AI-gestuurde kennisproductie, een onderzoek naar de impact van kunstmatige intelligentie op wetenschappelijke kennisproductie

Proefschrift (versie 3)

Remco Schimmel

15 maart 2026

Abstract

In dit proefschrift wordt onderzocht hoe generatieve kunstmatige intelligentie de productie van wetenschappelijke kennis beïnvloedt en welke epistemologische, methodologische en institutionele implicaties hieruit voortvloeien.

Het onderzoek is opgebouwd rond een analytisch raamwerk waarin drie dimensies van kennisproductie — de cognitief-methodologische, de pedagogisch-institutionele en de normatief-politieke dimensie — worden gekruist met vijf archetypen van wetenschapsbeoefening. Deze 5×3-matrix vormt het centrale analysekader van het proefschrift. In een eerste fase worden de vijf archetypen op basis van literatuurstudies langs deze drie dimensies geanalyseerd. Dit levert een voorlopig beeld op van de mogelijke invloed van generatieve AI op verschillende vormen van wetenschappelijk onderzoek.

In een tweede fase wordt dit beeld empirisch getoetst aan de hand van vijf praktijkstudies, telkens één binnen elk archetype van wetenschapsbeoefening. De casussen laten zien hoe generatieve AI daadwerkelijk functioneert in uiteenlopende onderzoekspraktijken, variërend van experimenteel en historisch onderzoek tot interpretatieve en ontwerpgerichte studies, evenals een reflexieve meta-casus waarin het onderzoeksproces zelf wordt geanalyseerd.

De synthese van deze analyses laat zien dat generatieve AI wetenschap niet simpelweg automatiseert, maar de architectuur van kennisproductie herconfigureert. Onder bepaalde voorwaarden kan AI onderzoeksprocessen versnellen, nieuwe combinaties van kennisdomeinen mogelijk maken en nieuwe vormen van verificatie ondersteunen. Tegelijkertijd worden nieuwe epistemische kwetsbaarheden zichtbaar, waaronder narratieve overcoherentie, semantische drift en afhankelijkheid van technologische infrastructuren. De kwaliteit van AI-gestuurde kennisproductie blijkt daarom in sterke mate afhankelijk van de manier waarop het generatieve proces wordt georganiseerd en bewaakt.

Het proefschrift concludeert dat generatieve AI leidt tot een hybride onderzoekspraktijk waarin menselijke curatie en machine-ondersteunde analyse structureel met elkaar verweven raken. Betrouwbare toepassing van AI in wetenschap vereist daarom expliciete onderzoeksarchitecturen, transparantie over AI-gebruik en nieuwe vormen van methodologische controle.

Management Samenvatting:

AI-gestuurde kennisproductie, een onderzoek naar de impact van kunstmatige intelligentie op wetenschappelijke kennisproductie

Dit proefschrift vorm het sluitstuk in een pentalogie van 5 academische monografieën. In de vier voorgaande monografieën werd het gebruik van AI in het natuurwetenschappelijk domein, het geschiedkundig domein, het antropologisch domein en het ontwerp-gerichte domein beschreven. AI werd daar gebruikt om een wetenschappelijk proefschrift te schrijven binnen de conventies van het betreffende domein. Dit proefschrift is een meta-proefschrift: Het is de paraplu boven de overige monografieën. Het ordent de wetenschap door 4+1 archetypes van wetenschapsbeoefening te benoemen om vervolgens te bekijken hoe het gebruik van AI deze vormen van wetenschapsbeoefening beïnvloedt. Aldus vormen de vier eerder genoemde monografieën het casus-materiaal van dit proefschrift.

Hoofdstuk 1 – Inleiding: AI en kennisproductie.

In dit proefschrift staat de fundamentele vraag hoe kunstmatige intelligentie (AI) de basis van wetenschappelijke kennisproductie verandert, centraal. Terwijl computers in het verleden voornamelijk als rekenwerktuigen fungeerden, zijn ze nu medeproducenten van kennis geworden. Deze ontwikkeling roept urgente vragen op over de manier waarop AI processen van data-verzameling, analyse, verklaring en presentatie beïnvloedt.

Om deze complexe materie te structureren introduceert het hoofdstuk vier archetypen van kennisproductie. Het eerste archetype betreft natuurwetenschappelijk onderzoek, dat gericht is op het ontdekken van universele wetmatigheden via metingen en experimenten. Het tweede archetype omvat historisch onderzoek, waarbij het begrijpen van unieke gebeurtenissen centraal staat door middel van bronnenkritiek en interpretatie. Het derde archetype behelst antropologisch onderzoek, waarin cultuur wordt begrepen via participerende observatie en 'thick description'. Het vierde archetype betreft ontwerpgericht onderzoek, waarbij kennisproductie plaatsvindt door het maken van oplossingen.

Naast deze vier archetypen wordt een belangrijk onderscheid gemaakt tussen generieke AI (grote modellen getraind op internetdata) en domein-specifieke AI (getraind op zorgvuldig geselecteerde datasets). Daarenboven krijgt het meta-domein van informatica en AI-ontwikkeling zelf speciale aandacht vanwege zijn reflexieve karakter, waarbij AI tegelijk instrument en object van onderzoek is.

Het analysekader dat in dit proefschrift wordt gehanteerd, omvat zes impactdimensies: cognitief, methodologisch, pedagogisch, institutioneel, macht/eigenaarschap, en betrouwbaarheid/normativiteit. Bijzondere aandacht gaat uit naar wat wordt aangeduid als de AI-paradox: de paradoxale situatie dat men AI pas vruchtbaar kan gebruiken wanneer men eerst zonder AI gevormd is.

Hoofdstuk 2 – Onderzoeksvragen.

In het tweede hoofdstuk worden de centrale onderzoeksvragen geformuleerd tegen de achtergrond van wat wordt beschreven als fragmentatie in de huidige academische cultuur. Hoewel AI

onderzoeksprocessen versnelt, dreigt het tegelijkertijd het regime van "postzegel-wijsheden" - kleine, afgebakende bijdragen - te versterken in plaats van echte synthese te bevorderen.

De hoofdvraag van het onderzoek luidt: Hoe beïnvloedt de inzet van kunstmatige intelligentie de processen, instituties en epistemologie van wetenschappelijke kennisproductie, en welke ruimte opent dit voor het doorbreken van fragmentatie en het bevorderen van multidisciplinaire originaliteit?

Deze hoofdvraag wordt uitgewerkt in drie deelvragen. Ten eerste wordt onderzocht hoe AI de cognitieve en methodologische dimensie beïnvloedt, specifiek de manier waarop kennis wordt gezocht, geordend en geverifieerd. Ten tweede komen de pedagogische en institutionele gevolgen aan bod, waarbij wordt gekeken naar de impact op opleidingspraktijken en wetenschappelijke erkenning. Ten derde wordt de macht- en normatieve dimensie bestudeerd, met aandacht voor hoe AI de verdeling van macht en eigenaarschap in de wetenschap beïnvloedt.

De wetenschappelijke relevantie van dit onderzoek ligt in het bieden van een geïntegreerd perspectief dat de fragmentatie in bestaande AI-literatuur doorbreekt. Vanuit maatschappelijk perspectief is het relevant omdat AI zowel de betrouwbaarheid van kennis als de vorming van nieuwe generaties beïnvloedt.

Hoofdstuk 3 - Methodologische verantwoording.

In dit hoofdstuk wordt uitgelegd uit hoe het onderzoek is vormgegeven. Omdat AI zowel het onderwerp van studie is als het instrument dat bij het onderzoek wordt gebruikt, ontstaat een bijzondere methodologische situatie die een hybride aanpak vereist.

Het onderzoek combineert verschillende methoden: literatuuronderzoek, analyse van bestaande AI-toepassingen in verschillende wetenschapsgebieden, en reflectie op eigen ervaringen met AI-gebruik. Deze combinatie is nodig omdat AI nog zo nieuw is dat er weinig gevestigde onderzoekstradities bestaan.

De onderzoeksstructuur is opgezet als een systematische vergelijking. Vier verschillende types wetenschap (natuurwetenschappen, geschiedenis, antropologie en ontwerpgericht onderzoek) worden elk onderzocht op drie aspecten: hoe AI het denken en werken verandert, hoe het onderwijs en instituties beïnvloedt, en welke machtsverhoudingen ontstaan. Deze systematische aanpak voorkomt dat het onderzoek verdwaalt in losse observaties. Voor elk wetenschapsgebied wordt gebruik gemaakt van bestaande literatuur over AI-toepassingen, aangevuld met voorbeelden uit de praktijk. Daarnaast worden drie concrete gevallen uitgewerkt waarin de onderzoeker zelf intensief met AI heeft gewerkt, om inzicht te krijgen in hoe AI-gestuurde kennisproductie in de praktijk verloopt.

Kwaliteitswaarborging gebeurt door verschillende bronnen en perspectieven te vergelijken, alle stappen in het onderzoek transparant te documenteren, en voortdurend kritisch te reflecteren op de beperkingen van zowel AI als de menselijke onderzoeker.

Hoofdstuk 4 - Cognitieve en methodologische impact.

In het vierde hoofdstuk wordt geanalyseerd hoe AI de fundamentele processen van kennisproductie herconfigureert. Een belangrijke verschuiving die wordt geïdentificeerd betreft de veranderende rol van onderzoekers: van data-verzamelaars naar curators van AI-gegenereerde inzichten. Terwijl AI grootschalige replicatie mogelijk maakt, ontstaat er een problematische verschuiving: AI kan vaak heel

accuraat voorspellen wat er zal gebeuren (zoals welke patiënt ziek wordt of welke eiwitstructuur ontstaat), maar wetenschappers begrijpen steeds minder waarom die voorspelling klopt. Dit ondermijnt het traditionele wetenschappelijke ideaal om niet alleen te weten wat er gebeurt, maar ook waarom.

Wat betreft de kansen wordt vastgesteld dat AI aanzienlijke versnelling mogelijk maakt van literatuuronderzoek en dataverwerking. Daarnaast worden grootschalige her-analyses en replicatiestudies haalbaar die voorheen ondenkbaar waren. AI schept ook cognitieve ruimte voor interpretatie door automatisering van repetitieve taken en maakt patroonherkenning over disciplines heen mogelijk.

Tegenover deze kansen staan echter substantiële risico's. Deze omvatten oversimplificatie en nuanceverlies, evenals een bijzonder problematisch fenomeen: de fabricage van literatuurverwijzingen die kunnen variëren van correct tot volledig fictief. Verder bestaat er bias-reproductie uit trainingsdata en blijft het probleem van black-box ondoorzichtigheid bestaan.

De evaluatie van deze ontwikkelingen laat zien dat de spanning tussen versnelling en verdieping cruciaal is voor de toekomst van kennisproductie. AI kan fragmentatie versterken door meer papers mogelijk te maken, maar tegelijkertijd synthese bevorderen door patronen zichtbaar te maken. De multiplicatie van AI-reviewers kan echter echokamers creëren als alle modellen op vergelijkbare data zijn getraind.

Hoofdstuk 5 – Pedagogische en institutionele impact.

In het vijfde hoofdstuk wordt hoe AI zowel het leren als de institutionele structuren van wetenschap verandert. Een fundamentale observatie is dat het schrijfproces transformeert tot een leeromgeving, terwijl traditionele rituelen zoals scripties en peer review onder druk komen te staan.

Op *pedagogisch gebied* worden verschillende kansen geïdentificeerd. AI maakt personalisering van leren mogelijk en verhoogt de toegankelijkheid voor studenten met beperkingen. Bovendien kan transparantie van leerprocessen worden bevorderd door automatische logboeken die het leertraject documenteren.

Institutionele verschuivingen manifesteren zich op verschillende niveaus. De publicatiecultuur wordt versneld maar mogelijk ook oppervlakkiger. Peer review kan worden geautomatiseerd maar verliest daarbij zijn menselijke dimensie. Daarnaast moeten criteria voor academische carrières worden herijkt om recht te doen aan nieuwe vormen van kennisproductie.

Drie fundamentele spanningen worden geïdentificeerd die de kern van deze verschuivingen raken. Ten eerste betreft dit de communicatieve functie van wetenschappelijke teksten: deze zijn niet alleen informatiedragers maar ook rituelen die gemeenschap vormen. Ten tweede speelt de AI-paradox een rol: studenten die met AI opgroeien missen het referentiekader voor kritische beoordeling. Ten derde bestaat er spanning tussen originaliteit en fragmentatie: AI kan zowel postzegel-wijsheden versterken als doorbreken.

Hoewel de Hong Kong Principles (integriteit, transparantie, open science) kaders bieden voor deze ontwikkelingen, moeten nieuwe rituelen voor een AI-tijdperk nog worden ontwikkeld.

Hoofdstuk 6 - Macht, normativiteit en de aard van AI-gestuurde kennisproductie.

In het zesde hoofdstuk wordt gefocussed op een cruciale vraag: wie bepaalt eigenlijk wat geldt als betrouwbare kennis wanneer AI wordt gebruikt? Het antwoord blijkt niet alleen bij wetenschappers te liggen, maar ook bij de bedrijven en instellingen die AI-systemen ontwikkelen en bezitten.

Macht en toegang spelen een centrale rol omdat AI-training extreem duur is. Alleen grote technologiebedrijven kunnen zich systemen zoals GPT of AlphaFold veroorloven. Dit betekent dat kleinere universiteiten en onderzoekers afhankelijk worden van wat deze bedrijven beschikbaar stellen. Hierdoor krijgen niet alle onderzoekers gelijke kansen om kennis te produceren.

Ingebouwde keuzes in AI-systemen bepalen mede wat als kennis geldt. Wanneer een AI-model wordt getraind, worden er keuzes gemaakt over welke data erin gaan en welke niet. Een model getraind op voornamelijk Engelstalige teksten zal andere patronen herkennen dan een model getraind op diverse talen. Deze keuzes worden gemaakt door ontwikkelaars, niet door de wetenschappers die het model later gebruiken.

Data als machtsmiddel wordt zichtbaar doordat goede datasets schaars zijn. Medische gegevens, historische archieven of sociale media-data zijn niet voor iedereen toegankelijk. Wie controle heeft over deze data, bepaalt mede welk onderzoek mogelijk is en welke vragen gesteld kunnen worden.

Het black-box probleem betekent dat onderzoekers vaak niet weten hoe een AI-systeem tot zijn conclusies komt. Ook bij gespecialiseerde systemen zoals AlphaFold is niet altijd duidelijk waarom het model een bepaalde voorspelling doet. Dit maakt het moeilijk om resultaten te controleren of fouten te herkennen.

Het hoofdstuk concludeert dat AI-gebruik in wetenschap nooit neutraal is, maar altijd beïnvloed wordt door wie toegang heeft tot technologie en data, en door keuzes die in de systemen zijn ingebouwd.

Tussenbalans 1 - Taxonomieën van kansen, risico's en uitdagingen.

Na de behandeling van de drie dimensiehoofdstukken worden de inzichten gebundeld in systematische taxonomieën die dienen als analytisch raster voor de behandeling van de archetypen.

De taxonomie van kansen omvat drie hoofdcategorieën. Ten eerste betreft dit versnelling en schaal, waarbij radicale verkleining van feedbackloops en cyclische processen mogelijk wordt. Ten tweede gaat het om ontdekking en synthese, waardoor verbanden over disciplines heen zichtbaar worden gemaakt. Ten derde behelst dit pedagogische en institutionele vernieuwing door gepersonaliseerd leren en nieuwe transparantievormen.

De taxonomie van risico's identificeert vijf kernproblemen. Het eerste betreft black-box en transparantieverlies, wat leidt tot ondermijning van reproduceerbaarheid. Het tweede omvat decontexualisatie en nuanceverlies, waarbij een universaliteitsschijn ontstaat van contextueel contingente patronen. Het derde behelst ritualisering en betekenisverlies, wat resulteert in uitholling van de gemeenschapsvormende functie. Het vierde betreft machtsconcentratie waarbij epistemische agenda's gestuurd worden door commerciële belangen. Het vijfde omvat fabricage van referenties, wat een bedreiging vormt voor bronverantwoording.

De taxonomie van uitdagingen identificeert drie kernproblemen: hoe bepalen we of AI-onderzoek betrouwbaar is als we de onderliggende processen niet kunnen controleren? Hoe behouden we de betekenisvolle rituelen van wetenschap (zoals scripties en peer review) terwijl we AI integreren? En hoe stellen verschillende disciplines hun eigen normen vast voor welke mate van AI-onzekerheid acceptabel is?

Hoofdstuk 7 - Natuurwetenschappelijk onderzoek en AI.

Het zevende hoofdstuk wordt geanalyseerd hoe AI het eerste archetype beïnvloedt. Natuurwetenschappen, traditioneel gericht op reproduceerbare metingen en falsificeerbare hypothesen, tonen de meest harmonieuze relatie met AI omdat beide waardesystemen convergeren rond empirische verificatie.

De kansen die AI biedt aan natuurwetenschappelijk onderzoek zijn substantieel. Versnelling van ontdekkingsprocessen wordt mogelijk in domeinen zoals moleculaire biologie en astronomie. Analyse van grote datasets opent nieuwe mogelijkheden in klimatologie en andere data-intensieve velden. Democratisering van kennis door publieke databanken vergroot de toegankelijkheid, terwijl geavanceerde simulaties en modellering de methodologische precisie verbeteren.

Tegenover deze kansen staan echter ook risico's. Het probleem van black-box verklaarbaarheid blijft bestaan, ook wanneer voorspellingen accuraat zijn. Afhankelijkheid van gesloten infrastructuur vormt een toenemende zorg. Bias in trainingsdata kan een objectieve schijn wekken die misleidend is. Daarnaast spelen hoge energie- en resourcekosten een rol in de toegankelijkheid.

De uitdagingen manifesteren zich op drie niveaus. Methodologisch rijst de vraag of accurate maar onverklaarbare voorspellingen wetenschappelijk geldig zijn wanneer het onderliggende mechanisme niet begrepen wordt. Institutioneel moeten ook kleinere onderzoeksgroepen toegang krijgen tot dure infrastructuur. Normatief moet per domein bepaald worden hoeveel onzekerheid acceptabel is.

De conclusie luidt dat AI in natuurwetenschappen vooral functioneert als versneller die bestaande waarden versterkt, maar wel nieuwe afhankelijkheden en normatieve keuzes introduceert die expliciet moeten worden geadresseerd.

Hoofdstuk 8 - Geschiedenis-onderzoek en AI.

In de geschiedwetenschap, waar authenticiteit en context centraal staan, roept AI specifieke spanningen op rond anachronisme en canonvorming. Deze spanning vormt de rode draad door de analyse.

De kansen die AI biedt zijn aanzienlijk. Toegang tot digitale archieven wordt sterk verbeterd, terwijl patroonherkenning onverwachte verbanden kan blootleggen. OCR en vertaaltools vergroten de toegankelijkheid van bronnen substantieel. Scholing en institutionele vernieuwing worden mogelijk door de actieve rol die archieven kunnen gaan spelen in kennisproductie.

Daartegenover staan significante risico's. Anachronistische interpretaties kunnen ontstaan door het toepassen van hedendaagse categorieën op historische gegevens. Verlies van authenticiteit en context dreigt bij dataficatie van bronnen. Black-box patronen kunnen kritiekloos worden overgenomen. Canonvorming door selectieve digitalisering vormt een bijzonder probleem.

De uitdagingen concentreren zich rond drie kernpunten. Methodologisch moeten historici AI-patronen altijd vertalen naar de historische context waarin bronnen ontstonden. Institutioneel moeten keuzes bij digitalisering expliciet worden gemaakt omdat deze bepalen welke verhalen zichtbaar worden. Normatief blijft de historicus verantwoordelijk voor het beoordelen of AI-analyses historisch geldig zijn.

De conclusie is dat geschiedwetenschap AI kan benutten voor toegang en patroonherkenning, maar dat menselijke interpretatie essentieel blijft om anachronismen te voorkomen en authenticiteit te waarborgen.

Hoofdstuk 9 - Antropologisch onderzoek en AI.

De analyse van het derde archetype laat zien dat antropologie, gebaseerd op relationele nabijheid en contextuele rijkdom, de grootste wrijving ervaart met AI's neiging tot dataficatie en patroonherkenning. Deze fundamentele spanning kleurt de gehele behandeling van dit archetype.

Kansen voor antropologisch onderzoek met AI omvatten verschillende innovatieve toepassingen. Digitale etnografie maakt schaalvergroting van analyse mogelijk, terwijl AI-ondersteuning bij transcriptie en thematische codering het onderzoeksproces kan versnellen. Documentatie van bedreigde talen en rituelen wordt technisch haalbaar gemaakt, en remote sensing opent nieuwe mogelijkheden voor landschapsanalyse en het bestuderen van mobiliteitspatronen.

De risico's die hiermee gepaard gaan zijn echter substantieel. Reductie van gemeenschappen tot datapoints bedreigt de kern van antropologisch werk, terwijl verlies van nabijheid en empathie de methodologische basis ondermijnt. AI kan classificaties opleggen die niet aansluiten bij hoe gemeenschappen zichzelf begrijpen - bijvoorbeeld rituele gebaren labelen als alledaags gedrag, of universele patronen herkennen waar lokale betekenissen heel anders liggen.

De uitdagingen voor antropologisch onderzoek raken de fundamenten van de discipline. Methodologisch moet AI functioneren als aanvulling en niet als vervanging van interpretatieve arbeid. Institutioneel is het essentieel dat data worden teruggekoppeld naar onderzoekssubjecten om relationele legitimiteit te behouden. Normatief dient consent relationeel en doorlopend gegarandeerd te worden.

De conclusie luidt dat antropologie expliciete verankering van AI vereist in de waarden van nabijheid, empathie en gedeelde verantwoordelijkheid om haar disciplinaire integriteit te bewaren.

Hoofdstuk 10 - Ontwerpgericht onderzoek en AI.

Binnen dit vierde archetype treedt AI op als co-designer, wat nieuwe mogelijkheden opent maar tegelijkertijd fundamentele vragen oproept over uitlegbaarheid en verantwoordelijkheid van artefacten.

De kansen die ontwerpgericht onderzoek biedt zijn indrukwekkend. AI als co-designer kan duizenden ontwerpvarianten genereren, waardoor de creatieve horizon wordt verbreed. Snellere en goedkopere simulatie en prototyping maken iteratieve ontwikkeling toegankelijker. Iteratieve cycli worden sterk verkort, wat efficiëntie verhoogt. Nieuwe pedagogiek wordt mogelijk door directe samenwerking tussen studenten en AI-tools.

Tegenover deze beloften staan echter ook risico's. Black-box-ontwerpen zijn moeilijk uitlegbaar, wat de legitimiteit van ontwerpkeuzes ondermijnt. Verlies van verantwoording over ontwerpkeuzes dreigt wanneer AI autonome beslissingen neemt. Eigenaarschap en auteurschap worden diffuus in hybride mens-machine processen. Bovendien kan optimalisatie door AI morele waarden veronachtzamen die niet expliciet zijn geprogrammeerd.

De uitdagingen voor ontwerpgericht onderzoek concentreren zich rond drie kerngebieden. Methodologisch moeten normen voor uitlegbaarheid worden ontwikkeld om transparantie te waarborgen. Institutioneel zijn nieuwe publicatie- en erkenningsvormen nodig die recht doen aan hybride auteurschap. Normatief dienen waarden en ethiek expliciet te worden ingebouwd in ontwerptrajecten.

De analyse concludeert dat ontwerpgericht onderzoek door AI kan worden versterkt, mits uitlegbaarheid, institutionele erkenning en collectieve verantwoording systematisch worden ingebouwd.

Hoofdstuk 11 - Het meta-domein van AI & informatica.

Het meta-domein onderscheidt zich omdat AI hier zowel instrument als object van onderzoek is (onderzoek naar AI-gestuurde kennisontwikkeling m.b.v. AI), wat reflexiviteit en circulaire validatie oplevert. Deze bijzondere positie maakt dit domein tot een interessante testcase voor de bredere analyse.

Het utopische scenario toont indrukwekkende mogelijkheden. AutoML en meta-learning creëren zelfverbeterende systemen die exponentiële ontwikkeling mogelijk maken. Open platforms zoals arXiv en Wikipedia bevorderen collectief eigenaarschap en democratiseren kennisproductie. Nieuwe pedagogiek wordt mogelijk waarbij leren met AI over AI plaatsvindt, wat reflexieve competenties ontwikkelt.

Het dystopische scenario waarschuwt echter voor ernstige gevaren. Circulaire validatie binnen zelf-gedefinieerde benchmarks kan vooruitgang illusoir maken. Extreme machtsconcentratie bij enkele bedrijven dreigt de autonomie van onderzoek te ondermijnen. Black-box-karakter wordt zo extreem dat zelfs ontwerpers hun systemen niet meer begrijpen. Fragmentatie van legitimiteit tussen verschillende validatiepraktijken ondermijnt wetenschappelijke consensus.

De uitdagingen voor het meta-domein zijn veelomvattend. Methodologisch moeten criteria worden ontwikkeld die circulaire validatie doorbreken en externe geldigheid waarborgen. Institutioneel dient een balans te worden gevonden tussen open en gesloten systemen. Normatief moet verantwoordelijkheid worden geherdefinieerd in een recursief ecosysteem waarin traditionele scheidslijnen vervagen.

Een belangrijke relativering wordt geboden door tegenvoorbeelden zoals Wikipedia en arXiv die aantonen dat alternatieve validatievormen mogelijk zijn. Oudere onderzoekstradities zoals hermeneutiek, auto-etnografie en actor-network-theorie bieden handvaten voor reflexieve legitimering die in het AI-tijdperk relevant blijven. Deze tradities leren respectievelijk: AI-resultaten hebben altijd interpretatie nodig (hermeneutiek), zelfonderzoek kan legitiem zijn mits transparant gereflecteerd (auto-etnografie) en kennis ontstaat altijd in netwerken van mensen en technologie samen (actor-network theory).

Hoofdstuk 12 - Tussenbalans 2 - Overkoepelende patronen en archetypen-vergelijking.

In hoofdstuk 12 wordt een tweede tussenbalans opgemaakt. In deze tussenbalans worden de inzichten uit de drie impactdimensies en de vier archetypen van wetenschap samengebracht. De vergelijking van deze domeinen laat zien dat AI in verschillende disciplines uiteenlopende toepassingen krijgt, maar dat zich tegelijkertijd enkele terugkerende spanningen aftekenen. De analyse maakt duidelijk dat deze spanningen geduid kunnen worden als drie fundamentele paradoxen van AI-gestuurde kennisproductie.

