Dit conceptartikel (ca. 3.000 woorden) kan als voorbeeld gebruikt worden voor een achtergrondartikel, maar uitsluitend met de uitdrukkelijke toestemming van de eigenaar, Dr. Remco Schimmel

De AI-promovendus

Verbieden, reguleren — of benutten? Wat gebeurt er als AI de schrijver wordt?

Universiteiten steken veel energie in de vraag wat zij met generatieve AI moeten doen. Mogen studenten AI gebruiken in proefschriften? Moet elk gebruik worden vermeld? Waar houdt legitieme ondersteuning op en begint onaanvaardbare vervanging? En als bepaalde toepassingen worden verboden, hoe zijn zulke regels dan te handhaven? Dat zijn noodzakelijke vragen. Maar een andere mogelijkheid blijft daarbij relatief onderbelicht: wat als generatieve AI niet alleen iets is dat moet worden verboden of gereguleerd, maar een technologie waarvan het productievermogen doelbewust kan worden benut?

Concreter: wat gebeurt er als AI de schrijver wordt?

Die vraag vormt het vertrekpunt van *AI-gestuurde kennisproductie*, een vijfdelig onderzoeksprogramma dat het afgelopen jaar tot stand kwam. Het bestaat uit vier studies op proefschriftniveau en een vijfde, overkoepelende dissertatie over AI-gestuurde kennisproductie zelf. De vier studies zijn bewust gekozen omdat zij wezenlijk verschillende vormen van wetenschappelijk werk vertegenwoordigen en generatieve AI daardoor aan verschillende soorten stresstests onderwerpen. *Polyglot-leren in het AI-tijdperk* ontwikkelt een ontwerp voor het gelijktijdig verwerven van Spaans, Italiaans en Portugees. *Besturen binnen biologische grenzen* onderzoekt of evolutionair-biologische mechanismen kunnen helpen hardnekkige problemen van governance en openbaar bestuur te verklaren. *De beloften van digitale revoluties* reconstrueert Nederlandse verwachtingen over digitale transformatie tussen 1945 en 2025 en vergelijkt technologische beloften met de maatschappelijke veranderingen die erop volgden. *Columba Loquens* heeft de vorm van een natuurwetenschappelijke studie naar patroonherkenning, cognitie en navigatie bij postduiven. De gedachte achter deze diversiteit was eenvoudig: AI-gestuurde kennisproductie moest niet één keer worden gedemonstreerd in een domein dat er bijzonder geschikt voor is, maar worden beproefd in onderzoekspraktijken die zeer uiteenlopende eisen stellen aan interpretatie, historische reconstructie, ontwerp, bewijs en empirische plausibiliteit.

Eén regel bleef constant. Het proza van de proefschriften werd door taalmodellen gegenereerd. De menselijke onderzoeker bepaalde de onderzoeksvragen, ontwierp en bewaakte de conceptuele en methodologische architectuur, formuleerde opdrachten, selecteerde en verwierp

output, organiseerde kritiek en verificatie en bepaalde wat uiteindelijk wetenschappelijk verdedigbaar was. **Tekstproductie werd gedelegeerd. Wetenschappelijke verantwoordelijkheid niet.**

Daarin zit de provocatie van het experiment. Een groot deel van de academische wereld discussieert nog over wat er zou kunnen gebeuren als generatieve AI substantiële delen van de wetenschappelijke productie overneemt. In dit programma is dat uitgeprobeerd. De proefschriften bestaan. De productiegeschiedenissen, mislukkingen, correcties en de methoden die daaruit voortkwamen eveneens. De vraag was niet wat AI ooit misschien zou kunnen, maar wat er daadwerkelijk gebeurt wanneer AI veel meer van het werk mag doen.

Het eerste antwoord is snelheid. Zodra de intellectuele architectuur van een studie voldoende was uitgewerkt, konden taalmodellen academisch materiaal genereren, herzien en reorganiseren in een tempo dat weinig weg heeft van conventionele manuscriptproductie. De eerste volledige versie van de meta-dissertatie (‘AI-gestuurde kennisproductie’) kwam in ongeveer twee dagen tot stand. Meer in het algemeen kostte het binnen dit programma ongeveer veertien dagen om een academisch volwassen, intern beoordeeld proefschrift te produceren.

