



Waarom we dit jaar AI inzetten bij de verwerking van onze papieren toetsen

De Leerling in beeld - doorstroomtoets bevat veel open vragen. Hoe blijf je die zorgvuldig beoordelen als het aantal papieren toetsen groeit? Een interview met Joanne van Emmerik, productmanager van doorstroomtoets Leerling in beeld, over de inzet van AI bij de voorbereiding van de beoordeling.

Joanne, vanaf eind januari wordt de doorstroomtoets weer afgenomen. Hoeveel open vragen moeten er dan worden beoordeeld?

“We verwachten dat komend jaar 70.000 tot 75.000 leerlingen de papieren Leerling in beeld - doorstroomtoets zullen maken. Elke toets bevat ongeveer 65 tot 70 open vragen. In totaal gaat het dan om zo’n 5 miljoen antwoorden. Een groot deel kijken we twee keer na, omdat een mens ook fouten maakt. En daar waar de twee oordelen verschillen, bepaalt een derde beoordelaar de score voor de leerling. Al die toetsen moeten zoveel mogelijk binnen tien werkdagen beoordeeld worden. Die scores heeft CvTE namelijk nodig voor de normering.”

Hoe gaat het beoordelen van open vragen in de papieren toets eigenlijk in zijn werk?

“De afgelopen jaren scanden we de papieren toetsboekjes in en maakten we van elk open antwoord een afzonderlijke, geanonimiseerde uitsnede: een zogenaamde ‘snippet’. Getrainde menselijke beoordelaars kregen die snippets in een nakijktool te zien en beoordeelden die per vraag, niet per leerling of boekje. Daardoor wordt het werk van één leerling verdeeld over zo’n 30 tot 40 beoordelaars. Maar we merkten dat we tegen de grenzen aanliepen van wat menselijke beoordelaars in die korte tijd kunnen verwerken.”



Welke opties had je om die nakijklast en werkdruk te verminderen?

“Een optie was om veel minder open vragen in de toets op te nemen. Maar dat zou voor mij echt een laatste redmiddel zijn. Scholen vinden het belangrijk dat we spelling met een dictee toetsen en werkwoordspelling met een open vraag. En bij rekenen lenen sommige domeinen zich niet goed voor meerkeuzevragen. Daarom hebben we gekeken hoe we de open vragen kunnen behouden en toch alle antwoorden op tijd kunnen beoordelen.”

Waar ben je op uitgekomen? En hoe heb je onderzocht of dat werkt?

“We hebben onderzoek gedaan naar hoe AI ons hierbij kan helpen. Daarbij hebben we gekeken of AI het handschrift op de gescande, geanonimiseerde antwoorden kan omzetten naar digitale tekst. Dit is vergelijkbaar met traditionele OCR-technologie (Optical Character Recognition), maar dan efficiënter en beter. Die digitale tekst wordt vervolgens met onze eigen scoringsmodellen nagekeken, zoals we dat ook al jaren doen met de digitale doorstroomtoets.

Het omzetten gaat erg goed. Voor een doorstroomtoets vinden we het echter belangrijk dat er geen twijfel mag bestaan, en daarom laten we ook alle leerlingantwoorden labelen waar gegumd of gestreept is door de leerlingen. Deze antwoorden laten we expliciet extra door mensen controleren. Zodat er geen enkel risico is dat zo’n antwoord verkeerd wordt gelezen en daarmee gevolgen zou hebben voor de kwaliteit en betrouwbaarheid van de beoordeling. Daarbij beslist altijd een menselijke beoordelaar. AI beslist dus nooit of een antwoord goed of fout is.”

Wat vinden scholen ervan dat je AI inzet in de voorbereiding op het beoordelen?

“Zij vinden het belangrijk dat open vragen in de toets blijven, en ook is de kwaliteit en betrouwbaarheid van de beoordeling heel belangrijk. Dat is voor ons niet anders. Tegelijkertijd vragen ze zich ook af wat er met de handschriften van kinderen gebeurt als we AI inzetten. We gaan daar heel zorgvuldig mee om. De antwoorden die we verwerken zijn geanonimiseerd en niet naar een leerling te herleiden. Het AI-model draait via een beveiligde verbinding op Europese servers. De data worden nooit gebruikt voor modeltraining, en AI speelt op geen enkele manier een rol bij het beoordelen van goede of foute antwoorden.”



Joanne van Emmerik, productmanager van doorstroomtoets Leerling in beeld