De eerste paradox betreft de *relatie tussen versnelling en legitimiteit*. In vrijwel alle onderzochte domeinen vergroot AI de snelheid en schaal van analyse. Tegelijkertijd zet die versnelling juist de mechanismen onder druk die wetenschappelijke legitimiteit traditioneel waarborgen. In de natuurwetenschappen kan AI bijvoorbeeld hypothesen genereren die pas later experimenteel kunnen worden gevalideerd; in historisch onderzoek kunnen grootschalige tekstanalyses patronen suggereren die historisch gezien anachronistisch zijn; en in ontwerpgericht onderzoek kan generatieve AI het creatieve proces versnellen zonder dat de onderliggende ontwerpkeuzes volledig transparant blijven.

De tweede paradox betreft *toegang en ongelijkheid*. AI verlaagt in veel gevallen de drempel tot onderzoek doordat digitale archieven, analysetools en generatieve systemen breed beschikbaar zijn. Tegelijkertijd ontstaan nieuwe vormen van afhankelijkheid van technische infrastructuur, specialistische expertise en representatieve datasets. Daardoor verdwijnen bestaande ongelijkheden niet, maar worden zij in nieuwe vormen gereproduceerd.

De derde paradox betreft *kennisvorming en afhankelijkheid*. Effectief gebruik van AI veronderstelt juist de analytische vaardigheden en methodologische discipline die AI-systemen gedeeltelijk overnemen. Wanneer onderzoekers deze vaardigheden niet zelfstandig ontwikkelen, ontstaat het risico dat AI-systemen wel productief worden ingezet, maar dat hun uitkomsten onvoldoende kritisch kunnen worden beoordeeld.

Deze tussenbalans laat daarmee zien dat AI niet alleen nieuwe mogelijkheden voor kennisproductie creëert, maar ook structurele spanningen introduceert. Deze paradoxen lijken geen tijdelijke overgangsproblemen maar structurele kenmerken van een onderzoekspraktijk waarin menselijke en artificiële intelligentie steeds nauwer verweven raken. In de hoofdstukken 13 t/m 17 wordt gecontroleerd of empirische ervaringen leiden tot bijstelling of verdieping van deze bevindingen.

Hoofdstuk 13 - Casus 'Aviaire Intelligentie & homing pigeons'.

In dit hoofdstuk wordt een natuurwetenschappelijk ogend onderzoek naar navigatiegedrag bij postduiven geanalyseerd. Het manuscript volgt de klassieke architectuur van een empirisch proefschrift met een literatuurreview, een methodologische verantwoording, meerdere datasets en een synthesehoofdstuk waarin de resultaten worden geïnterpreteerd. De studie concludeert dat duivennavigatie niet door één enkel mechanisme wordt verklaard, maar door een combinatie van overlappende oriëntatiestrategieën.

De casus laat zien dat generatieve AI in verschillende fasen van het onderzoeksproces een rol kan spelen. De grootste toegevoegde waarde blijkt te liggen in de vroege fasen van onderzoek, zoals literatuurverkenning, hypothesevorming en het systematisch verkennen van mogelijke verklaringsmodellen. AI kan grote hoeveelheden literatuur snel synthetiseren, alternatieve hypothesen genereren en helpen bij het structureren van onderzoeksontwerpen. Hierdoor kan het proces van probleemverkenning en conceptuele afbakening aanzienlijk worden versneld.

Tegelijkertijd laat de casus zien dat de epistemische verantwoordelijkheid voor de uiteindelijke onderzoeksarchitectuur bij de menselijke onderzoeker blijft liggen. De selectie van hypothesen, de interpretatie van resultaten en de integratie van verschillende verklaringslijnen vereisen nog steeds curatoriale regie en methodologische discipline. AI functioneert in deze configuratie primair als een instrument voor analyse en synthese, niet als autonome producent van wetenschappelijke kennis.

Wetenschappelijk illusionisme. In de slotparagraaf van het hoofdstuk wordt echter een cruciaal gegeven expliciet gemaakt: het experiment waarop het proefschrift berust, heeft in werkelijkheid nooit plaatsgevonden. De datasets waarop de analyse steunt zijn synthetisch gegenereerd met behulp van generatieve AI. De casus fungeert daarmee als een gecontroleerd experiment in wetenschappelijk illusionisme. Het laat zien dat generatieve AI in staat is om een empirisch ogende studie te produceren die methodologisch plausibel oogt en stilistisch aansluit bij academische conventies. De overtuigingskracht van het manuscript berust niet op spectaculaire claims, maar juist op methodologische soberheid, plausibele variatie in datasets en een interpretatieve stijl die onzekerheden en alternatieve verklaringen expliciet benoemt.

Deze exercitie maakt zichtbaar dat de vorm van wetenschappelijke argumentatie zelf een belangrijke bron van overtuigingskracht kan zijn. Wanneer datasets plausibel ogen en de argumentatie aansluit bij bestaande disciplinetradities, kan een studie overtuigend lijken, zelfs wanneer de empirische basis synthetisch is. De casus onderstreept daarmee het belang van transparantie over de herkomst van data en van methodologische controlemechanismen die verder reiken dan stilistische of retorische plausibiliteit

Hoofdstuk 14 - Casus 'Voorspellingen van digitale transformaties'.

In dit hoofdstuk wordt de totstandkoming van een historisch onderzoek naar voorspellingen over digitale technologie en hun feitelijke maatschappelijke uitkomsten in Nederland tussen 1945 en 2025 geanalyseerd. Het onderzoek reconstrueert een canon van ICT-ontwikkelingen die in politieke en maatschappelijke bronnen als potentieel transformerend werden gepresenteerd. Vanuit deze canon worden historische toekomstclaims systematisch vergeleken met de daadwerkelijke maatschappelijke ontwikkelingen die daarop volgden.

Om deze vergelijking analytisch hanteerbaar te maken wordt een 2×2-matrix ontwikkeld waarin twee dimensies worden gecombineerd: de accuraatheid van de oorspronkelijke voorspelling en de mate waarin de voorspelde maatschappelijke transformatie daadwerkelijk gerealiseerd is. Hierdoor ontstaat een typologie van historische gevallen waarin verwachtingen en uitkomsten uiteenlopende relaties vertonen.

Generatieve AI speelde in dit onderzoek vooral een rol in de exploratieve en analytische fasen van het onderzoeksproces. Taalmodellen werden gebruikt om mogelijke canon-episodes te identificeren, om grote hoeveelheden historisch en beleidsmateriaal te ordenen en om verschillende classificaties van technologische toekomstclaims te verkennen. Daarnaast werd AI ingezet bij het systematisch toepassen van drie analytische lenzen: een claimgerichte lens, een transformatielens en een causale lens die de relatie tussen beide onderzoekt.

De casus laat zien dat AI historisch onderzoek kan ondersteunen bij het structureren en analyseren van omvangrijke discursieve corpora. Tegelijkertijd blijft de beoordeling van historische adequaatheid afhankelijk van menselijke contextualisering en interpretatieve expertise. De analyse illustreert daarmee dat AI vooral waarde toevoegt in de ordening en vergelijking van historisch materiaal, terwijl de uiteindelijke historiografische interpretatie een menselijke verantwoordelijkheid blijft.

Hoofdstuk 15 – casus 'Besturen binnen biologische grenzen'.

In dit hoofdstuk wordt de totstandkoming van een proefschrift dat de invloed van evolutionair-biologisch gevormd op persistente problemen binnen de bestuurskunde beschrijft, onderzocht. Het vertrekpunt van dit onderzoek is de these dat verklaringen van beleidsfalen onvolledig blijven wanneer zij uitsluitend leunen op economische, juridische of religieus-culturele modellen. Daarom is een raamwerk ontwikkeld van tien evolutionair-biologische mechanismen. Het negeren van deze mechanismen bij de beleidsvorming, zou de hardnekkigheid van bepaalde bestuurskundige problemen kunnen verklaren. Empirisch is het onderzoek opgebouwd rond zes casussen, verdeeld over drie conflictvelden, en methodologisch uitgewerkt als kwalitatief en abductief pattern matching-onderzoek.

Generatieve AI speelde in deze studie een brede rol in de vroege en analytische fasen van het onderzoeksproces. AI werd gebruikt voor literatuurverkenning, conceptuele ordening, structurering van het theoretisch kader en het uitwerken van interpretatieve analyses binnen een interdisciplinair corpus. De casus laat daarmee zien dat AI vooral waarde toevoegt waar grote hoeveelheden verspreide literatuur moeten worden samengebracht tot een coherenter analytisch raamwerk.

Tegelijkertijd maakt het hoofdstuk een klassieke epistemische kwetsbaarheid van interpretatieve disciplines zichtbaar. Onderzoek dat werkt met betekenisgeving en narratieve reconstructie is traditioneel gevoelig voor selectieve interpretatie, narratieve coherentie en confirmation bias. Generatieve AI kan deze kwetsbaarheden versterken, omdat taalmodellen optimaliseren voor plausibele en consistente tekst en impliciete aannames gemakkelijk uitwerken tot een steeds coherenter narratief.

Juist deze kwetsbaarheid leidde in deze casus tot een methodologische innovatie. In plaats van AI alleen generatief in te zetten, werd een tweede model gebruikt in een adversariële en controlerende rol. Eén model genereerde tekstvoorstellen en analyses, terwijl een ander model systematisch werd ingezet om inconsistenties, zwakke plekken en alternatieve interpretaties te signaleren. Zo ontstond een vorm van rolgescheiden modellering waarin verschillende epistemische functies bewust van elkaar werden gescheiden. De casus laat daarmee zien dat AI in interpretatieve disciplines niet alleen risico's kan versterken, maar ook nieuwe vormen van methodologische tegencontrole mogelijk maakt.

Hoofdstuk 16 – Casus 'Polyglotte taalverwerving'.

In dit hoofdstuk wordt de totstandkoming van een ontwerpgericht proefschrift over polyglotte taalverwerving besproken. Het vertrekpunt van dit onderzoek betreft de observatie dat taalonderwijs traditioneel per taal wordt georganiseerd, terwijl veel meertalige sprekers gebruikmaken van structurele overeenkomsten tussen talen. In het proefschrift wordt onderzocht hoe deze overeenkomsten systematisch kunnen worden benut in een leerarchitectuur voor het parallel verwerven van meerdere talen. Het onderzoek combineert inzichten uit taalkunde, cognitiewetenschap en didactiek en ontwikkelt een ontwerpraamwerk waarin grammaticale structuren, woordvorming en syntactische patronen tussen talen expliciet met elkaar worden verbonden. Generatieve AI werd daarbij gebruikt om voorbeelden, oefenmateriaal en illustraties van dergelijke structuren te ontwikkelen en te verfijnen.

De casus laat zien dat AI in ontwerpgericht onderzoek vooral waarde heeft bij het uitwerken en operationaliseren van een conceptueel ontwerp**.** Waar traditionele ontwerptrajecten vaak worden begrensd door de tijd die nodig is om voorbeelden en onderwijsconcepten te ontwikkelen, maakt AI het mogelijk om structurele taalpatronen snel te concretiseren in didactische toepassingen. Tegelijkertijd maakt het hoofdstuk duidelijk dat een generatief proces nieuwe eisen stelt aan de conceptuele stabiliteit van het ontwerp**.** Wanneer voorbeelden en formuleringen op grote schaal worden gegenereerd, ontstaat het risico dat terminologie of didactische principes geleidelijk en subtiel gaan verschuiven (tot aan het punt waarop het gegenereerde oefenmateriaal compleet onbruikbaar is geworden). De casus maakt duidelijk dat het genereren en globaal uitwerken van nieuwe concepten met AI, iets totaal anders is dan het op grote schaal en stabiel produceren van artefacten m.b.v. AI.

De casus laat daarmee zien dat generatieve AI in ontwerpgericht onderzoek vooral functioneert als instrument voor concretisering en uitwerking van een bestaande ontwerpstructuur**,** terwijl de architectuur van het leerproces zelf door de onderzoeker wordt ontworpen en bewaakt.

Hoofdstuk 17 - Casus 'AI-gestuurde kennisproductie' (meta-casus).

In dit hoofdstuk word de totstandkoming van dit document geanalyseerd als reflexieve casus van AI-gestuurde kennisproductie. De studie fungeert daarmee als een mise-en-abyme: het onderzoeksobject – de impact van generatieve AI op wetenschappelijke kennisproductie – wordt zichtbaar in het proces waarin het proefschrift zelf tot stand kwam (het 'verpleegstertjes-effect' of 'Droste-effect). De analyse richt zich daarom niet op de inhoud, maar op de architectuur van het onderzoeksproces waarin deze teksten zijn geproduceerd.

De casus laat zien dat generatieve AI academische productie aanzienlijk kan versnellen. De eerste volledige versie van dit proefschrift werd binnen twee dagen gegenereerd in één lange interactie met een taalmodel en was onmiddellijk bruikbaar als academisch werkdocument. Deze uitzonderlijke doorlooptijd vormt echter niet het belangrijkste inzicht van de studie. De snelheid van het proces blijkt vooral te zijn verklaard door de wijze waarop het onderzoeksproces werd georganiseerd.

Twee vormen van disciplinering blijken daarbij cruciaal. Procesdiscipline hield het generatieve traject gericht op het onderzoeksdoel en voorkwam dat het schrijfproces ontspoorde in iteratieve herschrijvingen of zijpaden. Belangrijker nog was structuurdiscipline: de systematische opbouw van een analytisch framework, gevolgd door een consequente toepassing ervan in de casushoofdstukken en de beantwoording van de empirische onderzoeksvragen (passend binnen een vooraf ontworpen architectuur van het manuscript). Deze architectuur bepaalde de cadans van het gehele document en zorgde ervoor dat de interactie met het taalmodel gedurende een lange dialoog inhoudelijk stabiel bleef.

De casus maakt bovendien zichtbaar dat een expliciete onderzoeksarchitectuur circulaire validatie kan doorbreken. Door verschillende epistemische rollen – zoals genereren, analyseren en controleren – van elkaar te onderscheiden, wordt voorkomen dat modeloutput impliciet als eigen bewijs fungeert. Curatoriale regie over het proces blijkt daarmee een noodzakelijke voorwaarde voor betrouwbare AI-ondersteunde kennisproductie.

Naast de analyse van de totstandkoming van dit proefschrift brengt het hoofdstuk ook een aantal casus-overstijgende implicaties van AI-gestuurde kennisproductie in beeld. Een eerste implicatie betreft de rol van architectuur: taalmodellen produceren lokaal coherente tekst, maar bewaken niet zelfstandig de globale argumentatieve opbouw van een manuscript. Een tweede implicatie betreft curatoriale regie: door contextverlies en terminologische drift kan de samenhang van een betoog geleidelijk vervagen, menselijk curatie is daarom onontbeerlijk. Een derde implicatie betreft literatuurgebruik: generatieve AI bleek hier niet alleen een bron van risico, maar ook een hulpmiddel voor systematische controle op het juist gebruik van sleutelreferenties. Een vierde implicatie betreft de organisatie van het onderzoeksproces: het scheiden van producerende en evaluerende rollen bleek effectiever dan het laten samenvallen van beide functies in één model.

Daarnaast worden in het hoofdstuk een reeks praktische leerervaringen uit het bredere onderzoeksprogramma benoemd: regeneratiegedrag dat zorgvuldig gecureerde passages kan ondermijnen, modeloutput die vaak parafraserend in plaats van tekstueel exact is, gebruikers die niet kunnen vertrouwen op de leesvaardigheden van een taalmodel, beperkingen in context-vensters die leiden tot semantische drift. Juist door procesdiscipline en structuurdiscipline kon dat in deze casus – bij de vervaardiging van dit, voorliggende document dus - in grote mate gemitigeerd worden.

Hoofdstuk 18 - Synthese en definitieve beantwoording.

In dit hoofdstuk worden de bevindingen uit de vijf casussen bijeen gebracht en afgezet tegen de voorlopige conclusies uit de tweede tussenbalans (hoofdstuk 12). Daarbij is een expliciete delta-benadering gevolgd: per empirische onderzoeksvraag wordt nagegaan in hoeverre de casusanalyses het eerdere antwoord bevestigen, nuanceren of corrigeren.

De analyse laat zien dat generatieve AI de cognitief-methodologische dimensie van kennisproductie verandert. In verschillende casussen blijkt AI in staat om synthese en analyse aanzienlijk te versnellen, maar ook nieuwe epistemische kwetsbaarheden te introduceren. Taalmodellen produceren overtuigende en coherente narratieven, terwijl empirische onzekerheden of alternatieve interpretaties naar de achtergrond kunnen verdwijnen. Tegelijkertijd maken dezelfde systemen nieuwe vormen van verificatie mogelijk, bijvoorbeeld bij het systematisch controleren van literatuurgebruik en argumentatieve consistentie.

De pedagogisch-institutionele analyse laat zien dat AI het karakter van academisch werk verandert. In meerdere casussen verschuift de rol van de onderzoeker van lineaire auteur naar curator van een generatief proces. Onderzoek en schrijven worden daarmee minder een lineaire productie van tekst en meer een iteratieve organisatie van genereren, analyseren en valideren.

In de normatief-politieke dimensie blijkt de impact van AI sterk domein-specifiek. Nieuwe afhankelijkheden ontstaan vooral op het niveau van infrastructuur en toegang tot modellen, terwijl veel normatieve keuzes in de onderzochte casussen nog steeds verbonden blijven met traditionele onderzoekspraktijken.

Op basis van deze bevindingen wordt de centrale onderzoeksvraag als volgt beantwoord: generatieve AI verandert wetenschappelijke kennisproductie doordat zij de architectuur van het onderzoeksproces herconfigureert. Onder bepaalde voorwaarden kan dit onderzoek versnellen, interdisciplinair verbreden en methodologisch verrijken. Deze effecten zijn echter niet automatisch; zij hangen sterk af van de manier waarop het generatieve proces wordt georganiseerd.

De casussen laten zien dat de kwaliteit van AI-gestuurde kennisproductie niet primair wordt bepaald door de capaciteiten van individuele modellen, maar door de onderzoeksarchitectuur waarin zij worden ingezet. Expliciete rolscheiding, curatoriale regie en transparante methodologische procedures blijken daarbij cruciale voorwaarden. Het hoofdstuk sluit af met een bespreking van de methodologische waarborgen van het onderzoek en met aanbevelingen voor vervolgonderzoek naar discipline-specifieke effecten van AI en naar robuuste onderzoeksarchitecturen voor AI-ondersteunde wetenschap.

Epiloog.

In dit proefschrift is onderzocht in hoeverre kunstmatige intelligentie de kennisproductie verandert in haar aard, kwaliteit en legitimiteit. De vijf casussen hebben laten zien dat AI-gestuurde kennisproductie tegelijk beloftevol en kwetsbaar is. Beloftevol, omdat AI processen radicaal kan versnellen, patronen zichtbaar maakt die anders verborgen zouden blijven en nieuwe vormen van synthese mogelijk maakt. Kwetsbaar, omdat diezelfde kracht gepaard gaat met structurele risico's: black-box-karakter, bias, hallucinaties, en de verleiding van schijnbare meerstemmigheid.

De doorslaggevende factor is telkens de menselijke curator. Waar AI de rol van uitvoerder en patroonherkenner vervult, is het de mens die selecteert, corrigeert en legitimeert. Zonder deze rol vervallen de beloftes van AI in onbetrouwbaarheid; met deze rol kan AI juist de horizon van kennisproductie verbreden. Transparantie, methodologische zuiverheid en gemotiveerde curatie zijn hierbij niet optioneel, maar noodzakelijke voorwaarden.

Deze bevindingen leiden tot wat kan worden opgevat als een 'eed van Socrates' voor het AI-tijdperk. Evenals artsen gebonden zijn aan het principe eerst geen schade te berokkenen, zou elke onderzoeker die AI gebruikt moeten zweren: eerst geen valse kennis produceren. Deze eed houdt in dat men de grenzen van zowel menselijke als machinale kennis erkent, dat men transparant verantwoordt welke keuzes zijn gemaakt, en dat men de verantwoordelijkheid voor interpretatie nimmer aan de technologie delegeert.

De heilzaamheid van deze eed ligt in haar preventieve werking. Zij voorkomt dat de kracht van AI omslaat in intellectuele arrogantie en dat schijnbare efficiëntie belangrijker wordt dan epistemologische integriteit. Daarmee wordt zij een waarborg tegen de uitholling van wetenschappelijke legitimiteit.

Met bewuste toewijding aan de eed van Socrates kan AI-gestuurde kennisproductie een nieuwe, legitieme praktijk van weten worden. Zonder deze toewijding dreigt zij te verworden tot een instrument dat vertrouwen ondermijnt en de epistemische basis van wetenschap uitholt.

Inhoud:

Hoofdstuk 1 – Inleiding: AI en kennisproductie

1.1 Inleiding: AI en kennisproductie

De opkomst van kunstmatige intelligentie vormt een fundamentele uitdaging voor de wijze waarop wetenschap wordt bedreven. Sinds de jaren zestig wordt al nagedacht over de rol van computers in onderzoek, maar pas met de doorbraak van generatieve systemen en grootschalige taalmodellen is er een situatie ontstaan waarin de computer niet langer slechts een rekentool of simulatiemachine is, maar een medeproducent van kennis. Daarmee wordt de vraag urgent hoe AI de epistemische kern van wetenschap verandert: hoe wij data verzamelen, analyseren, verklaren en presenteren.

Het probleem bij veel discussies over AI is dat men spreekt over “de” impact van “de” AI, alsof het gaat om een homogeen fenomeen. In werkelijkheid is zowel de wetenschap als de technologie divers. Wetenschap kent uiteenlopende tradities en doelen; AI kent verschillende vormen, variërend van brede taalmodellen tot uiterst specifieke diagnostische systemen. Wie die verschillen niet onderkent, loopt het risico te vervallen in generalisaties die meer verhullen dan verhelderen.

Dit hoofdstuk zet daarom een kader neer dat de diversiteit van kennisproductie recht doet en tegelijk overzichtelijk blijft. Het onderscheid tussen vier archetypen van kennisproductie biedt een compacte maar veelzeggende ordening. Daarnaast wordt expliciet aandacht besteed aan de tweedeling tussen generieke en domein-specifieke AI, en wordt de bijzondere status van de informatica en AI-ontwikkeling zelf als meta-domein belicht. Samen vormt dit het fundament voor de verdere analyse van de impact van AI in dit proefschrift.

1.2 Vier archetypen van kennisproductie

1.2.1 Natuurwetenschappelijk onderzoek

De natuurwetenschappen streven naar universele wetmatigheden door middel van metingen, experimenten en wiskundige modellen (Popper, 1959). Betrouwbaarheid berust hier op reproduceerbaarheid en controleerbare causaliteit.

AI raakt dit domein op meerdere manieren. Systemen voor data-analyse en simulatie versnellen de ontdekking van patronen. Denk aan toepassingen in de klimaatwetenschap of moleculaire biologie (Jumper et al., 2021). Tegelijkertijd dreigt causaliteit te worden vervangen door correlatie: algoritmen herkennen patronen, maar verklaren ze niet altijd. Black box-modellen zijn moeilijk te controleren, wat wringt met het klassieke ideaal van transparantie (Burrell, 2016).

1.2.2 Historisch onderzoek

Geschiedenis is gericht op het begrijpen van unieke gebeurtenissen in hun context. Bronnenkritiek en hermeneutische interpretatie zijn essentieel (Gamer, 1960/2004). AI opent hier nieuwe mogelijkheden. Digitalisering van archieven en automatische tekstanalyse maken corpora toegankelijk die voorheen onhandelbaar waren (Berry, 2012). Maar er schuilen gevaren. Automatische clustering kan nuance reduceren. Omdat veel AI-systemen getraind zijn op hedendaagse taal, dreigt anachronisme: moderne betekenissen worden op het verleden geprojecteerd. Daarnaast kunnen systemen bronnen of citaten verzinnen, zogenaamde hallucinaties, wat de kernwaarde van authenticiteit bedreigt (Bender et al., 2021).

1.2.3 Antropologisch onderzoek

Antropologie draait om nabijheid: door participerende observatie en 'thick description' probeert de onderzoeker cultuur van binnenuit te begrijpen (Geertz, 1973). AI kan hier ondersteunen door veldnotities te transcriberen, interviews te coderen en vertalingen te verzorgen (Boellstorff, 2021). Dit versnelt analyse en maakt cross-culturele vergelijking mogelijk. Toch gaat er iets wezenlijks verloren. De relationele dimensie van veldwerk – vertrouwen, aanwezigheid, lichamelijkheid – laat zich niet automatiseren. Er dreigt oppervlakkigheid en homogenisering, omdat mondiale datapatronen lokale diversiteit kunnen overschrijven (Couldry & Mejias, 2019).

1.2.4 Ontwerpgericht onderzoek

Naast verklaren en begrijpen is er ook een vorm van kennisproductie die primair gericht is op het maken van oplossingen: ontwerpgericht onderzoek (Simon, 1969; Schön, 1983). Succes wordt hier vaak afgemeten aan bruikbaarheid en effectiviteit. AI blijkt een krachtige co-ontwerper. Het genereert varianten, optimaliseert ontwerpen en simuleert uitkomsten, van medicijnen tot computerchips (Dean et al., 2021). Maar dit roept nieuwe risico's op. Onderzoekers kunnen afhankelijk worden van AI, waardoor hun eigen vakkennis erodeert. Ontwerpen zijn vaak ondoorzichtig en moeilijk verantwoordbaar. En omdat AI vaak optimaliseert voor efficiëntie, dreigt het morele en sociale waarden te veronachtzamen (Zhang & Dafoe, 2019).

1.3 Generieke en domein-specifieke AI

Het is cruciaal te onderscheiden tussen generieke AI en domein-specifieke AI. Generieke AI verwijst naar grote modellen die getraind zijn op uiteenlopende internetbronnen. Hun kracht ligt in veelzijdigheid: samenvatten, ideeën genereren, teksten formuleren. Hun zwakte is dat deze internetdata vol ruis, bias en desinformatie zitten (Bender et al., 2021). Voor natuurwetenschappen kunnen ze inspirerend zijn, maar leveren ze geen valide data. Voor historici vergroten ze het risico op anachronisme en hallucinaties. Voor antropologen homogeniseren ze culturele verschillen. Domein-specifieke AI daarentegen wordt getraind op zorgvuldig geselecteerde datasets en richt zich op een smal toepassingsgebied. Voorbeelden zijn AlphaFold in de biologie of medische beeldherkenning. Zulke systemen blijken vaak betrouwbaarder dan menselijke analyse, maar missen de breedte en flexibiliteit van generieke modellen. In geschiedenis en antropologie spelen ze vooral een ondersteunende rol, bijvoorbeeld bij OCR of transcriptie. Dit onderscheid laat zien dat "AI" geen uniforme entiteit is. Impact verschilt per type technologie, per discipline en per kernwaarde van het vakgebied.

