context van de opdracht uit te leggen;
- controlevragen te stellen;
- voorbeelden en tegenvoorbeelden te geven;
- directe terugkoppeling te geven op een eerste poging;
- zichtbaar te maken welke informatie nog ontbreekt.

Deze ondersteuning is vooral waardevol wanneer zij de student aanzet tot een leerhandeling. Een uitleg heeft pas betekenis als de student die vervolgens moet toepassen. Een stappenplan draagt bij aan leren wanneer de student gaandeweg begrijpt waarom de stappen nodig zijn. En terugkoppeling helpt alleen als de student daarmee een nieuwe poging doet.

Docenten moeten daarom vanaf het begin weten welk deel van de ondersteuning later kan verdwijnen. Een hulpmiddel dat geen mogelijkheid biedt om structuur, aanwijzingen of directe terugkoppeling te verminderen, past slecht bij een leerlijn die op zelfstandigheid is gericht.

Daarna de denктаak teruggeven

Zodra studenten de basis beheersen, moet meer denkwerk bij hen komen te liggen. De toepassing kan dan stoppen met het voorstellen van stappen en in plaats daarvan vragen waarom de student voor een bepaalde aanpak kiest. Ook kan die onvolledige

informatie geven, meerdere verdedigbare alternatieven tonen of een eerdere aanname ter discussie stellen.

De verschuiving verloopt bijvoorbeeld van:

Vroege ondersteuning	Latere ondersteuning
De taak in stappen verdelen	De student zelf een aanpak laten formuleren
Uitleg en voorbeelden geven	Om een onderbouwing en afweging vragen
Fouten direct aanwijzen	Alleen signaleren dat er iets niet klopt
Een passende vervolgstap voorstellen	Meerdere opties met tegenstrijdige gevolgen tonen
Bevestigen dat een antwoord voldoet	Kritische tegenargumenten geven
Ontbrekende informatie aanvullen	De student laten bepalen welke informatie nodig is

Daarmee verandert AI van een gids in een kritische gesprekspartner. Het systeem helpt niet langer voornamelijk om verder te komen, maar om de kwaliteit van het eigen handelen te onderzoeken. De student moet keuzes toelichten, alternatieven afwegen en reageren op tegenspraak.

Ook kritische ondersteuning moet worden gedoseerd. Soms kan het systeem één bezwaar noemen en de student om twee andere bezwaren vragen. In een volgende fase krijgt de student alleen de opdracht om de eigen redenering te beproeven, zonder dat AI daarvoor inhoudelijke suggesties levert.

Simuleren zonder het handelen over te nemen

Bij complexere beroepssituaties verschuift de rol verder naar simulatie. AI kan optreden als cliënt, collega, opdrachtgever of belanghebbende en reageren op wat de student zegt en doet. De ondersteuning bestaat dan niet meer uit uitleg tijdens de taak, maar uit de mogelijkheid om een situatie te oefenen en achteraf terug te kijken.

Dat sluit beter aan bij professioneel handelen. In een echt gesprek ontvangt een student immers ook geen stappenplan dat precies op het juiste moment verschijnt. Hij moet luisteren, relevante signalen herkennen, reageren en zijn aanpak zo nodig aanpassen. Een simulatie kan die druk benaderen, mits de toepassing niet ondertussen alsnog antwoorden influistert.

Naarmate het vereiste niveau stijgt, kan de simulatie minder voorspelbaar worden. De gesprekspartner geeft niet alle informatie prijs, reageert kritisch of brengt een nieuw belang in. Terugkoppeling volgt pas na afloop en wordt beperkter: eerst uitgebreid en criteriumgericht, later vooral in de vorm van vragen die de student tot zelfbeoordeling aanzetten. Bij risicovolle of ethisch gevoelige situaties blijft bespreking met een docent nodig. AI kan een oefensituatie bieden, maar niet zelfstandig bepalen of het professionele oordeel van een student toereikend is.

Bewuste terughoudendheid ontwerpen

De laatste stap is niet een slimmere vorm van hulp, maar het bewust achterwege laten daarvan. Studenten voeren een taak uit zonder aanwijzingen, suggesties of onmiddellijke terugkoppeling. Pas daarna wordt zichtbaar welke kennis en werkwijzen zij zelfstandig kunnen inzetten.

Die terughoudendheid moet worden ontworpen. Alleen zeggen dat studenten AI niet mogen gebruiken, is daarvoor meestal onvoldoende. De opleiding moet bepalen tijdens welke onderdelen ondersteuning beschikbaar is, welke functies worden uitgeschakeld en wanneer studenten hun handelen zonder digitale hulp moeten aantonen. Ook moeten zij leren beoordelen wanneer het professioneel verantwoord is om AI juist wél te raadplegen. Zelfstandigheid betekent in de beroepspraktijk immers niet dat iemand nooit hulpmiddelen gebruikt, maar dat hij zelf verantwoordelijk blijft voor de keuze, de controle en de gevolgen.

Een werkbaar leerlijn kent daarom geen eenvoudige beweging van veel naar weinig AI. De ondersteuning verandert van structuur naar vragen, van vragen naar tegenspraak, van tegenspraak naar simulatie en van simulatie naar het doelbewust achterwege laten van hulp. Na een zelfstandige uitvoering kan AI opnieuw worden ingezet voor vergelijking of reflectie, zolang de student eerst een eigen analyse heeft gemaakt.

Voor opleidingsmanagers is vooral van belang dat deze afbouw niet wordt overgelaten aan afzonderlijke docenten of aan de standaardinstellingen van een toepassing. Per fase moet duidelijk zijn welke ondersteuning studenten krijgen, welke professionele bekwaamheid daarmee wordt opgebouwd en wanneer de steun verandert of verdwijnt. Daarbij hoort ook bewijs: leveren studenten zonder hulp betere analyses, beter onderbouwde besluiten of geloofwaardiger professioneel handelen dan vóór de oefenreeks?

Eigen analyse naast AI-feedback

Eerst zelf analyseren, dan AI-feedback



16.4 Besluiten over de leerlijn

Dat vraagt om een besluit op het niveau van de leerlijn. Niet iedere toepassing die in een vak goed werkt, verdient immers een vaste plaats in het curriculum. De relevante vraag is niet of studenten en docenten het hulpmiddel waarderen, maar welke professionele bekwaamheid ermee wordt opgebouwd, op welk moment dat gebeurt en wat studenten daarna zonder die ondersteuning kunnen.

Een opleiding moet daarvoor eerst de ontwikkellijn expliciet maken. Welke beroepssituaties moeten studenten achtereenvolgens leren hanteren? Welke afwegingen, handelingen en vormen van professioneel oordeel horen daarbij? En hoe neemt de complexiteit toe? Pas daarna komt de vraag welke rol AI in elke fase kan vervullen.

Daarmee verandert ook het niveau waarop de opleiding besluit. Het gaat niet langer om een verzameling bots, toepassingen of licenties, maar om een samenhangende reeks leerhandelingen. Een uitlegtoepassing kan bijvoorbeeld voorbereiden op een eerste analyse, een oefenbot kan variatie en herhaling bieden, een simulatie kan professioneel handelen onder druk uitlokken en een kritische gesprekspartner kan studenten dwingen hun keuzes te verantwoorden. Afzonderlijk zijn dat hulpmiddelen. Samen kunnen ze een leerlijn ondersteunen – mits iedere toepassing een aantoonbare functie heeft en de overgangen ertussen bewust zijn ontworpen.

Koppelen aan bekwaamheid en docentrol

Bij iedere toepassing moeten twee koppelingen helder zijn. De eerste is die met de beoogde professionele bekwaamheid. De tweede is die met de rol van de docent.

De koppeling met bekwaamheid voorkomt dat gebruiksactiviteit wordt aangezien voor leren. Veel gesprekken met een bot, veel gegenereerde oefenvragen of een snel verbeterd beroepsproduct zeggen op zichzelf weinig. Relevant is of de student beter leert analyseren, beslissen, communiceren, handelen en verantwoorden. Dat moet uiteindelijk blijken uit werk waarin de student zelf verantwoordelijkheid draagt.

De koppeling met de docentrol voorkomt dat begeleiding ongemerkt verschuift naar de toepassing. In een vroege fase kan de docent vooral ontwerpen en bewaken: de oefenomgeving afbakenen, geschikte voorbeelden kiezen en misvattingen signaleren. Later kan de docent juist observeren, doorvragen en het professionele oordeel aanscherpen. Bij complexe simulaties beoordeelt de docent niet alleen de uitkomst, maar ook hoe de student tot keuzes kwam en hoe die omging met onzekerheid, belangen en onverwachte reacties.

Dat leidt per curriculumfase tot een andere taakverdeling:

Fase in de leerlijn	Mogelijke bijdrage van AI	Rol van de docent	Wat de student moet laten zien
----------------------------	----------------------------------	--------------------------	---------------------------------------

Oriëntatie en basisvorming	Uitleg, structuur en laagdrempelige oefening	Begrippen afbakenen en misvattingen corrigeren	Basiskennis toepassen en eigen redeneringen formuleren
Gerichte beroepsuitoefening	Variatie, feedback en herhaling	Voortgang volgen en feedback helpen duiden	Een passende aanpak kiezen en verbeteren
Complexe beroepssituaties	Tegenspraak, rollen en simulaties	Professionele normen, dilemma's en gevolgen bespreken	Afwegen, handelen en keuzes onderbouwen
Zelfstandige uitvoering	Beperkte of geen ondersteuning	Observeren, beoordelen en verdiepend bevragen	Zonder sturing geloofwaardig professioneel handelen
Reflectie en transfer	Vergelijken, alternatieven onderzoeken en reflectievragen stellen	Generaliseren naar nieuwe situaties	Ervaringen vertalen naar een nieuwe context

Deze tabel is geen voorgeschreven volgorde van toepassingen. Wel maakt zij zichtbaar dat dezelfde toepassing niet gedurende de hele opleiding dezelfde onderwijswaarde behoudt. Een bot die eerstejaars houvast geeft, kan bij ouderejaars juist zelfstandigheid in de weg staan. Omgekeerd kan een kritische simulatie in het begin te veel vragen, terwijl die later precies de onzekerheid oproept waarmee studenten in hun beroep moeten leren omgaan.

Behouden, aanpassen of laten vervallen

Curriculumbesluiten vragen daarom meer dan een inventarisatie van wat er al wordt gebruikt. Per toepassing zijn ten minste vier vragen nodig:

1. Welke leerhandeling wordt ermee versterkt?

Als niet duidelijk is wat de student méér, vaker of beter doet, ontbreekt een didactische reden voor opname.

2. Aan welke professionele bekwaamheid draagt die leerhandeling bij?

De toepassing moet verbonden zijn met een herkenbare stap in de ontwikkeling, niet alleen met gemak of tevredenheid.

3. Welke menselijke begeleiding blijft nodig?

Er moet vaststaan wanneer de docent ontwerpt, volgt, ingrijpt, bespreekt of beoordeelt. Ook moet duidelijk zijn wat de student zelf moet doen voordat AI wordt ingezet.

4. Wat kan de student na afloop zonder deze ondersteuning?

Dit is de beslissende vraag. Een hulpmiddel dat langdurig gebruik uitlokt maar geen zelfstandiger handelen oplevert, creëert mogelijk afhankelijkheid in plaats van bekwaamheid.

Op basis hiervan kan een opleiding een toepassing behouden, aanpassen of laten vervallen. Behouden ligt voor de hand als de bijdrage aan de leerlijn duidelijk is, studenten aantoonbaar vooruitgaan en de docentrol uitvoerbaar blijft. Aanpassen is nodig wanneer het onderwijsdoel wel deugt, maar de timing, instructie, begeleidingsvorm of afbouw niet. Vervallen is verstandig wanneer een toepassing vooral productie versnelt, dezelfde functie vervult als een beter ingebed hulpmiddel of studenten structureel een stap uit handen neemt die zij zelf moeten leren beheersen.

Dat laatste besluit verdient nadrukkelijk ruimte. Experimenten stoppen is geen mislukking wanneer ze duidelijk hebben gemaakt wat niet in de leerlijn past. Integendeel, het voorkomt dat het curriculum dichtslibt met losse hulpmiddelen, elk met eigen instructies, verwachtingen en uitzonderingen. Voor studenten ontstaat dan een onoverzichtelijke verzameling digitale loketten; voor docenten een reeks toepassingen waarvan niemand precies weet welk leerdoel ze nog dienen.