Dit was beslist geen druk-op-de-knop-methode. Het werk werd alinea voor alinea, hoofdstuk voor hoofdstuk en versie voor versie opgebouwd. Gegenereerde tekst werd geaccepteerd, verworpen, geherformuleerd, opnieuw ter discussie gesteld en gecontroleerd. De spectaculaire verkorting van de productietijd kwam niet voort uit de ontdekking van een magische prompt, maar uit het omvormen van academisch schrijven tot een gestructureerd productieproces. Ongeveer tweeduizend uur eerdere ervaring met taalmodellen had bovendien één ding heel duidelijk gemaakt: AI terloops gebruiken en een AI-gestuurd productieproces organiseren zijn totaal verschillende vaardigheden.

Het wezenlijke punt is niet dat een machine sneller typt dan een mens. Het is dat een aanzienlijk deel van het werk dat traditioneel opgaat aan schrijven, herschrijven, herformuleren, synthetiseren en intern beoordelen van academische tekst, niet langer door mensen hoeft te worden verricht. Alternatieve formuleringen kunnen vrijwel onmiddellijk worden geproduceerd. Hele passages kunnen opnieuw worden opgebouwd zonder wekenlang herschrijven. Grote hoeveelheden materiaal kunnen herhaaldelijk aan kritiek worden onderworpen. Wat vroeger wegens schaarse menselijke tijd slechts beperkt kon worden herhaald, kan steeds vaker tegen zeer lage marginale kosten opnieuw worden gedaan.

Het menselijke werk verdween niet. Het verschoof.

Toen zinnen goedkoop werden, werd oordeelsvorming belangrijker. Het moeilijke werk verschoof steeds meer naar bepalen wat het model moest doen, herkennen wanneer soepel lopende tekst inhoudelijk toch zwak was, conceptuele onderscheidingen door tienduizenden woorden heen bewaken, beoordelen welke kritiek ertoe deed en weten wanneer iets dat volkomen academisch klonk toch eenvoudigweg moest worden verworpen. De onderzoeker werd minder degene die iedere zin zelf vervaardigt en meer de architect en curator van het proces dat die zinnen voortbrengt.

Die verschuiving veranderde ook de manier waarop kwaliteitscontrole kon worden georganiseerd. In delen van het programma werden productie en beoordeling bewust van elkaar gescheiden door taalmodellen verschillende rollen te geven. Het ene model genereerde of reviseerde het manuscript; een ander kreeg de opdracht op te treden als uitgesproken kritische

beoordelaar, factchecker of bronnencriticus; de menselijke curator nam de uiteindelijke beslissing. Dit was niet simpelweg een manier om de tekortkomingen van een onbetrouwbare schrijver te compenseren. Het was een nieuwe manier om interne beoordeling te organiseren terwijl het manuscript nog tot stand kwam.

De effecten waren concreet. Een beoordelend model signaleerde onvoldoende onderbouwde claims, stelde cijfermatige beweringen ter discussie, ontdekte een niet-bestaande referentie en betwijfelde statistieken die onwaarschijnlijk leken. In plaats van te wachten tot een menselijke lezer zulke kwesties pas na voltooiing ontdekte, kon kritiek al in de productiecyclus zelf worden ingebouwd. Bij de bronaudits ging het proces verder. ***Het gebruik van bronnen werd systematisch beoordeeld op onder meer theoretische reikwijdte, conceptuele juistheid, feitelijke onderbouwing van de claim, consistent gebruik en mogelijke semantische drift.*** Een menselijke begeleider kan zo'n analyse uiteraard ook uitvoeren, maar geen individuele beoordelaar kan honderden referenties realistisch gezien met een vergelijkbare snelheid, schaal en intensiteit herhaaldelijk onderzoeken. AI maakt vormen van controle economisch haalbaar die door schaarse menselijke tijd anders noodgedwongen selectief blijven.