1.4 Het meta-domein: informatica en AI zelf

De vier archetypen beschrijven verschillende manieren van kennisproductie. Maar er is één domein dat erbovenuit steekt: de informatica, en specifiek de ontwikkeling van AI zelf. Hier is sprake van reflexiviteit: AI is zowel object als instrument van onderzoek. AI-ontwikkeling versnelt zichzelf. Modellen helpen nieuwe modellen ontwerpen, algoritmen optimaliseren algoritmen, chips worden ontwikkeld met behulp van AI die later weer AI krachtiger maken (Real et al., 2019). Deze zelfversterkende cyclus vergroot innovatiekracht, maar roept ook vragen op over controle en afhankelijkheid. Fouten, bias of blinde vlekken kunnen exponentieel versterkt worden.

Daarbij komt dat AI-ontwikkeling niet neutraal is. Het is een geopolitiek strijdtoneel geworden: wie AI beheerst, beheerst ook de kennisproductie in andere domeinen. AI wordt ingezet in cyberoorlogsvoering,

surveillance en geopolitieke competitie (Kania, 2019). Daarmee raakt het meta-domein niet alleen de epistemische kern van wetenschap, maar ook de maatschappelijke en politieke ordening.

1.5 Impactdimensies en relevantie

De vraag hoe kunstmatige intelligentie de wetenschap beïnvloedt kan vanuit talloze invalshoeken worden gesteld: economisch, technologisch, sociologisch of filosofisch. Voor dit proefschrift is gekozen voor zes impactdimensies: cognitief, methodologisch, pedagogisch, institutioneel, macht en eigenaarschap, en betrouwbaarheid en normativiteit.

Deze keuze is niet willekeurig, maar berust op drie overwegingen. In de eerste plaats sluiten de dimensies direct aan bij de kernfuncties van wetenschap. Wetenschap is altijd een cognitieve onderneming waarin naar kennis wordt gezocht, een methodologische praktijk waarin onderzoek wordt gedaan, een pedagogische context waarin nieuwe generaties worden gevormd, een institutioneel bouwwerk dat bestaat uit universiteiten, tijdschriften en peer review, een systeem dat ingebed is in machtsrelaties die toegang en richting bepalen, en een normatief kader dat vastlegt wat telt als bewijs en als verantwoord. Andere invalshoeken – zoals economische effecten of juridische regulering – zijn zeker relevant, maar vloeien grotendeels voort uit of overlappen met deze kernfuncties.

In de tweede plaats maakt deze indeling vergelijkbaarheid mogelijk tussen disciplines. Natuurwetenschappen, geesteswetenschappen en ontwerpwetenschappen verschillen aanzienlijk in werkwijze en doel, maar in elk domein spelen cognitieve, methodologische en institutionele vragen. Door een vast raster van dimensies te hanteren, wordt zichtbaar dat AI overal ingrijpt, zij het op uiteenlopende manieren.

In de derde plaats zijn dit precies de dimensies waar de maatschappelijke en wetenschappelijke debatten zich concentreren. Discussies over hallucinaties en bias raken de betrouwbaarheid; zorgen van docenten over scripties betreffen de pedagogische dimensie; vragen over wie AI beheerst raken machtsverhoudingen; en de roep om transparantie in algoritmen raakt methodologische en normatieve kwesties. Door deze zes dimensies centraal te stellen, sluit dit onderzoek niet alleen aan bij academische discussies, maar ook bij urgente maatschappelijke zorgen.

1.5.1 Het pedagogische vraagstuk van de AI-generatie

Waar voor de huidige generatie onderzoekers kunstmatige intelligentie vooral een nieuwe aanvulling is, staat de volgende generatie kinderen en jongeren op het punt op te groeien in een wereld waarin AI vanaf het begin vanzelfsprekend aanwezig zal zijn. Dit verschil is niet louter temporeel, maar raakt de kern van menselijke vorming. De bestaande scheiding tussen menselijk redeneren en technologische ondersteuning, ooit zo bepalend voor leerprocessen, dreigt te verdwijnen. Het gevolg is dat de cognitieve ontwikkeling van deze nieuwe generatie zich voltrekt in voortdurende wisselwerking met systemen die direct antwoorden en oplossingen aanleveren.

Deze verschuiving roept fundamentele vragen op. Als basisvaardigheden zoals memoriseren, rekenen of taalbeheersing systematisch worden uitbesteed aan AI, welke vermogens blijven dan nog voorbehouden aan de mens? En welke vormen van autonomie worden ondermijnd wanneer een leerling niet langer leert omgaan met begrensde kennis, maar vanaf het begin toegang heeft tot een quasi-onbegrensde informatievoorziening? Het risico bestaat dat de vormende waarde van leren – de ervaring van inspanning, het maken van fouten, het langzaam opbouwen van inzicht – verloren gaat in een omgeving die voortdurend snelle en ogenschijnlijk sluitende antwoorden biedt.

Voor scholen, curricula en universiteiten betekent dit dat hun rol fundamenteel moet worden herzien. Hun taak verschuift van het overdragen van kennis die elders moeilijk te verkrijgen was naar het cultiveren van kritische vermogens, metacognitieve vaardigheden en een reflectieve houding tegenover technologie. Het pedagogische project van de eenentwintigste eeuw kan niet beperkt blijven tot instrumentele vaardigheid in het hanteren van AI-tools, maar moet gericht zijn op de vorming van zelfstandige denkers die in staat zijn deze technologie op waarde te schatten, te corrigeren en te begrenzen.

Hier treedt ook de zogenaamde AI-paradox op scherp naar voren. Om AI vruchtbaar te kunnen gebruiken, is een basis nodig die onafhankelijk van AI is gevormd. Alleen wie zelf beoordelingsvermogen en denkpatronen heeft ontwikkeld, kan technologie kritisch en creatief inzetten. Maar de generaties die nu opgroeien, missen juist die ervaring van een wereld zonder AI. Dat maakt het des te urgenter om in de primaire, secundaire en tertiaire vorming expliciet ruimte te scheppen voor activiteiten die niet direct door AI worden gemedieerd, maar waarin leerlingen de kans krijgen hun eigen oordeelsvermogen, creativiteit en doorzettingskracht te ontwikkelen.

1.5.2 Cognitieve dimensie

De wetenschap is allereerst een cognitieve onderneming: zij draait om het zoeken naar kennis, het herkennen van patronen en het vormen van verklaringen. Kunstmatige intelligentie grijpt hier direct in. Systemen die in staat zijn miljoenen teksten samen te vatten of onverwachte patronen in datasets te detecteren, veranderen de manier waarop wetenschappers denken en redeneren. Waar voorheen kennis langzaam werd opgebouwd door close reading en minutieuze analyse, kan AI in enkele ogenblikken een overzicht genereren. Daarmee verandert ook de verhouding tussen detailkennis en synthese: onderzoekers worden minder verzamelaars van gegevens en meer curatoren van AI-gegenereerde inzichten. De vraag is echter of dit nieuwe evenwicht leidt tot verdieping van begrip of tot een oppervlakkige versnelling, waarin het gevaar schuilt dat nuance en interpretatieve rijkdom verloren gaan.

1.5.3 Methodologische dimensie

Naast het denken over kennis is wetenschap altijd ook een praktijk van methoden en technieken geweest, en juist hier grijpt AI diep in. In de natuurwetenschappen betekent dit dat experimenten en simulaties door AI kunnen worden versneld en opgeschaald. In de geesteswetenschappen opent AI de mogelijkheid om archieven en corpora op ongekende schaal te doorzoeken. En in de sociale wetenschappen verandert de analyse van interviews of observaties fundamenteel door automatische transcriptie en patroonherkenning. Maar in al deze gevallen rijst de vraag of de methodologische kernwaarden behouden blijven: transparantie, reproduceerbaarheid en causaliteit. Als AI resultaten oplevert die niet of nauwelijks te verifiëren zijn, dreigt de methodologische basis van wetenschap te verschuiven van verklaarbaarheid naar louter bruikbaarheid, een ontwikkeling die zowel beloftevol als gevaarlijk is.

1.5.4 Institutionele dimensie

Wetenschap voltrekt zich nooit in een vacuüm, maar altijd binnen instituties: universiteiten, onderzoeksraden, tijdschriften en conferenties. Deze instituties zijn dragers van erkenning en selectie, ze bepalen wie toegang krijgt tot de gemeenschap van wetenschappers en wie niet. AI ondermijnt en versterkt deze instituties tegelijk. Enerzijds kan zij de peer review versnellen, de toegang tot literatuur verruimen en de samenwerking vergemakkelijken. Anderzijds roept zij vragen op over auteurschap, originaliteit en de waarde van publicaties die door niemand meer handmatig gelezen worden. Als AI het schrijven en beoordelen van artikelen overneemt, verliest de publicatiecultuur haar rituele en sociale betekenis, en moet zij opnieuw worden gelegitimeerd.

1.5.5 Macht en eigenaarschap

Geen van deze ontwikkelingen kan ten slotte begrepen worden zonder oog voor macht en eigenaarschap. De ontwikkeling en het gebruik van AI is geconcentreerd bij enkele grote technologiebedrijven en militaire of geopolitieke spelers. Daarmee is de toegang tot AI, en dus tot de kennis die ermee geproduceerd wordt, ongelijk verdeeld. Universiteiten en publieke instellingen raken afhankelijk van infrastructuren die zij zelf niet in handen hebben. Dit roept vragen op over soevereiniteit van kennis, over de mogelijkheid om eigen wetenschappelijke agenda's te bepalen, en over de kwetsbaarheid van wetenschap als publieke zaak. AI is daarmee niet alleen een instrument van kennis, maar ook een instrument van macht, dat de verhoudingen tussen staten, bedrijven en wetenschappelijke gemeenschappen opnieuw ordent.

1.5.6 Betrouwbaarheid en normativiteit

Al deze kwesties raken aan de laatste dimensie: betrouwbaarheid en normativiteit. Hier gaat het om de kern van wat wetenschap onderscheidt van andere vormen van kennis: haar oriëntatie op waarheid, bewijs en verantwoording. Hallucinaties, bias en black box-modellen ondermijnen de betrouwbaarheid van AI-gestuurde kennisproductie. Wat betekent peer review nog als de data en algoritmen waarop een conclusie berust, niet inzichtelijk zijn? Hoe kan authenticiteit worden gewaarborgd in disciplines die afhankelijk zijn van bronnen, wanneer AI plausibele maar fictieve citaten kan produceren? De uitdaging is niet alleen technisch, maar ook normatief: de wetenschappelijke gemeenschap moet opnieuw formuleren welke standaarden gelden voor bewijs en verantwoording in een tijdperk waarin kennis steeds vaker door hybride mens-machine-systemen wordt voortgebracht.

1.6 Synthese en vooruitblik

Met de vier archetypen en het meta-domein is een raamwerk neergezet dat de diversiteit van wetenschappelijke kennisproductie zichtbaar maakt, zonder te verzanden in een eindeloze opsomming van methodologische details. Deze archetypen laten zien dat wetenschap vele vormen kan aannemen: van het zoeken naar universele wetmatigheden in de natuurwetenschappen, tot het reconstrueren van unieke gebeurtenissen in de geschiedenis, het verdiepen in menselijke nabijheid en betekenis in de antropologie, en het ontwerpen van oplossingen in de ingenieurs- en ontwerpdisciplines. Daarboven bevindt zich de informatica, en in het bijzonder de ontwikkeling van AI zelf, die een aparte status heeft omdat zij de voorwaarden van kennisproductie in alle domeinen tegelijk verandert.

De keuze voor zes impactdimensies maakt het mogelijk om deze diversiteit te analyseren zonder haar uit het oog te verliezen. Cognitieve, methodologische, pedagogische, institutionele, machtsgerelateerde en normatieve aspecten raken de kern van wat wetenschap tot wetenschap maakt. Door ze te onderscheiden wordt zichtbaar dat AI in elk domein op eigen wijze ingrijpt: soms als versneller, soms als verstorende kracht, vaak als beide tegelijk. Wat in de natuurwetenschappen tot ongekende ontdekkingen kan leiden, kan in de geschiedschrijving resulteren in anachronismen of hallucinaties. Wat in de ontwerpwetenschappen innovatiekracht geeft, kan in de antropologie de relationele kern van veldwerk ondermijnen.

Het onderscheid tussen generieke en domein-specifieke AI onderstreept dit nog eens. Generieke modellen, getraind op een veelheid aan internetbronnen, zijn breed inzetbaar maar kwetsbaar door hun neiging tot vertekening en verzinsels. Domein-specifieke systemen, getraind op zorgvuldig gecureerde data, blijken vaak betrouwbaarder binnen hun niche, maar missen de brede toepasbaarheid die generieke modellen juist zo aantrekkelijk maakt. De impact van AI is dus nooit uniform, maar altijd afhankelijk van zowel het type technologie als het type wetenschap dat ermee werkt.

In die zin kan het hier gepresenteerde raamwerk worden opgevat als een matrix: de archetypen vormen de rijen, de impactdimensies de kolommen. Het is geen schema dat in één oogopslag alle antwoorden geeft, maar een structuur die uitnodigt tot analyse. In de volgende hoofdstukken zal langs de zes dimensies systematisch worden onderzocht hoe AI de wetenschap hertekent, waarbij de archetypen als illustratieve semi-casuïstiek dienen. Hun bespreking is volledig gebaseerd op openbare bronnen en laat de breedte van AI's

Literatuurlijst Hoofdstuk 1

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Berry, D. M. (2012). *Understanding Digital Humanities*. Palgrave Macmillan.

Boellstorff, T. (2021). *Virtual Society: Technology, Cyberculture, and Ethnography*. Princeton University Press.

Burrell, J. (2016). How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1), 1–12. https://doi.org/10.1177/2053951715622512

Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

Dean, J., Corrado, G., & Shazeer, N. (2021). The Deep Learning Revolution and Its Implications for AI. *Communications of the ACM*, 64(6), 58–65. https://doi.org/10.1145/3448250

Gadamer, H.-G. (2004). *Truth and Method* (2nd rev. ed., J. Weinsheimer & D. G. Marshall, Trans.). Continuum. (Original work published 1960).

Geertz, C. (1973). *The Interpretation of Cultures*. Basic Books.

Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kania, E. (2019). AI Weapons and China's Military Innovation. *Texas National Security Review*, 2(1), 42–53.

Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson.

Real, E., Liang, C., So, D. R., & Le, Q. V. (2019). AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. *Proceedings of the 37th International Conference on Machine Learning*, 8007–8019.

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Basic Books.

Simon, H. A. (1969). *The Sciences of the Artificial*. MIT Press.

Zhang, B., & Dafoe, A. (2019). Artificial Intelligence: American Attitudes and Trends. *Oxford University Pres*

Hoofdstuk 2 – Onderzoeksvragen

2.1 Inleiding: van kader naar vragen

In het vorige hoofdstuk is een raamwerk ontwikkeld dat de veelvormigheid van wetenschap recht doet en tegelijk houvast biedt om de impact van kunstmatige intelligentie te begrijpen. Vier archetypen van kennisproductie en een meta-domein werden onderscheiden, en zes impactdimensies werden benoemd die samen de kernfuncties van wetenschap raken. Daarmee is een structuur ontstaan die zichtbaar maakt waar AI ingrijpt, maar nog niet welke vragen we daarover moeten stellen. Het is nu zaak de stap te maken van beschouwing naar explicitering: welke kwesties moeten dit proefschrift leiden, en waarom zijn die juist nu relevant?

De noodzaak om deze vragen te formuleren wordt zichtbaar in de manier waarop AI de wetenschappelijke praktijk al verandert. Onderzoekers hebben eeuwenlang kennis in fragmenten geproduceerd, vaak in de vorm van kleine specialistische bijdragen. Dat is nuttig geweest, maar roept ook de vraag op of wetenschap zich in een tijdperk van AI nog kan veroorloven te blijven steken in zulke postzegel-wijsheden. Zoals Ioannidis (2005) aantoonde, zijn veel gepubliceerde onderzoeksresultaten mogelijk onjuist, wat de waarde van gefragmenteerde bijdragen verder in twijfel trekt. AI maakt het mogelijk verbanden te leggen die het menselijke blikveld overschrijden. Dat opent perspectieven op integratie, maar stelt ons ook voor de vraag hoe waardevol specialisatie nog is in een context waarin synthese door machines kan worden gegenereerd.

Daarbij komt dat het hart van veel onderzoek, het literatuuronderzoek, een ander karakter krijgt. Waar vroeger maanden nodig waren om de stand van zaken in een vakgebied te doorgronden, kunnen algoritmen in een paar minuten duizenden artikelen doorzoeken en patronen blootleggen. Dit roept de vraag op of de klassieke vorm van literatuurstudie nog relevant blijft, of dat zij transformeert in een proces van beoordelen en corrigeren van AI-gegenereerde overzichten.

Ook empirisch onderzoek wordt geraakt. Data science en machinaal leren maken het mogelijk patronen in grote datasets te herkennen met een snelheid die voorheen ondenkbaar was. Daarmee lijkt het tempo van ontdekking te versnellen, maar tegelijk verschuift de nadruk van verklaren naar voorspellen. Dat roept de vraag op hoe wetenschap haar belofte van causaliteit en inzicht overeind kan houden in een tijd waarin correlaties steeds dominanter worden.

Zelfs de sociale organisatie van wetenschap staat onder druk. Als AI kan optreden als peer reviewer – en dat desnoods in veelvoud, met meerdere modellen in verschillende rollen – dan wordt het beoordelingssysteem dat de academie sinds de Verlichting draagt fundamenteel herzien. Wat betekent erkenning nog als de toets van menselijke collega's vervangen of aangevuld wordt door machinale evaluatie? Daarbij is het niet alleen denkbaar dat AI de menselijke beoordelaar vervangt, maar ook dat meerdere systemen tegelijk verschillende rollen aannemen. Het ene model kan optreden als criticus, het andere als verdediger, en weer een derde als redacteur die de balans opmaakt. De klassieke situatie van één of enkele peers die een artikel beoordelen, verandert dan in een machinale simulatie van een debat. Wat blijft er in zo'n constellatie over van het idee dat peer review een gesprek is tussen vakgenoten die hun oordeel baseren op gedeelde tradities en menselijke ervaring?

Daarmee raakt de discussie ook de zinvolheid van de academische rituelen zelf. Als studenten AI gebruiken om scripties te schrijven en docenten AI om ze te controleren, dringt zich de vraag op welk doel het schrijven van scripties dan nog dient. En als wetenschappelijke artikelen grotendeels door machines worden gegenereerd en ge-peer-reviewd, wie leest ze dan nog en waarom? Het radicale scenario is dat wetenschappelijke teksten uiteindelijk nauwelijks meer door mensen gelezen worden. Wanneer artikelen door machines worden geschreven en door andere machines beoordeeld, ontstaat een

circulair systeem waarin de communicatieve functie van wetenschap – het uitwisselen van ideeën en het gezamenlijk zoeken naar waarheid – in feite wordt uitgehold.

Ten slotte is er de AI-paradox: om AI vruchtbaar te kunnen gebruiken, moet men eerst zonder AI gevormd zijn. Alleen dan beschikt men over de kritische vermogens om het instrument productief en verantwoord in te zetten. Maar de generaties die nu opgroeien, hebben die ervaring van een wereld zonder AI niet meer. Hoe kunnen zij leren zelfstandig te oordelen in een omgeving waar AI met de paplepel wordt ingegoten?

En zelfs wanneer men erin slaagt AI kritisch te hanteren, blijft de fundamentele onbetrouwbaarheid van deze systemen op de achtergrond meespelen. Zoals Bender et al. (2021) waarschuwen, zijn hallucinaties, bias en verzonnen referenties geen randverschijnselen, maar ingebakken risico's van grootschalige taalmodellen. Lawrence (2003) benadrukt bovendien dat publicatiedruk als politieke dynamiek in de academie al zorgt voor vertekeningen, en Fanelli (2010) toont empirisch aan dat deze druk systematische bias in wetenschappelijk onderzoek vergroot. De vraag is niet óf deze problemen optreden, maar hoe wetenschap daarmee kan omgaan zonder haar normatieve oriëntatie op waarheid en bewijs te verliezen.

2.2 Wetenschappelijke relevantie van de vragen

Deze observaties laten zien dat AI niet zomaar een instrument is dat de wetenschap efficiënter maakt, maar een kracht die de logica van kennisproductie zelf verandert. Wetenschappelijk relevant zijn de vragen omdat ze raken aan de kern van wat wetenschap definieert: de verhouding tussen data en theorie, tussen detail en synthese, tussen mens en machine.

Bestaande literatuur over AI in de wetenschap is vaak gefragmenteerd. Sommige studies richten zich op specifieke disciplines, zoals biologie of geschiedenis, en tonen hoe AI daar processen versnelt of ondersteunt. Andere bijdragen belichten ethische kwesties zoals bias en privacy. Wat ontbreekt, is een geïntegreerd perspectief dat de hele breedte van wetenschap omvat en de uiteenlopende effecten systematisch in kaart brengt. Juist het onderscheid tussen archetypen en impactdimensies maakt zo'n overzicht mogelijk.

Daarbij speelt ook de factor van versnelling. Empirisch onderzoek krijgt door AI en data science een enorme boost. Waar onderzoekers vroeger maanden of jaren nodig hadden om patronen te ontdekken in complexe datasets, leveren algoritmen nu in dagen of zelfs uren resultaten op. Deze versnelling verandert de logica van wetenschappelijke productie: de lengte van projecten, de timing van publicaties en zelfs de manier waarop carrières worden opgebouwd, verschuiven mee. Toch ontbreekt in de literatuur vaak een systematische reflectie op dit verschijnsel. Daarmee blijft onderbelicht hoe AI niet alleen de inhoud maar ook het ritme en tempo van wetenschap beïnvloedt.

Juist hier wordt zichtbaar waarom de relevantie van dit onderzoek verder reikt dan technische toepassingen. Het gaat niet alleen om de vraag wat AI kan, maar om de vraag hoe wetenschap zichzelf moet heruitvinden. Zoals Kuhn (2012) duidelijk maakte, bestaat wetenschap uit paradigma's, en fragmentatie of omslagpunten moeten altijd in historisch perspectief worden gezien. Originaliteit kan niet langer uitsluitend worden gedefinieerd door het aantal publicaties of citaties, maar vraagt om het vermogen barrières tussen disciplines te doorbreken en verbanden te leggen die eerder onzichtbaar bleven. In deze context bieden de interpretatieve benaderingen die Geertz (1973) introduceerde – met nadruk op thick description – een alternatief voor fragmentatie, terwijl Gadamer's (2004) hermeneutische perspectief de legitimiteit van geesteswetenschappelijk onderzoek onderstreept die verder reikt dan positivistische methodologie.

2.3 Maatschappelijke relevantie van de vragen

De vraag naar de impact van kunstmatige intelligentie op kennisproductie is niet alleen een wetenschappelijke kwestie. Zij raakt ook aan de manier waarop samenlevingen omgaan met kennis, autoriteit en vorming. Wetenschap is nooit een onderneming geweest die zich louter binnen de muren van universiteiten afspeelde. Haar resultaten vinden hun weg naar beleid, onderwijs, gezondheidszorg en cultuur, en bepalen mede hoe samenlevingen zichzelf begrijpen. Wanneer AI deze kennisproductie fundamenteel verandert, heeft dit dus directe maatschappelijke consequenties.

Een eerste zorg betreft de betrouwbaarheid van kennis waarop burgers, beleidsmakers en professionals vertrouwen. Wanneer AI betrokken is bij het genereren of beoordelen van wetenschappelijke inzichten, maar tegelijkertijd vatbaar blijft voor hallucinaties, bias en ondoorzichtige besluitvorming, dan staat de geloofwaardigheid van wetenschap op het spel. Zoals Burrell (2016) uitlegt, zijn black box-modellen fundamenteel ondoorzichtig en ondermijnen daarmee de betrouwbaarheid. In een tijd waarin desinformatie en wantrouwen tegenover experts al wijdverbreid zijn, kan een te kritiekloze inzet van AI dit vertrouwen verder ondermijnen.

Een tweede maatschappelijk vraagstuk betreft het onderwijs. Als leerlingen en studenten AI vanaf hun vroege vorming vanzelfsprekend gebruiken, verandert dit hun verhouding tot kennis. De klassieke notie dat leren een proces van oefening, geheugenopbouw en geleidelijke beheersing is, komt onder druk te staan wanneer antwoorden direct beschikbaar zijn. Voor de samenleving als geheel is dit geen randverschijnsel: het raakt aan de vraag welk type burgers, professionals en denkers er in de toekomst gevormd worden. Hier wordt zichtbaar dat de pedagogische dimensie niet slechts een interne zorg van de wetenschap is, maar een maatschappelijke kwestie die de kwaliteit van democratische cultuur en burgerschap raakt.

Daarnaast is er de kwestie van macht en eigenaarschap. AI-systemen worden grotendeels ontwikkeld en beheerd door een beperkt aantal technologiebedrijven en grootmachten. Dit betekent dat de infrastructuur waarop kennisproductie in toenemende mate berust, in private en geopolitieke handen ligt. Voor een samenleving die wetenschap traditioneel als publieke zaak ziet, roept dit vragen op over autonomie en soevereiniteit. Wie toegang heeft tot krachtige AI-modellen bepaalt mede wie kan deelnemen aan de productie van kennis en wie niet.

Ten slotte speelt er een culturele dimensie. Als wetenschappelijke publicaties grotendeels door machines worden geproduceerd en beoordeeld, en als studenten hun scripties schrijven met behulp van AI die docenten op hun beurt weer inzetten om die scripties te controleren, ontstaat het gevaar dat de symbolische waarde van deze praktijken verdwijnt. Scripties en artikelen zijn niet alleen instrumenten om kennis over te dragen of te toetsen, maar ook rituelen die betekenis geven aan de status van wetenschap en onderwijs. Wanneer zulke rituelen bovendien worden uitgehold doordat wetenschappelijke teksten in toenemende mate door machines voor andere machines worden geproduceerd, dreigt ook het communicatieve weefsel van wetenschap te rafelen. Artikelen en scripties functioneren traditioneel niet alleen als bewijs van bekwaamheid of dragers van data, maar ook als uitnodiging tot dialoog. Als niemand ze nog leest omdat de rol van schrijver én lezer door AI wordt overgenomen, verliest wetenschap een van haar meest wezenlijke sociale praktijken: het gezamenlijk construeren van betekenis door middel van taal.

