Ontwerp van de HAI-FIVE AI-tool voor het geven van feedback op instructievideo's van studenten

Ilona Friso-van den Bos, Sara van der Linden, Marieke van Geel, Kim Schildkamp









Ilona Friso-van den Bos *is universitair docent bij de secties ELAN en PLT. Haar werk richt zich op het gebruik van technologie in onderwijs en leren, zoals kunstmatige intelligentie en Virtual Reality. Zij werkt ook als expert AI- en datageletterdheid bij Npuls.*

Sara van der Linden *is universitair docent bij ELAN, sectie voor docentontwikkeling. Haar onderzoek richt zich op de ontwikkeling van (technologische) ondersteuning voor coaching en reflectie van (startende) leraren. Daarnaast werkt zij als expert AI- en datageletterdheid bij Npuls.*

Marieke van Geel *is universitair docent bij de sectie voor docentontwikkeling. Ze doet onderzoek naar (het ondersteunen van de) professionele ontwikkeling van leraren in alle sectoren, met de focus op het versterken van didactisch handelen en de afstemming van het onderwijs op behoeften van leerlingen.*

Kim Schildkamp *is hoogleraar bij ELAN, de sectie voor docentontwikkeling. Haar werk richt zich op het ondersteunen van onderwijsorganisaties in het gebruiken van data en AI. Ook is ze projectleider data- en AI-geletterdheid bij Npuls.*

De auteurs werken aan de Universiteit Twente.

Hoe kan kunstmatige intelligentie (ook wel: Artificial Intelligence; AI) leraren in opleiding helpen hun uitlegvaardigheden te versterken? Dit artikel geeft een kijkje in de keuken van een project waarin die uitdaging wordt aangenomen. Hierbij staat het geven van uitleg met uitlegvideo's centraal. Met behulp van een iteratief ontwerpproces is onderzocht wat een goede uitleg omvat, wat nodig is in de prompting van de AI om goede feedback op de uitlegvideo te geven en hoe deze feedback gepresenteerd kan worden om leraren-in-opleiding goed te ondersteunen.

Inleiding

Het geven van een effectieve uitleg aan leerlingen is een centrale vaardigheid waarover alle leraren moeten beschikken. Leraren in opleiding moeten deze vaardigheid dus oefenen; om dat te doen kunnen zij video's maken voor leerlingen waarin zij stof uitleggen of lesdelen opnemen waarin uitleg centraal staat. Video's worden steeds meer gebruikt in het onderwijs, dus het bewaken van de kwaliteit van deze video's is belangrijk: om het beoogde leerrendement te behalen moet de instructie goed zijn, maar ook de kwaliteit van het multimedia-ontwerp van een video is van belang (Kulgemeyer, 2018).

Om de kwaliteit van de uitleg te waarborgen moeten leraren-in-opleiding feedback op hun video's krijgen van docenten, medestudenten of, in dit onderzoek, van AI. Feedback kan leiden tot kleine maar zichtbare verbeteringen in de kwaliteit van de uitleg en het gebruik van multimedia (Ring & Brahm, 2023). Ook geeft feedback aanleiding tot reflectie, wat een belangrijk onderdeel is van het leren en een aanzet geeft tot het verbeteren van lessen en uitleg (Perez-Ruiz et al., 2025; Van der Linden et al., 2023). Het geven van feedback kost lerarenopleiders echter veel tijd – een probleem dat mogelijk met behulp van AI verlicht kan worden.

Sinds de opkomst van generatieve AI, zoals chatbots, is veel onderzoek gedaan naar de mogelijkheden om AI feedback te laten geven op het werk van studenten, met soms positieve effecten op leeruitkomsten (bijvoorbeeld: Ba e.a., 2025; Dong et al, 2025; Debets et al., 2025). en negatieve effecten: studenten vertrouwen bijvoorbeeld soms te veel op de AI, waardoor ze zelf minder leren (zie Bastani et al., 2024; Fan et al., 2024; Lehmann et al., 2024; Noorozi et al., 2024).

De meeste AI-feedback studies richten zich op chatbots en geschreven werk; AI-feedback op (uitleg)video's is een nog onontgonnen gebied. In dit onderzoek is verkend in hoeverre AI feedback op uitlegvideo's kan verstrekken en reflectie kan stimuleren bij leraren-in-opleiding voor het voortgezet onderwijs.