Eén herkenbare ontwikkellijn

Samenhang ontstaat wanneer studenten begrijpen waarom zij AI in de ene fase ruim mogen gebruiken, in een andere fase alleen onder voorwaarden en op een later moment helemaal niet. Die keuzes moeten geen toevallige verschillen tussen vakken lijken, maar herkenbare stappen in hun ontwikkeling naar zelfstandige beroepsbeoefening.

Daarvoor hoeft niet iedere docent met hetzelfde hulpmiddel te werken. Wel moeten docenten dezelfde leerlijn ondersteunen: dezelfde oplopende verwachtingen, duidelijke afspraken over zelfstandigheid en een gedeeld beeld van professioneel handelen. De toepassing is vervangbaar; de beoogde bekwaamheid en de onderwijslogica zijn dat niet.

Het uiteindelijke curriculumbesluit luidt dan ook niet: *welke AI willen we gebruiken?* De opleiding besluit welke beroepsbekwaamheden studenten achtereenvolgens moeten ontwikkelen, welke oefening daarvoor nodig is en waar ondersteuning waarde toevoegt of juist moet verdwijnen. Alleen toepassingen die binnen die redenering een duidelijke plaats hebben, verdienen een plek in het curriculum. De rest kan worden beëindigd – ook als het experiment technisch geslaagd was.

17 Opschalen met blijvende menselijke regie

Wat voor één opleiding een verdedigbare keuze is, is nog niet vanzelfsprekend geschikt voor de hele instelling. Bij opschaling veranderen immers niet alleen de aantallen, maar ook de gevolgen. Een onduidelijke afspraak, beperkte controle of onderschatte beheerstaak die in een pilot nog kan worden opgevangen, kan op grotere schaal de onderwijskwaliteit, werkdruk en positie van studenten en medewerkers raken.

Daarom is opschalen geen technisch vervolg op een geslaagd experiment. Het is een nieuw bestuurlijk besluit. De vraag is niet alleen of een toepassing waarde oplevert, maar ook wie de koers mede bepaalt, wie verantwoordelijk blijft wanneer het systeem tekortschiet en of de instelling voldoende in staat blijft om bij te sturen of te stoppen.

Blijvende menselijke regie vraagt daarbij meer dan een docent die de uitkomsten controleert. Die regie moet herkenbaar zijn in het ontwerp, de besluitvorming en het dagelijks beheer. Pas wanneer invloed, verantwoordelijkheid en toezicht vooraf zijn georganiseerd, kan een instelling opschalen zonder onderweg de zeggenschap over haar onderwijs uit handen te geven.

Menselijke regie vóór opschaling



17.1 Medezeggenschap vóór de koers vastligt

Een opleiding kan dat besluit niet eerst nemen en het daarna ter instemming of advies voorleggen alsof de wezenlijke keuzes al vaststaan. Zodra doelen, werkprocessen en technische uitgangspunten zijn bepaald, blijft voor studenten en medewerkers vaak alleen nog de vraag hoe zij met de toepassing zullen werken. Hun invloed beperkt zich dan tot de uitvoering, terwijl juist in het ontwerp wordt bepaald welke taken veranderen, welke gegevens worden gebruikt en hoeveel ruimte er overblijft voor menselijk oordeel.

Medezeggenschap hoort daarom vóór de koersbepaling te beginnen. Niet omdat ieder experiment eerst een lang formeel traject moet doorlopen, maar omdat betrokkenen invloed moeten kunnen uitoefenen op keuzes die hun onderwijs, werk en positie raken.

Begin bij het vraagstuk

Een gesprek over een concreet onderwijsprobleem levert meer op dan een gesprek over de invoering van een AI-toepassing. Vraag studenten bijvoorbeeld waar zij te weinig kunnen oefenen, op welke feedback zij lang wachten en bij welke opdrachten onduidelijk blijft wat er van hen wordt verwacht. Vraag docenten welke begeleiding veel tijd vraagt, waar zij terugkerende fouten zien en bij welke taken hun deskundige oordeel werkelijk nodig is.

Daarmee verschuift het gesprek van acceptatie naar ontwerp. De centrale vraag is dan niet of studenten en medewerkers ‘mee willen’ met AI, maar welke verbetering zij nodig vinden en onder welke voorwaarden een toepassing daaraan mag bijdragen.

Dat onderscheid is bestuurlijk belangrijk. Wie medezeggenschap pas betreft bij de keuze tussen twee leveranciers, heeft de belangrijkste vraag namelijk al overgeslagen: is een digitale toepassing voor dit vraagstuk überhaupt de beste oplossing? Misschien blijkt extra oefencapaciteit waardevol, maar is een onderwijsassistent, een andere roosterindeling of beter herbruikbaar lesmateriaal doelmatiger. Ook dat is een legitieme uitkomst.

Bespreek de verandering van taken

AI verandert zelden slechts één afgebakende handeling. Als een toepassing eerste feedback geeft op conceptwerk, verandert ook wat de student vóór een gesprek met de docent doet, wat de docent tijdens dat gesprek nog moet beoordelen en welke uitzonderingen extra aandacht vragen. Het zichtbare voordeel – sneller feedback ontvangen – kan dus samengaan met nieuw werk.

Studenten moeten bijvoorbeeld leren hoe zij een antwoord controleren, wanneer zij aanvullende uitleg vragen en hoe zij voorkomen dat zij correcties zonder begrip overnemen. Docenten moeten mogelijk steekproeven uitvoeren, foutpatronen analyseren, instructies bijstellen en meldingen afhandelen. Ondersteunende diensten krijgen er beheer, toegangscontrole en leverancierscontact bij.

Breng deze taakverschuiving vroeg in kaart. Bespreek daarbij ten minste:

- welke handelingen vervallen, veranderen of erbij komen;
- wie verantwoordelijk blijft wanneer het systeem een fout maakt;
- wanneer een student of medewerker moet kunnen terugvallen op een mens;
- welke deskundigheid nodig is om uitkomsten te controleren;
- of de beloofde tijdwinst daadwerkelijk bruikbare ruimte oplevert;
- hoe de gevolgen voor werkdruk en toegankelijkheid worden gevolgd.

Juist medewerkers die het werk uitvoeren, zien vaak eerder dan projectteams waar een ogenschijnlijk kleine uitzondering tot structureel extra werk leidt. Studenten merken op hun beurt waar een toepassing hulp biedt, maar ook waar zij zich gecontroleerd, gestuurd of ongelijk behandeld voelen. Die kennis is geen sluitend effectbewijs, maar wel onmisbare ontwerp-informatie.

Maak gegevensgebruik bespreekbaar

Gegevensgebruik mag geen bijlage bij de technische uitwerking worden. Studenten en medewerkers moeten vooraf kunnen begrijpen welke gegevens worden verwerkt, met welk doel dat gebeurt, wie toegang krijgt en hoelang de gegevens worden bewaard. Ook moet duidelijk zijn of interacties worden gebruikt om prestaties te beoordelen, begeleiding te sturen, modellen te verbeteren of gebruiksprofielen op te bouwen.

Daarbij gaat het niet alleen om formele toestemming of een privacyverklaring. In een onderwijsrelatie is vrije keuze immers niet altijd vanzelfsprekend. Een student kan zich moeilijk aan een toepassing onttrekken wanneer die feitelijk nodig is om een vak te volgen. Een docent kan formeel keuzevrijheid hebben, maar toch druk ervaren wanneer het management een toepassing als nieuwe standaard presenteert.

Een reëel alternatief is daarom een belangrijke ontwerpvoorwaarde. Wie een toepassing niet kan of wil gebruiken, moet weten wat de gevolgen zijn en hoe deelname aan het onderwijs mogelijk blijft. Dat alternatief hoeft niet identiek te zijn, maar het mag studenten of medewerkers ook niet stilzwijgend benadelen.

Bespreek ook toezicht en evaluatie

Medezeggenschap krijgt pas betekenis wanneer betrokkenen invloed hebben op de criteria waarmee een toepassing wordt beoordeeld. Een projectteam kan tevreden zijn omdat het gebruik hoog is, terwijl studenten vooral veel inloggen omdat het systeem verplicht is. Een manager kan tijdwinst rapporteren, terwijl docenten die tijd kwijt zijn aan controle en herstelwerk.

Spreek daarom vóór de proef af welke uitkomsten ertoe doen. Denk aan de kwaliteit van studentwerk, het aantal betekenisvolle oefenmomenten, de snelheid én bruikbaarheid van feedback, de ervaren werkdruk, toegankelijkheid, foutpercentages en het aantal situaties waarin menselijke tussenkomst nodig blijkt. Leg bovendien vast welke signalen aanleiding geven om de toepassing aan te passen of te stoppen.

Studenten en medewerkers moeten tijdens de proef niet alleen ervaringen kunnen

delen, maar ook kunnen zien wat daarmee gebeurt. Dat vraagt om een herkenbare terugkoppeling: welke signalen zijn ontvangen, welke veranderingen volgen daaruit en welke punten worden bewust niet overgenomen? Zonder die terugkoppeling wordt participatie al snel een raadpleging zonder aantoonbare invloed.

Combineer formele en directe betrokkenheid

Formele medezeggenschapsorganen hebben een eigen positie en moeten tijdig beschikken over begrijpelijke informatie over doelen, gevolgen, risico's en voorgenomen besluiten. Alleen formele behandeling is echter niet genoeg. Een centrale raad kan niet alle gevolgen overzien voor een specifieke opleiding, stagecontext of studentengroep.

Combineer daarom verschillende vormen van betrokkenheid. Laat een opleidingscommissie meedenken over de plaats in het curriculum, betrek docenten en ondersteuners bij het herontwerp van taken en nodig verschillende groepen studenten uit om de toepassing onder realistische omstandigheden te beproeven. Besteed daarbij bewust aandacht aan studenten met een functiebeperking, studenten die minder digitaal vaardig zijn en studenten die afhankelijk zijn van hulpmiddelen of aangepaste begeleiding.

Voorkom wel dat een kleine groep enthousiaste voorlopers als vertegenwoordiging van de hele gemeenschap gaat gelden. Zij zijn waardevol voor verkenning, maar hun bereidheid om fouten en ongemak te verdragen zegt weinig over de geschiktheid voor regulier gebruik.

Bestuurlijke toetsvragen

Voordat een koers wordt vastgesteld, moet het bestuur ten minste antwoord kunnen geven op de volgende vragen:

1. Op welk onderwijs- of organisatievraagstuk hebben studenten en medewerkers invloed gehad?
2. Welke veranderingen in taken, verantwoordelijkheden en werkdruk zijn met hen onderzocht?
3. Welke keuzes over gegevensgebruik en toezicht lagen nog daadwerkelijk open?
4. Welke groepen kunnen nadelige gevolgen ondervinden en hoe zijn zij betrokken?
5. Welke criteria bepalen of de toepassing doorgaat, wordt aangepast of stopt?
6. Hoe wordt zichtbaar gemaakt wat er met inbreng uit de medezeggenschap is gedaan?
7. Op welk moment volgt het formele advies- of instemmingstraject?

Vroege medezeggenschap hoeft een zorgvuldig traject niet te vertragen. Zij kan juist voorkomen dat een technisch geslaagde toepassing later vastloopt op werkbaarheid, vertrouwen of publieke waarden. Voorwaarde is wel dat de instelling vooraf duidelijk maakt waar invloed mogelijk is, wie uiteindelijk besluit en wie verantwoordelijk blijft

voor de gevolgen.

17.2 Institutionele verantwoordelijkheid organiseren

Die verantwoordelijkheid eindigt niet bij het formele besluit. Zodra een AI-toepassing deel uitmaakt van onderwijs of bedrijfsvoering, moet duidelijk zijn wie de werking bewaakt, wie fouten herstelt en wie mag besluiten het gebruik te beperken of te stoppen. Zonder die verdeling ontstaat al snel een bestuurlijk niemandsland: de leverancier verwijst naar de instelling, de centrale dienst naar de opleiding en de opleiding naar de individuele docent.