Zoals Müller (2020) in zijn overzicht van ethische en normatieve vragen rond AI-systemen benadrukt, hebben deze ontwikkelingen directe maatschappelijke relevantie die verder reikt dan technische overwegingen. Om al deze redenen is het maatschappelijk noodzakelijk de vragen over AI en kennisproductie scherp te stellen. Het gaat niet om een technische discussie over hulpmiddelen, maar om een bredere reflectie op de toekomst van kennis als publieke waarde, op de vorming van nieuwe generaties en op de verdeling van macht in een digitale wereld.

2.4 Afleiding van de onderzoeksvragen

De voorafgaande analyse maakt duidelijk dat kunstmatige intelligentie niet slechts een instrument is dat bestaande wetenschappelijke praktijken efficiënter maakt, maar een kracht die de logica van kennisproductie zelf herijkt. De vragen die dit proefschrift stelt, kunnen daarom niet beperkt blijven tot technische of disciplinaire details: zij moeten gericht zijn op de fundamentele verschuivingen die optreden wanneer AI het zoeken, ordenen, legitimeren en communiceren van kennis binnendringt.

Een belangrijk vertrekpunt is de constatering dat de academische cultuur al geruime tijd wordt gekenmerkt door fragmentatie. Onderzoekers worden vaak beoordeeld op hun vermogen kleine, afgebakende bijdragen te leveren die vooral scoren in kwantitatieve termen van publicaties en citaties. Zoals Moosa (2018) uitgebreid analyseert, heeft dit 'publish or perish'-systeem geleid tot een proliferatie van wat hier "postzegel-wijsheden" genoemd kan worden: inzichten die misschien correct en netjes verantwoord zijn, maar die weinig bijdragen aan het grotere geheel. Dit systeem heeft ertoe geleid dat juist die onderwerpen die om brede, multidisciplinaire benaderingen vragen, vaak buiten beeld blijven.

AI maakt dit probleem scherper zichtbaar. Enerzijds kan zij de productie van zulke fragmenten enorm versnellen: artikelen kunnen sneller worden geschreven, beoordeeld en gepubliceerd, waardoor de publicatiecultuur nog verder wordt opgejaagd. Anderzijds opent AI de mogelijkheid om barrières tussen disciplines te slechten. Algoritmen die corpora over de grenzen van vakgebieden heen analyseren, maken verbanden zichtbaar die in een gefragmenteerd systeem verborgen zouden blijven. Hier ligt een alternatieve norm voor wetenschappelijke originaliteit: niet de kwantiteit van bijdragen, maar de kwaliteit van synthese, de verbeeldingskracht om bruggen te slaan waar disciplines elkaar tot nu toe niet vonden.

De eigen casuïstiek die in dit proefschrift wordt ingebracht – eerdere onderzoeksprojecten over polyglot leren en over biologie & social governance – illustreert dit spanningsveld. Het zijn thema's die onder het regime van postzegel-wijsheden moeilijk tot wasdom komen, omdat zij meerdere vakgebieden tegelijk raken en daardoor niet eenvoudig in de bestaande publicatiekanalen passen. Juist AI heeft het mogelijk gemaakt om deze barrières te slechten en zulke vragen alsnog vruchtbaar te verkennen.

Tegen deze achtergrond worden de onderzoeksvragen van dit proefschrift geformuleerd. Zij zijn breed genoeg om de veelvormige impact van AI te omvatten, maar richten zich expliciet op de kernfuncties van wetenschap en op de noodzaak multidisciplinaire barrières te doorbreken.

Hoofdvraag Hoe beïnvloedt de inzet van kunstmatige intelligentie de processen, instituties en epistemologie van wetenschappelijke kennisproductie, en welke ruimte opent dit voor het doorbreken van fragmentatie en het bevorderen van multidisciplinaire originaliteit?

Deelvragen

1. **Cognitief en methodologisch** – Hoe verandert AI de manier waarop kennis wordt gezocht, geordend en geverifieerd, in vergelijking met traditionele vormen van literatuuronderzoek en empirische methoden, en hoe verhoudt dit zich tot het spanningsveld tussen fragmentatie en synthese?
2. **Pedagogisch en institutioneel** – Welke gevolgen heeft AI voor de opleidingspraktijken, de rol van scripties en publicaties, en de instituties die wetenschappelijke erkenning en selectie dragen, en in hoeverre versterken of corrigeren deze veranderingen het regime van postzegel-wijsheden?
3. **Macht en normativiteit** – Hoe beïnvloedt AI de verdeling van macht en eigenaarschap in de wetenschap, en welke nieuwe waarborgen zijn nodig om betrouwbaarheid, transparantie en

ethiek te handhaven, juist wanneer originaliteit niet langer in kleine fragmenten maar in brede syntheses gezocht moet worden?

2.5 Plaats van de onderzoeksvragen in dit proefschrift

De formulering van de hoofdvraag en de drie deelvragen vormt niet het eindpunt van de inleiding, maar juist het begin van de verdere uitwerking van dit proefschrift. Zij bepalen de structuur en richting van de analyse. De hoofdvraag kan pas aan het slot worden beantwoord, wanneer de afzonderlijke lijnen samenkomen. De drie deelvragen geven richting aan de tussenstappen. Zij corresponderen met drie clusters van thema's die in de opeenvolgende hoofdstukken worden behandeld: de cognitieve en methodologische verschuivingen in kennisproductie, de pedagogische en institutionele veranderingen die deze verschuivingen begeleiden, en ten slotte de machts- en normatieve vragen die daaruit voortvloeien. Elk hoofdstuk draagt bij aan de beantwoording van een of meer deelvragen, en de volgorde weerspiegelt de beweging van denken (cognitief), via vorming en organisatie (pedagogisch en institutioneel), naar macht en betrouwbaarheid (normatief).

Deze vragen hebben niet alleen betrekking op de manier waarop AI de bestaande praktijken beïnvloedt, maar ook op de mogelijkheid dat AI dwingt tot een heroverweging van wat wetenschappelijke originaliteit betekent. In een academisch systeem dat vaak wordt gedomineerd door fragmentatie en publicatiedruk, maakt AI zowel de beperkingen van dit regime zichtbaar als de kans om erdoorheen te breken. Het gaat er niet om de waarde van specialistische bijdragen te ontkennen, maar om te onderzoeken hoe wetenschap kan voorkomen dat zij verengt tot postzegel-wijsheden, en hoe zij ruimte kan scheppen voor bredere perspectieven en multidisciplinaire verbindingen.

De gekozen aanpak weerspiegelt dit spanningsveld. De vier archetypen van wetenschap en het meta-domein van informatica functioneren als semi-casuïstiek: zij zijn gebaseerd op bestaande literatuur en laten zien hoe AI op uiteenlopende manieren kernfuncties van wetenschap raakt. Deze archetypen maken zichtbaar dat fragmentatie niet het enige perspectief is: door vergelijking en contrast worden verbanden duidelijk die in afzonderlijke vakgebieden vaak verborgen blijven. Daarnaast worden twee concrete casussen ingebracht, de eerdere onderzoeksprojecten over polyglot leren en over biologie en social governance. Zij laten zien hoe AI in de praktijk kan bijdragen aan onderzoek dat zich juist aan de grenzen van disciplines beweegt en daardoor buiten de radar van het traditionele publicatiesysteem blijft.

Daarmee wordt ook de afbakening van dit proefschrift zichtbaar. Het is geen technisch verslag van algoritmische innovaties, geen discipline-gebonden bijdrage die zich beperkt tot één vakgebied, en geen beleidsadviesdocument dat directe oplossingen voorschrijft. Het doel is beschouwend en kader stellend: een analyse die de veelvormigheid van wetenschappelijke kennisproductie in kaart brengt, laat zien hoe AI die veelvormigheid hertekent, en verkent welke ruimte dit opent voor een heroriëntatie op originaliteit die niet wordt gedefinieerd door kwantiteit van publicaties, maar door de kwaliteit van verbinding en synthese.

Literatuur bij Hoofdstuk 2.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?*. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Burrell, J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*. Big Data & Society, 3(1), 1–12. https://doi.org/10.1177/2053951715622512

Fanelli, D. (2010). *Do pressures to publish increase scientists' bias? An empirical support from US States Data*. PLoS ONE, 5(4), e10271. https://doi.org/10.1371/journal.pone.0010271

Gadamer, H.-G. (2004). *Wahrheit und Methode* (6e druk). Tübingen: Mohr Siebeck. (Oorspr. werk gepubliceerd 1960).

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Ioannidis, J. P. A. (2005). *Why most published research findings are false*. PLoS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kuhn, T. S. (2012). *The structure of scientific revolutions* (50th anniversary ed.). Chicago: University of Chicago Press.

Lawrence, P. A. (2003). *The politics of publication*. Nature, 422(6929), 259–261. https://doi.org/10.1038/422259a

Moosa, I. A. (2018). *Publish or perish: Perceived benefits versus unintended consequences*. Cheltenham: Edward Elgar.

Müller, V. C. (2020). *Ethics of artificial intelligence and robotics*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/

Hoofdstuk 3 – Methodologische verantwoording

3.1 Inleiding: de methodologische uitdaging

Het formuleren van scherpe onderzoeksvragen is slechts de eerste stap; minstens zo belangrijk is de vraag hoe deze vragen op een verantwoorde wijze kunnen worden beantwoord. De uitdaging is hier bijzonder, omdat de inzet van kunstmatige intelligentie zowel onderwerp van analyse is als instrument dat het onderzoek mede mogelijk maakt. Dat betekent dat de methodologische keuzes niet louter technisch of procedureel zijn, maar zelf deel uitmaken van de problematiek die onderzocht wordt.

Traditionele methodologische kaders bieden hier slechts gedeeltelijk houvast. Kwantitatieve benaderingen zijn te smal om de volle reikwijdte van de vragen te omvatten, terwijl puur kwalitatieve strategieën tekortschieten om de schaal en snelheid van AI-toepassingen te vangen (Creswell & Creswell, 2018). Ook discipline-gebonden methoden – historisch, sociologisch, natuurwetenschappelijk – sluiten onvoldoende aan bij de ambitie om over de grenzen van vakgebieden heen te kijken. Het gevolg is dat dit onderzoek zich moet positioneren als een beschouwende en systematische verkenning die verschillende bronnen en strategieën combineert.

Deze hybride positie brengt verantwoordelijkheden met zich mee. Enerzijds moet vermeden worden dat het onderzoek vervalt in vrijblijvende speculatie; anderzijds mag het zich niet laten insnoeren door methodologische orthodoxie die onvoldoende recht doet aan de aard van de vragen. De centrale methodologische opgave is daarom: een kader ontwerpen dat valide en betrouwbare inzichten oplevert, maar tegelijk open genoeg is om de complexiteit van AI's impact op kennisproductie recht te doen.

3.2 Positionering van het onderzoek

Om deze positie te verhelderen kan aansluiting worden gezocht bij het model van Saunders et al. (2019), vaak aangeduid als de *research onion*. Dit model laat zien dat onderzoek zich altijd beweegt langs verschillende lagen van keuzes: van filosofische grondslag via strategie en methoden naar concrete technieken voor dataverzameling en analyse. Binnen dat model kiest dit proefschrift niet voor één eenduidige route, maar voor een combinatie die past bij de aard van de onderzoeksvragen.

De filosofische grondslag is in essentie kritisch-realistisch (Bhaskar, 2008). Er wordt verondersteld dat wetenschappelijke kennispraktijken reële structuren en effecten hebben, maar dat onze toegang daartoe altijd bemiddeld is – en in dit geval in toenemende mate door AI. Dit impliceert dat waarheid en betrouwbaarheid niet eenvoudig gegeven zijn, maar voortdurend opnieuw bevraagd en verantwoord moeten worden.

Op strategisch niveau kiest het onderzoek voor een beschouwende en systematisch vergelijkende aanpak. Het raamwerk van archetypen en impactdimensies fungeert hierbij als analytisch instrument: het maakt het mogelijk de impact van AI breed in kaart te brengen zonder te vervallen in onoverzichtelijke detailstudies. De bespreking van deze archetypen is te begrijpen als semi-casuïstiek: zij berust op bestaande literatuur en laat zien hoe AI verschillende kernfuncties van wetenschap raakt (Knorr-Cetina, 1999).

Daarnaast worden twee concrete casussen ingebracht: eerdere onderzoeksprojecten waarin AI intensief werd gebruikt, respectievelijk bij het navigatiegedrag van duiven, bij het voorspellen van IT-successen, bij biologie en social governance en bij polyglotte taalverwerving. Deze casussen zijn niet bedoeld als ontwerpinterventies of als strikt auto-etnografisch materiaal, maar als reflexieve bronnen die laten zien

hoe AI in de praktijk de productie van kennis mede vormgeeft. Zij bieden inzicht in de kansen en beperkingen die AI oplevert wanneer men daadwerkelijk probeert met behulp van deze technologie nieuwe kennis te genereren.

Deze combinatie van conceptuele analyse, literatuuronderzoek, archetypische vergelijking en casuïstisch materiaal maakt het mogelijk de onderzoeksvragen in hun volle breedte te adresseren. Zij legitimeert bovendien een aanpak die zich niet in één traditionele methode laat vangen, maar toch wetenschappelijk verantwoord blijft door transparantie in herkomst van bronnen, triangulatie van inzichten en expliciete reflectie op validiteit en betrouwbaarheid.

3.3 Onderzoeksaanpak

3.3.1 Architectuur van het onderzoek

De methodologische kern van dit proefschrift bestaat uit een onderzoeksarchitectuur die de impact van kunstmatige intelligentie systematisch in kaart brengt. Deze architectuur is opgezet als een matrix die twee assen met elkaar verbindt. De eerste as bestaat uit de drie eerder onderscheiden impactdimensies: cognitief en methodologisch, pedagogisch en institutioneel, en macht en normativiteit. De tweede as bestaat uit de vier archetypen van wetenschap en het meta-domein van informatica: natuurwetenschappen, geschiedenis, antropologie, ontwerpgericht onderzoek en informatica/AI zelf.

Door deze twee assen te kruisen ontstaat een raster waarin elk wetenschapsgebied wordt onderzocht langs dezelfde drie lenzen. Dit voorkomt dat de analyse fragmentarisch of impressionistisch wordt, en maakt vergelijking tussen de archetypen mogelijk. Tegelijkertijd wordt zo recht gedaan aan de eigenheid van ieder archetype: de vragen rond causaliteit en reproduceerbaarheid in de natuurwetenschappen zijn fundamenteel anders dan de vragen rond authenticiteit en anachronisme in de geschiedenis, of de relationele kern van veldwerk in de antropologie. De matrix dwingt ertoe om deze verschillen expliciet te maken, terwijl de parallelle structuur ervoor zorgt dat de resultaten uiteindelijk kunnen worden samengebracht in een synthese.

De rol van het meta-domein van informatica is hier bijzonder. Omdat dit domein niet alleen een archetype naast de andere is, maar tegelijk de technologie levert die alle andere domeinen raakt, fungeert het als spiegel en toetssteen (Floridi, 2014). Waar de natuurwetenschappen of de geschiedenis laten zien hoe AI ingrijpt in bestaande praktijken, maakt de informatica zichtbaar hoe AI zichzelf voortbrengt en versterkt. Daarmee vormt dit domein een kritische randvoorwaarde voor de hele analyse.

De architectuur bepaalt ook de hoofdstukstructuur van het proefschrift. Eerst worden de drie impactdimensies afzonderlijk besproken, waarbij hun betekenis en onderlinge relaties worden uitgewerkt. Daarna volgen de hoofdstukken waarin de archetypen centraal staan. Elk van deze archetypen-hoofdstukken wordt systematisch langs de drie dimensies geanalyseerd. Zo ontstaat een dubbel perspectief: eerst verdiepend per dimensie, vervolgens verbredend per wetenschapsgebied. Het resultaat is een gelaagde en consistente aanpak die zowel theoretische scherpte als praktische toepasbaarheid waarborgt.

3.3.2 Data-collectie en data-analyse per deelvraag

De drie deelvragen die in hoofdstuk 2 zijn geformuleerd, geven richting aan de verdere uitwerking van dit proefschrift. Elk van deze vragen vereist een specifieke combinatie van bronnen en analysetechnieken, maar alle drie worden ingebed in de onderzoeksarchitectuur die de impactdimensies kruist met de wetenschappelijke archetypen. Op die manier blijft de behandeling van de deelvragen consistent en onderling vergelijkbaar.

De eerste deelvraag, gericht op de cognitieve en methodologische dimensie, vraagt om inzicht in de manier waarop AI processen van kennis zoeken, ordenen en verifiëren verandert. De dataverzameling hiervoor bestaat voornamelijk uit systematische literatuurstudie: academische artikelen en meta-studies over de inzet van AI bij literatuuronderzoek, dataverwerking en patroonherkenning (Cooper, 2016). Deze literatuur wordt aangevuld met secundaire voorbeelden van AI-toepassingen in verschillende vakgebieden, zoals moleculaire biologie of klimatologie. Daarnaast worden observaties uit de eigen eerdere projecten gebruikt als reflexief materiaal om de spanning tussen fragmentatie en synthese concreet te maken.

De tweede deelvraag, die de pedagogische en institutionele dimensie bestrijkt, vraagt om een bredere verzameling bronnen. Naast wetenschappelijke literatuur wordt hier gebruikgemaakt van beleidsdocumenten en richtlijnen van universiteiten en onderzoeksorganisaties over AI in onderwijs en publicatie (Selwyn, 2019). Ook ervaringen met scripties en artikelen die (deels) tot stand kwamen met AI worden betrokken, om inzicht te krijgen in de veranderende rol van academische rituelen.

De derde deelvraag, die draait om macht en normativiteit, vergt een meer kritische selectie van bronnen. Hier gaat het om literatuur over AI-governance, eigenaarschap van modellen en data, en ethische reflecties op betrouwbaarheid en transparantie (Crawford, 2021). Empirische voorbeelden van machtsconcentratie, zoals de rol van grote technologiebedrijven bij de ontwikkeling van generatieve modellen, worden aangevuld met beleidsdocumenten van overheden en internationale organisaties. Ook worden eigen ervaringen met afhankelijkheid van commerciële tools meegenomen om zichtbaar te maken hoe deze machtsstructuren in de praktijk doorwerken.

Door deze aanpak ontstaat per deelvraag een combinatie van literatuur, secundaire voorbeelden en reflexieve casussen. De systematiek van de matrix waarborgt dat deze bronnen niet willekeurig worden gebruikt, maar telkens worden geordend naar dimensie en archetype. Daarmee krijgt het onderzoek zowel breedte als diepte: breedte doordat de hele wetenschappelijke diversiteit wordt bestreken, diepte doordat elke deelvraag in samenhang met de andere wordt uitgewerkt.

3.3.3 Evaluatiepunten per wetenschapsgebied

De matrix die de drie impactdimensies kruist met de vier archetypen en het meta-domein biedt een stevige structuur, maar een puur schema zou onvoldoende recht doen aan de complexiteit van AI's impact. Daarom zijn er enkele evaluatiepunten geformuleerd die in elk wetenschapsgebied systematisch terugkeren. Zij functioneren als kritische lenzen die de analyse scherpen en zorgen dat essentiële kwesties niet onderbelicht blijven.

Een eerste evaluatiepunt betreft de versnelling die AI teweegbrengt. In vrijwel elk domein zien we dat dataverwerking, literatuuronderzoek en zelfs publicatiecycli aanzienlijk sneller verlopen wanneer AI wordt ingezet (Kitchin, 2014). Deze versnelling biedt kansen voor nieuwe ontdekkingen, maar roept ook vragen op over de waarde van kennis die in steeds kortere tijd wordt geproduceerd.

Een tweede evaluatiepunt betreft de multiplicatie van stemmen in evaluatieprocessen. Waar peer review traditioneel een gesprek was tussen enkele vakgenoten, maakt AI het mogelijk meerdere modellen tegelijk in te zetten, elk in een andere rol. Dit roept de vraag op of zo'n machinaal gegenereerde dialoog de kwaliteit van beoordeling verhoogt of juist het menselijke karakter van wetenschappelijke dialoog ondermijnt.

Een derde evaluatiepunt is de communicatieve functie van wetenschappelijke teksten. Scripties, artikelen en monografieën zijn meer dan instrumenten om data of analyses over te dragen; zij zijn rituelen die betekenis geven aan wetenschappelijke gemeenschap en dialoog (Becher & Trowler, 2001). Als AI niet alleen teksten kan genereren maar ook beoordelen, ontstaat het gevaar dat deze teksten circuleren tussen machines zonder dat mensen ze nog lezen.

Een vierde evaluatiepunt betreft de spanning tussen postzegel-wijsheden en synthese. De publicatiecultuur beloont nog te vaak kleine, smalle bijdragen, terwijl AI juist in staat is om patronen over disciplines heen zichtbaar te maken. De vraag is of dit leidt tot een verdieping van synthese of slechts tot een versnelling van fragmentatie (Huutoniemi et al., 2010).

Ten slotte is er de AI-paradox: de notie dat men AI pas kritisch en vruchtbaar kan inzetten wanneer men is gevormd in een context zonder AI. Voor de generaties die AI vanaf het begin als vanzelfsprekend ervaren, rijst de vraag hoe zij de nodige kritische afstand kunnen ontwikkelen.

Door deze evaluatiepunten in elk wetenschapsgebied systematisch terug te laten keren, wordt de analyse niet alleen breder maar ook dieper. Zij waarborgen dat de bespreking van de impactdimensies niet blijft steken in algemene observaties, maar telkens wordt getoetst aan kwesties die de kern van de wetenschappelijke praktijk raken.

3.4 Validiteit en betrouwbaarheid

Een methodologische verantwoording kan niet volstaan met het schetsen van een architectuur en het uitwerken van analyseplannen. Even belangrijk is de vraag hoe de kwaliteit van dit onderzoek kan worden gegarandeerd. Validiteit en betrouwbaarheid zijn hierbij geen vanzelfsprekende grootheden, omdat de inzet van kunstmatige intelligentie als onderzoeksinstrument en als onderzoeksobject nieuwe onzekerheden introduceert.

De validiteit van dit onderzoek berust op de mate waarin het erin slaagt de kernfuncties van kennisproductie onder AI zichtbaar te maken. Dit vraagt niet alleen om een correcte weergave van de stand van de literatuur, maar ook om een doordachte toepassing van de impactdimensies en evaluatiepunten op de verschillende archetypen. Interne validiteit wordt versterkt doordat dezelfde lenzen consequent op alle wetenschapsgebieden worden toegepast, waardoor vergelijking en synthese mogelijk worden. Externe validiteit wordt gewaarborgd door de diversiteit van de archetypen (Maxwell, 2012).

De betrouwbaarheid van dit onderzoek vergt bijzondere aandacht omdat AI-systemen zelf vatbaar zijn voor hallucinaties, bias en ondoorzichtige besluitvorming (Bender et al., 2021). Het gebruik van AI als instrument voor literatuursearch of tekstproductie wordt daarom steeds transparant verantwoord: bronnen worden expliciet genoemd, gegenereerde passages worden kritisch getoetst aan bestaande literatuur, en verzonnen referenties worden systematisch uitgesloten. Waar AI een rol heeft gespeeld in eerdere casusprojecten, wordt dat niet verhuld maar expliciet beschreven, inclusief de beperkingen die dit opleverde.

Betrouwbaarheid wordt verder versterkt door triangulatie: inzichten uit literatuuronderzoek, archetypen en casusmateriaal worden naast elkaar gelegd en tegen elkaar afgewogen. Een bijkomende maatregel is het hanteren van de eerder geformuleerde evaluatiepunten als kwaliteitscriteria.

Tot slot is er de kwestie van reflexiviteit. Omdat de onderzoeker zelf ervaring heeft met AI-gestuurde kennisproductie, is er een voortdurende kans op blinde vlekken of bevestigingsbias. Dit wordt ondervangen door deze ervaringen niet als neutrale data te presenteren, maar als reflexief materiaal dat expliciet wordt ingebed in een bredere context van literatuur en archetypische analyse. Het doel is niet om een subjectieve ervaring te veralgemeniseren, maar om die ervaring als prisma te gebruiken waardoor bredere patronen zichtbaar worden.

3.5 Begrippen en definities

Een analyse die zich uitstrekt over de volle breedte van wetenschappelijke kennisproductie kan niet zonder een zorgvuldig begrippenapparaat.

Kunstmatige intelligentie (AI). AI wordt in dit proefschrift opgevat als een brede familie van systemen die uiteenlopende taken uitvoeren: patroonherkenning, simulatie, spraak- en beeldverwerking, en tekstgeneratie. Tot deze familie behoren machine learning, deep learning, natural language processing, computer vision en generatieve modellen (Russell & Norvig, 2021). De diversiteit is cruciaal omdat de impact per wetenschapsgebied verschilt: simulatie in natuurwetenschappen, digitale archieven in de geesteswetenschappen, transcriptietools in de antropologie, optimalisatie in ontwerpgerichte disciplines en zelfontwikkeling in de informatica. In dit hoofdstuk staat vooral de technische en functionele veelvormigheid centraal; de epistemische gevolgen worden in latere hoofdstukken uitgewerkt.

Kennisproductie. Met kennisproductie wordt in dit proefschrift het geheel bedoeld van processen waarmee nieuwe inzichten worden gegenereerd, gevalideerd en verspreid. Het gaat daarbij niet alleen om dataverzameling en analyse, maar ook om de institutionele en communicatieve praktijken die bepalen welke kennis als legitiem wordt erkend (Knorr-Cetina, 1999).

Originaliteit. Originaliteit verwijst niet primair naar het afbakenen van een niche of het leveren van incrementen, maar naar het vermogen disciplines en perspectieven met elkaar te verbinden. Het doorbreken van barrières en het zichtbaar maken van verbanden die in een gefragmenteerd systeem verborgen blijven, is de kern van originaliteit (Huutoniemi et al., 2010).

Postzegel-wijsheden. Dit begrip verwijst naar het verschijnsel dat onderzoekers carrière maken door kleine, sterk afgebakende bijdragen te publiceren, die vooral waarde hebben binnen de logica van publicatiedruk en citatie-indicatoren (Becher & Trowler, 2001). Deze bijdragen zijn vaak correct en zorgvuldig, maar leveren slechts beperkte bijdrage aan bredere samenhang of maatschappelijk betekenisvolle kennis.

Synthese. Synthese wordt hier gedefinieerd als het vermogen om inzichten uit verschillende disciplines of onderzoekstradities samen te brengen tot een geheel dat meerwaarde heeft boven de som der delen (Repko & Szostak, 2021). Waar multidisciplinariteit vaak juxtapositie blijft, verwijst synthese naar integratie en betekenisvolle verbinding.

Multidisciplinariteit. Met multidisciplinariteit wordt het betrekken van meerdere disciplines bij een vraagstuk bedoeld, maar niet in de minimale zin van juxtapositie. Multidisciplinariteit wordt hier gezien als een opstap naar integratie en synthese (Frodeman, 2017).