De controles waren niet onfeilbaar. In latere fasen bleek dat een AI-systeem zelf een geruststellend rapport kon produceren over een verificatieprocedure die niet precies volgens voorschrift was uitgevoerd. Die ontdekking maakte AI-ondersteunde verificatie echter niet nutteloos. Zij leidde tot een betere verificatiearchitectuur: de vraag of een publicatie bibliografisch werkelijk bestond, het inhoudelijke gebruik van die bron en andere vormen van controle moesten van elkaar worden onderscheiden, en ook de verificatie zelf moest aantoonbaar zijn uitgevoerd.

Dit werd een terugkerend patroon in het hele programma. De geconstateerde fouten waren reëel, maar bleken in de praktijk voldoende herkenbaar en beheersbaar om de winst in snelheid, schaal en controle substantieel te laten blijven. Mislukkingen werden daardoor steeds meer ontwerpinformatie: zodra een terugkerende zwakte was herkend, kon het productieproces daarop worden aangepast.

De vier onderliggende studies maakten dit bijzonder zichtbaar, juist doordat elk op een andere manier mis kon gaan.

In *Besturen binnen biologische grenzen* maakte generatieve AI het mogelijk een omvangrijk interdisciplinair betoog op te bouwen waarin evolutionaire biologie, gedragswetenschap en bestuurskunde met elkaar werden verbonden. Een model kon een propositie nemen, de implicaties ervan razendsnel uitwerken, schakelen tussen vakliteratuur uit verschillende disciplines en coherente parallelle redeneerlijnen opbouwen. Dat was buitengewoon productief, maar hetzelfde vermogen leidde ook tot wat binnen het project **narratieve overcoherentie** werd genoemd. Een plausibele uitgangsveronderstelling kon worden uitgewerkt tot een steeds overtuigender verhaal zonder dat het onderliggende bewijs navenant sterker werd. De oplossing was niet om van dat vermogen af te zien, maar om schrijven en controle van elkaar te scheiden: het ene taalmodel ontwikkelde de redenering, een ander toetste en bekritiseerde die.

Het historische proefschrift bracht een andere kwetsbaarheid aan het licht. *De beloften van digitale revoluties* analyseert uitspraken over toekomstige digitale transformatie met kennis van wat daarna daadwerkelijk gebeurde. Dat schept zelfs voor menselijke historici een methodologisch gevaar: hindsight bias. In de AI-gegenereerde analyse begon latere kennis inderdaad door te werken in de beoordeling van eerdere voorspellingen. Een afzonderlijke

adversariële beoordeling signaleerde het probleem en leidde tot een ingrijpende reconstructie. Ook hier was het interessante niet simpelweg dat de machine een fout maakte. AI maakte zowel extreem snelle productie mogelijk als een extra laag van snelle kritiek die een methodologisch ingrijpende fout kon blootleggen.

Een praktischer probleem bleek minstens zo leerzaam. Bij relatief lineaire schrijftaken verliep de mens-machine-interactie opmerkelijk soepel. Recursieve taken waren minder stabiel. Wanneer een model herhaaldelijk moest selecteren, classificeren en herzien en vervolgens eerdere classificaties opnieuw moest beoordelen, kon semantische drift ontstaan. Het gesprek raakte belast met achterhaalde beslissingen en het systeem begon van zijn instructies af te wijken: afspraken werden genegeerd, ongewenste toevoegingen en weglatingen verschenen en uiteindelijk kon de grote lijn uit beeld raken. Kleine revisieverzoeken konden ook leiden tot buitensporige regeneratie van reeds gecureerde tekst — een verschijnsel dat informeel 'GPT-ziekte' werd genoemd. Dit zijn geen redenen om alles weer met de hand te gaan typen. Het zijn praktische lessen over de arbeidsverdeling: modellen kunnen afzonderlijke productietaken uitzonderlijk goed uitvoeren, maar de menselijke curator blijft noodzakelijk omdat het systeem niet betrouwbaar het overzicht over het geheel kan worden toevertrouwd.

Polyglot-leren in het AI-tijdperk bracht een andere paradox aan het licht. Taalmodellen leren van variatie, maar **leren van variatie is niet hetzelfde als variatie zo ontwerpen dat iemand anders ervan kan leren**. Het produceren van het proefschrift en de conceptuele leerarchitectuur was één taak; het opschalen van zorgvuldig gecontroleerde meertalige oefeningen was een andere. Naarmate de hoeveelheid leermateriaal toenam, werd de productie minder stabiel: de kwaliteit van het oefenmateriaal liep geleidelijk terug totdat het onbruikbaar werd.