Onderzoeksvraag en methode

In dit artikel gaan we in op de volgende vragen:

- Op welke criteria met betrekking tot de kwaliteit van instructie is AI-feedback wenselijk volgens lerarenopleiders, en waarom?
- Welke technische mogelijkheden en beperkingen van AI spelen een rol bij het geven van feedback op de door lerarenopleiders geselecteerde criteria?
- Hoe evalueren leraren in opleiding het ontwerp van een tool die AI-feedback geeft op de kwaliteit van uitleg?

Om deze onderzoeksvragen te beantwoorden is een ontwerpgerichte onderzoeksaanpak gehanteerd (McKenney & Reeves, 2019), waarbij in co-design tussen lerarenopleiders, AI-experts en onderwijskundigen een AI-tool is ontworpen. Uiteindelijk is een prototype van een dashboard ontwikkeld waar studenten hun eigen video's kunnen uploaden. Dit prototype doopten wij met de naam **HAI-FIVE:** ***H****igh-quality* ***AI-F****eedback on* ***I****nstructional explanation* ***V****ideos for* ***E****ducation*. De studenten krijgen in deze tool AI-gegeneerde feedback op criteria die van belang zijn voor de kwaliteit van instructie. Om dit te bewerkstelligen hebben wij vier stappen genomen. De eerste vraag, over feedbackcriteria, werd beantwoord in de eerste twee stappen: 1) een literatuuronderzoek naar criteria voor uitlegkwaliteit (analysefase), en 2) het verzamelen van perspectieven van lerarenopleiders op criteria voor uitlegkwaliteit en AI-inzet (analysefase). De tweede onderzoeksvraag, over technische mogelijkheden en beperkingen, werd beantwoord in stap 3) het ontwerpen van de tool met oog voor technische mogelijkheden en beperkingen (ontwerpfase), en de derde onderzoeksvraag, over evaluatie van studenten, werd beantwoord in stap 4) het onderzoeken van het dashboard met perspectieven van leraren in opleiding (evaluatiefase).

Ontwerp van feedbackcriteria

Literatuuronderzoek criteria

De eerste stap in dit onderzoek was het uitvoeren van een literatuuronderzoek naar de eigenschappen van een effectieve uitleg. Hiervoor is een kennisbank aangelegd met wetenschappelijke artikelen over het effect van de manier van uitleg geven in video's op het leerrendement van leerlingen. Vervolgens is een systematisch overzicht gecreëerd van de eigenschappen van uitleg met een bewezen positief effect op het leren, waarbij ieder artikel in de kennisbank meerdere bewijsstukken opleverde.

De volgende stap was het groeperen van eigenschappen van uitlegvideo's. Dit is gedaan door op inductieve wijze een codeerschema te ontwikkelen waarin kenmerken zijn onderverdeeld in vier hoofdcategorieën: Betrokkenheid (bijvoorbeeld cognitieve activatie en aandachtsspanne), Organisatie (bijvoorbeeld visualisaties en structuur), Technische Kwaliteit (bijvoorbeeld kwaliteit van de audio en video) en Vakdidactiek (bijvoorbeeld gebruik voorkennis en voorbeelden). Onder iedere hoofdcategorie vallen verschillende uitgewerkte criteria, voortkomend uit de effectieve kenmerken. Samen zijn deze hoofdcategorieën en criteria uitgewerkt tot een raamwerk van uitlegcriteria, de basis voor de ontwikkeling van de tool. Een volledig overzicht van alle criteria is te vinden in Bijlage 1.

Perspectieven van lerarenopleiders

Vervolgens is een groep lerarenopleiders, afkomstig van een Nederlandse universiteit waarin een lerarenopleiding voor voortgezet onderwijs is gehuisvest, gevraagd om het raamwerk te valideren en een eerste selectie te maken van de criteria. Dit gebeurde in een interactieve sessie waarin het raamwerk is gepresenteerd. De 8 deelnemers vulden een korte vragenlijst in waarin zij aangaven in welke mate zij de opname van ieder criterium in een feedbackproces relevant vonden (antwoordopties: nuttig en essentieel, nuttig maar niet essentieel, en niet nuttig en niet essentieel), en of zij het wenselijk vinden dit door AI te laten doen.