3.6 Synthese en vooruitblik

Met de voorgaande paragrafen is het methodologische fundament van dit proefschrift gelegd. Eerst is de bijzondere positie van dit onderzoek verduidelijkt: beschouwend van aard, ingebed in een kritische analyse van de rol van AI in wetenschappelijke kennisproductie, en gevoed door literatuur en casuïstiek. Vervolgens is de onderzoeksarchitectuur uitgewerkt als een matrix die dimensies en archetypen kruist. Daarna is concreet gemaakt hoe de deelvragen worden beantwoord, met evaluatiepunten als kritische lenzen. Validiteit en betrouwbaarheid zijn gewaarborgd door transparantie, triangulatie en reflexiviteit. Tot slot zijn de kernbegrippen gedefinieerd die de analyse dragen.

Het vervolg van dit proefschrift werkt eerst de drie impactdimensies afzonderlijk uit (H4–H6), daarna de archetypen (H7–H11), gevolgd door een synthese (H12) en vijf casussen (H13–H17). Hoofdstuk 18 sluit af met beantwoording van de hoofdvraag

Literatuurlijst Hoofdstuk 3

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2e ed.). Open University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bhaskar, R. (2008). *A realist theory of science* (rev. ed.). Routledge.

Cooper, H. (2016). *Research synthesis and meta-analysis: A step-by-step approach* (5e ed.). Sage.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5e ed.). Sage.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Frodeman, R. (2017). *Sustainable knowledge: A theory of interdisciplinarity*. Palgrave Macmillan.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, 39(1), 79–88. https://doi.org/10.1016/j.respol.2009.09.011

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Maxwell, J. A. (2012). *A realist approach for qualitative research*. Sage.

Pickstone, J. V. (2000). *Ways of knowing: A new history of science, technology, and medicine*. University of Chicago Press.

Repko, A. F., & Szostak, R. (2021). *Interdisciplinary research: Process and theory* (4e ed.). Sage.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4e ed.). Pearson.

Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8e ed.). Pearson.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press

Hoofdstuk 4 – Cognitieve en methodologische impact

4.1 Inleiding

De eerste deelvraag van dit proefschrift luidt: *Hoe beïnvloedt kunstmatige intelligentie de cognitieve en methodologische kernprocessen van kennisproductie?* Deze vraag richt zich op de fundamenten van wetenschappelijk werk: de manieren waarop data worden verzameld, geanalyseerd en geïnterpreteerd, en hoe dit doorwerkt in hypothesevorming, toetsing en validatie. In dit hoofdstuk staat de cognitieve en methodologische dimensie centraal. Er wordt onderzocht hoe AI kennispraktijken herdefinieert, welke kansen dit biedt voor versnelling en vernieuwing, maar ook welke risico's ontstaan door bias, oversimplificatie en verlies van nuance.

4.2 AI en het herdefiniëren van kennispraktijken

Kunstmatige intelligentie heeft zich in korte tijd ontwikkeld van hulpmiddel tot structureel element in het wetenschappelijke proces. In de natuurwetenschappen is dit het meest zichtbaar. Erduran (2024) laat zien hoe AI-modellen als AlphaFold de biologie hebben veranderd: niet langer vormt de onderzoeker eerst een hypothese om die vervolgens te toetsen, maar patronen die de machine aanreikt worden het uitgangspunt van verdere interpretatie. Dit betekent een verschuiving in de cognitieve logica van onderzoek, waarbij algoritmische voorspellingen deel gaan uitmaken van theorievorming.

In andere disciplines doet zich een vergelijkbare herordening voor. Fabiano (2024) bespreekt de inzet van AI bij systematische literatuurreviews. Het gaat daarbij niet slechts om het versnellen van een tijdrovend proces. Wat verandert, is de aard van de vraag die wordt gesteld. Waar een review traditioneel bedoeld is om overzicht te bieden over wat er bekend is, kan AI op grote schaal patronen en lacunes identificeren die voorheen onzichtbaar bleven. Daarmee verandert de methodologische functie van de review: van een samenvatting van de stand van zaken naar een algoritmisch gegenereerd meta-overzicht.

Ook in bibliometrische toepassingen wordt deze verschuiving zichtbaar. Saiednia et al. (2024) tonen hoe AI kan worden ingezet om publicatietrends en citation networks in kaart te brengen. De methodologische vraag verschuift daarmee van de betrouwbaarheid van afzonderlijke studies naar de betrouwbaarheid van de algoritmen die deze studies ordenen. Bernard (2025) maakt dit concreet in zijn bespreking van de AI-tool *Elicit*, die systematische reviews ondersteunt. Hij benadrukt dat de selectiecriteria die het algoritme hanteert vaak minder transparant zijn dan traditionele zoekstrategieën, waardoor de betrouwbaarheid van de resultaten ter discussie staat.

4.3 Kansen en mogelijkheden

De cognitieve en methodologische voordelen van AI lijken evident. Hao et al. (2024) tonen empirisch aan dat onderzoekers die AI benutten vaak een grotere impact genereren. Dat komt doordat zij sneller verbanden leggen tussen datasets en disciplines. In methodologische zin gaat het hier om een radicale verbreding: AI maakt het mogelijk om patronen te ontdekken die in afzonderlijke disciplines verborgen blijven.

Een tweede kans betreft de replicatieproblematiek. Ioannidis (2005) stelde scherp dat een groot deel van de gepubliceerde bevindingen waarschijnlijk onjuist is, mede door selectieve rapportage en

methodologische tekortkomingen. AI kan dit probleem deels helpen oplossen door grootschalige her-analyses mogelijk te maken. Open Science Collaboration (2015) liet zien hoe laag de reproduceerbaarheid in de psychologie uitviel. Waar zulke studies traditioneel jaren in beslag nemen, kan AI re-analyses versnellen en opschalen. Dit opent de mogelijkheid tot een collectieve kwaliteitscontrole die voorheen onhaalbaar was.

Daarnaast biedt AI cognitieve ruimte. Fabiano (2024) benadrukt dat het automatiseren van repetitieve onderdelen van systematische reviews onderzoekers in staat stelt zich te richten op interpretatie en theorievorming. Dat betekent dat AI niet alleen bijdraagt aan de snelheid van onderzoek, maar ook aan de herwaardering van de menselijke rol: het kritisch beoordelen van relevantie en het synthetiseren van inzichten.

4.4 Uitdagingen en risico's

Tegenover deze kansen staan aanzienlijke risico's. Messeri en Crockett (2024) waarschuwen dat AI de productie van kennis weliswaar versnelt, maar dat de diepgang van wetenschappelijke reflectie eronder kan lijden. Onderzoekers "doen meer, maar leren minder": de intensieve omgang met bronnen en data maakt plaats voor algoritmische tussenstappen, waardoor de cognitieve binding met het materiaal verzwakt.

Een tweede risico betreft oversimplificatie. Recente evaluaties van LLM-systemen laten zien dat zij wetenschappelijke teksten vaak reduceren tot te simpele samenvattingen, waarbij cruciale details verdwijnen (LiveScience, 2025). Wat in eerste instantie efficiënt lijkt, kan leiden tot een verlies van methodologische precisie. Het gevaar is dat onderzoekers conclusies trekken uit algoritmische samenvattingen zonder de onderliggende complexiteit te doorgronden.

Een derde risico betreft de fabricage van literatuurreferenties. Generatieve AI-modellen kunnen literatuurverwijzingen produceren die variëren van volledig correct tot geheel fictief, een verschijnsel dat fundamentele problemen oplevert voor wetenschappelijke betrouwbaarheid. Dit probleem manifesteert zich in een spectrum van vormen: aan de ene kant kunnen modellen volkomen juiste referenties genereren wanneer bepaalde standaardwerken zeer frequent voorkomen in hun trainingsdata. Aan de andere kant produceren zij volledig verzonnen artikeltitels met niet-bestaande auteurs. Tussen deze uitersten ligt een grijs gebied van gedeeltelijk correcte referenties: juiste auteursnamen gekoppeld aan fictieve publicaties, bestaande tijdschriften met verzonnen artikelen, correcte bibliografische gegevens die echter naar verkeerde pagina's of secties verwijzen, of accurate titels met onjuiste publicatiejaren.

De oorzaak ligt in de aard van taalmodellen zelf. Deze systemen leren patronen herkennen in miljoenen tekstcorpora en ontwikkelen daarbij het vermogen te voorspellen hoe een literatuurreferentie eruit hoort te zien: welke auteursnamenstijlen gebruikelijk zijn, hoe tijdschrifttitels worden geformatteerd, welke jaargangen plausibel klinken. Voor het model zijn referenties echter geen verwijzingen naar werkelijke bronnen, maar louter tekstpatronen die statistisch waarschijnlijk zijn. Het systeem heeft geen toegang tot werkelijke bibliografische databases en kan dus niet verifiëren of een geconstrueerde referentie daadwerkelijk bestaat of naar de juiste locatie verwijst. Over de precieze verhoudingen tussen correcte, gedeeltelijk correcte en volledig gefabriceerde referenties is nog relatief weinig systematisch onderzoek beschikbaar. De bestaande studies suggereren dat het probleem substantieel is, maar de omvang varieert sterk afhankelijk van het specifieke model, de prompt en het vakgebied. Dit gebrek aan systematische kennis maakt het risico des te problematischer: onderzoekers kunnen niet inschatten hoe waarschijnlijk fouten zijn en moeten daarom elke AI-gegenereerde referentie individueel verifiëren.

Dit probleem is bijzonder verraderlijk omdat gedeeltelijk correcte referenties vaak overtuigend accuraat lijken en moeilijk te onderscheiden zijn van echte bronnen zonder systematische controle. Voor

onderzoekers die AI inzetten bij literatuuronderzoek ontstaat daarmee een structurele bedreiging van de bronverantwoording, een kernpraktijk van wetenschappelijke legitimiteit.

Daarnaast is er de structurele kwestie van bias. Kahneman en Tversky (1979) lieten zien hoe heuristieken systematische vertekeningen veroorzaken in menselijke besluitvorming. Nickerson (1998) maakte duidelijk dat confirmation bias alomtegenwoordig is. Fanelli (2010) toonde aan dat positieve resultaten vaker worden gepubliceerd dan negatieve. AI dreigt al deze biases te reproduceren of zelfs te versterken, omdat modellen worden getraind op bestaande data waarin deze vertekeningen ingebakken zitten. In methodologische zin betekent dit dat AI de illusie van objectiviteit kan wekken, terwijl het in werkelijkheid bestaande vooroordelen reproduceert.

4.5 Evaluatiepunten

De cognitieve en methodologische impact van AI kan het scherpst worden beoordeeld door enkele terugkerende evaluatiepunten systematisch te betrekken. Deze punten fungeren als lenzen die telkens blootleggen waar kansen omslaan in risico's.

Een eerste lens betreft de **versnelling van onderzoek**. Dataverwerking en literatuur-searches die voorheen weken of maanden duurden, kunnen door AI in enkele dagen of zelfs uren worden uitgevoerd. In de natuurwetenschappen levert dit reële voordelen op voor experimenten met enorme datasets, zoals in de genomica. Tegelijkertijd bestaat het gevaar dat snelheid wordt verward met kwaliteit. Wanneer de tijdsdruk om te publiceren toeneemt, kan een versneld proces leiden tot oppervlakkige interpretatie. De vraag is dan of AI bijdraagt aan verdieping of juist aan een verschraling van wetenschappelijke zorgvuldigheid.

Een tweede lens richt zich op de **spanning tussen fragmentatie en synthese**. De publicatiecultuur beloont nog steeds kleine, incrementele bijdragen, de zogenaamde postzegel-wijsheden. AI kan deze tendens versterken door het eenvoudiger te maken steeds nieuwe, afgebakende papers te genereren. Maar hetzelfde instrument kan ook patronen en verbindingen zichtbaar maken die juist uitnodigen tot synthese. Een bibliometrische analyse met AI kan bijvoorbeeld wijzen op een onverwachte samenhang tussen klimaatonderzoek en epidemiologie. De vraag is of wetenschappers die verbanden benutten voor integratie, of dat de institutionele druk hen dwingt patronen op te splitsen in nieuwe niches.

Een derde lens betreft de **multiplicatie van reviewers**. AI maakt het mogelijk dat verschillende modellen tegelijk fungeren als virtuele peer reviewers, elk met een andere rol: kritische lezer, editor, methodologisch adviseur. Dat lijkt een verrijking van de kwaliteitscontrole, maar roept de vraag op of deze stemmen werkelijk onafhankelijk zijn. Als alle modellen immers getraind zijn op grotendeels hetzelfde corpus, is er eerder sprake van echo's dan van pluraliteit. Voor de cognitieve en methodologische dimensie betekent dit dat een vermeende uitbreiding van checks and balances ook kan resulteren in een versterking van bestaande biases.

Samen genomen maken deze evaluatiepunten duidelijk dat de impact van AI zich niet laat vangen in simpele tegenstellingen van kans of bedreiging. Het gaat steeds om spanningsvelden waarin de richting die wetenschap kiest bepalend wordt voor de waarde van AI.

4.6 Conclusie

De eerste deelvraag luidde: *Hoe beïnvloedt kunstmatige intelligentie de cognitieve en methodologische kernprocessen van kennisproductie?* De analyse in dit hoofdstuk laat zien dat AI deze processen fundamenteel herconfigureert. Aan de ene kant vergroot de technologie de mogelijkheden: versnelling maakt grootschalige replicatie haalbaar, synthese komt dichterbij doordat patronen over disciplines heen zichtbaar worden, en nieuwe vormen van kwaliteitscontrole lijken mogelijk via virtuele reviewers.

Aan de andere kant brengen dezelfde ontwikkelingen risico's met zich mee. De versnelling kan leiden tot oppervlakkigheid; synthese kan door de bestaande publicatiecultuur omslaan in verdere fragmentatie; en multiplicatie van reviewers kan een illusie blijken wanneer alle stemmen dezelfde trainingsbias delen. Wat op cognitief en methodologisch niveau dus zichtbaar wordt, is geen lineaire vooruitgang maar een set spanningsvelden die de toekomst van wetenschap zullen bepalen.

Het voorlopige antwoord op de eerste deelvraag luidt dan ook genuanceerd: kunstmatige intelligentie vergroot de mogelijkheden voor kennisproductie aanzienlijk, maar de waarde daarvan hangt volledig af van de mate waarin wetenschappers deze evaluatiepunten onderkennen en er methodologische zorgvuldigheid tegenover stellen. Alleen wanneer versnelling samengaat met verdieping, fragmentatie wordt omgezet in synthese en meerstemmigheid werkelijk kritisch van aard is, kan AI vruchtbaar bijdragen aan cognitieve en methodologische vooruitgang.

Literatuurlijst Hoofdstuk 4

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), e35179. https://doi.org/10.7759/cureus.35179

Bernard, C. (2025). Using AI for systematic review: the example of Elicit. *BMC Medical Research Methodology, 25*(1), 1–10. https://doi.org/10.1186/s12874-025-02528-y

Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv preprint*, arXiv:2302.03494.

Erduran, S. (2024). The impact of artificial intelligence on scientific practices. *International Journal of Science Education, 46*(3), 325–343. https://doi.org/10.1080/09500693.2024.2306604

Fabiano, C. (2024). How to optimize the systematic review process using AI tools. *JCPP Advances, 4*(1), e12234. https://doi.org/10.1002/jcv2.12234

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE, 5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine*, 6(1), 75. https://doi.org/10.1038/s41746-023-00819-6

Hao, D., Li, Y., & Wang, X. (2024). AI expands scientists' impact but contracts science's focus. *arXiv preprint*. https://arxiv.org/abs/2412.07727

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185

LiveScience. (2025, January 15). AI chatbots oversimplify scientific studies and gloss over critical details — the newest models are especially guilty. https://www.livescience.com/technology/artificial-intelligence/ai-chatbots-oversimplify-scientific-studies-and-gloss-over-critical-details-the-newest-models-are-especially-guilty

Messeri, L., & Crockett, Z. (2024, March 7). Doing more, learning less: The risks of AI in research. *Yale News*. https://news.yale.edu/2024/03/07/doing-more-learning-less-risks-ai-research

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Hoofdstuk 5 – Pedagogische en institutionele impact

5.1 Inleiding

De tweede deelvraag van dit proefschrift luidt: *Hoe beïnvloedt kunstmatige intelligentie de pedagogische en institutionele praktijken van kennisproductie?* Deze vraag richt zich op twee nauw verweven domeinen. Enerzijds gaat het om de pedagogische dimensie: de vorming van studenten en onderzoekers in een academische cultuur die steeds meer door AI wordt doordrongen. Anderzijds betreft het de institutionele dimensie: de structuren van erkenning, publicatie en peer review die de wetenschappelijke praktijk vormgeven. In dit hoofdstuk wordt onderzocht hoe AI in beide domeinen kansen biedt voor innovatie en inclusiviteit, maar ook risico's creëert van oppervlakkigheid, afhankelijkheid en verlies van legitimiteit.

5.2 AI in onderwijs en training

Het idee dat AI het onderwijs fundamenteel kan veranderen, heeft de afgelopen jaren veel aandacht gekregen. Luckin et al. (2016) benadrukken dat AI de potentie heeft om leren te personaliseren: systemen kunnen zich aanpassen aan het niveau, tempo en interesse van individuele studenten. Daarmee zou AI traditionele didactische modellen kunnen doorbreken en meer ruimte scheppen voor differentiatie. Holmes et al. (2021) wijzen in hun UNESCO-rapport op een vergelijkbare belofte, maar voegen daaraan toe dat deze technologie tegelijk nieuwe vormen van ongelijkheid kan creëren. Niet elke student beschikt immers over dezelfde toegang tot hoogwaardige digitale infrastructuur.

Kritische stemmen benadrukken dat AI in onderwijs vaak past in een bredere neoliberale logica. Knox (2020) laat zien dat AI-systemen niet neutraal zijn: ze leggen accenten op meetbaarheid en efficiëntie, en reduceren daarmee onderwijs tot een reeks outputs die goed te registreren zijn maar niet noodzakelijk de volle breedte van leren dekken. Selwyn (2019) sluit daarbij aan met zijn stelling dat AI-onderwijs niet alleen een pedagogisch vraagstuk is, maar ook een politiek en ethisch vraagstuk. De vraag is niet alleen wat studenten leren, maar ook wie bepaalt welke kennis en vaardigheden als relevant worden beschouwd.

De pedagogische dimensie raakt ook aan de AI-paradox. Jongere generaties groeien op in een context waarin AI vanzelfsprekend aanwezig is, maar missen daardoor wellicht de kritische distantie om de beperkingen en biases van deze technologie te doorzien. Voor studenten die nooit een scriptie zonder AI hebben geschreven, kan het moeilijk zijn te onderscheiden wat eigen werk is en wat door de machine is geproduceerd. Dit roept vragen op over de vorming van academische oordeelsvorming en de overdracht van onderzoeksvaardigheden.

5.3 AI en academische instituties

Niet alleen het onderwijs, maar ook de instituties van wetenschap ondergaan een verschuiving onder invloed van AI. De publicatiecultuur is daarvan een eerste voorbeeld. Bastow, Dunleavy en Tinkler (2014) laten zien dat de sociale wetenschappen steeds sterker zijn gaan werken met impact-metrics die kwantitatieve zichtbaarheid belangrijker maken dan kwalitatieve inhoud. In een dergelijke context kan AI de productie van artikelen en rapporten aanzienlijk versnellen, maar ook bijdragen aan de proliferatie van wat eerder als postzegel-wijsheden werd getypeerd.

Het peer review-systeem, traditioneel het hart van wetenschappelijke kwaliteitsborging, staat eveneens onder druk. Biagioli (2002) herinnert ons eraan dat peer review historisch niet louter een kwaliteitscontrole is, maar ook een machtsinstrument dat bepaalt welke stemmen toegang krijgen tot de wetenschappelijke gemeenschap. AI dreigt deze machtsdynamiek te versterken door beoordelingsprocessen te automatiseren. Müller-Birn et al. (2020) beschrijven hoe AI-tools al worden ingezet om manuscripten te screenen en reviewers te ondersteunen. Dat kan de efficiëntie vergroten, maar roept fundamentele vragen op over legitimiteit: wie beoordeelt de beoordelaar, wanneer die zelf een algoritme is?

Een tweede institutionele verschuiving betreft de criteria voor academische carrières. Moher et al. (2020) pleitten in hun Hong Kong Principles voor een heroriëntatie op integriteit, transparantie en open science. AI zou hier als katalysator kunnen werken: het kan hergebruik en open data bevorderen en meer ruimte scheppen voor samenwerking. Maar er schuilt ook een risico in de afhankelijkheid van commerciële tools. Wanneer toegang tot geavanceerde AI-systemen ongelijk verdeeld is, kan dit bestaande machtsverhoudingen versterken en nieuwe vormen van uitsluiting scheppen.

5.4 Kansen en mogelijkheden

Wanneer we de pedagogische en institutionele rol van AI beoordelen vanuit de geest van de Hong Kong Principles, valt op dat AI in potentie kan bijdragen aan de integriteit en openheid van wetenschap. Pedagogisch kan AI studenten ondersteunen bij het transparanter maken van hun leerproces. Denk aan automatische logboeken van schrijf- en zoekactiviteiten, waardoor zichtbaar wordt welke stappen in een literatuuronderzoek zijn gezet en welke keuzes daarbij zijn gemaakt. Dit zou docenten in staat stellen om niet alleen de uitkomst, maar ook het proces van studenten te beoordelen, en zo integriteit al in de opleiding te verankeren.

Op institutioneel niveau kan AI de ambities rond open science versterken. Automatische annotatie en indexering van datasets maken het eenvoudiger onderzoeksresultaten herbruikbaar te maken. Vertaal- en samenvattingstools kunnen kennis toegankelijker maken voor een breder publiek, waardoor het ideaal van maatschappelijke impact dichterbij komt (Holmes et al., 2021). Ook in peer review kan AI bijdragen aan transparantie, bijvoorbeeld door het signaleren van tekstoverlap of onregelmatigheden in data-analyse. Zo kan AI de menselijke reviewer ontlasten van routinematig speurwerk en ruimte scheppen voor diepere inhoudelijke beoordeling.

Daarnaast is er een inclusiviteitsperspectief. Luckin et al. (2016) benadrukken dat AI-onderwijs niet alleen gaat over efficiëntie, maar ook over toegankelijkheid. Voor studenten met uiteenlopende achtergronden of beperkingen kan AI drempels verlagen door teksten automatisch aan te passen, uit te spreken of te vertalen. Dit sluit aan bij het streven naar rechtvaardige en inclusieve leeromgevingen.

5.5 Uitdagingen en risico's

Tegenover deze kansen staan echter substantiële risico's die evenzeer vanuit de Hong Kong Principles te begrijpen zijn. Transparantie en integriteit zijn immers niet vanzelfsprekend wanneer AI wordt ingezet. Integendeel: AI kan juist een illusie van transparantie creëren. Een automatisch gegenereerd rapport over hergebruikte data kan de indruk wekken dat open science wordt nageleefd, terwijl de onderliggende dataset incompleet of slecht gedocumenteerd is. Hier dreigt een paradox: hoe meer AI processen automatiseert, hoe minder duidelijk wordt of onderzoekers zelf werkelijk begrijpen en verantwoorden wat er gebeurt.

Bovendien kan de inzet van AI bestaande ongelijkheden in de academie versterken. Toegang tot krachtige AI-modellen is vaak afhankelijk van commerciële aanbieders. Universiteiten en onderzoekers met beperkte middelen kunnen daardoor uitgesloten raken van de nieuwste tools, waardoor een kloof ontstaat tussen instellingen met en zonder technologische infrastructuur. In plaats van open science kan zo een nieuw regime van gesloten science ontstaan, waarin machtsconcentratie bij technologiebedrijven de spelregels bepaalt (Crawford, 2021).

Ook pedagogisch zijn er risico's. Wanneer studenten te sterk leunen op AI-gegenereerde teksten, kan het leerproces verschralen tot het consumeren van outputs. Selwyn (2019) en Knox (2020) waarschuwen dat AI-onderwijs te vaak wordt ingericht rond meetbaarheid en standaardisatie, waardoor juist de kritische en creatieve dimensies van academische vorming verloren gaan. De AI-paradox komt hier opnieuw naar voren: studenten die nooit zonder AI hebben gewerkt, missen mogelijk het referentiekader om de beperkingen van de technologie te doorzien.

5.6 Evaluatiepunten

De pedagogische en institutionele impact van kunstmatige intelligentie kan niet worden begrepen zonder drie onderliggende spanningsvelden te adresseren. Deze evaluatiepunten zijn geen losse observaties, maar kritische lenzen die zichtbaar maken hoe AI de fundamenten van academische cultuur raakt.

De communicatieve functie van wetenschappelijke teksten. Wetenschappelijke teksten zijn nooit louter informatiedragers geweest. Zij functioneren als rituelen die de academische gemeenschap bijeenhouden: het schrijven en lezen van artikelen, scripties en monografieën is ook een proces van socialisatie, waarin studenten en onderzoekers leren wat telt als bewijs, welke argumentatiestijlen legitiem zijn, en hoe kennis wordt ingebed in bredere tradities (Becher & Trowler, 2001). Wanneer AI zowel de productie als de beoordeling van teksten overneemt, dreigt deze rituele functie uitgehold te raken. Als artikelen voornamelijk door algoritmen worden geschreven, en vervolgens door andere algoritmen worden beoordeeld en geclassificeerd, ontstaat een circulatie die de mens buitenspel zet. De vraag is dan niet of teksten nog worden geproduceerd, maar of zij nog functioneren als instrumenten van academische dialoog en vorming.

Originaliteit versus postzegel-wijsheden. Een tweede spanningsveld betreft de spanning tussen originaliteit en fragmentatie. AI kan in beginsel bijdragen aan synthese: bibliometrische analyses en literatuurtools maken patronen zichtbaar die onderzoekers over disciplinaire grenzen heen kunnen verbinden. Maar de institutionele logica van 'publish or perish' werkt in tegenovergestelde richting. Bastow, Dunleavy en Tinkler (2014) lieten zien hoe carrière- en erkenningssystemen steeds sterker gekoppeld zijn aan metrics zoals impactfactoren en h-index. Binnen zo'n regime kan AI juist fungeren als een productiemachine voor kleine, afgebakende bijdragen die het systeem in stand houden. Of AI originaliteit werkelijk bevordert, hangt dus niet af van de technologie zelf, maar van de institutionele context waarin zij wordt ingezet. Pas wanneer universiteiten en financieringsinstanties, in de geest van

de Hong Kong Principles (Moher et al., 2020), meer nadruk leggen op integriteit en samenwerking dan op kwantiteit, kan AI zijn potentieel als synthese-instrument waarmaken.