Institutionele verantwoordelijkheid betekent daarom niet dat het bestuur iedere toepassing inhoudelijk beoordeelt. Het betekent wel dat de instelling kaders vaststelt, verantwoordelijkheden toewijst en controleert of toepassingen binnen die kaders blijven. Opleidingen en diensten moeten vervolgens aantoonbaar kunnen uitleggen waarom zij een toepassing gebruiken, welke waarde zij ermee willen bereiken en hoe zij de kwaliteit bewaken en de risico's beheersen.

Verantwoordelijkheid kan niet aan het systeem worden overgedragen

Een AI-toepassing kan uitleg geven, feedback formuleren, patronen signaleren of een voorstel doen. Zij kan echter niet verantwoordelijk zijn voor de onderwijsinhoud, de behandeling van studenten of de gevolgen van een beslissing. Ook wanneer een leverancier spreekt over een 'autonome assistent', blijft de verantwoordelijkheid bij mensen en organisaties liggen.

Dat uitgangspunt lijkt vanzelfsprekend, maar wordt in de praktijk gemakkelijk vertroebeld. Een docent kan een onjuist advies overnemen omdat het systeem doorgaans goede antwoorden geeft. Een student kan een automatisch gegenereerde terugkoppeling als formeel oordeel opvatten. Een manager kan veronderstellen dat een centrale dienst de kwaliteit controleert, terwijl die dienst alleen de technische beschikbaarheid beheert.

Daarom moet iedere toepassing ten minste een herkenbare eigenaar hebben. Die eigenaar hoeft niet al het werk zelf uit te voeren, maar is wel aanspreekbaar op het geheel: doel, werking, kwaliteit, gegevensgebruik, evaluatie en beëindiging. Bij een onderwijsbot ligt dat eigenaarschap doorgaans bij de opleiding of het onderwijsonderdeel, niet bij de leverancier en evenmin uitsluitend bij de docent die ermee experimenteert.

Drie niveaus van verantwoordelijkheid

Een werkbare verdeling onderscheidt drie niveaus.

Niveau

**Primaire
verantwoordelijkheid**

Kernvragen

Bestuur en instelling	Publieke waarden, risicogrenzen, basisvoorzieningen en toezicht	Wat staan we toe, onder welke voorwaarden en hoe grijpen we in?
Opleiding of dienst	Inhoud, werkproces, onderwijskwaliteit en lokale invoering	Welk probleem lossen we op, wie controleert de uitkomsten en wat betekent dit voor studenten en medewerkers?
Medewerker of team	Zorgvuldig dagelijks gebruik en professioneel oordeel	Wanneer wordt de toepassing gebruikt, wat controleert de medewerker en wanneer neemt die het over?

Deze verdeling voorkomt twee ongewenste uitersten. Bij volledige centralisatie moet een centrale commissie beslissen over toepassingen waarvan zij de onderwijscontext onvoldoende kent. Dat vertraagt niet alleen, maar kan ook schijnzekerheid opleveren. Bij volledige decentralisatie maakt ieder team eigen keuzes over leveranciers, gegevens, controle en evaluatie. Dan raakt het overzicht zoek en worden dezelfde risico's telkens opnieuw onderzocht.

Het bestuur stelt dus de grenzen en de gemeenschappelijke eisen. De opleiding of dienst blijft verantwoordelijk voor de inhoudelijke rechtvaardiging. Medewerkers behouden de professionele ruimte én de plicht om in concrete situaties zelf te oordelen.

Publieke waarden vertalen naar uitvoerbare eisen

Begrippen als autonomie, rechtvaardigheid, toegankelijkheid, privacy en transparantie krijgen pas betekenis wanneer ze zichtbaar worden in ontwerp- en beheerkeuzes. Een algemene verklaring dat de instelling 'verantwoord met AI omgaat' is daarvoor niet genoeg.

Rechtvaardigheid vraagt bijvoorbeeld dat een opleiding onderzoekt of bepaalde studentgroepen structureel minder bruikbare feedback krijgen. Toegankelijkheid betekent dat studenten niet worden uitgesloten door taalniveau, bediening, kosten of een verplichte technische voorziening. Autonomie houdt onder meer in dat studenten weten wanneer zij met AI te maken hebben en waar zij terecht kunnen voor menselijke begeleiding. Transparantie vraagt dat helder is waarvoor gegevens worden gebruikt, hoe lang ze worden bewaard en of een leverancier ze inzet om eigen modellen verder te ontwikkelen.

Ook de mogelijkheid om een menselijke beoordeling te vragen hoort daarbij. Die mogelijkheid moet daadwerkelijk bruikbaar zijn. Een bezwaarroute die wel op papier bestaat maar weken duurt, biedt weinig bescherming wanneer een student direct studievertraging oploopt.

Instellingsbrede kaders moeten daarom concreet genoeg zijn om handelen te sturen. Ze leggen ten minste vast:

- welke toepassingen of beslissingen nooit zonder menselijke controle mogen

plaatsvinden;

- welke soorten gegevens wel en niet mogen worden verwerkt;
- welke informatie studenten en medewerkers vooraf ontvangen;
- welke minimale eisen gelden voor toegankelijkheid en gelijke behandeling;
- hoe fouten, incidenten en vermoedelijke benadeling worden gemeld;
- hoe vaak de werking en meerwaarde opnieuw worden beoordeeld;
- wie een toepassing tijdelijk kan stilleggen of definitief kan beëindigen.

Kwaliteitsbewaking omvat meer dan technische werking

Een toepassing die beschikbaar is en snel reageert, werkt technisch. Daarmee is nog niet vastgesteld dat zij onderwijsinhoudelijk deugt. Kwaliteitsbewaking moet daarom meerdere vragen tegelijk beantwoorden.

Geeft de toepassing inhoudelijk verantwoorde uitkomsten? Ondersteunt zij de bedoelde leerhandeling? Blijven studenten zelf analyseren, afwegen en handelen? Is de vrijgekomen tijd werkelijk beschikbaar voor begeleiding of onderwijsontwerp? En wegen de opbrengsten op tegen controlewerk, licentiekosten, scholing, onderhoud en afhankelijkheid van de leverancier?

De opleiding verzamelt daarvoor niet alleen tevredenheidsscores. Ook gebruiksgegevens, steekproeven van uitkomsten, ervaringen van verschillende studentgroepen, werkdruk van medewerkers en signalen uit bezwaar- en incidentprocedures zijn nodig. Bij toepassingen met grotere gevolgen – bijvoorbeeld rond studievoortgang, toetsing of persoonlijke begeleiding – moet de bewijslast zwaarder zijn en de menselijke controle intensiever.

Kwaliteitsbewaking is bovendien geen eenmalige toelatingstest. Modellen, bronbestanden, instellingen en gebruikersgedrag veranderen. Een toepassing die bij de start voldeed, kan enkele maanden later andere antwoorden geven of voor een ander doel worden gebruikt. Periodieke herbeoordeling hoort daarom bij het gewone beheer.

Richt ook de tegenmacht in

Verantwoordelijkheid werkt alleen wanneer kritische signalen de besluitvormers kunnen bereiken. Docenten moeten een toepassing kunnen betwisten zonder als tegenstander van innovatie te worden weggezet. Studenten moeten fouten of ongelijke behandeling veilig kunnen melden. Privacy-, beveiligings- en informatiespecialisten moeten vroeg kunnen adviseren, maar ook kunnen ingrijpen wanneer afgesproken grenzen worden overschreden.

Dat vraagt om een herkenbare escalatieroute. Niet ieder fout antwoord hoeft naar het college van bestuur, maar ernstige of terugkerende problemen mogen evenmin binnen een projectteam blijven hangen. Leg daarom vooraf vast welke incidenten lokaal worden afgehandeld, welke centraal worden gemeld en wanneer het gebruik onmiddellijk wordt opgeschort.

Een onafhankelijk oordeel kan nodig zijn wanneer de toepassing grote gevolgen heeft, wanneer de eigenaar zelf belang heeft bij voortzetting of wanneer studenten moeilijk kunnen aantonen dat zij zijn benadeeld. Tegenmacht is dan geen teken van wantrouwen, maar een normale voorziening voor kwaliteitszorg.

Van kader naar aantoonbare beheersing

Het bestuur hoeft niet te bewijzen dat AI nooit een fout maakt. Dat is niet haalbaar. Het moet wel kunnen aantonen dat de instelling haar verantwoordelijkheid heeft georganiseerd. Daarbij moet zichtbaar zijn wie eigenaar is, welke risico's zijn aanvaard, welke controles plaatsvinden en wat er met bevindingen gebeurt.

Een bestuurlijk dossier bevat daarom geen verzameling algemene beginselen, maar een samenhangend beeld van doel, baten, lasten, gegevensstromen, menselijke controle, evaluatieresultaten en resterende risico's. Ook moet duidelijk zijn welke aanpassing volgt wanneer de baten achterblijven of de risico's toenemen.

Pas wanneer die basis op orde is, kan een instelling verantwoord verder kijken dan afzonderlijke toepassingen. Instellingsbrede opschaling vraagt immers niet alleen dat een pilot kansrijk is, maar ook dat verantwoordelijkheid, toezicht en beheer op grotere schaal blijven werken.

17.3 Besluiten over instellingsbrede opschaling

Instellingsbrede opschaling is pas verantwoord wanneer een toepassing niet alleen in een pilot werkt, maar ook onder gewone omstandigheden beheersbaar blijft. Dat vraagt meer dan een positief evaluatierapport. De eerder vastgelegde verantwoordelijkheden voor eigenaarschap, toezicht en beheer vormen daarbij de basis. Voor het opschalingsbesluit moet de instelling vooral aannemelijk maken dat de baten reproduceerbaar zijn, dat opleidingen de toepassing in hun eigen onderwijs kunnen inpassen en dat uitbreiding stapsgewijs en omkeerbaar kan plaatsvinden.

Opschalen betekent dan ook niet: dezelfde toepassing zo snel mogelijk aan iedereen beschikbaar stellen. Het betekent dat de instelling bewust meer studenten, medewerkers of opleidingen laat deelnemen, terwijl de kwaliteit van het onderwijs en het vermogen tot bijsturen behouden blijven.

Van pilotsucces naar reproduceerbaar bewijs

Een geslaagde pilot kan sterk afhankelijk zijn van omstandigheden die bij uitbreiding verdwijnen. Misschien werkte een bijzonder gemotiveerde docent mee, was de studentengroep klein of loste het projectteam problemen direct op. Ook kan de leverancier tijdens de proef extra ondersteuning hebben geboden die in de structurele dienstverlening niet terugkomt.

Voor een opschalingsbesluit moet daarom duidelijk zijn welke onderdelen van het resultaat overdraagbaar zijn. Daarbij zijn ten minste vijf vragen van belang:

1. **Blijft de onderwijsopbrengst zichtbaar?** Studenten oefenen aantoonbaar meer of gericht, verwerken feedback beter of ontwikkelen een relevante professionele bekwaamheid.
2. **Blijft de toepassing bruikbaar in verschillende omstandigheden?** De werking is niet beperkt tot één docent, één klas of één zorgvuldig geselecteerde groep studenten.
3. **Zijn de volledige lasten bekend?** Naast licenties en invoering zijn ook controle, ondersteuning, scholing, gegevensbeheer, kwaliteitsbewaking en aanpassing van onderwijs meegenomen.
4. **Blijft de uitvoering haalbaar?** De eerder vastgelegde controles en beheertaken passen ook bij grotere aantallen binnen de beschikbare tijd en deskundigheid.
5. **Is uitbreiding omkeerbaar?** Aanpassing, tijdelijke beperking en stopzetting zijn technisch, organisatorisch en contractueel mogelijk.

Hoe groter de gevolgen van fouten, hoe overtuigender het bewijs moet zijn. Voor een vrijwillige oefenbot buiten de formele beoordeling kunnen ervaringen met het gebruik in enkele opleidingen voldoende zijn om gecontroleerd uit te breiden. Voor een toepassing die studieadvies geeft, prestaties beoordeelt of beslissingen over studenten voorbereidt, is veel sterker bewijs nodig. Daar moet niet alleen blijken dat het systeem meestal bruikbaar is, maar ook dat fouten tijdig worden ontdekt en hersteld.

Een gemeenschappelijke basis, lokaal toegepast

Opschaling wordt onnodig duur wanneer iedere opleiding zelfstandig leveranciers selecteert, contracten beoordeelt, gegevensstromen uitzoekt en ondersteuningsmateriaal ontwikkelt. Een gemeenschappelijke basisvoorziening kan dat werk beperken. Denk aan afspraken over toegang, gegevensbescherming, beveiliging, leveranciersbeheer, technische koppelingen, gebruiksregistratie, ondersteuning en minimale kwaliteitseisen.