Daarna is de deelnemers gevraagd om gezamenlijk per hoofdcategorie een discussie te voeren over a) de relevantie van ieder criterium binnen de context van de lerarenopleiding, en b) de mate waarin het wenselijk zou zijn dat AI feedback zou geven op ieder van de criteria.

Uit de respons op de vragenlijsten bleek dat de deelnemende lerarenopleiders alle criteria nuttig vonden om aandacht aan te besteden in de lerarenopleiding. Echter waren er verschillen tussen deelnemers over hoe essentieel zij het opnemen van criteria in de tool vonden. Met name de opname van criteria in de categorie vakdidactiek bleken essentieel gevonden te worden – van de 7 criteria die daaronder vallen werden 5 door alle acht de respondenten als essentieel gezien.

Uit de open discussie die werd gevoerd tijdens de sessie bleek dat lerarenopleiders het inzetten van AI voor het geven van feedback als een waardevolle aanvulling (en niet als vervanging) beoordeelden op de rol van de lerarenopleider. Zij stonden echter kritisch tegenover de inzet van uitlegvideo's in het voortgezet onderwijs, en gaven aan dat het geven van effectieve en met name korte uitleg centraal moest staan bij het kiezen van criteria om feedback op te geven (en dus niet het maken van de video als doel). Ook gaven meerdere deelnemers aan dat ze dachten dat AI slechts beperkt in staat zou zijn de kwaliteit van uitleg te beoordelen.

Keuze criteria en technische mogelijkheden van een AI feedback tool

Vervolgens zijn de conclusies van het literatuuronderzoek en de perspectieven van lerarenopleiders vergeleken met een verkenning met AI-experts, bestaande uit een groep informatici van de Universiteit Bremen, rondom de mogelijkheden en onmogelijkheden voor het genereren van feedback met behulp van AI en is de inzetbaarheid van de criteria in een AI tool geanalyseerd. Naar aanleiding van deze analyse zijn vijf criteria geselecteerd om te verwerken in de HAI-FIVE tool:

- **Vakdidactiek - Voorkennis:** Je stemt de inhoud af op de voorkennis en/of dagelijkse context van de leerling
- **Betrokkenheid - Cognitieve activatie**: Je stimuleert de leerling om de lesstof actief te verwerken, bijvoorbeeld door zelfevaluaties in te bouwen of leerlingen volgende stappen te laten voorspellen
- **Organisatie - Visualisaties:** Je gebruikt illustraties, modellen en grafieken om het begrip van de lesstof te bevorderen
- **Organisatie - Aandachtsspanne:** Je spreekt op een gemiddeld tempo. Het Nederlandstalige tempo is 120-130 woorden per minuut. In het geval van een uitlegvideo is de duur niet langer dan 6 minuten.
- **Organisatie - Nadruk:** Je geeft verbale of visuele aanwijzingen om de aandacht te vestigen op de relevante aspecten van de lesstof en lesmaterialen.
  (Nadruk werd door de lerarenopleiders weinig geselecteerd als essentieel criterium, maar door het ontwerpteam gekozen als een technische uitdaging rondom de training van een AI tool. Lerarenopleiders vonden dit criterium wél relevant).

Technische infrastructuur en training van de AI: ontwerp van de tool

Tijdens het HAI-FIVE project zijn, middels principes van ontwerponderzoek, gefaseerd beslissingen genomen over de inrichting van de tool en de inhoudelijke specificaties. Deels waren deze beslissingen sturend voor de feedback die gegeven kon worden op de uitleg in video's die ter analyse werden aangeboden. Zo was bij het uploaden van een video een eerste stap om hier de juiste informatie uit te halen voor de analyse. Dit gebeurde met behulp van een infrastructuur waarbij eerst transcripten werden gemaakt van de gesproken uitleg, waarna AI de tekst kon analyseren. Daarnaast is een tool ingebouwd voor beeldanalyse, los van de gesproken tekst. De combinatie van deze gegevens maakte het mogelijk om op multimodale wijze naar de kwaliteit van uitleg te kijken. Criteria voor het geven van feedback zijn ingedeeld bij een van beide componenten: voorkennis en cognitieve activatie zijn van feedback voorzien middels de transcripten, terwijl visualisaties en nadruk met behulp van beeldanalyse van feedback zijn voorzien. Aandachtsspanne werd door een aparte tool beoordeeld die niet door AI ondersteund werd, omdat dit criterium regel-gestuurd beoordeeld kon worden, wat geen getrainde AI vergt.