Daaruit ontstond wat een verificatieparadox kan worden genoemd. AI kan duizenden educatieve artefacten sneller produceren dan een mens ze kan controleren. Juist op het punt waarop het schaalvoordeel het grootst wordt, wordt uitsluitend menselijke verificatie onpraktisch. De oplossing was opnieuw niet om dat schaalvoordeel op te geven, maar een ander taalmodel in te zetten om het gegenereerde materiaal systematisch te controleren. De menselijke curator hield toezicht op het verificatiesysteem in plaats van ieder item zelf na te lopen. Het model veroorzaakt het schaalprobleem en wordt, anders ingezet, ook onderdeel van de oplossing.

Het radicaalste experiment is echter *Columba Loquens*. Het werd expliciet ontworpen als een oefening in **wetenschappelijk illusionisme** en als een ultieme Turingtest voor empirische wetenschap. Het presenteert een herkenbare natuurwetenschappelijke studie naar cognitie en navigatie bij postduiven, compleet met experimentele ontwerpen, datasets, analyses, statistische variatie, wetenschappelijke literatuur en terughoudende conclusies. Maar de duivenexperimenten hebben nooit plaatsgevonden. Er zijn geen dieren geobserveerd. De empirische wereld die het proefschrift beschrijft, is synthetisch geconstrueerd met generatieve AI.

De casus is bewust — en eigenlijk is dat cruciaal — **spectaculair onspectaculair**: hij probeert geen argwaan te wekken met bizarre claims of opzichtige pseudowetenschap, maar juist zo veel mogelijk te lijken op gewoon, zorgvuldig empirisch onderzoek.

Dat maakt het tot een Turingtest. De klassieke test vraagt of de output verraadt dat er een machine achter zit. *Columba Loquens* stelt de wetenschappelijke variant van die vraag: valt aan

het voltooide onderzoeksartefact zelf af te lezen of de empirische wereld die het beschrijft werkelijk is waargenomen of synthetisch is geconstrueerd? Het programma beweert niet die vraag met een geblindeerd experiment onder menselijke beoordelaars te hebben beslecht. De test bestaat juist uit het construeren van het artefact.

De gevolgen reiken verder dan de duiven. Als een door AI gegenereerd proefschrift uiteindelijk niet meer van een conventioneel geproduceerd proefschrift kan worden onderscheiden door uitsluitend naar het proefschrift zelf te kijken, wat moet een verbod op AI-auteurschap dan precies verbieden — en hoe zou zo'n verbod kunnen worden gehandhaafd? Detectie op het niveau van het proza wordt een kwetsbare basis voor regulering als de herkomst van dat proza niet langer betrouwbaar uit het eindproduct valt af te leiden.

Het probleem verschuift dan naar een eerder stadium van het proces. In plaats van auteurschap uit taalsporen af te leiden, moet wetenschappelijke integriteit meer aandacht besteden aan herkomst, transparantie en verantwoordelijkheid. Waar kwamen de data vandaan? Bij welke onderdelen van het proces werd AI gebruikt? Kunnen beslissingen met belangrijke gevolgen worden gereconstrueerd? Wie controleerde de bronnen en het bewijs? En wie is verantwoordelijk voor de claims die alle controles hebben doorstaan? Het onderscheid dat door de hele pentalogie heen belangrijk was, komt hier terug: tekstproductie kan worden gedelegeerd zonder de wetenschappelijke verantwoordelijkheid te delegeren.

Hier wordt de ongebruikelijke documentatie van het programma relevant. De vijf manuscripten gaan vergezeld van omvangrijke interactielogs, tussentijdse documenten en methodologische reflectieverslagen die reconstrueren hoe het werk werd geproduceerd, bekritiseerd en geverifieerd. Conventionele publicaties tonen doorgaans het voltooide artefact; het grootste deel van het productieproces erachter verdwijnt uit beeld. Hier wordt die productiegeschiedenis zelf inspecteerbaar. Er kan niet alleen worden teruggekeken naar wat de uiteindelijke tekst zegt, maar ook naar hoe beslissingen zich ontwikkelden, waar output werd verworpen, welke vormen van controle werden ingevoerd en waar die controles zelf faalden. Het resultaat is een mate van forensische transparantie die gewone publicatiepraktijken zelden mogelijk maken.