Die centrale basis mag echter niet uitmonden in één standaardtoepassing voor al het onderwijs. Opleidingen verschillen immers in leerdoelen, beroepscontext, toetsvormen, studentpopulatie en aanvaardbare risico's. Wat bij een opleiding communicatie een bruikbare gesprekssimulatie is, kan in een zorgopleiding extra toezicht vereisen omdat studenten er professioneel handelen in risicovolle situaties mee oefenen.

De instelling organiseert daarom wat gezamenlijk moet, terwijl de opleiding beslist hoe de toepassing didactisch wordt ingebed. De verdeling van verantwoordelijkheden is in de voorgaande paragraaf vastgelegd; bij opschaling gaat het erom of de bijbehorende bevoegdheden, gegevens en middelen ook daadwerkelijk beschikbaar zijn. Dat onderscheid voorkomt twee bekende mislukkingen: een centraal aangeboden hulpmiddel waar docenten nauwelijks gebruik van maken en een verzameling lokale oplossingen die niet meer verantwoord te beheren is.

Scholing en beheer zijn onderdeel van de businesscase

Bij een pilot wordt veel nieuw werk opgevangen door het projectteam. Tijdens

opschaling moet zichtbaar worden hoeveel van dat werk structureel terugkeert. Medewerkers hebben instructie nodig, gebruikers stellen vragen, bronnen moeten worden bijgewerkt en wijzigingen in het model of bij de leverancier moeten worden beoordeeld. Daarnaast zijn er incidenten, klachten en uitzonderingssituaties die niet in de oorspronkelijke proef voorkwamen.

De businesscase moet daarom niet alleen de technische kosten bevatten, maar ook het werk voor onderwijsinhoud, controle, ondersteuning en periodieke beoordeling van de baten. Worden deze taken niet begroot, dan lijkt de toepassing in eerste instantie doelmatig, terwijl het werk feitelijk naar docenten en ondersteuners wordt verschoven.

Scholing moet bovendien aansluiten op de rol van de medewerker. Niet iedereen hoeft te begrijpen hoe een taalmodel technisch is opgebouwd. Een docent moet wel kunnen beoordelen waarvoor de toepassing geschikt is, welke fouten kunnen optreden, wat studenten zelf moeten blijven doen en wanneer ingrijpen nodig is. Beheerders en inhoudelijk eigenaren hebben weer andere kennis nodig. Een algemene bewustwordingsbijeenkomst is daarvoor zelden genoeg.

Gefaseerd uitbreiden en kunnen terugkeren

Tussen pilot en instellingsbrede invoering ligt doorgaans een verstandige tussenstap: gecontroleerde uitbreiding. De toepassing wordt dan beschikbaar voor meer teams of opleidingen, maar nog niet als onomkeerbare standaard ingevoerd. In deze fase wordt zichtbaar of de aangetoonde opbrengsten standhouden en of ondersteuning, beheer en toezicht ook bij grotere aantallen uitvoerbaar blijven.

Het bestuur kan de uitbreiding koppelen aan vooraf bepaalde beslismomenten. Daarbij worden niet alleen gebruik en tevredenheid bekeken, maar ook onderwijsopbrengsten, werkdruk, incidenten, kosten en verschillen tussen studentengroepen. Op basis daarvan kan de instelling verder opschalen, de toepassing beperken, het ontwerp aanpassen of stoppen.

Dat laatste moet een reële mogelijkheid blijven. Een toepassing is niet beheersbaar wanneer gegevens niet kunnen worden geëxporteerd, onderwijsprocessen volledig afhankelijk zijn geworden van één leverancier of medewerkers geen werkbaar alternatief meer hebben. Omkeerbaarheid vraagt dus om contractuele afspraken, overdraagbare gegevens, documentatie en een alternatief werkproces. Ook moet duidelijk zijn wat er gebeurt met eerder opgeslagen gegevens en ontwikkeld onderwijsmateriaal.

Het bestuurlijke opschalingsbesluit

Een positief opschalingsbesluit steunt uiteindelijk op een samenhangend oordeel. De toepassing draagt aantoonbaar bij aan een omschreven onderwijs- of organisatie-doel en die bijdrage blijft buiten de oorspronkelijke pilot overeind. De baten rechtvaardigen de volledige lasten, de gemeenschappelijke voorzieningen zijn beschikbaar en opleidingen kunnen de toepassing verantwoord in hun eigen onderwijs inpassen. Bovendien kan de

instelling gefaseerd uitbreiden, tussentijds bijsturen en de invoering zo nodig terugdraaien.

Daarmee zijn de kernvragen voor het bestuur:

- Welke waarde blijft aantoonbaar overeind buiten de oorspronkelijke pilot?
- Voor welke opleidingen, studenten en medewerkers geldt die waarde wel en niet?
- Onder welke omstandigheden nemen de opbrengsten af of lopen de risico's op?
- Welke taken worden centraal georganiseerd en welke blijven bij de opleiding?
- Wie krijgt nieuw werk door controle, ondersteuning en beheer?
- Welke deskundigheid en tijd zijn daarvoor beschikbaar?
- Welke meetpunten en beslismomenten gelden tijdens de uitbreiding?
- Bij welke signalen wordt de toepassing aangepast, beperkt of stopgezet?
- Kan de instelling zonder onaanvaardbare gevolgen terugkeren of overstappen?

Wie deze vragen niet kan beantwoorden, is nog niet toe aan instellingsbrede opschaling. Dan is verdere beproeving, aanpassing of begrenzing verstandiger. Opschaling is immers geen beloning voor een geslaagde pilot, maar een nieuw besluit met grotere gevolgen. Juist daarom moet het bestuur niet alleen beoordelen of uitbreiding wenselijk is, maar ook of de bewezen waarde tijdens die uitbreiding zichtbaar blijft en het besluit omkeerbaar blijft.

Het bestuurlijke opschalingsbesluit

Vóór instellingsbrede opschaling



18 Van onderwijsbot naar HBO-opleiding

Die organisatie van menselijke regie is niet alleen nodig wanneer een instelling AI opschaaft. Ze is al bepalend bij het eerste kleine experiment. Terugkijkend op vier jaar werken met onderwijsbots zie ik dan ook geen rechte ontwikkellijn van een eenvoudige toepassing naar een steeds geavanceerder systeem. Ik zie een reeks pogingen om studenten gericht te laten leren – met successen, beperkingen en ontwerpen die moesten worden herzien.

De technologie veranderde onderweg, maar de belangrijkste ontwikkeling vond elders plaats. Losse bots werden onderdeel van een samenhangend didactisch concept. Daarin kregen studenten meer mogelijkheden om te oefenen, werden feedbackcycli korter en nam de complexiteit van de beroepshandelingen stap voor stap toe. Tegelijk bleef de docent nodig om richting te geven, kwaliteit te beoordelen en te bepalen wanneer een student zonder ondersteuning moest kunnen handelen.

De eerste bot loste geen onderwijsprobleem op

De eerste experimenten begonnen klein. Een bot kon uitleg geven, vragen beantwoorden of studenten helpen zich op een onderwerp te oriënteren. Dat was aantrekkelijk, omdat studenten niet hoefden te wachten tot een docent beschikbaar was. Ze konden verder wanneer het hun uitkwam en desgewenst nog een vervolgvraag stellen.

Mijn eerste ontwerpveronderstelling was dan ook eenvoudig: als uitleg vaker en sneller beschikbaar kwam, zouden studenten beter worden ondersteund. In de praktijk bleek dat te grof. Ik zag wel dat studenten de bot gebruikten en ermee verder konden, maar daaruit kon ik niet afleiden wat zij leerden. Een antwoord kon helder klinken zonder dat de student het onderwerp werkelijk begreep. Sommige studenten gebruikten de bot om hun denken te verdiepen, andere vooral om snel tot een bruikbaar product te komen. Dezelfde toepassing ondersteunde dus verschillende leerhandelingen – en niet al die handelingen waren wenselijk.

Dat was het eerste keerpunt. Beschikbaarheid, gebruik en een overtuigend resultaat bleken onvoldoende gegevens om onderwijswaarde vast te stellen. Er moest zichtbaar worden wat de student tijdens de interactie deed en wat hij daarna beter of zelfstandiger kon. Daarmee verschoof de ontwerpvrage. Niet langer stond centraal welke antwoorden een bot kon geven, maar wat een student door de interactie moest leren doen. Moest de student begrippen kunnen uitleggen, argumenten beoordelen, een aanpak kiezen, een professioneel gesprek voeren of een beslissing verantwoorden? Pas toen die leerhandeling nauwkeuriger werd omschreven, kon ook de rol van AI worden begrensd.

Van uitleg naar actief handelen

De praktijkmomenten in dit boek laten samen een verbreding zien. Na bots voor uitleg en oriëntatie kwamen toepassingen waarmee studenten konden oefenen, feedback verwerken, gesprekken simuleren en hun keuzes verantwoorden. Daarmee veranderde ook de aard van het leren.

Die verbreding was geen vanzelfsprekende uitbreiding van een geslaagd eerste ontwerp. Ze volgde juist uit de beperking daarvan. Een bot die vooral antwoorden gaf, maakte het te gemakkelijk om activiteit met leren te verwarren. Daarom verschoof het ontwerp naar opdrachten waarin de student zelf eerst iets moest doen: een analyse maken, een keuze onderbouwen, een reactie formuleren of een gesprek voeren. De bot werd minder een antwoordmachine en meer een oefenomgeving.

Ook dat werkte niet zonder meer. Bij een uitlegbot ligt het risico vooral in onjuiste of oppervlakkige informatie. Bij een oefenbot komt daar de vraag bij of de student werkelijk actief nadenkt. In een simulatie moet bovendien worden beoordeeld of de situatie beroepsmatig geloofwaardig is. Zodra AI feedback geeft op een conceptproduct, wordt het belangrijk om vast te stellen of de student beter leert of alleen sneller herziet. Bij reflectie, samenwerking en toetsing komen daar vragen over auteurschap, verantwoordelijkheid en zelfstandig handelen bij.

De beschikbare aanwijzingen verschilden per toepassing. Gebruik liet zien dat studenten de bots wisten te vinden. Hun producten maakten zichtbaar dat zij suggesties konden verwerken. De interacties boden aanknopingspunten om hun aanpak te bespreken. Maar geen van die gegevens bewees op zichzelf dat hun professionele bekwaamheid toenam. Daarvoor moest de student in een volgende taak laten zien dat hij de redenering, keuze of handeling ook werkelijk beheerste.

De bots werden dus niet eenvoudigweg beter. De onderwijsontwerpen werden veeleisender. Iedere nieuwe toepassing maakte zichtbaar dat meer interactie ook nieuw begeleidingswerk oproept. Studenten moeten leren hoe zij antwoorden controleren, feedback beoordelen en afwijkende adviezen herkennen. Docenten moeten kunnen zien hoe een student tot een resultaat is gekomen. Opleidingen moeten bepalen waar zelfstandig gebruik passend is en waar controle of directe begeleiding nodig blijft.

Korte feedbackcycli als didactisch beginsel

Een volgend keerpunt ontstond toen niet het afzonderlijke antwoord, maar de herhaling waardevol bleek. AI kreeg vooral betekenis wanneer studenten meerdere keren konden handelen, terugkoppeling ontvingen en hun aanpak vervolgens verbeterden. De winst zat niet in één goed antwoord, maar in de mogelijkheid om een cyclus vaker te doorlopen.

Zo'n korte feedbackcyclus kan er eenvoudig uitzien: een student maakt een analyse, ontvangt gerichte vragen, past de analyse aan en licht de gemaakte keuzes toe. Bij een gesprekssimulatie kan de student een aanpak uitproberen, de reactie van de gesprekspartner onderzoeken en het gesprek opnieuw voeren. De docent hoeft daarbij niet iedere tussenstap zelf van commentaar te voorzien, maar blijft wel

verantwoordelijk voor het ontwerp van de oefening en voor de beoordeling van wezenlijke vooruitgang.