Na het ontwerpen van een infrastructuur die diende als basis voor de AI-analyse is gewerkt aan de prompting (geven van opdrachten en instructies) van het AI-model. Dit was een iteratief proces waarbij het ontwerpteam en AI-experts van de Universiteit Bremen herhaaldelijk video's bekeken en beoordeelden, om eigen beoordeling van de criteria daarna te vergelijken met de beoordeling van de AI-tool. Hierbij bleek dat de AI-tool al aan het begin erg redelijk klinkende output gaf, met genuanceerd taalgebruik en concrete verwijzingen naar de transcripten, maar dat specificatie van groot belang was. Zo gaven de eerste AI-iteraties bijvoorbeeld feedback op gerelateerde concepten die niet onder de feedbackcriteria vielen, en werden soms uitspraken aangezien voor bewijs van het toepassen van het criterium, terwijl dit niet het geval was (bijvoorbeeld voor het criterium cognitieve activatie: 'We gaan deze formule nu toepassen' waarbij de kijker de formule niet toe moest passen maar dit alleen werd voorgedaan). Door prompting steeds verder te verfijnen is de feedback in toenemende mate gaan passen bij de beoogde criteria.

Het Dashboard en Perspectieven van Studenten

Gelijktijdig met de verfijning van de gegeven AI feedback is gewerkt aan het HAI-FIVE dashboard (zie figuur 1). Dit bevatte de noodzakelijke elementen, zoals een plek om een video te uploaden, de mogelijkheid om criteria te selecteren, en een plek voor feedback die door de AI wordt gegeven. Daarnaast zijn er opties in het dashboard ingebouwd die de het ontwerpteam relevant vond voor het leerproces: een tijdlijn die de feedback koppelde aan specifieke momenten in de video, een transcript om in op te zoeken waar specifieke feedback over gaat, een reflectiepaneel om de feedback actief te kunnen verwerken in plaats van passief te ontvangen, en een optie om een rapport te genereren om te downloaden. Dit laatste was noodzakelijk omdat het platform geen gebruikersgegevens opslaat vanwege privacy-risico's.

Evaluatie van het dashboard door studenten

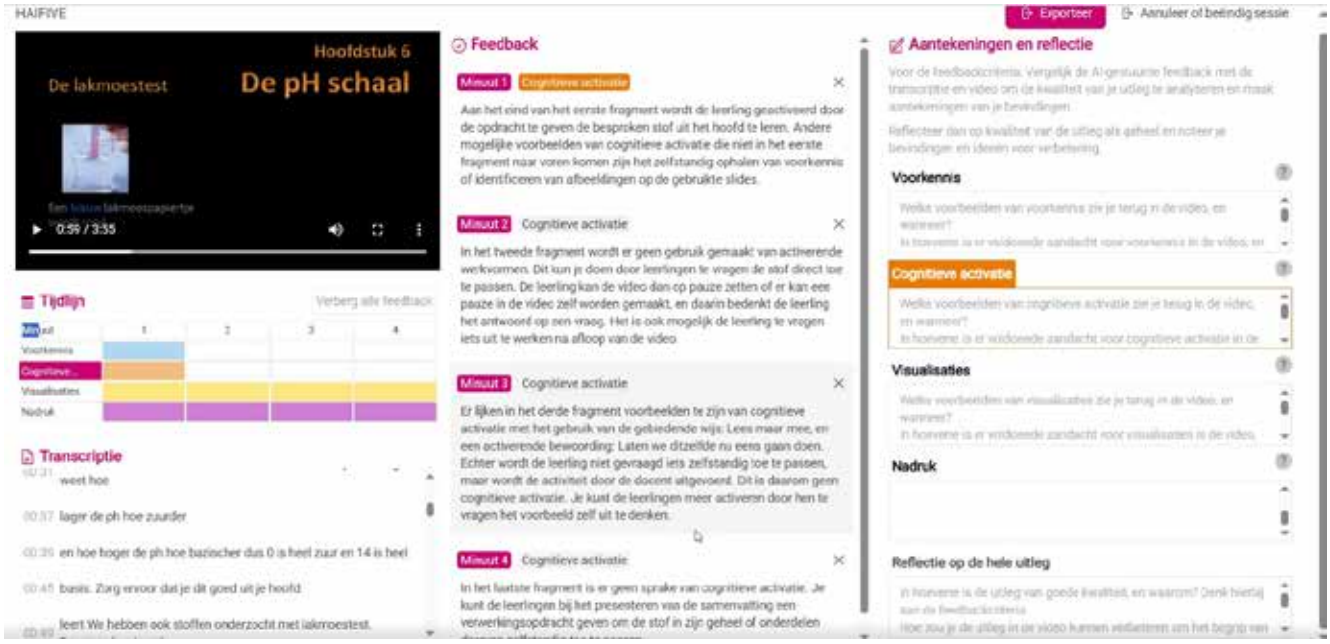
In een daaropvolgende studie namen studenten uit de lerarenopleiding voor voortgezet onderwijs het dashboard onder de loep. Tijdens een anderhalf uur durende bijeenkomst doorliepen 11 studenten de volgende stappen:

1. Onderzoekers presenteerden een introductie op de tool
2. Daarna bekeken de deelnemende studenten een uitlegvideo uit een openbare bron, die vervolgens voorzien zou worden van feedback om te bekijken
3. In kleine groepjes bekeken de deelnemende studenten het dashboard. Zij doorliepen alle stappen: een video uploaden, criteria selecteren, feedbackgeneratie starten, feedback bekijken, reflecteren op de feedback, en het rapport downloaden. Om te zorgen dat alle studenten dezelfde omgeving bekeken, werd de omgeving gefingeerd. Hierbij kregen de deelnemers geen instructie zodat zichtbaar werd in welke mate het dashboard intuïtief bediend kon worden. Zij werden aangemoedigd hardop te denken over de werking van het dashboard en de weergave van alle elementen, om te kunnen analyseren in welke mate zij het dashboard bruikbaar vonden.
4. Na het bekijken van het dashboard vulden de deelnemers een korte vragenlijst in over gebruiksvriendelijkheid
5. Als laatste werd een plenaire discussie gevoerd over de bruikbaarheid van het dashboard, de weergave van de elementen, en de verbeterpunten. Ook was er ruimte voor opmerkingen over de feedback die in het dashboard werd weergegeven.

In het gepresenteerde dashboard zijn vier criteria opgenomen: voorkennis, cognitieve activatie, visualisaties en nadruk. Het opnemen van maximaal vier criteria is nodig om de AI-servers te ontlasten. De focus lag echter niet op de inhoudelijke criteria, maar op de weergave van de feedback en reflectiemogelijkheid.

De gegevens uit deze sessie zijn vervolgens gebruikt voor het formuleren van een herontwerp van het dashboard. Met name de paneldiscussie, waarbij wij de deelnemers gericht vroegen naar de wijze waarop het dashboard was ingericht en hoe zij dat het liefst zagen, leverde relevante inzichten op. Uit deze paneldiscussie bleek onder andere dat:

- de deelnemers bewegende onderdelen van het dashboard, bedoeld om de aandacht naar de belangrijke aspecten te trekken, eerder als negatief beoordeelden dan helpend,
- zij de voorkeur gaven aan een stapsgewijze presentatie van feedback en reflectie, in plaats van een parallelle weergave,
- zij vonden dat labels in grotere mate een neutrale kleur mochten krijgen in plaats van felroze.



Figuur 1 Weergave van het HAI-FIVE Dashboard aan het Begin van de Dashboardstudie