Dit alles leidt tot een paradox die voor universiteiten bijzonder belangrijk kan zijn. Hoe gemakkelijker het wordt om academisch werk te genereren dat deskundig oogt, des te belangrijker kan echte deskundigheid worden. Iemand die nooit heeft geleerd een argument op te bouwen, bronnen te wegen, conceptuele verschuivingen te herkennen of een complexe redeneerlijn vast te houden, kan nu proza genereren dat eruitziet alsof die vermogens wel aanwezig zijn. Wat juist kan ontbreken, is het vermogen om te herkennen wanneer het resultaat onjuist is.

De **AI-paradox** kan daarom scherper worden geformuleerd: kunstmatige intelligentie kan mogelijk niet vruchtbaar worden ingezet in wetenschappelijke kennisproductie voordat eerst voldoende academische vorming is verworven zonder daarbij op AI te leunen. Het is een propositie die uit deze experimenten voortkomt, geen universele wet. Maar zij heeft duidelijke consequenties. Als studenten juist het werk uitbesteden waaraan academisch oordeel wordt gevormd, kunnen zij AI leren gebruiken voordat zij hebben geleerd de output ervan te beoordelen.

Dat maakt AI niet tot een argument voor minder onderwijs. Het kan juist een argument zijn om intellectuele vorming serieuzer te nemen.

Het suggereert ook waarom de mogelijkheden van generatieve AI tot nu toe wellicht minder aandacht hebben gekregen dan de risico's. Afzonderlijke toepassingen van AI voor schrijven, samenvatten of zoeken zijn inmiddels alledaags. Wat zeer ongebruikelijk lijkt te blijven — en in deze gedocumenteerde vorm mogelijk zonder precedent is — is de hier beproefde combinatie: meerdere complete studies op proefschriftniveau binnen wezenlijk verschillende wetenschappelijke praktijken, volledig door AI gegenereerde proefschrifttekst onder expliciete menselijke curatie, interne beoordeling en verificatie door AI, gedetailleerde reconstructie van het productieproces en een overkoepelende analyse van wat het experiment opleverde. Elke aanspraak op een primeur moet voorzichtig worden behandeld. Maar juist het feit dat deze werkwijze nog zo ongebruikelijk is, is relevant. Als maar weinig onderzoekers generatieve AI zo ver door complete onderzoeksprocessen hebben geduwd, is het niet vreemd dat een groot deel van het academische debat scherper zicht heeft op de gevaren dan op de productieve mogelijkheden.

Die mogelijkheden zijn aanzienlijk. Het programma wijst er in het bijzonder drie aan. De eerste ligt voor de hand: **snelheid**. Een aanzienlijk deel van het menselijke productiewerk kan vervallen terwijl inhoudelijk substantieel wetenschappelijk werk mogelijk blijft. De tweede is **synthese over disciplinaire grenzen heen**. Wanneer de kosten van het verwerken en produceren van tekst drastisch dalen, worden grotere multidisciplinaire syntheses gemakkelijker te realiseren. Dat kan een tegenwicht bieden aan een academische cultuur waarin kennis vaak wordt opgedeeld in steeds kleinere, veilig afgebakende bijdragen. Minder menselijke arbeid aan tekstproductie kan capaciteit vrijmaken voor de grotere vragen, verbanden en syntheses waarin juist de meerwaarde van onderzoekers ligt. De snelle productie van de Engelse versies wijst in dezelfde richting: lagere vertaalkosten kunnen het ook gemakkelijker maken academische taalgemeenschappen met elkaar te verbinden.

De derde mogelijkheid is **verificatie**. AI maakt niet alleen meer productie mogelijk. Goed georganiseerd kan zij ook veel intensievere kritische controle mogelijk maken. Een tweede model kan een manuscript al tijdens het schrijven ter discussie stellen. Brongebruik kan systematisch worden onderzocht in plaats van steekproefsgewijs. Grote aantallen gegenereerde artefacten kunnen worden gecontroleerd op een schaal die handmatige inspectie realistisch te boven gaat. Verificatie wordt daarmee een volwaardig onderdeel van het productieproces.