Daarmee veranderde het onderwijsconcept aantoonbaar. De bot stond niet meer naast het onderwijs als extra informatiekanaal, maar kreeg een afgebakende plaats tussen handelen, terugkoppeling en herziening. Het leerproces bestond niet langer uit één opdracht en één beoordelingsmoment. Studenten konden vaker proberen, fouten onderzoeken en een verbeterde aanpak toelichten. De docent verschoof daarbij van het geven van alle tussentijdse standaardreacties naar het ontwerpen en bewaken van de cyclus en het bespreken van de beslissende momenten.

Die verandering leverde niet automatisch tijdswinst op. Ontwerpen waarin de bot algemene opmerkingen gaf, werkten onvoldoende: studenten konden de reactie dan wel verwerken, maar hoefden niet noodzakelijk te begrijpen waarom een wijziging beter was. Bruikbare feedback moest aansluiten op de leerhandeling, de student aan het denken zetten en ruimte laten voor een eigen afweging. Waar dat niet gebeurde, moest het ontwerp worden aangepast.

Juist dat onderscheid is bestuurlijk van belang. Tijd die een docent niet meer besteedt aan telkens dezelfde standaardfeedback is nog geen gerealiseerde besparing. Die tijd krijgt pas waarde wanneer ze beschikbaar komt voor bijvoorbeeld complexe begeleiding, gezamenlijke beoordeling of verbetering van het onderwijsontwerp. Bovendien ontstaan er nieuwe taken: de werking controleren, instructies actualiseren, afwijkende uitkomsten bespreken en studenten leren verantwoord met de toepassing om te gaan.

Oplopende professionele complexiteit

Gaandeweg werd duidelijk dat afzonderlijke bots niet het passende ordeningsprincipe vormden. Een opleiding moet immers geen verzameling hulpmiddelen aanbieden, maar een samenhangende ontwikkeling van professionele bekwaamheid organiseren. Dat inzicht vormde het volgende keerpunt: niet het bottype, maar het ontwikkelpad van de student moest de samenhang bepalen.

Aan het begin van zo'n ontwikkeling kan een student veel ondersteuning krijgen. De kennisbron is afgebakend, de opdracht is overzichtelijk en de bot stelt vragen die de student op weg helpen. Later worden situaties minder voorspelbaar, bronnen minder eenduidig en belangen tegenstrijdiger. Dan moet de student zelf bepalen welke informatie relevant is, onzekerheid herkennen, keuzes onderbouwen en verantwoordelijkheid nemen voor de uitkomst.

AI kan in beide fasen een rol spelen, maar niet dezelfde. Eerst kan de toepassing het leren structureren en extra oefengelegenheid bieden. Later kan zij juist tegenspraak organiseren, alternatieven aandragen of een complexe beroepssituatie simuleren. Uiteindelijk moet de student ook zonder ongepaste ondersteuning kunnen aantonen wat hij zelfstandig weet en kan.

Deze opbouw vormde een belangrijk uitgangspunt voor de hbo-opleiding Toegepaste AI. De ontwikkeling van de opleiding was daarmee geen opschaling van één succesvolle bot, maar een herordening van de eerdere experimenten rond professionele ontwikkeling. Uitleg, oefening, simulatie, feedback en reflectie kregen pas samen betekenis doordat zij werden verbonden aan opeenvolgende leerhandelingen en een toenemende mate van zelfstandigheid.

AI is in de opleiding zowel onderwerp van professionele deskundigheid als hulpmiddel bij het leren. Studenten leren toepassingen ontwerpen en beoordelen, maar gebruiken AI ook om analyses te oefenen, oplossingen te beproeven, feedback te verwerken en hun professionele oordeel te ontwikkelen. Dat kan alleen geloofwaardig wanneer de opleiding steeds zichtbaar maakt welke bijdrage van de student zelf wordt verwacht. Daarom horen bij de inzet van AI ook momenten waarop studenten hun keuzes moeten toelichten, hun werkwijze moeten verantwoorden en zelfstandig moeten aantonen wat zij beheersen.

Geen bewijs uit één ontwikkelreis

Deze ontwikkellijn is geen universeel bewijs dat onderwijsbots vanzelf tot beter onderwijs leiden. Mijn praktijkervaring laat zien wat mogelijk wordt wanneer toepassingen gedurende langere tijd worden beproefd en aangepast. Ze maakt ook zichtbaar wat niet volstaat: enthousiaste reacties, veelvuldig gebruik of sneller geproduceerde opdrachten bewijzen nog niet dat studenten beter leren.

De beschikbare gegevens hadden dan ook grenzen. Waarnemingen lieten zien hoe studenten de toepassingen gebruikten. Producten en gesprekken maakten zichtbaar welke feedback zij verwerkten en hoe zij hun keuzes toelichtten. De opeenvolgende ontwerpen toonden dat uitleg plaatsmaakte voor actief oefenen, herhaalde feedback en oplopende complexiteit. Maar zonder vergelijking, heldere beoordelingscriteria en vervolgmetingen kan uit die ontwikkelreis geen algemeen effect worden afgeleid.

De betekenis van de casus ligt daarom niet in de bots afzonderlijk. Ze ligt in de concrete opeenvolging van onderwijskeuzes. De eerste bot maakte uitleg beschikbaar, maar liet onvoldoende zien wat studenten leerden. Oefenbots dwongen tot meer activiteit, maar riepen de vraag op of studenten zelf nadachten. Feedbacktoepassingen verkortten de verbetercyclus, maar moesten zo worden ontworpen dat sneller herzien niet werd verward met beter leren. Simulaties brachten het beroepshandelen dichterbij, maar vergrootten tegelijk het belang van geloofwaardige situaties en menselijke beoordeling. Uiteindelijk werden deze lessen samengebracht in een opleiding waarin de ondersteuning afneemt naarmate professioneel oordeel en zelfstandigheid zwaarder wegen.

Voor managers en bestuurders volgt daaruit een nuchtere les. Investeer niet in een verzameling toepassingen zonder didactische samenhang. Vraag welk ontwikkelpad studenten doorlopen, hoe iedere toepassing daaraan bijdraagt en waar menselijke begeleiding het verschil moet blijven maken. Leg bovendien vast welke gegevens

beschikbaar zijn, wat die wel en niet aantonen en hoe opbrengsten, lasten en risico's tijdens dat ontwikkelpad worden gevolgd.

Pas dan groeit een onderwijsbot uit tot meer dan een interessant experiment. Niet doordat de technologie steeds meer overneemt, maar doordat de opleiding steeds preciezer weet welk menselijk handelingsvermogen zij wil ontwikkelen – en welke rol AI daarin wel en niet mag krijgen. De eindafweging is daarmee scherp: zonder aantoonbare bijdrage aan dat ontwikkelpad, uitvoerbare menselijke regie en zicht op de volledige lasten wordt een toepassing niet opgeschaald. Dan moet de instelling haar aanpassen, begrenzen of stoppen.

19 Praktijkcatalogus voor ontwerp en besluitvorming

Die afbakening maakt het mogelijk onderwijsbots niet als losse producten te beoordelen, maar als onderdelen van een onderwijsontwerp. De naam van een bot zegt daarbij weinig. Een ‘coachbot’ kan studenten uitleg geven, vragen stellen, feedback geven of een beroepsgesprek simuleren. Voor een verantwoorde keuze moet daarom eerst duidelijk zijn **wat de student ermee doet, welke professionele bekwaamheid daarmee wordt ontwikkeld en onder welke voorwaarden de toepassing wordt gebruikt.**

Bottypen naar leerhandeling

Onderstaande indeling is bedoeld als startpunt. Bottypen kunnen worden gecombineerd, maar iedere extra functie vergroot ook de complexiteit van het ontwerp, het toezicht en het beheer.

Bottype	Leerhandeling	Beoogde bekwaamheid	Mogelijke baten	Belangrijkste
Diagnosebot	Voorkennis tonen, vragen beantwoorden, aanpak toelichten	Eigen niveau herkennen en vervolgstappen kiezen	Snellere signalering, gerichtere oefenroutes	Schijnnaauwkeurige verkeerde indeling, etikettering
Uitlegbot	Begrippen laten uitleggen, voorbeelden opvragen, doorvragen	Kennis begrijpen en in eigen woorden toepassen	Uitleg op het moment dat die nodig is	Onjuiste uitleg, gebruik, overschot begrip
Cursus- of boekbot	Vragen stellen over afgebakend studiemateriaal	Bronnen raadplegen en vakkennis verbinden	Contextgebonden ondersteuning, minder zoektijd	Verouderde bronnen, onduidelijke bronverwijzingen
Oefenbot	Opgaven maken, redeneringen formuleren, fouten herstellen	Kennis actief gebruiken en aanpakken automatiseren	Meer oefenmomenten, directe terugkoppeling	Oppervlakkige herhaling, onjuiste voorzegging

Rollenspel- of simulatiebot	Beslissen, adviseren, onderhandelen of een gesprek voeren	Professioneel handelen in herkenbare situaties	Veilige oefenomgeving, herhaalbaarheid	Onrealistische reacties, stereotypering emotionele be
Feedbackbot	Een conceptproduct vergelijken met criteria en verbeteren	Kwaliteit beoordelen en feedback verwerken	Kortere verbetercycli, vaker herzien	Generieke feedback, afhankelijkheid uniformering
Reflectiebot	Keuzes, bronnen en werkwijze verantwoorden	Kritisch reflecteren en professioneel oordeel vormen	Meer structuur en diepgang in reflectie	Sociaal wettelijke antwoorden, nagebootste reacties
Samenwerkingsbot	Taken ordenen, argumenten vergelijken, peerfeedback voorbereiden	Samen analyseren, afstemmen en besluiten	Betere voorbereiding en zicht op bijdragen	Ongelijke deelname, onduidelijk auteurschap
Toetsvoorbereidingsbot	Proefvragen beantwoorden en prestaties bespreken	Zelfstandig hiaten herkennen en aanpakken	Gerichte voorbereiding, tijdige bijsturing	Trainen op heren in plaats van leren
Praktijkcoach	Een beroepsvraag analyseren en handelingsopties afwegen	Kennis overdragen naar nieuwe situaties	Ondersteuning dicht bij de beroepshandeling	Onverantwoordelijkheid, advies, gebruik gevoelige gegevens

Een bottype is dus nooit op zichzelf kansrijk. De waarde ontstaat pas door de combinatie van leerhandeling, kennisbron, interactievorm, onderwijscontext en begeleidingsniveau.

Ontwerpkeuzes in samenhang

Gebruik voor iedere toepassing dezelfde ontwerpkaart. Daarmee wordt voorkomen dat een aantrekkelijke demonstratie al als onderwijsconcept wordt behandeld.

Ontwerpdimensie	Kernvraag	Voorbeelden
Onderwijsdoel	Wat moet de student beter leren doen?	Analyseren, feedback verwerken, gesprekken voeren

Leerhandeling	Welk zichtbaar gedrag lukt de toepassing uit?	Uitleggen, oefenen, beslissen, herzien, verantwoorden
Kennisbron	Waarop mag de toepassing haar antwoorden baseren?	Cursusmateriaal, leeromgeving, boeken, beroepsstandaarden
Interactievorm	Welke vorm past bij de handeling?	Tekst, spraak, beeld of een combinatie daarvan
Onderwijscontext	Waar wordt de toepassing gebruikt?	Cursus, project, oefenomgeving, toetsvoorbereiding, praktijk
Begeleidingsniveau	Hoeveel menselijke controle is nodig?	Zelfstandig, achteraf gecontroleerd, alleen onder begeleiding
Beoogde baten	Wie merkt welke verbetering?	Student, docent, opleiding of ondersteunende dienst
Bewijs	Hoe wordt de verbetering vastgesteld?	Gebruik, kwaliteit van werk, voortgang, evaluatie, vergelijking
Volledige lasten	Welk nieuw werk ontstaat?	Controle, bronbeheer, ondersteuning, scholing, incidentafhandeling
Randvoorwaarden	Wat moet vooraf geregeld zijn?	Privacy, beveiliging, toegankelijkheid, medezeggenschap, eigenaarschap

Vul deze kaart niet alleen met verwachtingen. Leg vooraf ook de uitgangssituatie vast. Zonder gegevens over bijvoorbeeld het huidige aantal oefenmomenten, de duur van feedbackcycli of de docenttijd voor begeleiding kan achteraf nauwelijks worden vastgesteld of er werkelijk waarde is ontstaan.