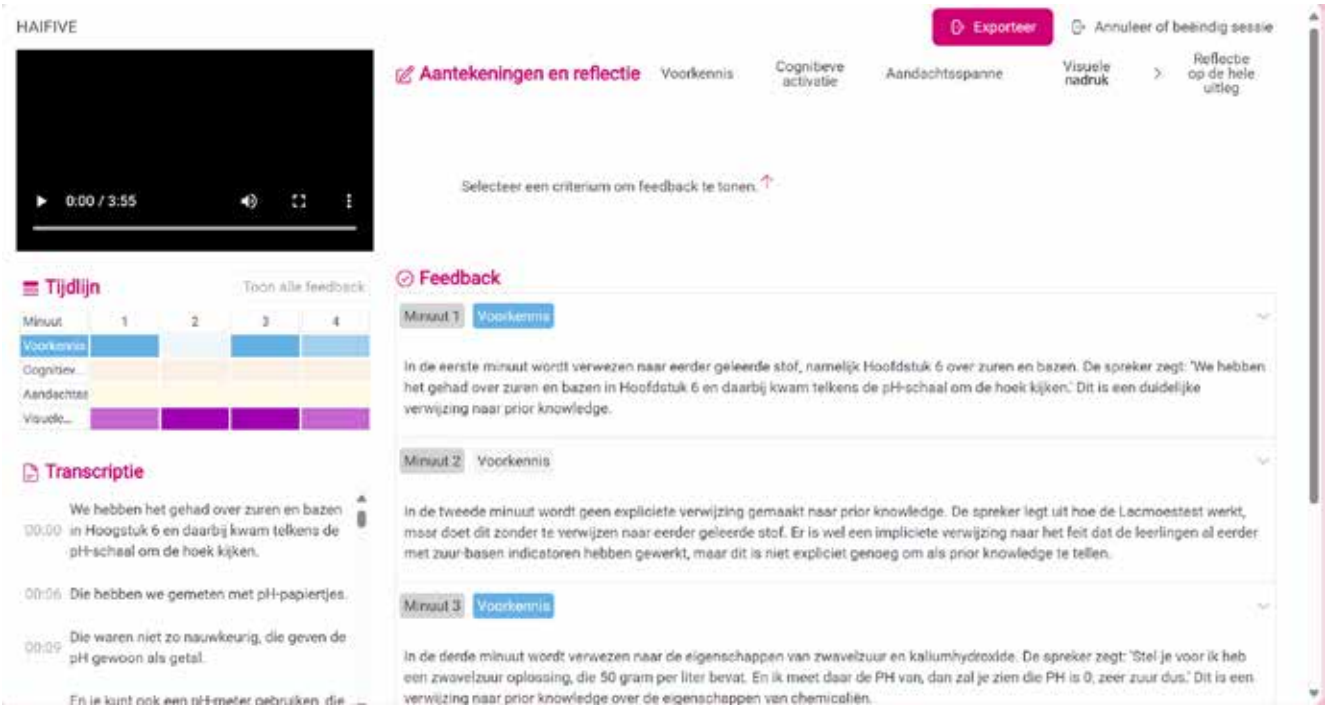
Deze gegevens werden aangevuld met de hardop-denken reacties van de deelnemende studenten op het dashboard tijdens hun verkenning van het dashboard. Deze reacties gaven nauwkeurigere invulling aan de manier waarop het dashboard direct ontvangen werd en wat daaraan gewijzigd moest worden. De bevindingen zijn vervolgens ingezet voor het herontwerp van het dashboard, te vinden in Figuur 2. Bij de huidige weergave is ook gekozen voor een subset van criteria, wat nodig is om de AI-servers te ontlasten, waarbij visualisaties zijn vervangen voor aandachtsspanne. Beide getrainde modellen zijn op de achtergrond beschikbaar voor verdere ontwikkeling.

Conclusie

Het HAI-FIVE project, dat als doel had te onderzoeken of en hoe een AI-tool feedback kan geven op uitlegvideo's, heeft een aantal belangrijke conclusies opgeleverd. Allereerst kan het geven van feedback met behulp van AI waardevol zijn voor studenten en lerarenopleiders. Dit bleek uit het feit dat het binnen een aantal iteraties gelukt is om de AI-tool te trainen om feedback te geven die van acceptabele kwaliteit is (dat is: feedback die inhoudelijk deels of geheel overeenkomt met observaties van het ontwerpteam), en uit het feit dat de lerarenopleiders positief stonden tegenover de mogelijkheid om AI in te zetten voor dit doeleinde.

Daarnaast is gebleken dat afstemming op wensen van de gebruiker van groot belang is. De resultaten van de dashboardstudie wezen uit dat gebruikers zich overweldigd kunnen voelen bij het presenteren van veel gegevens ineens, wat de verwerking van de feedback mogelijk in de weg kan staan (Twabu, 2025). De gebruikersadviezen die voor het dashboard geformuleerd werden draaiden met name om het versimpelen van de visuele weergave om de lerende optimaal te ondersteunen en in de gelegenheid te stellen gericht te reflecteren op de feedback. Toch konden de studenten al tijdens de dashboardstudie met het dashboard uit de voeten – zij wisten alle mogelijkheden betrekkelijk snel te vinden. De verbeteringen die uit de studie volgden hebben geleid tot het ontwerp van een nog beter bruikbaar en overzichtelijker dashboard.