De AI-paradox. Het meest pregnante evaluatiepunt in dit hoofdstuk is de AI-paradox: het idee dat vruchtbaar gebruik van AI juist vereist dat men gevormd is in een academische context zonder AI. Studenten die vanaf het begin van hun opleiding leunen op AI-systemen missen vaak de ervaring van het moeizame proces van literatuur zoeken, bron-kritiek en argumentatieopbouw. Juist die ervaring is nodig om te kunnen beoordelen wat AI overslaat of vervormt. Institutioneel speelt een vergelijkbare dynamiek: het schrijven van een scriptie of het doorlopen van peer review fungeert traditioneel als rite de passage, een proces waarin onderzoekers niet alleen kennis, maar ook academische normen en waarden internaliseren. Wanneer AI deze taken grotendeels overneemt, verliest de academische cultuur een cruciale vormingsfunctie.

De paradox wordt nog scherper wanneer men de intergenerationele dimensie betrekt. Onderzoekers die zonder AI zijn gevormd beschikken over een referentiekader om de beperkingen van de technologie kritisch te beoordelen. Jongere cohorten, die AI vanaf het begin als vanzelfsprekend ervaren, dreigen afhankelijk te worden van een instrument dat zij onvoldoende begrijpen of bevragen. Dit kan leiden tot spanningen binnen de academische gemeenschap: wie geldt als gezaghebbend, wie als te afhankelijk, en wie bepaalt de standaard van kritische distantie?

De AI-paradox is daarmee niet slechts een didactisch vraagstuk, maar een structureel probleem voor de toekomst van academische legitimiteit. Zij dwingt ons na te denken over nieuwe vormen van AI-literacy en metacognitieve training, maar ook over institutionele waarborgen die rituelen van kritische vorming behouden.

5.7 Conclusie

De tweede deelvraag luidde: *Hoe beïnvloedt AI de pedagogische en institutionele praktijken van kennisproductie?* De analyse laat zien dat AI op beide terreinen kansen en risico's met zich meebrengt. Pedagogisch kan AI leren inclusiever en persoonlijker maken, maar ook leiden tot verlies van kritische vorming en nieuwe vormen van surveillance. Institutioneel kan AI publicaties transparanter en peer review efficiënter maken, maar ook de fragmentatie en instrumentalisering van kennis versterken.

Het voorlopige antwoord op deze tweede deelvraag is dan ook dat AI de pedagogische en institutionele praktijken van kennisproductie niet alleen versnelt en efficiënter maakt, maar tevens de voorwaarden ondermijnt waaraan deze praktijken hun legitimiteit ontlenen. Alleen wanneer teksten hun communicatieve functie behouden, originaliteit voorrang krijgt boven fragmentatie, en de AI-paradox expliciet wordt onderkend en geadresseerd, kan AI vruchtbaar bijdragen aan een academische cultuur die zowel lerend als legitiem is.

5.8 Nieuwe rituelen in een AI-tijdperk

De analyse tot nu toe benadrukte vooral hoe AI bestaande pedagogische en institutionele rituelen onder druk zet. Scripties, peer review en wetenschappelijke publicaties verliezen immers hun traditionele betekenis wanneer productie en beoordeling grotendeels door algoritmen worden overgenomen. Toch zou het te kort door de bocht zijn om dit louter als verlies te beschouwen. De geschiedenis van de academie laat zien dat rituelen steeds veranderen. Peer review, nu een hoeksteen van legitimiteit, ontstond pas in de twintigste eeuw uit een mengsel van boekcensuur en collegiale controle (Biagioli, 2002). Evenzo waren promotieceremonies oorspronkelijk religieus van aard en kregen zij pas later hun

moderne academische vorm. Rituelen zijn dus geen vaste grootheden, maar dynamische praktijken die zich aanpassen aan veranderende omstandigheden.

Tegen die achtergrond kan AI niet alleen gezien worden als een bedreiging, maar ook als katalysator van nieuwe rituelen. Eén mogelijkheid is de opkomst van **collaboratieve transparantie**. In plaats van een afgewerkte tekst die door één auteur wordt gepresenteerd, kan de weg naar een publicatie voortaan zichtbaar worden gemaakt in een logboek waarin menselijke keuzes en AI-interventies zijn gedocumenteerd. Het ritueel verschuift dan van het enkelvoudige eindproduct naar het proces van gezamenlijke totstandkoming.

Een tweede mogelijkheid is **community review**. Waar peer review traditioneel anoniem en gesloten plaatsvindt, kan AI worden ingezet om grotere groepen onderzoekers in staat te stellen een tekst te annoteren, analyseren en bediscussiëren. Menselijke en machinefeedback kunnen zo samen een nieuw ritueel vormen waarin transparantie en pluraliteit belangrijker worden dan formele accreditatie door enkele anonieme reviewers.

Ook in het onderwijs kunnen nieuwe rituelen ontstaan. Het schrijven van een scriptie kan zijn vormende kracht verliezen, maar dat betekent niet dat academische vorming zelf verdwijnt. In plaats daarvan kan het **kritisch beoordelen van AI-output** uitgroeien tot een nieuwe rite de passage. Studenten zouden niet langer uitsluitend worden getoetst op hun vermogen een tekst te produceren, maar op hun vaardigheid te onderscheiden wat betrouwbaar, valide en relevant is in een wereld vol automatisch gegenereerde kennis.

Deze mogelijkheden wijzen erop dat de weerstand tegen AI niet enkel verklaard kan worden door verlies, maar ook door onzekerheid over wat ervoor in de plaats komt. Als AI nieuwe rituelen kan helpen vormen die beter aansluiten bij de waarden van integriteit, transparantie en inclusiviteit, dan kan de pedagogische en institutionele cultuur van de academie er zelfs bij winnen. Het vraagt echter om bewuste keuzes: rituelen ontstaan niet vanzelf, maar worden geconstrueerd en gelegitimeerd door gemeenschappen.

Literatuurlijst Hoofdstuk 5

Bastow, S., Dunleavy, P., & Tinkler, J. (2014). *The impact of the social sciences: How academics and their research make a difference*. SAGE. https://doi.org/10.4135/9781473957947

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2e ed.). Open University Press.

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures, 12*(1), 11–45. https://doi.org/10.1080/1045722022000003430

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Holmes, W., Bialik, M., & Fadel, C. (2021). *Artificial intelligence in education: Promises and implications for teaching and learning*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000376709

Knox, J. (2020). *Artificial intelligence and education futures: Critical perspectives on learning, knowledge and culture*. Routledge.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., … & Dirnagl, U. (2020). The Hong Kong principles for assessing researchers: Fostering research integrity. *PLoS Biology, 18*(7), e3000737. https://doi.org/10.1371/journal.pbio.3000737

Müller-Birn, C., Dobusch, L., & Seitz, S. (2020). Peer review in the age of artificial intelligence. *Science and Public Policy, 47*(6), 749–759. https://doi.org/10.1093/scipol/scaa039

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The case for critical research in learning analytics. *Learning Analytics: Fundaments, Applications, and Trends*, 105–128. Springer. https://doi.org/10.1007/978-3-319-52977-6_6

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press

Hoofdstuk 6 – Macht, normativiteit en de aard van AI-gestuurde kennisproductie

6.1 Inleiding

De derde deelvraag in dit proefschrift luidt: Hoe herconfigureert kunstmatige intelligentie de machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar voor AI-gestuurde kennisproductie? Deze vraag raakt aan de kern van de wetenschappelijke praktijk. Waar macht en normativiteit traditioneel vaak als externe condities worden beschouwd — contexten waarin wetenschap plaatsvindt maar die de kern van kennisproductie zelf niet veranderen — blijkt bij AI dat deze dimensies constitutief zijn. Toegang tot infrastructuur, de verdeling van data, en de wijze waarop normen worden ingebouwd in systemen bepalen niet alleen wie kennis kan produceren, maar ook welke kennis als geldig wordt beschouwd. Dit hoofdstuk laat zien dat macht en normativiteit de aard en kwaliteit van AI-gestuurde kennisproductie fundamenteel mede vormgeven.

6.2 Macht als voorwaarde voor kennisproductie

AI-gestuurde kennisproductie is afhankelijk van grootschalige infrastructuren: rekenkracht, energie, cloud-voorzieningen en gespecialiseerd talent. Het trainen van foundation models vereist investeringen die slechts een handvol actoren zich kan veroorloven. Zoals Brookings en GovAI (2023) hebben aangetoond, leiden deze hoge vaste kosten tot schaalvoordelen en dus tot concentratie van macht bij enkele bedrijven en onderzoeksconsortia. De Britse mededingingsautoriteit CMA waarschuwde dat verticale allianties tussen cloudproviders, chipfabrikanten en modelontwikkelaars de toegang nog verder beperken.

Deze machtsconcentratie heeft directe gevolgen voor de aard van kennisproductie. Onderzoekers die niet over toegang tot dergelijke infrastructuur beschikken, zijn aangewezen op reeds getrainde modellen. Daarmee verliezen zij invloed op cruciale vroege fasen van onderzoek: de formulering van vragen, de selectie van datasets en de keuzen die in het modelontwerp worden ingebouwd. Kennisproductie wordt daardoor conditioneel en relationeel: niet iedereen kan deelnemen, en deelname is afhankelijk van machtsposities binnen een technologisch ecosysteem.

6.3 Normativiteit ingebouwd

Normativiteit in AI is niet slechts een externe toetssteen, maar een constitutief element van de technologie zelf. Waar traditionele wetenschap kan terugvallen op gedeelde methodologische standaarden die achteraf worden getoetst (via peer review, replicatie of ethische commissies), zijn bij AI de normen vaak vooraf ingebouwd in datasets en modellen. Zoals Biagioli (2002) heeft laten zien, vervullen academische communicatie en beoordelingsprocessen niet alleen een informatieve, maar ook een rituele en legitimerende functie in de kennisproductie. Dit maakt dat AI-gestuurde kennisproductie vanaf de basis gekleurd is door impliciete en expliciete keuzes.

Een eerste probleem betreft de selectie en uitsluiting in datasets. Elke dataset is een reductie van de werkelijkheid: bepaalde bronnen, talen of groepen worden meegenomen, andere niet. Als een model voornamelijk is getraind op Engelstalige wetenschappelijke literatuur, leert het vooral die perspectieven te reproduceren en te legitimeren. Dit is niet alleen een kwestie van bias, maar ook een normatieve keuze over wat als kennis telt en wie in dat proces representatie krijgt. De claim van universaliteit die veel AI-modellen uitstralen, maskeert het feit dat hun basis fundamenteel contingent is.

Een tweede probleem ligt in de definitie van wat geldt als legitieme output. Documentatiepraktijken zoals Model Cards (Mitchell et al., 2019) en Datasheets for Datasets (Gebru et al., 2018) leggen vaak vast waarvoor een model bedoeld is en hoe het verantwoord kan worden gebruikt. Daarmee lijken ze transparantie te bevorderen, maar in feite verschuift de normatieve macht: ontwikkelaars bepalen welk gebruik "gepast" is en welke toepassingen buiten de kaders vallen. Voor kennisproductie betekent dit dat wetenschappers vaak binnen die door anderen getrokken grenzen moeten opereren, zonder de mogelijkheid die grenzen zelfstandig te heronderhandelen.

Een derde probleem betreft epistemische afhankelijkheid. Onderzoekers die een bestaand model gebruiken, nemen de ingebouwde aannames onvermijdelijk over. Vaak is niet meer te achterhalen welke selectiecriteria golden, welke data zijn verwijderd, of welke optimalisaties zijn toegepast. Hierdoor wordt kennisproductie afhankelijk van normen die de onderzoeker niet zelf heeft gekozen, noch kan controleren. Wat traditioneel gold als verantwoordelijkheid van de onderzoeker, verschuift naar onzichtbare actoren en processen in de keten van AI-ontwikkeling.

Concrete voorbeelden laten zien hoe dit problematisch kan uitpakken. In de natuurwetenschappen kan een AI-systeem dat klimaatmodellen analyseert bepaalde historische datasets uitsluiten, waardoor patronen ontstaan die robuust lijken, maar in feite contingent zijn. In de geesteswetenschappen kan een taalmodel impliciet besluiten dat bepaalde dialecten of genres niet relevant zijn, waarmee een canon wordt bevestigd zonder dat onderzoekers dit expliciet beogen. In de sociale wetenschappen kan sentimentanalyse op Twitter worden gepresenteerd als objectieve kennis over "de samenleving", terwijl de onderliggende normatieve claim is dat Twitter representatief is voor die samenleving.

De kern is dat AI-gestuurde kennisproductie zijn legitimiteit niet alleen ontleent aan klassieke criteria zoals transparantie en reproduceerbaarheid, maar ook aan keuzes die op onzichtbare wijze in data en modellen zijn ingebouwd. Zolang die keuzes niet expliciet worden gemaakt en besproken, blijft de kwaliteit van de geproduceerde kennis fundamenteel onzeker.

6.4 Decontexualisatie en de schaarste van data

Een van de meest fundamentele uitdagingen voor de kwaliteit van AI-gestuurde kennisproductie ligt in de manier waarop data worden verzameld en hergebruikt. Data die ooit in een specifieke context zijn geproduceerd — bijvoorbeeld medische dossiers, sociale mediaberichten of etnografische observaties — verliezen hun oorspronkelijke betekenis zodra zij in een ander kader worden ingezet. Dit proces van decontexualisatie is niet nieuw, maar krijgt in AI een systematisch en massaal karakter. Modellen worden getraind op datasets die hun herkomst grotendeels verbergen, waardoor cruciale informatie over omstandigheden, doel en interpretatiekaders verloren gaat.

De gevolgen hiervan zijn direct merkbaar voor de kwaliteit van kennisproductie. Zonder context wordt de betrouwbaarheid van patronen die AI herkent onzeker: correlaties kunnen ontstaan die in de oorspronkelijke setting betekenisvol waren, maar die in een nieuw domein misleidend blijken. De claim van universaliteit die AI-modellen vaak uitstralen, maskeert dus de fragiliteit van hun basis.

Daarbij komt dat data zelf steeds meer een schaarse grondstof zijn. Hoogwaardige, goed gedocumenteerde en representatieve datasets zijn zeldzaam. De concurrentie om toegang tot zulke data leidt tot machtsconcentratie bij partijen die erin slagen deze grondstof te verzamelen, te kopen of af te schermen. Zoals Couldry en Mejías (2019) aantonen, vormen data-infrastructuren niet alleen een technische kwestie, maar creëren zij nieuwe vormen van extractie en toe-eigening die direct samenhangen met mondiale machtsrelaties. Macht over data betekent daarmee macht over kennis: wie toegang heeft, kan kennis produceren; wie buitengesloten is, blijft afhankelijk van andermans modellen en aannames.

De combinatie van onvermijdelijke decontexualisatie en de schaarste van bruikbare data maakt duidelijk dat AI-gestuurde kennisproductie kwetsbaar is. Niet alleen omdat modellen vertekeningen of biases kunnen bevatten, maar ook omdat de voorwaarden waaronder data beschikbaar komen bepalend zijn voor de vragen die überhaupt gesteld kunnen worden. Macht over data vertaalt zich zo direct in macht over kennis.

6.5 Het black-box-karakter en normatieve spanningsvelden

Het meest fundamentele probleem voor de legitimiteit van AI-gestuurde kennisproductie is het black-box-karakter van de technologie. Zowel gebruikers als wetenschappers hebben vaak geen inzicht in de datasets waarop modellen zijn getraind, de precieze algoritmische keuzes die daarin zijn ingebouwd, of de mechanismen waarmee een output tot stand komt. Dit gebrek aan transparantie maakt dat elke toepassing gepaard gaat met onzekerheid: niet alleen over de uitkomst, maar over de epistemische grond waarop die uitkomst rust.

De normatieve spanningen die in dit hoofdstuk zijn beschreven zijn in feite manifestaties van dit black-box-karakter. Het conflict tussen transparantie en bedrijfsgeheimen vloeit voort uit het feit dat veel modellen gesloten blijven onder verwijzing naar intellectueel eigendom en concurrentievoordeel. Hierdoor is het onmogelijk om systematisch te controleren of outputs betrouwbaar zijn of welke bias aanwezig is.

Het debat rond open versus closed source raakt eveneens aan de vraag hoe ver de doos open kan worden gemaakt. Open-source modellen lijken toegankelijker, maar ook hier blijft de training vaak gebaseerd op datasets waarvan de herkomst en selectie niet volledig te achterhalen zijn. Openheid garandeert dus niet noodzakelijk inzicht, maar legt wel nieuwe risico's bloot, zoals misbruik door kwaadwillenden.

Zelfs de discussie over duurzaamheid laat zien hoe het black-box-karakter doorwerkt. Zoals Strubell, Ganesh en McCallum (2019) empirisch hebben aangetoond, blijft het energie- en grondstoffengebruik van AI-trainingen grotendeels verborgen achter technische en commerciële muren en vormt daarmee onderdeel van de epistemische ondoorzichtigheid. Gebruikers zien de outputs, maar niet de ecologische prijs die eraan verbonden is. Dit creëert een vals gevoel van neutraliteit: kennisproductie lijkt immaterieel, terwijl zij in werkelijkheid stevig verankerd is in materiële kosten en machtsstructuren. Zuboff (2019) plaatst deze ontwikkelingen in een bredere context van surveillance capitalism, waarin data-infrastructuren nieuwe vormen van afhankelijkheid en extractie creëren met directe gevolgen voor kennisproductie.

Het gevolg is dat elke poging om AI-gestuurde kennisproductie te evalueren onvermijdelijk worstelt met dit black-box-karakter. Zolang de doos niet open kan, blijven claims van validiteit, betrouwbaarheid en universaliteit fundamenteel onzeker. Transparantie-instrumenten en reguleringspogingen zoals de Artificial Intelligence Act van de Europese Unie (2024) en de OECD Principles on Artificial Intelligence (2019) kunnen de doos deels lichten, maar veranderen niets aan het gegeven dat de epistemische kern

van AI-systemen grotendeels ontoegankelijk blijft. Dit is de reden dat AI-gestuurde kennisproductie, hoe veelbelovend ook, steeds onderhevig blijft aan fundamentele twijfel over haar aard en kwaliteit.

6.6 Generatieve en dedicated AI: verschillende implicaties voor kennisproductie

Tot dusver is in dit hoofdstuk vaak verwezen naar generatieve AI-modellen die getraind zijn op enorme tekst- en beeldcorpora. Zulke modellen, zoals GPT-systemen, maken het black-box-karakter en de problematiek van ingebouwde normativiteit bijzonder zichtbaar. Hun trainingsdata zijn omvangrijk maar ongedocumenteerd, hun aannames grotendeels ondoorzichtig, en hun outputs lijken universeel terwijl zij feitelijk gebaseerd zijn op een oncontroleerbare selectie. Voor de kwaliteit van AI-gestuurde kennisproductie betekent dit dat generaliseerbaarheid en validiteit altijd omgeven zijn door twijfel: de gebruiker kan vaak niet reconstrueren hoe een bepaalde conclusie tot stand is gekomen.

Dit beeld is echter niet representatief voor alle AI-toepassingen. In veel domeinen wordt gewerkt met dedicated AI-systemen, die zijn getraind op specifieke en afgebakende datasets, zoals medische beeldvorming, astronomische observaties of experimentele laboratoriumdata. Hier is de context van de data vaak beter gedocumenteerd, zijn de normatieve keuzes bij selectie en annotatie explicieter, en kan de black box in principe kleiner en beter te auditen zijn. De problematiek verschuift: minder nadruk op radicale ondoorzichtigheid, meer op toegang tot schaarse datasets en de macht van instellingen die zulke data bezitten of reguleren.

Deze vergelijking laat zien dat AI-gestuurde kennisproductie niet eenduidig kan worden beoordeeld. Bij generatieve modellen ligt de kwetsbaarheid vooral in de epistemische ondoorzichtigheid van de onderliggende corpora. Bij dedicated modellen ligt zij in de machtsconcentratie rond data-infrastructuren en de normatieve keuzes die in labeling en interpretatie worden gemaakt. Beide configuraties hebben dus hun eigen problematiek, en beide moeten expliciet worden betrokken bij het beoordelen van de kwaliteit en legitimiteit van AI-gestuurde kennisproductie.

6.7 Voorlopige conclusie

De analyse van macht en normativiteit toont dat AI-gestuurde kennisproductie een fundamenteel andere aard heeft dan traditionele vormen van wetenschap. Macht is niet louter contextueel, maar bepaalt wie kan deelnemen en welke vragen gesteld worden. Normativiteit is niet extern, maar ingebouwd in data, modellen en infrastructuren. Universaliteit is niet vanzelfsprekend, maar fragiel, omdat representatie en legitimiteit scheef verdeeld zijn. Kennisproductie is niet autonoom, maar relationeel en conditioneel: afhankelijk van toegang, inclusie en duurzaamheid.

Zoals UNESCO (2021) in haar aanbevelingen over de ethiek van kunstmatige intelligentie benadrukt, gaan normatieve vragen rond AI-gestuurde kennisproductie verder dan louter academische of technische overwegingen en vormen zij onderdeel van mondiale governance-discussies. Het voorlopige antwoord op de derde deelvraag luidt dan ook dat AI-gestuurde kennisproductie slechts legitiem kan zijn wanneer de constitutieve rol van macht en normativiteit expliciet wordt onderkend en gereguleerd. Kwaliteit kan niet langer uitsluitend worden beoordeeld aan de hand van methodologische zorgvuldigheid of peer review, maar moet ook langs de lijnen van infrastructuurtoegang, pluraliteit van perspectieven, en transparantie over ingebouwde normen. Alleen door deze bredere criteria te hanteren, kan kennisproductie in een AI-tijdperk zowel robuust als rechtvaardig zijn.

6.8 Relativering en mitigatie

De analyse in dit hoofdstuk heeft benadrukt dat macht, normativiteit en het black-box-karakter constitutief zijn voor AI-gestuurde kennisproductie. Daarmee lijken de fundamenten van wetenschappelijke legitimiteit voortdurend onder druk te staan. Toch is het van belang te relativeren. De vraag is immers niet alleen welke problemen er bestaan, maar ook hoe ernstig zij zijn in de praktijk en in welke mate zij kunnen worden opgevangen.

Beleidsmakers en onderzoeksinstellingen zullen onvermijdelijk pragmatische afwegingen maken. Als AI in 99,5% van de gevallen betrouwbare resultaten oplevert, terwijl slechts 0,5% problematisch is, kan de verleiding groot zijn om de resterende onzekerheid te accepteren. In de geschiedenis van de wetenschap zijn er immers altijd foutmarges geweest: instrumenten zijn niet perfect, metingen zijn niet feilloos, en ook menselijke onderzoekers maken interpretatieve keuzes die tot vertekening kunnen leiden. De kernvraag is dus niet of AI foutloos moet zijn, maar welke foutmarge maatschappelijk en wetenschappelijk aanvaardbaar wordt geacht. Voor experimentele natuurwetenschappen kan dit een kleine statistische onzekerheid zijn, terwijl in medische of juridische toepassingen zelfs zeldzame fouten grote consequenties hebben.

Daarnaast zijn er inmiddels uiteenlopende strategieën om de risico's van AI-gestuurde kennisproductie te mitigeren. Op technisch niveau bestaan instrumenten zoals Model Cards (Mitchell et al., 2019) en Datasheets for Datasets (Gebru et al., 2018), die systematisch documenteren welke data zijn gebruikt en met welke aannames. Bias-detectie en fairness-metrics kunnen helpen om vertekeningen zichtbaar te maken. Institutioneel zijn er vormen van auditing in ontwikkeling, waarbij modellen periodiek extern worden gecontroleerd. Het NIST Artificial Intelligence Risk Management Framework (2023) biedt beleidsmatige instrumenten die betrouwbaarheid en risico's van AI-gestuurde kennisproductie systematisch proberen te adresseren. Ook hybride vormen van samenwerking tussen mens en machine kunnen de betrouwbaarheid vergroten: AI genereert patronen, terwijl menselijke onderzoekers deze kritisch evalueren en contextualiseren.

Een bijzonder punt van aandacht betreft het gebruik van gelabelde data. Veel dedicated AI-systemen zijn gebaseerd op supervised learning: modellen worden getraind met datasets die door experts van labels zijn voorzien, bijvoorbeeld medische scans die door radiologen als 'tumor aanwezig' of 'afwezig' zijn geclassificeerd. Dit lijkt transparanter dan het trainen op ongedocumenteerde corpora, maar ook hier spelen normatieve keuzes een rol: wie bepaalt wat als correcte label geldt, hoe consequent worden die labels toegepast, en wat gebeurt er met gevallen die niet in de vooraf gedefinieerde categorieën passen? Mitigatie betekent in dit geval vooral: labels expliciet maken, intersubjectief toetsen, en ruimte laten voor herziening.

Deze relativering laat zien dat de problemen die in dit hoofdstuk zijn geschetst ernstig zijn, maar niet fataal. AI-gestuurde kennisproductie zal zich ondanks deze bezwaren waarschijnlijk breed verspreiden. De vraag is niet óf de technologie wordt gebruikt, maar onder welke voorwaarden en met welke waarborgen. De uitdaging ligt erin een evenwicht te vinden: foutmarges aanvaarden waar dit verantwoord is, terwijl tegelijk mechanismen worden ingebouwd om structurele vertekening, machtsmisbruik en black-box-onzekerheid te beperken. Alleen zo kan AI-gestuurde kennisproductie zich ontwikkelen tot een praktijk die niet alleen efficiënt en krachtig is, maar ook wetenschappelijk legitiem en maatschappelijk verantwoord.

.

Literatuurlijst Hoofdstuk 6

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences*, 12(1), 11–45.

Brookings Institution / GovAI. (2023). *Analyses of foundation model concentration and market power* [Policy reports].

Couldry, N., & Mejías, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

European Union. (2024). *Artificial Intelligence Act*. Official Journal of the European Union.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.

NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology.

OECD. (2019). *OECD Principles on Artificial Intelligence*. Organisation for Economic Co-operation and Development.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the ACL*, 3645–3650.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization.

Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs

Intermezzo – Tussenbalans en agenda

Na de bespreking van de drie impactdimensies — cognitief en methodologisch (hoofdstuk 4), pedagogisch en institutioneel (hoofdstuk 5), en macht en normativiteit (hoofdstuk 6) — is het zinvol een voorlopige balans op te maken. De deelvragen zijn in eerste aanleg behandeld, maar nog niet in hun volle breedte beantwoord. Wat wel duidelijk is, is dat AI-gestuurde kennisproductie zich aandient als een hybride fenomeen: tegelijk beloftevol en problematisch, tegelijk verrijkend en ondermijnend. In de voorgaande hoofdstukken zijn deze spanningen telkens vanuit één dimensie belicht. Het doel van dit intermezzo is om die inzichten samen te brengen in drie taxonomieën — van kansen, risico's en uitdagingen — die gezamenlijk de analytische basis vormen voor de uitwerking in de archetypen-hoofdstukken die volgen.