Drie manieren om de samenhang te zien

Dezelfde ontwikkeling kan vanuit drie gezichtspunten worden weergegeven. Geen daarvan is de enige juiste weergave; de keuze hangt af van de bestuurlijke vraag.

De cirkel toont leren als een terugkerend proces:

diagnosticeren → begrijpen → oefenen → handelen → feedback
verwerken → reflecteren → aantonen → opnieuw diagnosticeren

Deze weergave past wanneer een opleiding wil beoordelen of de volledige leercyclus wordt ondersteund. Ze maakt snel zichtbaar waar veel toepassingen zijn ontstaan en waar juist een leemte zit.

De spiraal voegt oplopende complexiteit toe. Studenten doorlopen vergelijkbare leerhandelingen meerdere keren, maar in steeds minder voorspelbare beroepssituaties. Eerst oefenen zij met afgebakende casussen en duidelijke criteria, later met tegenstrijdige belangen, onvolledige informatie en grotere gevolgen. Deze weergave helpt opleidingen te voorkomen dat veel oefenen slechts meer van hetzelfde oplevert.

De dubbele lijn plaatst twee ontwikkelingen naast elkaar. Op de eerste lijn groeit het

handelingsvermogen van de student. Op de tweede lijn veranderen begeleiding, toezicht en institutionele verantwoordelijkheid mee. Naarmate studenten zelfstandiger handelen, hoeft een docent niet iedere tussenstap te controleren. De opleiding blijft echter verantwoordelijk voor de kwaliteit van bronnen, de inrichting van toetsing, gegevensbescherming en de grenzen van verantwoord gebruik.

De instrumenten in dit boek horen bij verschillende fasen en zijn dus geen concurrerende checklists:

Fase	Instrument en doel	Primaire gebruiker	Samenhang
Verkennen	Acht bestuurlijke vragen om het vraagstuk en de gezochte waarde af te bakenen	Manager of bestuurder	Vormt de eerste selectie
Beoordelen	Vier beoordelingsaspecten om kansrijke toepassingen te vergelijken	Opdrachtgever of managementteam	Ordent de uitkomsten van de verkenning
Ontwerpen	Ontwerpkaart en uitgebreid canvas om de gekozen toepassing uit te werken	Projectleider en ontwerpteam	Verdiepen de gekozen richting
Besluiten	Twaalf toetsvragen voor het besluit over een pilot	Eigenaar en besluitvormer	Bundelen eerdere antwoorden tot een beslismoment

De instrumenten vervangen elkaar niet. De acht vragen en vier beoordelingsaspecten ondersteunen de selectie, de ontwerpkaart en het canvas werken die keuze uit en de snelle toets bepaalt vervolgens of het voorstel rijp is voor een pilot.

Snelle toets voor een voorgenomen toepassing

Een voorstel is pas rijp voor een pilot wanneer de volgende vragen concreet kunnen worden beantwoord:

1. Welk bestaand onderwijs- of organisatievraagstuk wordt opgelost?
2. Welke leerhandeling moet vaker, beter of gericht plaatsvinden?
3. Welke professionele bekwaamheid wordt daardoor ontwikkeld?
4. Waarom is AI hiervoor geschikter dan een eenvoudiger alternatief?
5. Op welke afgebakende en actuele kennisbronnen berust de toepassing?
6. Wat mag de bot zelfstandig doen en wanneer moet een medewerker ingrijpen?
7. Welke gegevens worden verwerkt en welke gegevens mogen niet worden ingevoerd?
8. Hoe worden toegankelijkheid, bronvermelding en correctiemogelijkheden geregeld?
9. Welke baten worden verwacht voor student, medewerker en instelling?
10. Welke controle-, beheer- en ondersteuningstaken komen erbij?
11. Hoe wordt vastgesteld of studenten beter leren in plaats van alleen sneller

produceren?

12. Wie beslist na de pilot over stoppen, aanpassen, voortzetten of opschalen?

Een toepassing die op meerdere vragen nog geen duidelijk antwoord heeft, is niet per definitie onbruikbaar. Ze is eenvoudigweg nog niet besluitrijp. Juist dat onderscheid helpt experimenteerruimte te behouden zonder een onbewezen toepassing voortijdig structureel te maken.

20 Addenda

20.1 Toepassingscanvas voor onderwijs-AI

Met die twaalf antwoorden ontstaat nog geen businesscase, maar wel een voldoende scherp vertrekpunt. De volgende stap is om de toepassing zo vast te leggen dat een opleiding haar kan ontwerpen, beproeven en beoordelen zonder dat doel, baten en verantwoordelijkheden onderweg verschuiven. Daarvoor kan een compact toepassingscanvas worden gebruikt.

Toepassingscanvas voor onderwijs-AI

Onderdeel	Vast te leggen
Vraagstuk	Het bestaande probleem, de omvang ervan en de groep die ermee te maken heeft
Gewenste verandering	Wat studenten, medewerkers of de instelling na invoering aantoonbaar beter moeten kunnen of doen
Leerhandeling of taak	De concrete activiteit die vaker, gericht, sneller of met hogere kwaliteit moet plaatsvinden
Professionele bekwaamheid	Het vermogen waaraan de activiteit bijdraagt, zoals analyseren, adviseren, ontwerpen, handelen of verantwoorden
Rol van AI	Uitleggen, vragen stellen, laten oefenen, simuleren, feedback geven, ondersteunen of signaleren
Kennisbron	De afgebakende bronnen waarop de toepassing berust en wie die actueel houdt
Interactievorm	Tekst, spraak, beeld of een combinatie daarvan
Onderwijscontext	Cursus, project, toetsvoorbereiding, formele beoordeling, stage of beroepspraktijk
Menselijk toezicht	Wat zelfstandig mag, wat wordt gecontroleerd en wanneer een docent of andere deskundige moet ingrijpen
Verwachte baten	De beoogde opbrengsten voor student, medewerker en instelling
Volledige lasten	Ontwikkeling, licenties, invoering, ondersteuning, controle, beheer, scholing en evaluatie
Risico's en publieke waarden	Mogelijke schade, ongelijkheid, afhankelijkheid, ondoorzichtigheid en ongewenst gegevensgebruik
Bewijsaanpak	Welke gegevens nodig zijn om gebruik, resultaat, kwaliteit en neveneffecten vast te stellen

Besluitmoment Wanneer en door wie op basis van de pilotresultaten een besluit wordt genomen

Het canvas dwingt niet tot lange beschrijvingen. Integendeel: als het vraagstuk niet in enkele zinnen kan worden afgebakend, is de toepassing vermoedelijk nog te breed. Hetzelfde geldt voor de rol van AI. ‘Studenten ondersteunen’ is geen bruikbare omschrijving. ‘Studenten drie keer laten oefenen met een intakegesprek en na iedere poging gerichte aanwijzingen geven over vraagstelling, samenvatting en doorvragen’ is dat wel.

Zo’n concrete formulering maakt ook zichtbaar wat niet door AI hoeft te worden gedaan. Misschien kan een bestaand oefenformulier hetzelfde doel bereiken. Misschien is extra docenttijd effectiever. Misschien ontbreekt het studenten vooral aan een helder beoordelingskader. De vergelijking met een eenvoudiger alternatief hoort daarom bij iedere serieuze afweging. Een toepassing rendeert immers niet doordat zij werkt, maar doordat zij onder de gegeven omstandigheden meer waarde oplevert dan redelijke alternatieven.

Baten zichtbaar maken op drie niveaus

Dezelfde toepassing kan voor studenten waardevol zijn en voor medewerkers nauwelijks tijd opleveren. Andersom kan een administratieve ondersteuning de werkdruk verlagen zonder rechtstreeks bij te dragen aan het leren. Dat hoeft geen bezwaar te zijn, zolang duidelijk blijft voor wie de baten bedoeld zijn.

Niveau	Mogelijke baten	Voorbeelden van aanwijzingen
Student	Meer oefenkansen, snellere feedback, betere toegankelijkheid, gerichtere ondersteuning en groei van professionele bekwaamheid	Aantal oefenpogingen, kwaliteit van opeenvolgende versies, beheersing in een nieuwe situatie, gebruik door verschillende studentgroepen
Medewerker	Minder herhaalwerk, beter zicht op voortgang en meer ruimte voor complexe begeleiding, onderwijsontwerp en deskundig oordeel	Bestede tijd per taak, aard van de begeleiding, bereikbaarheid voor studenten en nieuwe controlewerkzaamheden
Instelling	Betere kwaliteit, toegankelijker onderwijs, beheersbare groei, minder uitval of doelmatiger ondersteuning	Studieresultaten, verschillen tussen groepen, kosten per gebruiker, continuïteit, beheerlast en afhankelijkheid van leveranciers

Niet iedere pilot hoeft al deze baten tegelijk aan te tonen. Wel moet vooraf duidelijk zijn welke baten centraal staan. Zonder die keuze ontstaat achteraf al snel gelegenheidsbewijs: een team noemt dan het hoge gebruik, de positieve reacties of de snelle productie als succes, terwijl de oorspronkelijke bedoeling misschien was dat studenten beter zouden leren redeneren.

Ook de uitgangssituatie moet worden vastgelegd. Wie tijdwinst wil aantonen, moet weten hoeveel tijd de taak vóór de pilot kostte. Wie meer oefening verwacht, moet vaststellen hoe vaak studenten aanvankelijk oefenden. Wie betere feedback nastreeft, moet omschrijven wat goede feedback in deze context inhoudt. Zonder nulmeting kan een verschil wel worden ervaren, maar moeilijk geloofwaardig worden toegeschreven aan de toepassing.

Een beperkte maar bruikbare meetset

Een pilot hoeft geen omvangrijk onderzoeksprogramma te worden. Er zijn doorgaans vier soorten informatie nodig:

1. **Gebruik:** wie gebruikt de toepassing, hoe vaak, op welke momenten en waarvoor?
2. **Ervaring:** hoe beoordelen studenten en medewerkers de bruikbaarheid, begrijpelijkheid en toegankelijkheid?
3. **Resultaat:** verandert de kwaliteit van het leren, handelen of uitvoeren van de taak?
4. **Lasten en neveneffecten:** hoeveel ontwikkeling, controle, ondersteuning en herstelwerk zijn nodig en welke ongewenste effecten treden op?

Gebruiksgegevens laten bijvoorbeeld zien dat studenten veel gesprekken met een oefenbot voeren. Dat bewijst nog niet dat zij beter leren. Een vragenlijst kan laten zien dat studenten de bot prettig vinden, maar tevredenheid is evenmin een maat voor professionele bekwaamheid. Daarvoor moet worden gekeken naar hun handelen: stellen zij betere vragen, herkennen zij meer relevante informatie, onderbouwen zij hun keuzes sterker en kunnen zij dat ook zonder hulp in een nieuwe situatie?

Een combinatie van gegevens is meestal overtuigender dan één enkele maat. Denk aan gebruiksgegevens, beoordeling van opeenvolgende studentproducten, een korte vergelijking met de uitgangssituatie, tijdregistratie door medewerkers en gesprekken over onverwachte effecten. Daarbij hoeft niet alles kwantitatief te worden gemaakt. Een zorgvuldig beoordeelde beroepsprestatie kan meer zeggen dan een groot aantal oppervlakkige gebruikscijfers.

De meetaanpak moet wel in verhouding staan tot het risico en de voorgenomen schaal. Voor een vrijwillige oefenvoorziening binnen één cursus kan een lichte evaluatie volstaan. Een toepassing die studieadvies beïnvloedt, onderdeel is van formele beoordeling of door duizenden studenten wordt gebruikt, vraagt om sterker bewijs, intensiever toezicht en duidelijkere mogelijkheden om beslissingen te corrigeren.

De volledige lasten meenemen

Bij een businesscase voor AI wordt de prijs van de programmatuur vaak wel opgenomen, terwijl het werk eromheen buiten beeld blijft. Juist daar kunnen de grootste lasten ontstaan. Denk aan het selecteren en opschonen van bronnen, het

schrijven en testen van instructies, het controleren van antwoorden, het afhandelen van vragen, het ondersteunen van docenten, het bewaken van privacy en beveiliging en het aanpassen van de toepassing wanneer het onderwijs verandert.