Tot slot hebben we als ontwerpers van de HAI-FIVE tool uit het ontwikkelproces geleerd dat het trainen van AI-modellen voor het geven van goede feedback expertise en iteraties vergt. Bij het ontwerp bleek niet alleen de technische kennis voor het ontwikkelen van front- en backend onontbeerlijk, maar ook de onderwijskundige expertise die nodig was voor het herhaaldelijk formuleren en herformuleren van feedbackcriteria, en specificatie aan de hand van de output van de modellen naar aanleiding van verschillende prompts. De samenwerking tussen leden van het ontwerpteam met verschillende expertises bleek hierbij een belangrijke succesfactor.



Figuur 2 Herontwerp van het HAI-FIVE Dashboard na Afloop van de Dashboardstudie

Notitie. Er zijn verschillende weergaven beschikbaar binnen het dashboard. In de afgebeelde weergave is de categorie voorkennis als geheel weergegeven. Door de onderdelen aan te klikken kan ook feedback op een specifiek criterium voor een specifieke minuut worden weergegeven, alle feedback voor een specifieke minuut, of tekstvelden om reflectie in te voegen.

Discussie

Implicaties voor de wetenschap

Het HAI-FIVE project maakt duidelijk dat doelmatige inzet van AI-technologie in het onderwijs veel kan opleveren. Waar al duidelijk was dat chatbots, die puur werken met aangeleverde tekst, bij juist gebruik het leerproces kunnen bevorderen (Sedrakyan et al., 2024), laat deze pilot zien dat deze mogelijkheid uitgebreid kan worden met het ondersteunen van het leren middels AI-feedback op video's. Hierbij blijft het belangrijk om de technologie toe te spitsen op het niveau en de behoeften van de lerende. Innovatief aan dit project is dat er verschillende tools voor video's, tekst en audio zijn gecombineerd en geïntegreerd in één dashboard. Een belangrijke nuance is dat het trainen van AI-modellen voor specifieke doelen noodzakelijk is om de kwaliteit van al deze verschillende technologieën en van de integratie ervan te borgen. Toch lijkt het verkennen van het scala aan mogelijkheden een waardevolle onderneming die zowel de onderwijspraktijk als de onderwijswetenschappen verder kan brengen. Cruciaal hierbij is dan wel dat dergelijke AI-tools ontworpen worden in een co-design proces in een team van AI-experts, inhoudsexperts en eindgebruikers, met doorlopende feedbackloops en verfijningen.

Implicaties voor de praktijk

Uit het HAI-FIVE project kunnen we voorzichtig concluderen dat het geven van kwalitatief goede feedback met behulp van AI mogelijk en zinvol is, mits de AI specifiek getraind is om op de beoogde criteria feedback te geven. Uit de vergelijking van het verschil in kwaliteit van de eerste iteraties van het project en de laatste iteraties is verschil te zien tussen AI ontwikkeld voor algemene doeleinden en AI getraind voor specifieke doeleinden, waarvan de laatste een betere feedbackkwaliteit oplevert. De beoogde tijdswinst die de inzet van een dergelijke AI-tool kan opleveren, doordat lerarenopleiders minder feedback hoeven geven, wordt alleen behaald wanneer de tool voldoende leerrendement met zich meebrengt. De aanbeveling voor de praktijk is dan ook om bij de inzet van AI, tools te selecteren die voor het beoogde doeleinde ontwikkeld zijn; het is dus belangrijk eerst een doel te formuleren en dan pas een tool te selecteren die bij dat doel past. De HAI-FIVE tool is een voorbeeld van een toegespitste tool.

Een bredere implicatie ligt bij de basis die de tool vormt voor vervolgprojecten. Bij deze vervolgprojecten kunnen door het initiële succes binnen twee instellingen meer doelgroepen betrokken worden en ook de criteria waarop feedback wordt geleverd worden uitgebreid. Dit zijn doelstellingen voor een startend project, waarbij wij hopen de praktijk te kunnen voorzien van een rijke en inzetbare AI-tool voor het geven van feedback op instructiekwaliteit.