I. Taxonomie van kansen. De eerste taxonomie betreft de kansen die AI-gestuurde kennisproductie in algemene zin biedt. Deze kansen zijn niet louter bijvangst van technologische vooruitgang, maar raken aan kernprocessen van wetenschap.

1. Versnelling en schaal. AI maakt het mogelijk kennis te produceren in een tempo en op een schaal die voorheen ondenkbaar waren. Dataverwerking en literatuuronderzoek verlopen niet langer lineair en traag, maar cyclisch en dynamisch: hypotheses kunnen sneller worden gegenereerd, getoetst en bijgesteld. Deze versnelling opent niet alleen de weg naar efficiëntie, maar ook naar nieuwe vormen van wetenschappelijke praktijk waarin feedbackloops korter zijn en ontdekkingen sneller circuleren.

2. Ontdekking en synthese. AI fungeert daarnaast als epistemisch hulpmiddel dat onverwachte patronen zichtbaar maakt. Waar menselijke onderzoekers gebonden zijn aan disciplinaire kaders en cognitieve beperkingen, kan AI in heterogene datasets verbanden blootleggen die anders verborgen zouden blijven. In biologische wetenschappen zijn reeds doorbraken gerealiseerd in de voorspelling van eiwitstructuren, terwijl in de geesteswetenschappen nieuwe verbanden in omvangrijke corpora aan het licht zijn gekomen. Voor AI-gestuurde kennisproductie betekent dit dat fragmentatie kan worden overstegen en synthese mogelijk wordt gemaakt op een schaal die voorheen onbereikbaar was.

3. Pedagogische en institutionele vernieuwing. Een derde cluster van kansen ligt in de pedagogische en institutionele sfeer. AI kan fungeren als gepersonaliseerd leer- en begeleidingsinstrument en als multiplicator van beoordelingsperspectieven. Traditionele vormen van peer review kunnen worden verrijkt door parallelle bijdragen van AI-systemen die de argumentatieve consistentie of methodologische robuustheid analyseren. Daarnaast maken instrumenten zoals *model cards* en *datasheets* nieuwe vormen van transparantie en systematische verantwoording mogelijk. In termen van kennisproductie betekent dit dat ook de instituties en praktijken waarin wetenschap geworteld is, ruimte krijgen voor vernieuwing.

II. Taxonomie van risico's. Tegenover deze kansen staan evenzeer structurele risico's. Deze risico's raken niet aan de marge, maar aan de kern van wat wetenschappelijke kennis legitiem maakt.

1. Black-box en verlies van transparantie. AI-modellen functioneren vaak als black boxes. Gebruikers weten doorgaans niet welke data exact zijn gebruikt of hoe algoritmen beslissingen nemen. Dit ondermijnt reproduceerbaarheid, een fundamenteel criterium voor wetenschappelijke validiteit. Voor AI-gestuurde kennisproductie betekent dit dat de verantwoording van resultaten afhankelijk wordt van infrastructuren en datasets die vaak ontoegankelijk of commercieel gesloten zijn.

2. Decontexualisatie en verlies van nuance. Data die in specifieke contexten zijn verzameld verliezen hun betekenis zodra zij worden ingezet in nieuwe omgevingen zonder voldoende duiding. AI versterkt dit proces door op grote schaal patronen te construeren uit data die losgemaakt zijn van hun oorspronkelijke setting. Uitkomsten krijgen daardoor een schijn van universele toepasbaarheid, terwijl zij in werkelijkheid contextueel contingent zijn. Dit vormt een bedreiging voor de betrouwbaarheid van kennis die door AI wordt geproduceerd.

3. Ritualisering en betekenisverlies van communicatie. Wetenschappelijke communicatie vervult naast een informatieve ook een rituele functie: zij legitimeert onderzoekers als leden van een gemeenschap en draagt waarden over zoals zorgvuldigheid, transparantie en dialoog (Biagioli, 2002). Wanneer AI zowel de productie als de beoordeling van scripties en artikelen overneemt, dreigt deze rituele dimensie uitgehold te raken. Teksten circuleren dan vooral tussen mens en machine, zonder dat zij nog dezelfde gemeenschapsvormende kracht uitoefenen. Voor AI-gestuurde kennisproductie betekent dit een verschuiving van legitimiteit: van intersubjectieve toetsing naar technologische output.

4. Machtsconcentratie en afhankelijkheid. Omdat hoogwaardige data en rekenkracht schaars en kostbaar zijn, komen zij terecht bij een beperkt aantal bedrijven en instellingen. Deze machtsconcentratie betekent dat epistemische agenda's impliciet of expliciet worden gestuurd door commerciële of geopolitieke belangen. Onderzoekers die afhankelijk zijn van externe modellen, verliezen zo een deel van hun autonomie.

5. Fabricage en betrouwbaarheid van referenties. Een bijzonder risico dat nauw verweven is met het black-box-karakter, maar apart benoemd moet worden, betreft de fabricatie van literatuurverwijzingen. Grote taalmodellen kunnen overtuigende literatuur-reviews genereren, maar produceren daarbij vaak gedeeltelijk of geheel verzonnen referenties. Voor AI zijn bronnen geen betekenisvolle markers van wetenschappelijk gesprek, maar patronen in tekst. Dit ondermijnt een kernpraktijk van de wetenschap: de legitimering van nieuwe kennis door haar expliciet in relatie te brengen met bestaande literatuur. Voor AI-gestuurde kennisproductie vormt dit een directe bedreiging van de betrouwbaarheid en geloofwaardigheid van literatuuronderzoek.

III. Taxonomie van uitdagingen. De derde taxonomie bundelt de fundamentele uitdagingen die de combinatie van kansen en risico's aan het licht brengt. Het zijn vraagstukken die de toekomstige inrichting van AI-gestuurde kennisproductie bepalen.

1. Herijking van validiteit en betrouwbaarheid. De klassieke kwaliteitscriteria van wetenschap zijn niet langer vanzelfsprekend. Validiteit en betrouwbaarheid moeten worden geherdefinieerd in een context waarin instrumenten zelf epistemisch ondoorzichtig zijn. Nieuwe vormen van verantwoording, zoals auditprocedures en metadata-documentatie, zijn nodig om het vertrouwen in AI-gestuurde kennisproductie te waarborgen.

2. Institutionele integratie zonder erosie van kernrituelen. AI moet worden geïntegreerd in onderwijs, begeleiding en publicatie zonder dat de sociale en rituele functies van deze praktijken verdwijnen. De uitdaging is niet alleen technisch, maar ook cultureel: hoe kan de academische gemeenschap haar vormende rituelen behouden wanneer AI steeds meer taken overneemt?

3. Contextuele normstelling van foutmarges. Geen enkel systeem is foutloos, en AI evenmin. De vraag is dus niet óf fouten optreden, maar welke foutmarge in welke context aanvaardbaar is. In fundamenteel onderzoek kan enige onzekerheid acceptabel zijn, zolang robuuste trends zichtbaar blijven. In medische of juridische contexten daarentegen kunnen zelfs uiterst geringe fouten onaanvaardbaar zijn. Deze differentiatie vereist expliciete normstelling per domein en toepassing.

IV. Scharnierfunctie. Met deze drie taxonomieën is de tussenbalans compleet. Zij bundelen de inzichten uit de drie dimensiehoofdstukken en brengen de spanningen die AI-gestuurde kennisproductie kenmerken scherp in beeld. De taxonomieën fungeren als analytisch raster voor de komende hoofdstukken, waarin de archetypen van wetenschap centraal staan. Per archetype zal systematisch worden onderzocht welke kansen concreet tot hun recht komen, welke risico's zich het scherpst manifesteren, en welke uitdagingen het meest urgent zijn.

Dit intermezzo is daarmee meer dan een samenvatting: het markeert een overgang. Waar de voorgaande hoofdstukken de dimensies afzonderlijk hebben behandeld, wordt hier het geheel zichtbaar in een gestructureerde vorm. Tegelijk wijst dit intermezzo vooruit: de vraag is niet langer óf AI-gestuurde kennisproductie de wetenschap verandert, maar hóe deze veranderingen zich in verschillende kennispraktijken concretiseren.

Hoofdstuk 7 – Natuurwetenschappelijk onderzoek en AI-gestuurde kennisproductie

7.1 Inleiding

De natuurwetenschappen worden traditioneel gekenmerkt door hun sterke nadruk op empirische verificatie, reproduceerbaarheid en causaliteit. Juist deze kenmerken maken het domein bijzonder interessant om de rol van kunstmatige intelligentie (AI) te analyseren. AI wordt hier op grote schaal ingezet, variërend van moleculaire biologie tot astronomie en klimaatwetenschap. Anders dan in meer interpretatieve disciplines is de inzet doorgaans gericht op patroonherkenning, data-analyse en hypothesevorming, vaak in de vorm van dedicated AI-systemen die getraind zijn op domein-specifieke datasets.

De centrale vraag in dit hoofdstuk is hoe de kansen, risico's en uitdagingen van AI-gestuurde kennisproductie zich manifesteren binnen de natuurwetenschappen. Daarbij gaat het niet alleen om de praktische bruikbaarheid van AI, maar vooral om de mate waarin de resultaten wetenschappelijk legitiem en betrouwbaar zijn.

Om deze vraag empirisch te funderen, wordt in dit hoofdstuk gekozen voor een asymmetrische aanpak. AlphaFold in de moleculaire biologie fungeert als primaire casus en wordt diepgaand uitgewerkt over alle onderdelen van de taxonomie. Deze keuze is niet willekeurig: AlphaFold belichaamt bij uitstek de spanningen en mogelijkheden van dedicated AI-systemen in de natuurwetenschappen en biedt daarmee een rijk empirisch anker voor conceptuele analyse. Voorbeelden uit astronomie en klimaatwetenschap worden aanvullend ingezet om de generaliseerbaarheid van de bevindingen te ondersteunen, zonder de pretentie deze domeinen met gelijke diepgang te behandelen.

De analyse maakt gebruik van de taxonomieën die in het intermezzo zijn ontwikkeld en illustreert hoe de cognitieve, institutionele en normatieve dimensies van AI-gestuurde kennisproductie in dit archetype tot uitdrukking komen.

7.2 Kansen

Versnelling en schaal. Een van de meest tastbare kansen van AI in de natuurwetenschappen is de versnelling van dataverwerking. Het meest sprekende voorbeeld is *AlphaFold*, een systeem van DeepMind dat in staat is driedimensionale eiwitstructuren te voorspellen met opmerkelijke nauwkeurigheid (Jumper et al., 2021). Voorheen vergden zulke voorspellingen jaren van experimenteel onderzoek, terwijl AlphaFold binnen enkele maanden honderdduizenden structuren beschikbaar maakte. Deze versnelling heeft niet alleen geleid tot een grotere efficiëntie, maar ook tot een democratisering van kennis: de resulterende databanken zijn publiek toegankelijk, zodat onderzoekers wereldwijd toegang hebben tot informatie die eerder voorbehouden was aan gespecialiseerde laboratoria.

De doorbraak van AlphaFold verdient nadere analyse. Jumper et al. (2021) schrijven in Nature: "The ability to predict a protein's shape computationally from its amino acid sequence would be a boon to life sciences and medicine. It would vastly accelerate efforts to understand the building blocks of cells and enable quicker and more advanced drug discovery." Dit is niet zomaar retoriek. Hun systeem bereikte

een mediaanscore van 92.4 GDT op de CASP14-test[1], waar de beste vorige methodes onder de 60 bleven steken. Maar Jumper et al. zijn ook eerlijk over de beperkingen. Zij stellen: "AlphaFold produces a per-residue estimate of its confidence on a scale from 0-100. This confidence measure is calibrated to correspond approximately to the model's expected per-residue accuracy" (p. 587). Dit is cruciaal: het systeem *weet* waar het onzeker is. Dit contrasteert met de black-box-kritiek die vaak op AI wordt geuit. AlphaFold is weliswaar complex, maar niet volledig ondoorzichtig - het communiceert zijn eigen onzekerheid.

Ook in de astronomie is de bijdrage van AI evident. Moderne telescopen genereren enorme hoeveelheden data die onmogelijk handmatig te analyseren zijn. AI-modellen worden ingezet om patronen in lichtcurves te identificeren en exoplaneten te detecteren (Valenti et al., 2021). Hiermee wordt de schaal van observatie en analyse radicaal vergroot: waar vroeger enkele tientallen planeten per jaar konden worden ontdekt, maakt AI het mogelijk duizenden potentiële kandidaten te evalueren.

In de klimaatwetenschap biedt AI vergelijkbare voordelen. Machine learning-modellen worden ingezet om weersvoorspellingen te verfijnen en klimaatprojecties te verbeteren (Rolnick et al., 2019). Doordat deze modellen sneller kunnen worden herberekend dan traditionele simulaties, ontstaat een dynamischer proces van hypothesevorming en scenario-analyse. Dit stelt onderzoekers in staat sneller te reageren op nieuwe data en zo de empirische cyclus te verkorten.

Ontdekking en synthese. Naast versnelling biedt AI ook nieuwe vormen van ontdekking en synthese. AlphaFold levert niet alleen sneller resultaten, maar genereert ook hypothesen over structuren die nog niet empirisch zijn vastgesteld. Dit vergroot het bereik van wetenschappelijke verbeelding: AI fungeert hier als een bron van nieuwe vragen die vervolgens experimenteel kunnen worden getoetst.

In de astronomie geldt hetzelfde. AI-systemen hebben zwaartekrachtslenzen en patronen in kosmologische datasets geïdentificeerd die voorheen niet zichtbaar waren (Fluke & Jacobs, 2020). Deze ontdekkingen verschuiven de horizon van onderzoek, omdat zij nieuwe mogelijkheden bieden om theorieën over donkere materie en kosmologische expansie te toetsen.

De klimaatwetenschap laat zien hoe AI kan bijdragen aan synthese. Door uiteenlopende datasets — van satellietbeelden tot historische meetreeksen — te integreren, kunnen modellen robuustere voorspellingen genereren. AI fungeert hier niet alleen als versneller, maar ook als instrument dat fragmenten van data bijeenbrengt tot een samenhangender geheel.

Pedagogische en institutionele vernieuwing. Ten slotte biedt AI kansen voor pedagogische en institutionele innovatie. Databanken zoals de *Protein Data Bank*, die door AlphaFold aanzienlijk is uitgebreid, fungeren als onderwijs- en onderzoeksinstrumenten voor studenten en wetenschappers wereldwijd. Ook de kwaliteitsbewaking profiteert: AI-systemen kunnen datasets screenen op inconsistenties of anomalieën, waardoor menselijke beoordelaars zich kunnen concentreren op inhoudelijke evaluatie. Dit leidt tot een vorm van gelaagde kwaliteitsborging die zowel efficiëntie als betrouwbaarheid vergroot.

[1] Een *mediaanscore van 92,4 GDT op de CASP14-test* verwijst naar de prestatie van AlphaFold bij de internationale competitie *Critical Assessment of protein Structure Prediction* (CASP). De **GDT_TS-score** (*Global Distance Test – Total Score*) meet de structurele overeenkomst tussen een voorspelde en een experimenteel bepaalde eiwitstructuur, op een schaal van 0 tot 100. Een score boven **90** wordt beschouwd als *experimenteel nauwkeurig*, wat betekent dat de berekende driedimensionale vorm van het eiwit nauwelijks nog afwijkt van de structuur die met fysische methoden — zoals röntgendiffractie of cryo-elektronenmicroscopie — in het laboratorium is vastgesteld. Met een mediaanscore van 92,4 over alle testeiwitten behaalde AlphaFold dus een precisie die vrijwel gelijkstaat aan laboratoriumnauwkeurigheid, een sprong van eerdere topmodellen (onder 60 GDT) naar bijna perfecte structurele voorspellingen — algemeen erkend als een doorbraak in de structurele biologie.

7.3 Risico's

De natuurwetenschappen tonen onmiskenbaar de kracht van AI, maar zij zijn niet immuun voor risico's. Binnen dit domein verschuift de aandacht van interpretatieve onzekerheden naar vragen rond infrastructuur, data en transparantie.

Black-box en beperkte verklaarbaarheid. Zelfs wanneer modellen dedicated en domeinspecifiek zijn, blijft het probleem van de black box aanwezig. Bij AlphaFold is de nauwkeurigheid indrukwekkend, maar de weg naar het resultaat is slechts ten dele te doorgronden (Jumper et al., 2021). Ook in de astronomie herkennen AI-systemen patronen in lichtcurves die onderzoekers niet altijd fysisch kunnen verklaren (Valenti et al., 2021). Dit roept de vraag op of kennis zonder verklaarbaarheid wel dezelfde status kan hebben als kennis die door klassieke methoden tot stand komt.

Jenna Burrell's invloedrijke artikel "How the machine 'thinks'" (2016) onderscheidt drie vormen van algoritmische ondoorzichtigheid: "opacity as intentional corporate or state secrecy", "opacity as technical illiteracy", en "opacity from the mismatch between mathematical optimization and human semantic interpretation" (p. 3)[2]. Voor AlphaFold geldt vooral de derde vorm. Burrell: "When we are seeking to understand a decision made by a machine learning algorithm, the challenge is not only in the inscrutability of any given operation but in the interplay of weights and connections across the model" (p. 10). Dit is precies wat we bij AlphaFold zien: de architectuur is gepubliceerd, de code is open source, maar de wijze waarop 21 miljoen parameters samenwerken om tot een eiwitstructuur te komen, blijft voor menselijke interpretatie grotendeels gesloten. Toch moet Burrell's analyse worden genuanceerd. Zij schrijft haar stuk in 2016, vóór de ontwikkeling van attention mechanisms en interpretability tools die nu standaard zijn. AlphaFold's confidence scores zijn een vorm van wat Burrell "the provision of a rationale" noemt - niet perfect, maar substantieel beter dan pure black-box-systemen.

[2] Burrell (2016) onderscheidt drie bronnen van algoritmische ondoorzichtigheid. Ten eerste is er **opzettelijke geheimhouding** (*intentional corporate or state secrecy*), wanneer bedrijven of overheden de werking van hun algoritmen bewust afschermen om commerciële of strategische redenen. Ten tweede speelt **technische ongeletterdheid** (*technical illiteracy*): de meeste gebruikers, beleidsmakers en zelfs onderzoekers beschikken niet over de wiskundige of computationele expertise om complexe modellen te begrijpen. Ten derde is er een **structurele betekenis-kloof** (*the mismatch between mathematical optimization and human semantic interpretation*): algoritmen optimaliseren numerieke functies, terwijl mensen betekenis toekennen via taal en context. Zelfs met volledige inzage in de code blijft daardoor een deel van de werking principieel moeilijk te vertalen naar menselijke begrippen.

Decontexualisatie en datakwaliteit. De betrouwbaarheid van AI in de natuurwetenschappen staat of valt met de kwaliteit van de datasets. Klimaatmodellen bijvoorbeeld zijn robuuster in regio's waar de meetdichtheid hoog is, maar vertonen grotere onzekerheden in gebieden met beperkte datacoverage (Rolnick et al., 2019). De schijn van universaliteit maskeert zo ongelijkheden in dataverzameling en kan leiden tot overschatting van de precisie van voorspellingen.

Machtsconcentratie en afhankelijkheid. Een ander risico betreft de toegang tot rekenkracht en infrastructuur. Het trainen van geavanceerde AI-modellen vereist investeringen die slechts enkele partijen kunnen opbrengen. Brookings en GovAI (2023) waarschuwen dat hierdoor oligopolistische structuren ontstaan waarin een beperkt aantal bedrijven de epistemische agenda mede bepaalt. Voor natuurwetenschappelijke onderzoekers betekent dit afhankelijkheid: wie geen toegang heeft tot deze infrastructuur, kan slechts gebruikmaken van reeds getrainde modellen, met alle beperkingen en aannames die daarin ingebouwd zijn.

Fabricatie van referenties. Hoewel in de natuurwetenschappen de kern van kennisproductie doorgaans berust op empirische data, is ook hier waakzaamheid geboden voor het gebruik van generatieve modellen bij literatuuronderzoek. Studies tonen aan dat taalmodellen geregeld fictieve of foutieve referenties produceren (Borji, 2023). Wanneer zulke instrumenten onkritisch worden ingezet in secundaire processen, zoals het schrijven van reviews of het opstellen van onderzoeksvoorstellen, kan dit de betrouwbaarheid van de wetenschappelijke communicatie aantasten.

7.4 Uitdagingen

De analyse van kansen en risico's maakt duidelijk dat de natuurwetenschappen een bijzonder geval vormen binnen de bredere context van AI-gestuurde kennisproductie. Enerzijds laat dit domein overtuigend zien hoe AI kan bijdragen aan versnelling, ontdekking en synthese. Anderzijds komen specifieke uitdagingen naar voren die niet louter incidenteel zijn, maar structureel samenhangen met de aard van dit archetype. Drie clusters van uitdagingen verdienen bijzondere aandacht: methodologisch, institutioneel en normatief.

Methodologische uitdagingen: validiteit en verklaarbaarheid. De kernwaarden van de natuurwetenschappen zijn transparantie, reproduceerbaarheid en causaliteit. AI-toepassingen sluiten daar slechts gedeeltelijk bij aan. Resultaten van systemen zoals AlphaFold zijn reproduceerbaar, maar hun verklaarbaarheid blijft beperkt (Jumper et al., 2021). In de astronomie herkennen algoritmen patronen die consistent zijn, maar waarvan de fysische grond niet altijd duidelijk is (Fluke & Jacobs, 2020). Dit roept de vraag op hoe validiteit moet worden gedefinieerd: is een voorspelling die empirisch klopt, maar niet theoretisch begrepen wordt, legitiem wetenschappelijk kennis? Het antwoord op deze vraag bepaalt in hoge mate hoe AI zich in dit domein kan vestigen.

Epistemologische spanning - expertise versus automatisering. Harry Collins en Robert Evans analyseren in "Rethinking Expertise" (2007) wat zij "interactional expertise" noemen: "enough expertise to interact interestingly with participants" zonder zelf practitioner te zijn. Dit concept is hoogst relevant voor AI in de natuurwetenschappen. Zij waarschuwen: "The danger is that we come to treat the computer's correlations as understanding". Dit raakt de kern van het AlphaFold-dilemma. Het systeem heeft geen "understanding" van biochemie in de menselijke zin, maar produceert wel bruikbare voorspellingen. Collins & Evans zouden vragen: is dit contributory expertise (echte bijdrage aan kennis) of slechts een vorm van 'mimicry'? Hun antwoord zou genuanceerd zijn. Zij erkennen dat "expertise is not just about possessing information but about being able to make judgments" (p. 45). AlphaFold maakt geen judgments in de traditionele zin, maar de onderzoekers die het gebruiken wel. Dit suggereert een nieuwe vorm van hybride expertise waarin mens en machine complementair opereren.

Institutionele uitdagingen: infrastructuur en toegang. Een tweede cluster van uitdagingen betreft de toegang tot infrastructuur. De rekenkracht en datasets die nodig zijn voor grootschalige AI-toepassingen zijn vaak in handen van een beperkt aantal bedrijven of onderzoeksconsortia. Brookings en GovAI (2023) wijzen op de oligopolistische structuur die hierdoor ontstaat. Voor onderzoekers betekent dit dat autonomie wordt beperkt: zij kunnen resultaten toetsen of repliceren, maar hebben zelden de middelen om een model van vergelijkbare schaal zelf te trainen. Institutioneel gezien vraagt dit om mechanismen die gelijkwaardige toegang waarborgen, bijvoorbeeld door publieke investeringen in data-infrastructuren of internationale afspraken over datadeling.

Normatieve uitdagingen: foutmarges en legitimiteit. Tot slot zijn er normatieve vraagstukken rond foutmarges en legitimiteit. In klimaatmodellen, bijvoorbeeld, kan een geringe foutmarge grote consequenties hebben voor beleidsbeslissingen (Rolnick et al., 2019). Dit dwingt tot een expliciete normstelling: welke mate van onzekerheid is aanvaardbaar, en wie bepaalt dat? Het probleem wordt nog complexer wanneer AI wordt ingezet om voorspellingen te genereren die verder reiken dan de bestaande empirische basis. De legitimiteit van zulke kennis berust niet alleen op statistische robuustheid, maar ook op de acceptatie van de wetenschappelijke gemeenschap. Dit maakt normatieve afwegingen onvermijdelijk: betrouwbaarheid is niet enkel een technische eigenschap, maar ook een sociaal erkend kwaliteitskenmerk (Collins & Evans, 2007).

7.5 Conclusie

De analyse van de natuurwetenschappen maakt duidelijk hoe kunstmatige intelligentie de kennisproductie in dit archetype beïnvloedt. Om dit scherp te houden, worden de drie deelvragen expliciet herhaald en voorlopig beantwoord.

Deelvraag 1: Hoe verandert AI de cognitieve en methodologische dimensie van kennisproductie?
In de natuurwetenschappen versterkt AI bestaande methodologische kernwaarden zoals reproduceerbaarheid en empirische toetsing. Voorbeelden als AlphaFold (Jumper et al., 2021) en AI-toepassingen in astronomie (Fluke & Jacobs, 2020; Valenti et al., 2021) tonen aan dat voorspellingen herhaalbaar en toetsbaar zijn, maar vaak beperkt verklaarbaar blijven. De methodologische uitdaging ligt daarom niet in een verlies van reproduceerbaarheid, maar in een verschuiving van de status van verklarende naar voorspellende kennis. Dit betekent dat de legitimiteit van wetenschappelijke kennis wordt hergedefinieerd: waar natuurwetenschappen traditioneel streven naar causaal begrip ("waarom gebeurt X?"), accepteert AI-gestuurde kennisproductie vaak instrumentele voorspelbaarheid ("wat gebeurt er als X?"). Accurate voorspelling zonder mechanistisch begrip ondermijnt niet de betrouwbaarheid van resultaten, maar wel de aard van wat als 'wetenschappelijke verklaring' geldt. Het voorlopige antwoord op de eerste deelvraag is dat AI in de natuurwetenschappen de empirische cyclus versnelt en uitbreidt, maar dat de legitimiteit van kennis verschuift van causaal begrip naar voorspellende accuratesse als primaire legitimatiegrond.