Maak daarom ten minste onderscheid tussen:

- **eenmalig ontwikkelwerk**, zoals ontwerp, koppelingen, toetsing en scholing;
- **terugkerend onderwijswerk**, zoals toezicht, begeleiding en correctie;
- **technisch en functioneel beheer**, waaronder actualisering, toegangsbeheer en storingsafhandeling;
- **bestuurlijke en juridische lasten**, zoals inkoop, gegevensbescherming, medezeggenschap en leveranciersbeoordeling;
- **overgangskosten**, bijvoorbeeld tijdelijk dubbel werk zolang de oude en nieuwe werkwijze naast elkaar bestaan;
- **afbouwkosten**, wanneer de toepassing moet worden vervangen of beëindigd.

Tijdwinst verdient dezelfde nauwkeurigheid. Niet iedere bespaarde minuut wordt bruikbare professionele ruimte. Als een docent minder eerste feedback geeft maar meer tijd kwijt is aan het controleren van onbetrouwbare aanwijzingen, kan de nettowinst gering of zelfs negatief zijn. En als er wel tijd vrijkomt, moet vooraf worden bepaald waarvoor die wordt ingezet. Pas wanneer de ruimte terechtkomt bij persoonlijke begeleiding, onderwijsontwerp, complexe beoordeling of andere waardevolle taken, krijgt zij bestuurlijke betekenis.

Besluitmoment na een pilot

Leg in het canvas alleen vast wanneer het besluit valt, wie het neemt en welke gegevens daarvoor beschikbaar moeten zijn. De vier mogelijke uitkomsten – stoppen, aanpassen, voortzetten of opschalen – en de criteria die daarbij horen, zijn uitgewerkt in de beslismatrix in het hoofdstuk ‘Opnieuw diagnosticeren’.

Drie manieren om de samenhang te bekijken

De eerder beschreven cirkel, spiraal en dubbele lijn bieden elk een ander perspectief op de samenhang: de plaats van een toepassing in de leercyclus, de oplopende professionele complexiteit en het leerproces van de opleiding zelf. Samen voorkomen deze weergaven dat een bot als los hulpmiddel wordt beoordeeld.

De catalogus biedt daarmee geen ranglijst van goede en slechte bottypen. Dezelfde uitleg-, oefen- of simulatiebot kan in de ene opleiding veel waarde toevoegen en in een andere opleiding vooral extra werk veroorzaken. Doorslaggevend zijn het onderwijsdoel, de gekozen leerhandeling, de kwaliteit van de kennisbron, het beroepsrisico, het bewijs van opbrengst en de inrichting van menselijk toezicht.

Pas wanneer die onderdelen met elkaar kloppen, wordt een technisch werkende toepassing ook een verdedigbare onderwijskeuze.

20.2 De bottypen

De bottypen

Drie samenhangende assen

Een bot wordt in deze catalogus niet gekozen uit één vaste lijst, maar samengesteld uit drie onafhankelijke assen:

1. welke didactische functie vervult de bot;
2. verloopt de interactie via tekst, spraak of beide;
3. welke cursusgegevens en kennisbronnen gebruikt de bot.

Door deze assen te combineren, zijn geen tientallen bijna identieke bottypen nodig en blijft het ontwerp voor docenten en managers uitlegbaar.

Een gesproken oefenmondeling voor Onderzoeksvaardigheden dat de HBO-I-competentiematrix gebruikt, is bijvoorbeeld een mondelingbot met spraak als interactievorm en HBO-I als bron – geen afzonderlijk bottype. Elke bot hoort bij één vakaanbieding en heeft een eigen systeemprompt: de vaste instructie die bepaalt wat de bot doet en waar zijn grenzen liggen. De docent herziet die instructie, bewaart wijzigingen als afzonderlijke versies en bepaalt wanneer de bot voor studenten beschikbaar komt.

20.3 Twintig botypen voor de student

Daarmee ontstaat een beheersbaar ontwerp: niet één alleskunner die op iedere vraag een aannemelijk antwoord probeert te geven, maar een verzameling bots met elk een duidelijke didactische taak. Juist die beperking maakt een bot bruikbaar. De student weet waarvoor hij bedoeld is, de docent kan gericht controleren of hij goed werkt en de opleiding kan per toepassing bepalen welke gegevens, instructies en waarborgen nodig zijn.

Een bottype beschrijft alleen **wat de bot didactisch doet**. De modaliteit en de kennisbron worden daar afzonderlijk aan toegevoegd:

- **Modaliteit:** tekst, spraak of beide.
- **Kennisbron:** bijvoorbeeld Canvas, boeken, de HBO-I-competenties of een afgebakende omgeving in SharePoint.
- **Bottype:** de leerhandeling die wordt ondersteund, zoals oefenen, uitleg krijgen, feedback verwerken of reflecteren.

Een mondelingbot kan dus met spraak werken en de HBO-I-competenties als bron gebruiken. Een theoriebot kan via tekst uitleg geven op basis van een studieboek. Dat zijn geen nieuwe bottypen, maar andere uitvoeringen van hetzelfde type.

Elke bot heeft bovendien een expliciete grens. Die grens voorkomt dat ondersteuning ongemerkt verandert in overname van het leerwerk.

Oriëntatie en diagnose

Aan het begin van een blok moet eerst duidelijk worden wat een student al beheerst, waar begrippen door elkaar lopen en welke ondersteuning nodig is.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
Intakebot	Peilt voorkennis, stelt diagnostische vragen en spoort mogelijke misconcepties op. De uitkomst helpt student en docent bepalen waar extra uitleg of oefening nodig is.	Geeft geen cijfer. De uitkomst is formatief en telt niet mee voor de beoordeling.

Een intakebot is daarmee geen digitale toelatingstoets. Hij maakt de uitgangssituatie zichtbaar, zodat de daaropvolgende leeractiviteiten beter kunnen aansluiten.

Uitleg en begrip

Wie vastloopt op een begrip, cursusonderdeel of instructie hoeft niet te wachten tot het volgende contactmoment. Wel moet helder blijven of de student uitleg zoekt of probeert het denkwerk uit te besteden.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
Cursusbot	Beantwoordt praktische en inhoudelijke vragen over één cursus, zoals de opzet, planning, materialen en werkwijze. Verwijst door wanneer een docent nodig is.	Doet geen uitspraken over cijfers, beoordelingen of persoonlijke omstandigheden.
Theoriebot	Legt een vakconcept uit via een intuïtieve uitleg, een vakdefinitie, een voorbeeld en een tegenvoorbeeld. Past het niveau van de uitleg aan.	Werkt geen opdrachten of toetsvragen uit. De student moet het concept zelf toepassen.
Taalbot	Herformuleert cursusmateriaal in begrijpelijk Nederlands of biedt aanvullende meertalige uitleg. Licht moeilijke vaktermen toe.	Vereenvoudigt de taal, maar niet de vakinhoud. Noodzakelijke begrippen worden niet weggelaten.

Vooraf bij taalbots is die laatste grens belangrijk. Toegankelijker onderwijs mag niet leiden tot een lager inhoudelijk niveau. De bot helpt de student de vaktaal binnen te komen, niet eraan te ontsnappen.

Oefenen

AI kan het aantal oefenmomenten sterk vergroten. Waar een docent niet twintig keer hetzelfde gesprek kan voeren of voor iedere student dagelijks nieuwe vragen kan opstellen, kan een bot dat wel. Meer oefenen levert echter alleen iets op als de student actief moet ophalen, redeneren, beslissen of formuleren.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
Socratesbot	Helpt de student zelf tot inzicht te komen door telkens één gerichte vervolgvraag te stellen.	Geeft het antwoord niet, ook niet bij tijdsdruk of herhaald aandringen.
Casusbot	Laat een student een beroepscasus stap voor stap analyseren en biedt oplopende hints wanneer dat nodig is.	Geeft per casus maximaal drie hints. Daarna volgt verwijzing naar docent of medestudent.
Rollenspelbot	Speelt bijvoorbeeld een klant, patiënt, burger of opdrachtgever, zodat de student een beroepsgesprek kan oefenen.	Blijft in zijn rol totdat de student expliciet aangeeft dat het rollenspel stopt.
Ophaalbot	Biedt korte ophaal oefeningen aan over de huidige en eerdere blokstof, verspreid over de tijd.	Toont een antwoord pas nadat de student zelf een poging heeft gedaan. 'Ik weet het niet' geldt daarbij ook als poging.
Debatbot	Verdedigt consequent een tegenpositie en vraagt door op aannames, bewijs en redeneringen.	Neemt geen eigen moreel standpunt in en maakt duidelijk dat de positie onderdeel is van de oefening.

Deze bots produceren dus niet vooral méér inhoud. Ze organiseren méér leerhandelingen: herinneren, uitleggen, afwegen, reageren en opnieuw proberen.

Feedback op werk

Feedback heeft de meeste waarde zolang een student zijn werk nog kan verbeteren. Daarvoor moet de bot niet het product repareren, maar zichtbaar maken waar de student opnieuw moet kijken.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
Verbeterbot	Benoemt bij een conceptversie wat in de volgende versie beter kan en koppelt dat aan concrete passages of onderdelen.	Geeft geen cijfer of cijfervoorspelling en herschrijft het werk niet.
Rubricbot	Besprekt het werk per criterium en verwijst naar bewijs in het ingeleverde product.	Geeft geen eindcijfer en geen samenvattend totaaloordeel.

Peerfeedbackbot	Besprekt de feedback die een student aan een medestudent heeft gegeven en helpt die specifieker en bruikbaar te maken.	Beoordeelt niet het onderliggende werk van de medestudent.
Codebeoordelingsbot	Besprekt programmacode op correctheid, leesbaarheid, structuur en randgevallen. Geeft aanwijzingen over de plaats en aard van een probleem.	Levert geen werkende vervangende code aan.

De grens tussen feedback en productie moet hier technisch én didactisch worden bewaakt. Anders wint de student misschien tijd, maar verliest hij precies de oefening waarvoor de opdracht bedoeld was.

Toetsing en verantwoording

Studenten moeten de vorm en het niveau van een toets kunnen oefenen. Tegelijk mag een oefenomgeving geen voorspeller van officiële resultaten worden en evenmin de indruk wekken dat de bot beoordelingsbevoegd is.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
Toetsbot	Biedt oefenvragen aan die aansluiten bij het niveau en de vraagvorm van de toets.	Gebruikt geen echte toetsvragen en voorspelt niet of de student zal slagen.
Mondelingbot	Simuleert een mondelinge toets of een criteriumgericht interview, inclusief vervolgvragen en tijdsdruk.	Geeft geen officieel resultaat; het blijft een oefening.
Verdedigingsbot	Laat de student een eigen product of aanpak verdedigen en vraagt door op keuzes, onderbouwing en zwakke plekken.	Beschuldigt de student niet van fraude of ongeoorloofd AI-gebruik.

Reflectie en begeleiding

Niet al het leren speelt zich af in een opdracht of toets. Studenten moeten ook hun ontwikkeling kunnen duiden, een stage bespreekbaar maken en verantwoordelijkheid nemen voor samenwerking.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
----------------	-------------------------	-------------------------------------

Reflectiebot	Brengt reflectie op gang, bijvoorbeeld met STARR of Kolb, en vraagt naar concrete ervaringen, keuzes, gevolgen en leerpunten.	Schrijft de reflectie niet en vult geen ervaringen, gevoelens of motieven in.
Stagebot	Helpt een stageopdracht afbakenen, bereidt gesprekken voor en maakt vragen voor de begeleider concreter.	Vervangt de stagebegeleider niet en verwijst door bij conflicten of onveilige situaties.
Teambot	Ondersteunt een projectgroep bij rolverdeling, werkafspraken, overlegvormen en beginnende samenwerkingsproblemen.	Kiest geen partij en bepaalt niet wie er gelijk heeft.

De twintig bottypen vormen geen rangorde en hoeven ook niet allemaal binnen iedere opleiding beschikbaar te zijn. De passende keuze begint bij de leerhandeling: wat moet de student vaker, beter of zelfstandiger leren doen? Daarna volgen pas de modaliteit en de kennisbron.

Zo kan een opleiding klein beginnen. Niet met een algemeen loket voor alle studentvragen, maar bijvoorbeeld met één intakebot, één rollenspelbot of één verbeterbot waarvan doel, grens en menselijk toezicht vooraf duidelijk zijn. Pas wanneer die combinatie aantoonbaar bijdraagt aan het leren en beheersbaar blijft voor docenten en ondersteuning, is uitbreiding naar andere modaliteiten, bronnen of cursussen aan de orde.