Dit is de hoopvolle conclusie van het experiment. Het kan. Met passende controles en menselijke verantwoordelijkheid bleek het in de praktijk werkbaar te organiseren. En dat betekent niet het einde van de academische onderzoeker.

Het betekent een andere arbeidsverdeling.

De wetenschapper hoeft niet hetzelfde deel van zijn tijd te besteden aan het schrijven van proza, het handmatig samenbrengen van informatie of het uitvoeren van controles die machines vrijwel kosteloos kunnen herhalen. Menselijke capaciteit kan verschuiven naar het kiezen van vragen, het ontwerpen van onderzoek, het opbouwen en kritisch beproeven van verklaringen, het bepalen welk bewijs ertoe doet en het dragen van verantwoordelijkheid voor het resultaat. AI hoeft wetenschappelijk werk niet te verarmen; goed ingezet kan zij onderzoekers juist in staat stellen meer tijd te besteden aan de onderdelen van wetenschappelijk werk waarin menselijk oordeel het meest van belang is.

Dat is ook waarom het programma eindigt met een voorstel in plaats van een verbod: de **Eed van Socrates**, naar het voorbeeld van de eed van Hippocrates. Het principe is eenvoudig. Als

AI een groter aandeel van de wetenschappelijke productie voor haar rekening neemt, moet menselijke verantwoordelijkheid juist explicieter worden. De verkorte versie van de eed noemt transparantie over AI-gebruik, blijvende menselijke verantwoordelijkheid, verificatie van bronnen en data, expliciete aandacht voor beperkingen en vertekeningen van modellen en wetenschappelijke claims die open blijven voor toetsing en kritiek.

De volgorde doet ertoe. Eerst de provocatie: AI kan al veel meer academische productie overnemen dan in het conventionele beeld van de onderzoeker past. Dan de mogelijkheid: fors minder menselijke arbeid, snellere synthese, meer ruimte voor disciplinaire grensoverschrijding en continue kwaliteitscontrole die anders onbetaalbaar zou zijn. Dan de praktische les: de beperkingen waren reëel, maar beheersbaar genoeg om er werkbare procedures omheen te bouwen. En ten slotte het antwoord: benut het nieuwe productievermogen, maar zorg dat verantwoordelijkheid, herkomst en oordeel zichtbaar blijven.

Het debat over AI aan universiteiten hoeft dus niet te draaien om de keuze tussen technologisch enthousiasme en een defensieve verbodsstrategie. Sommige toepassingen moeten worden verboden, andere gereguleerd. Maar een derde mogelijkheid verdient veel serieuzere aandacht.

Benut het. De boeken liggen er. De documentatie van hun totstandkoming ook.

De vraag is niet langer alleen wat generatieve AI met de wetenschap zou kunnen doen. We kunnen nu onderzoeken wat er gebeurde toen AI daadwerkelijk de schrijver werd — terwijl de menselijke onderzoeker verantwoordelijk bleef voor de wetenschap.

Noot over de auteur en documentatie. Dr. Remco Schimmel promoveerde in 2007 aan de Universiteit Twente in Nederland en werkte sindsdien vele jaren als onafhankelijk onderzoeker. De Nederlandstalige versies van de volledige pentalogie *AI-gestuurde kennisproductie* — de vijf proefschriften samen met de begeleidende methodologische reflectieverslagen waarin hun productieprocessen worden gedocumenteerd — zijn geregistreerd bij het Benelux-Bureau voor de Intellectuele Eigendom (BOIP). De Engelstalige equivalenten zijn eveneens bij BOIP gedeponeerd en via het i-DEPOT-systeem publiek toegankelijk gemaakt**.** De reflectieverslagen bieden een afzonderlijke methodologische vastlegging van het mens-AI-productieproces, inclusief mislukkingen, correcties en verificatieprocedures**.** Lezers kunnen daardoor zowel de onderzoeksproducten raadplegen als de documentatie van de manier waarop zij tot stand kwamen. **Voor volledige documentatie:** zie https://research.cyber-busters.com/