20.4 5 Docent bottypen

Vijf bottypen voor docenten

Voor docenten geldt dezelfde ontwerpregel. Ook hier is een algemene alleskunner zelden de beste keuze. Een docent heeft meer aan een afgebakende bot die één ontwerptaak ondersteunt en duidelijk maakt waar het professionele oordeel begint.

Deze bots zijn niet bedoeld om het werk van de docent over te nemen. Ze leveren een eerste ontwerp, controleren de samenhang of brengen patronen in beeld. De docent beoordeelt de uitkomst, corrigeert waar nodig en blijft verantwoordelijk voor de uiteindelijke onderwijskeuze.

Bottype	Wat doet de bot?	Wat doet de bot bewust niet?
----------------	-------------------------	-------------------------------------

Toetsontwerpbots	Ontwerpt conceptvragen en opdrachten die aansluiten bij een leerdoel, toetsvorm en beoogd niveau binnen de taxonomie van Bloom. Kan varianten maken en helpen om de spreiding over onderwerpen en niveaus te controleren.	Stelt geen definitieve toets samen. De docent blijft verantwoordelijk voor validiteit, representativiteit, normering, beveiliging en inzet.
Rubricontwerpbots	Zet leerdoelen om in conceptcriteria en onderscheidende niveaubeschrijvingen. Controleert of formuleringen observeerbaar zijn en of de niveaus werkelijk van elkaar verschillen.	Bepaalt niet welk niveau voldoende is en stelt geen cesuur of normering vast.
Curriculumbots	Vergelijkt leerdoelen, leeractiviteiten en toetsvormen en signaleert waar de onderlinge aansluiting ontbreekt. Legt uit waarom er mogelijk sprake is van een leemte, doublure of onlogische opbouw.	Beslist niet welke curriculumwijziging nodig is en herschrijft het curriculum niet zelfstandig.
Lesontwerpbots	Maakt een concept voor een les of onderwijsblok, met passende werkvormen, fasering en tijdsindeling. Houdt rekening met het leerdoel, de beginsituatie en de beschikbare contacttijd.	Verwijst niet naar bronnen, literatuur of hulpmiddelen die niet zijn gecontroleerd. Alleen dit bottype mag daarvoor op het web zoeken, zodat een bron vóór vermelding kan worden geverifieerd.
Analysebots	Vat veelgestelde vragen en terugkerende knelpunten uit chatgesprekken samen. Zo ontstaat een lessignaal: welke uitleg ontbreekt, waar lopen studenten vast en welk onderwerp verdient opnieuw aandacht?	Rapporteert uitsluitend op geaggregeerd niveau en toont niets wat tot een individuele student te herleiden is.

Bij bestandsgebruik is terughoudendheid nodig. Alleen bots die een ingeleverd product of onderwijsmateriaal daadwerkelijk moeten bespreken, krijgen een uploadfunctie. Aan docentzijde betreft dat de rubricontwerpbots, curriculumbots en analysebots. Een upload is dus geen standaardvoorziening, maar een bewuste keuze die extra eisen stelt aan gegevensbescherming, bewaartermijnen en toegangsbeheer.

De bestuurlijke vraag blijft daarmee dezelfde als bij studentbots: welk afgebakend werk moet beter of doelmatiger worden uitgevoerd? Pas daarna komt de techniek. De opbrengst zit bovendien niet in het aantal gegenereerde toetsvragen, rubrics of lesplannen, maar in aantoonbaar beter ontwerpwerk én in de professionele ruimte die de docent met de gewonnen tijd werkelijk weet te benutten.

20.5 Methodiek

Diezelfde redenering geldt niet alleen voor afzonderlijke bots, maar voor de opleiding als geheel. Zodra AI het beroep ingrijpend verandert, volstaat het immers niet meer om bestaand onderwijs doelmatiger uit te voeren. Dan moet eerst opnieuw worden bepaald voor welk beroep de opleiding studenten voorbereidt.

De aanleiding: de hybride wereld van mens én agent

Onder invloed van AI – en vooral van systemen die zelfstandig deeltaken kunnen plannen en uitvoeren – verandert de rol van medewerkers. Professionals voeren niet langer vanzelfsprekend alle werkzaamheden zelf uit. Zij geven opdrachten aan digitale agents, verdelen taken, beoordelen resultaten, grijpen in wanneer het misgaat en blijven verantwoordelijk voor wat het gezamenlijke systeem oplevert.

Dat is meer dan een bestaande functie met een extra hulpmiddel. Het is een andere inrichting van het werk. Professionals moeten niet alleen vakinhoudelijk bekwaam zijn, maar ook kunnen bepalen:

- welke taken zij zelf uitvoeren;
- welke taken zij verantwoord aan een agent kunnen overlaten;
- welke informatie en grenzen zij daarbij meegeven;
- hoe zij de kwaliteit van het resultaat beoordelen;
- wanneer menselijk ingrijpen noodzakelijk is;
- wie verantwoordelijk en aanspreekbaar blijft.

Een opleiding die uitsluitend een extra AI-vak aan een bestaand programma toevoegt, loopt daardoor al snel achter. Studenten leren dan wel met een nieuw hulpmiddel werken, maar worden nog steeds voorbereid op het oude beroepsbeeld.

Daarom hebben we niet één vak toegevoegd, maar de hele keten opnieuw doordacht: van de vraag welke professionals we willen opleiden, via wat studenten daarvoor moeten leren en hoe docenten die verandering kunnen dragen, tot het onderwijsmateriaal waarmee zij voor de klas staan. Voor elk van die vier stappen is een eigen methode ontwikkeld.

In deze bijlage volstaat de hoofdlijn. De methoden zijn geen afzonderlijke AI-toepassingen, maar ontwerpstappen waarmee een opleiding de gevolgen van de veranderende beroepspraktijk in samenhang kan onderzoeken.

dit vak aantoonbaar bij aan het nieuwe uitstroomprofiel?

Wat wezenlijk is, blijft behouden. Wat vooral door historie, gewoonte of organisatorische verkaveling is blijven bestaan, kan worden aangepast of losgelaten. Tegelijk worden nieuwe leerhandelingen toegevoegd. Studenten moeten bijvoorbeeld leren een agent doelgericht aan te sturen, tussenproducten te controleren, bronnen te verifiëren, afwijkingen te herkennen en verantwoordelijkheid te nemen voor een resultaat dat zij niet volledig zelf hebben geproduceerd.

Zo verschuift het curriculum zonder zijn fundament kwijt te raken. Niet de beschikbaarheid van een AI-toepassing bepaalt wat er in het programma komt, maar het beroepsprofiel waarvoor de opleiding verantwoordelijkheid draagt.

Stap 3 · De docenten

Een herontworpen curriculum krijgt pas waarde wanneer docenten het begrijpen, kunnen uitvoeren en mede-eigenaar van de verandering worden. In de derde stap wordt daarom het veranderproces voor docenten ingericht. Daarbij gaat het niet alleen om scholing in het gebruik van AI. Minstens zo belangrijk zijn de vragen wat er met het vak van docenten gebeurt, welke deskundigheid belangrijker wordt en welke vertrouwde werkzaamheden een andere betekenis krijgen.

Per docent of docententeam wordt bepaald:

- wat men al beheerst;
- wat men als eerste moet leren;
- welke ondersteuning en oefenruimte nodig zijn;
- welke onzekerheden of bezwaren moeten worden besproken;
- hoe ervaringen uit de uitvoering worden teruggebracht naar het curriculum;
- welke tijd en professionele ruimte voor de verandering beschikbaar zijn.

Juist deze stap wordt bij onderwijsvernieuwing gemakkelijk onderschat. Een nieuw curriculum op papier kan snel worden vastgesteld, maar de dagelijkse uitvoering hangt af van docenten die hun onderwijs opnieuw moeten ontwerpen terwijl hun eigen vak onder hun handen verandert.

Stap 4 · Het materiaal

Readers, opdrachten, casussen, presentaties en tentamens zijn vaak ontwikkeld vóór de huidige hybride beroepswerkelijkheid. Daarmee zijn ze niet automatisch inhoudelijk onjuist, maar ze kunnen wel uitgaan van verouderde taakverdelingen. Een opdracht kan bijvoorbeeld uitsluitend beoordelen wat studenten zelf produceren, terwijl de beroepspraktijk juist vraagt dat zij werk verdelen, resultaten controleren en keuzes verantwoorden.

In de vierde stap herschrijven docenten het bestaande materiaal vanuit de nieuwe

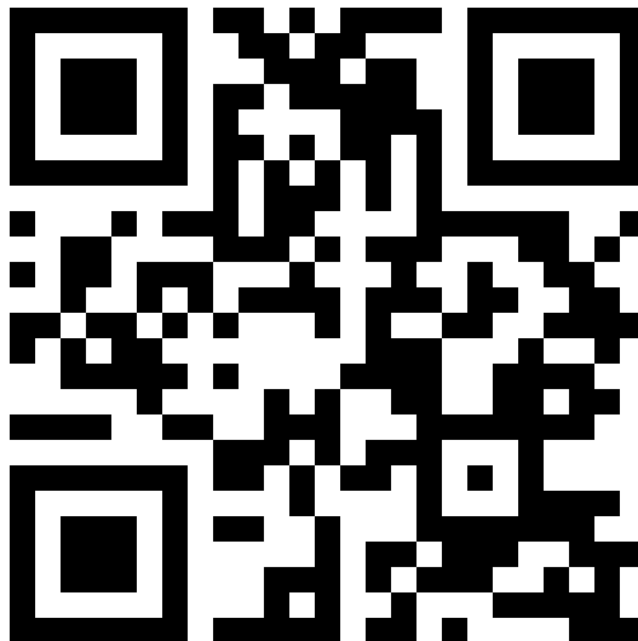
profielen en het herontworpen curriculum. Zij brengen de bronnen, structuur en inhoudelijke uitgangspunten in. AI kan vervolgens helpen om bij grotere hoeveelheden materiaal de begrippen, opbouw en onderlinge samenhang te bewaken. De docent blijft verantwoordelijk voor de inhoud, didactische keuzes, bronkwaliteit en uiteindelijke inzet.

De uitkomst is bovendien niet één vast bestand. Dezelfde inhoud kan worden uitgewerkt als reader, presentatie, werkboek, casus of toetsmateriaal, afhankelijk van wat de les en de leerhandeling vragen. Zonder deze laatste vertaalslag blijft curriculumvernieuwing al gauw op papier staan.

De vier ontwerpstappen vormen daarmee één keten: eerst wordt bepaald voor welk beroep wordt opgeleid, daarna wordt dat vertaald naar het curriculum, worden docenten in staat gesteld de verandering te dragen en krijgt die verandering vorm in het concrete onderwijs. Of die keten werkelijk waarde oplevert, moet vervolgens blijken uit de aansluiting op het beroepsprofiel, de uitvoerbaarheid voor docenten en de kwaliteit van het onderwijs – niet uit de methode of de naam op zichzelf.

Niet een vak erbij. Een ander beroep om voor op te leiden.

Toegepasteai.nl



<https://toegepasteai.nl/>

Waar voegt AI werkelijk waarde toe aan het hoger beroepsonderwijs? Niet iedere veelbelovende proef leidt immers tot beter onderwijs, meer professionele ruimte of een doelmatiger organisatie. Dit boek helpt managers en bestuurders om voorbij de hype te kijken en te beginnen bij het onderwijsdoel, de beoogde baten en de mensen voor wie die baten moeten ontstaan.

Aan de hand van vier jaar onderwijspraktijk laat de auteur zien hoe AI studenten gericht kan laten oefenen, feedbackcycli kan verkorten en professionele bekwaamheid kan ondersteunen. Ook de mislukkingen en het verborgen werk komen aan bod. Een concreet batenmodel en een besliskader maken zichtbaar hoe opbrengsten zich verhouden tot kosten, risico's, beheerlast, publieke waarden en menselijk toezicht.

Zo ontstaat een praktische route van eerste experiment naar verantwoorde opschaling. Met besluitvragen, praktijklessen en een catalogus van bottypen biedt dit boek een nuchter kompas voor wie wil bepalen waar AI rendoert, voor wie en onder welke voorwaarden.