Deelvraag 2: Hoe beïnvloedt AI de pedagogische en institutionele dimensie van kennisproductie?
AI maakt natuurwetenschappelijke kennis toegankelijker en institutioneel breder inzetbaar. De verrijking van publieke databanken, zoals de *Protein Data Bank* door AlphaFold, democratiseert de toegang tot kennis en vergroot de onderwijskundige waarde van AI-systemen. Institutioneel zien we echter ook nieuwe afhankelijkheden ontstaan. De toegang tot rekenkracht en hoogwaardige datasets concentreert zich bij een beperkt aantal bedrijven en consortia (Brookings Institution / GovAI, 2023). Dit betekent dat democratisering van kennis voorwaardelijk is: terwijl AI-gegenereerde datasets breed beschikbaar zijn, blijft de capaciteit om nieuwe kennis te produceren beperkt tot actoren met toegang tot grootschalige infrastructuur. Institutioneel verschuift de macht van individuele onderzoekers en kleinere instituten naar grote consortia en tech-bedrijven die de infrastructuur bezitten. Het voorlopige

antwoord op de tweede deelvraag luidt dat AI de pedagogische mogelijkheden versterkt en kennisdeling faciliteert, maar dat deze voordelen gepaard gaan met structurele afhankelijkheden die de autonomie van onderzoekers beperken. Toegankelijkheid van resultaten is niet hetzelfde als democratische participatie in kennisproductie.

Deelvraag 3: Hoe herconfigureert AI machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar? In de natuurwetenschappen manifesteert de machtsdimensie zich vooral in de toegang tot infrastructuur en datasets. Betrouwbaarheid en legitimiteit zijn niet langer uitsluitend methodologische kwesties, maar worden mede bepaald door wie toegang heeft tot data en rekenkracht. Normatief dwingt dit tot expliciete keuzes over foutmarges, legitimiteit van voorspellingen en criteria voor toegankelijke infrastructuur. Dit betekent dat macht verschuift van epistemische naar infrastructurele controle. Wie bepaalt welke datasets verzameld worden, wie toegang heeft tot rekenkracht, en welke vragen computationeel behandelbaar zijn, bepaalt mede de grenzen van wat geweten kan worden. De normatieve uitdaging is niet alleen om criteria voor 'goede AI-wetenschap' te definiëren, maar om de voorwaarden te scheppen waaronder diverse actoren kunnen deelnemen aan kennisproductie.Het voorlopige antwoord op de derde deelvraag is dat macht en normativiteit in dit archetype primair worden vormgegeven door infrastructuurtoegang en institutionele regelingen, en dat legitimiteit pas kan worden gegarandeerd wanneer deze randvoorwaarden expliciet worden geregeld.

Samenvattend. Voor de natuurwetenschappen geldt dat AI kennisproductie diepgaand transformeert, maar op een andere locus dan vaak wordt verondersteld. De kernwaarden van empirische toetsing en reproduceerbaarheid blijven in stand, en worden zelfs versterkt. De kwetsbaarheid ligt eerder in de verschuiving van verklaarbaarheid naar voorspelbaarheid, in de institutionele afhankelijkheid van infrastructuren en in de normatieve keuzes rond toegang en foutmarges. AI-gestuurde kennisproductie in de natuurwetenschappen is daarmee zowel krachtig als conditioneel: de kwaliteit van de kennis is sterk, maar zij staat of valt met de institutionele en normatieve randvoorwaarden die dit domein mogelijk maken.

Literatuuroverzicht hoofdstuk 7.

Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv preprint arXiv:2302.03494.*

Brookings Institution / GovAI. (2023). *Analyses of foundation model concentration and market power.*

Collins, H., & Evans, R. (2007). *Rethinking expertise.* University of Chicago Press.

Fluke, C. J., & Jacobs, C. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(2), e1349.

Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.

Rolnick, D., et al. (2019). Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433.*

Valenti, S., et al. (2021). Machine learning applications for exoplanet detection. *Astronomy & Astrophysics*, 645, A62. https://doi.org/10.1051/0004-6361/202039448

Hoofdstuk 8 – Geschiedenis-onderzoek en AI-gestuurde kennisproductie

8.1 Inleiding

De geschiedwetenschap wijkt fundamenteel af van de natuurwetenschappen. Waar daar de nadruk ligt op reproduceerbaarheid en causale verklaringen, draait historisch onderzoek om authenticiteit en context. Een document, getuigenis of artefact krijgt pas betekenis wanneer het zorgvuldig wordt ingebed in een groter verhaal. De legitimiteit van kennis in dit domein berust daarom op bronkritiek, interpretatie en narratieve verantwoording.

Juist hier roept de inzet van kunstmatige intelligentie (AI) nieuwe vragen op. AI kan grote archieven ontsluiten en patronen zichtbaar maken, maar doet dat door bronnen te dataficeren: teksten, beelden en geluiden worden omgezet in datapoints die losgemaakt zijn van hun oorspronkelijke betekenis. Dit biedt grote kansen, maar ook risico's. De centrale vraag is of de betrouwbaarheid en legitimiteit van historische kennisproductie behouden kan blijven wanneer AI steeds meer taken overneemt die traditioneel door historici werden uitgevoerd.

In dit hoofdstuk worden de kansen, risico's en uitdagingen van AI in de geschiedwetenschap geanalyseerd. Daarbij staan drie concrete toepassingsgebieden centraal: de digitalisering en analyse van Holocaust-archieven, de ontsluiting van koloniale archieven, en het fenomeen van 'deepfake history'. Deze casussen illustreren hoe AI de voorwaarden van historische kennisproductie verandert

8.2 Kansen

Versnelling en schaal. AI kan het werk van historici aanzienlijk versnellen. Handgeschreven documenten zijn eeuwenlang slechts toegankelijk geweest voor specialisten, maar met systemen zoals *Transkribus* kunnen miljoenen pagina's automatisch worden getranscribeerd (Kahle et al., 2017). In het geval van Holocaust-archieven betekent dit dat dagboeken, brieven en getuigenverklaringen die voorheen verspreid lagen over talloze archieven nu doorzoekbaar en vergelijkbaar zijn. Voor historici wordt het mogelijk patronen te ontdekken in vervolging, verzet of overleving die voorheen verborgen bleven. Deze schaalvergroting is niet alleen kwantitatief maar ook kwalitatief. Waar een individuele onderzoeker in zijn carrière misschien enkele duizenden pagina's kan bestuderen, kan AI miljoenen documenten doorzoeken op specifieke namen, plaatsen of gebeurtenissen. Dit opent nieuwe mogelijkheden voor prosopografisch onderzoek en netwerkanalyse, waarbij relaties tussen personen en gebeurtenissen grootschalig in kaart worden gebracht.

Ontdekking en synthese. AI maakt het mogelijk verbanden te leggen tussen verspreide en uiteenlopende bronnen. In koloniale archieven kunnen algoritmen bijvoorbeeld namen of plaatsen herkennen die in verschillende documenten terugkeren, zelfs wanneer ze in andere talen of spellingvarianten voorkomen. Dit kan leiden tot nieuwe historische inzichten over handelsnetwerken, familiebanden of bestuurlijke structuren die anders gefragmenteerd zouden blijven. De kracht van patroonherkenning ligt in het vermogen om grote hoeveelheden materiaal te overzien. Jockers (2013) laat in zijn werk over macroanalyse in de literatuurwetenschap zien hoe grootschalige tekstanalyse

patronen zichtbaar maakt die bij close reading verborgen blijven. Voor de geschiedwetenschap betekent dit dat AI trends kan identificeren in bijvoorbeeld krantenarchieven, pamfletten of correspondentie die de aandacht van individuele onderzoekers zou zijn ontgaan. Hoewel Jockers' werk zich primair richt op literaire teksten, is zijn methodologische argument – dat grootschalige patronen narratieve interpretatie niet kunnen vervangen maar wel kunnen verrijken – direct van toepassing op historisch onderzoek.

Toegankelijkheid en democratisering. Digitalisering en AI-gestuurde ontsluiting kunnen archieven toegankelijk maken voor onderzoekers die er voorheen geen toegang toe hadden. Terras (2011) toont aan hoe zelfs amateur-digitalisering van teksten kan bijdragen aan bredere toegankelijkheid. Voor de geschiedwetenschap is dit cruciaal: niet alleen professionele historici, maar ook lokale gemeenschappen, familieleden van slachtoffers of geïnteresseerde burgers kunnen nu bronnen raadplegen die voorheen alleen fysiek in specifieke archieven te bekijken waren. Deze democratisering heeft ook een institutionele dimensie. Kleinere universiteiten of onderzoekers in het mondiale Zuiden die geen budget hebben voor langdurige archiefbezoeken in Europa of Noord-Amerika, kunnen nu via digitale platformen toegang krijgen tot bronnenmateriaal. Dit verschuift potentieel de machtsbalans in historisch onderzoek en maakt meer diversiteit in perspectieven mogelijk.

8.3 Risico's

Contextverlies en anachronisme. Het grootste risico van AI in de geschiedwetenschap is het verlies van context. Wanneer een document wordt gedigitaliseerd en als datapoint verwerkt, verliest het zijn materiële en institutionele context. Een brief krijgt betekenis door te weten wie hem schreef, aan wie, onder welke omstandigheden, en met welk doel. AI-systemen kunnen wel de tekst transcriberen, maar missen het historische besef om deze contextuele lagen te reconstrueren. Zwierlein (2016) waarschuwt voor het gevaar van anachronisme wanneer hedendaagse categorieën worden geprojecteerd op historisch materiaal. AI-modellen zijn getraind op eigentijdse taal en concepten, en kunnen daarom verbanden leggen die historisch gezien geen betekenis hadden. Bijvoorbeeld, een algoritme dat zoekt naar 'revolutie' in achttiende-eeuwse documenten, herkent het woord wel maar begrijpt niet dat de betekenis toen fundamenteel anders was dan in de eenentwintigste eeuw. Underwood (2019) benadrukt in zijn werk over distant reading dat patroonherkenning alleen zinvol wordt wanneer ze wordt ingebed in interpretatieve kaders. Voor historici betekent dit dat AI-gegenereerde patronen in kranten, pamfletten of correspondentie suggestief kunnen zijn, maar niet zonder grondige historische duiding kunnen worden geaccepteerd. Het risico is dat onderzoekers onder tijdsdruk of door een gebrek aan kritische reflectie deze patronen te snel als feiten behandelen.

Canonvorming en machtsverhoudingen. Niet alle archieven worden gelijk behandeld in digitaliseringsprojecten. Vaak worden eerst de best bewaarde en gemakkelijkst toegankelijke collecties verwerkt – veelal Europese archieven. Putnam (2016) benadrukt dat digitalisering altijd 'schaduwen' werpt: wat wel digitaal beschikbaar is, wordt dominant, wat niet gedigitaliseerd is, raakt uit beeld. In koloniale archieven is dit bijzonder problematisch. De perspectieven van koloniale machthebbers zijn vaak goed gedocumenteerd en bewaard, terwijl lokale stemmen fragmentarisch of helemaal afwezig zijn. Wanneer AI wordt ingezet om deze archieven te ontsluiten, dreigt een canonisering van Europese perspectieven, tenzij bewust wordt geïnvesteerd in het digitaliseren en contextualiseren van bronnen uit voormalige koloniën. Ketelaar (2001) spreekt in dit verband van 'tacit narratives': impliciete verhalen die archieven door hun ordening en beschrijving altijd al meedragen. Met AI wordt dit effect versterkt,

omdat algoritmische keuzes in metadata en zoekstructuren bepalen welke geschiedenissen naar voren komen en welke uit beeld raken.

Fabricatie van referenties en deepfake history. Generatieve AI brengt een nieuw en fundamenteel risico met zich mee: het aanmaken van fictieve bronnen of foutieve citaten (Borji, 2023). Dit speelt niet alleen bij secundaire processen zoals literatuurreviews, maar raakt de kern van historische betrouwbaarheid. Een discipline die haar legitimiteit juist ontleent aan bronverantwoording wordt existentieel bedreigd wanneer bronnen niet meer te verifiëren zijn. Nog ernstiger is het fenomeen van 'deepfake history': hyperrealistische reconstructies van stemmen, gezichten of historische scènes gemaakt met AI. Weliswaar kunnen zulke reconstructies educatieve waarde hebben – bijvoorbeeld om historische figuren 'tot leven te brengen' – maar ze scheppen ook een gevaarlijke ruimte waarin het onderscheid tussen feit en fictie vervaagt. Voor de geschiedwetenschap, die al worstelt met ontkenning van gebeurtenissen als de Holocaust, is dit geen theoretisch risico maar een acute bedreiging voor de maatschappelijke legitimiteit van historische kennis.

8.4 Uitdagingen

Methodologische uitdagingen: authenticiteit en narrativiteit. Historisch onderzoek vraagt meer dan het identificeren van patronen: het vereist het construeren van een narratief dat bronnen betekenisvol met elkaar verbindt. AI kan wel correlaties vinden, maar heeft geen historisch besef of interpretatieve intuïtie. De uitdaging is dat onderzoekers AI-resultaten niet verwarren met kant-en-klare kennis, maar steeds kritisch vertalen naar authentieke, contextueel verantwoorde verhalen. Jockers (2013) en Underwood (2019) laten beide zien dat distant reading patronen blootlegt die literair interessant zijn, maar alleen betekenis krijgen wanneer ze worden ingebed in interpretatieve kaders. Deze methodologische spanning is evenzeer van toepassing op historisch onderzoek. Patronen in krantenarchieven, pamfletten of correspondentie kunnen suggestief zijn, maar historici moeten deze patronen steeds vertalen naar verhalen die recht doen aan de complexiteit en contingentie van het verleden. De uitdaging is om AI in te zetten als hulpmiddel bij patroonherkenning, zonder dat de hermeneutische taak van de historicus – het begrijpen van betekenis in context – wordt uitgehold. Dit vereist methodologische zorgvuldigheid en expliciete verantwoording van de rol die AI speelt in het onderzoeksproces.

Institutionele uitdagingen: representatie en toegang. De digitalisering en ontsluiting van archieven gebeurt niet in een vacuüm. De selectie van welke archieven worden verwerkt, welke metadata worden toegevoegd en wie toegang krijgt tot de resultaten, bepaalt de horizon van historisch onderzoek. Terras (2011) laat zien hoe amateur-digitalisering canonvorming kan beïnvloeden, doordat sommige teksten een digitale tweede leven krijgen en andere verdwijnen. In koloniale archieven geldt dit op grotere schaal. Europese instellingen bepalen vaak welke bronnen worden ontsloten en hoe ze worden beschreven. Ketelaar's (2001) concept van 'tacit narratives' is hier bijzonder relevant: archieven dragen door hun ordening en beschrijving altijd impliciete verhalen mee. Met AI wordt dit effect versterkt: algoritmische keuzes in metadata en zoekstructuren bepalen welke geschiedenissen naar voren komen en welke uit beeld raken. De institutionele uitdaging is om digitalisering en AI-ontsluiting zo te organiseren dat diversiteit van perspectieven behouden blijft. Dit vraagt om internationale samenwerking, waarbij niet alleen Europese en Noord-Amerikaanse instellingen de agenda bepalen, maar ook archieven en onderzoekers uit voormalige koloniën actief betrokken zijn bij het ontwerp van metadataschema's en zoekalgoritmen.

Normatieve uitdagingen: betrouwbaarheid en verantwoordelijkheid. De inzet van AI roept vragen op over verantwoordelijkheid. Als een AI-systeem een Holocaust-getuigenis verkeerd transcribeert, of een koloniale bron met verkeerde metadata categoriseert, wie draagt dan de verantwoordelijkheid voor die fout? Verhoeven (2016) benadrukt dat digitale methoden normatieve verplichtingen scheppen voor onderzoekers en instellingen. Het is niet voldoende om te zeggen dat 'de technologie het deed' – de keuze om AI in te zetten en de wijze waarop dat gebeurt, ligt bij mensen en instituties. Voor de geschiedwetenschap is dit bijzonder gevoelig, omdat fouten directe consequenties kunnen hebben voor historische waarheid en maatschappelijk debat. Een verkeerd getranscribeerde datum in een document over de Holocaust kan worden misbruikt door ontkenners. Een koloniale bron die door verkeerde metadata wordt gekoppeld aan de verkeerde plaats of periode, kan lokale gemeenschappen vervreemden van hun eigen geschiedenis. De normatieve uitdaging is daarom tweeledig: enerzijds moet worden gewaarborgd dat AI-systemen zorgvuldig worden ontworpen en gevalideerd, anderzijds moet duidelijk zijn dat eindverantwoordelijkheid altijd bij menselijke historici ligt. Dit vereist transparantie over het gebruik van AI, documentatie van keuzes en fouten, en openheid over de beperkingen van algoritmische methoden.

8.5 Conclusie

De analyse van de geschiedwetenschap laat zien dat AI dit archetype op fundamentele wijze raakt. Om de bevindingen scherp te formuleren, worden de drie deelvragen expliciet beantwoord voor dit domein.

Deelvraag 1: Hoe verandert AI de cognitieve en methodologische dimensie van kennisproductie?
AI verandert de cognitieve praktijk van historisch onderzoek door versnelling, schaalvergroting en patroonherkenning. Systemen zoals Transkribus maken miljoenen pagina's doorzoekbaar, en algoritmen kunnen verbanden leggen tussen verspreide bronnen. Dit vergroot de horizon van wat historici kunnen onderzoeken en maakt nieuwe vormen van netwerkanalyse en prosopografie mogelijk. Methodologisch ontstaat echter een fundamentele spanning tussen patroonherkenning en interpretatie. Waar AI correlaties ziet, moet de historicus betekenis construeren. De uitdaging is om AI in te zetten zonder dat authenticiteit en context verloren gaan. Jockers (2013) en Underwood (2019) laten zien dat grootschalige analyse alleen zinvol wordt wanneer ze wordt ingebed in hermeneutische interpretatie – een inzicht dat direct van toepassing is op historisch onderzoek.Het voorlopige antwoord luidt dat AI de cognitieve capaciteit van historici vergroot, maar dat de methodologische legitimiteit van kennisproductie afhankelijk blijft van kritische bronkritiek en contextuele verantwoording door de onderzoeker.

Deelvraag 2: Hoe beïnvloedt AI de pedagogische en institutionele dimensie van kennisproductie?
Pedagogisch biedt AI nieuwe mogelijkheden voor het onderwijzen van geschiedenis. Studenten kunnen via digitale platforms grote archieven verkennen en zelf patronen ontdekken die voorheen verborgen bleven. Dit kan leiden tot een actievere en meer onderzoeksgerichte vorm van geschiedenisonderwijs. Institutioneel verschuift de machtsbalans echter op problematische wijze. Terras (2011) en Putnam (2016) laten zien hoe digitalisering canonvorming beïnvloedt: wat gedigitaliseerd is, wordt zichtbaar, wat niet gedigitaliseerd is, raakt uit beeld. In koloniale archieven betekent dit dat Europese perspectieven dominant blijven tenzij actief wordt geïnvesteerd in het digitaliseren van bronnen uit voormalige koloniën De uitdaging is om AI-gestuurde ontsluiting zo te organiseren dat representatie en toegang rechtvaardig zijn. Dit vraagt om internationale samenwerking en expliciete aandacht voor wie de agenda bepaalt in digitaliseringsprojecten. Het voorlopige antwoord is dat AI institutionele vernieuwing en inclusie kan bevorderen, maar dat dit alleen houdbaar is wanneer er actief gestuurd wordt op representatie en evenwichtige toegang.

Deelvraag 3: Hoe herconfigureert AI machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar? Macht manifesteert zich in de keuzes van digitalisering, metadata en toegang. Deze bepalen welke verhalen circuleren en welke verdwijnen. Ketelaar (2001) laat zien hoe archieven door hun ordening altijd impliciete narratieven meedragen. Met AI wordt dit effect versterkt: algoritmische keuzes in zoekstructuren bepalen welke geschiedenissen naar voren komen. Normatief is het essentieel dat transparantie en verantwoordelijkheid centraal staan. Verhoeven (2016) benadrukt dat digitale methoden normatieve verplichtingen scheppen voor onderzoekers en instellingen. Het is niet voldoende om AI in te zetten zonder expliciete verantwoording van keuzes en beperkingen. Het voorlopige antwoord luidt dat AI slechts legitiem kan functioneren in dit domein wanneer de constitutieve rol van macht expliciet wordt erkend en gereguleerd, en wanneer de verantwoordelijkheid voor interpretatie ondubbelzinnig bij de historicus blijft. Dit vereist transparantie over het gebruik van AI, documentatie van keuzes en openheid over beperkingen.

Samenvattend. Voor de geschiedwetenschap biedt AI enorme kansen voor schaalvergroting, ontdekking en toegankelijkheid. Maar de risico's liggen scherp op het terrein van contextverlies, canonvorming en betrouwbaarheid. De uitdagingen draaien om het bewaren van authenticiteit en narrativiteit in een onderzoeksomgeving die steeds meer wordt gestuurd door patronen en algoritmen. AI kan de geschiedwetenschap verrijken, maar de legitimiteit van historische kennisproductie blijft afhankelijk van menselijke interpretatie en kritische verantwoording.

Literatuuroverzicht hoofdstuk 8.

Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv preprint arXiv:2302.03494.*

Jockers, M. (2013). *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press.

Kahle, P., Colutto, S., Hackl, G., & Fiel, S. (2017). Transkribus – A Service Platform for Transcription, Recognition and Retrieval of Historical Documents. *ICDAR 2017 Proceedings*, 19–24.

Ketelaar, E. (2001). Tacit Narratives: The Meanings of Archives. *Archival Science*, 1, 131–141.

Putnam, L**.** (2016). The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast. *American Historical Review*, 121(2), 377–402.

Terras, M. (2011). Digital Curiosities: Resource Creation Via Amateur Digitisation. *Literary and Linguistic Computing*, 26(4), 425–438.

Underwood, T**.** (2019). *Distant Horizons: Digital Evidence and Literary Change.* University of Chicago Press.

Verhoeven, D. (2016). Doing Digital Humanities: An Extended Review. *Digital Scholarship in the Humanities*, 31(1), 1–15.

Wevers, M., & Smits, T. (2020). The politics of the machine: Digital humanities and the automation of text analysis. *Digital Scholarship in the Humanities*, 35(1), 194–214.

Zwierlein, C. (2016). *Anachronism and Antiquarianism: Historical Knowledge and its Uses.* Routledge.

Hoofdstuk 9 –Antropologisch onderzoek en AI-gestuurde kennisproductie

9.1 Inleiding

Na de analyse van de geschiedwetenschap, waarin de focus lag op archieven, canonvorming en de interpretatie van bronnen uit het verleden, verplaatst dit hoofdstuk zich naar een ander archetype van kennisproductie: de antropologie. Anders dan de geschiedenis richt de antropologie zich niet primair op documenten of artefacten, maar op levende gemeenschappen en hun culturele praktijken. Waar het historisch onderzoek afhankelijk is van bronnen die reeds bestaan, is antropologisch onderzoek methodologisch ingebed in veldwerk, participatie en langdurige interactie met mensen in hun sociale context.

Deze verschuiving naar een onderzoekstraditie die relationeel en ervaringsgericht is, maakt dat de inzet van kunstmatige intelligentie (AI) hier andere kansen en risico's genereert. In de geschiedenis lag de kwetsbaarheid vooral in de representatie van bronnen en de canonisering van bepaalde perspectieven. In de antropologie verschuift de kernvraag naar de relatie tussen onderzoeker, gemeenschap en technologie: hoe verandert de kwaliteit en legitimiteit van kennisproductie wanneer AI bemiddelt tussen onderzoekers en de mensen die zij bestuderen?

AI wordt in dit domein al op uiteenlopende manieren ingezet. Digitale etnografie gebruikt AI om patronen in sociale media te analyseren. Spraak- en beeldherkenning ondersteunen het documenteren van bedreigde talen en rituelen. Remote sensing en geografische informatiesystemen (GIS) helpen bij het bestuderen van landschappen en gemeenschappen. Deze toepassingen bieden onmiskenbaar nieuwe mogelijkheden, maar brengen ook spanningen met zich mee. Worden gemeenschappen gereduceerd tot datapoints? Welke machtsrelaties sluipen mee in de manier waarop AI informatie verzamelt en structureert? En in hoeverre kan een discipline die haar legitimiteit ontleent aan nabijheid en empathie zich baseren op technologie die juist afstand creëert?

Dit hoofdstuk onderzoekt de kansen, risico's en uitdagingen die AI in de antropologie met zich meebrengt. De centrale vraag blijft of, en onder welke voorwaarden, AI-gestuurde kennisproductie in dit archetype wetenschappelijk legitiem kan zijn.

9.2 Kansen

Digitale etnografie en schaalvergroting. Antropologisch veldwerk is van oudsher kleinschalig en arbeidsintensief. Een onderzoeker verblijft maanden of jaren in een gemeenschap, maakt aantekeningen, voert interviews en observeert rituelen. Deze methode levert rijke, contextuele kennis op, maar is beperkt in schaal. AI maakt het mogelijk om deze beperkingen deels te overbruggen door *digitale etnografie*: het bestuderen van online gemeenschappen en sociale interacties via platforms als Reddit, WhatsApp of Twitter.Pink et al. (2016) tonen in hun standaardwerk over digitale etnografie aan hoe online communities nieuwe velden van betekenis vormen die even legitiem zijn als fysieke gemeenschappen. AI kan deze digitale ruimtes analyseren op patronen in taalgebruik, netwerkstructuren en collectieve betekenisvorming. Varis en Blommaert (2015) laten zien hoe sociale media nieuwe vormen van conviviality en collectieve structuren creëren, die met behulp van AI-analyse zichtbaar gemaakt kunnen worden. Deze schaalvergroting betekent dat antropologen niet langer beperkt zijn tot

één dorp, wijk of gemeenschap, maar kunnen kijken naar transnationale netwerken, diasporagemeenschappen of online subculturen die voorheen methodologisch ontoegankelijk waren.

Transcriptie en thematische codering. Een groot deel van antropologisch veldwerk bestaat uit het transcriberen van interviews en het coderen van observaties. Dit is tijdrovend en arbeidsintensief werk. AI kan dit proces aanzienlijk versnellen. Spraakherkenningssoftware kan interviews automatisch transcriberen, en natural language processing kan thematische patronen identificeren in grote hoeveelheden tekst. Dit betekent niet dat AI de interpretatieve taak van de antropoloog overneemt, maar wel dat meer tijd beschikbaar komt voor de analyse zelf. Boellstorff et al. (2012) benadrukken in hun werk over etnografie in virtuele werelden dat technologische hulpmiddelen de onderzoeker niet vervangen, maar wel in staat stellen om grotere hoeveelheden data te verwerken zonder de diepgang van interpretatie te verliezen.

Documentatie van bedreigde talen en rituelen. Voor taalkundig-antropologisch onderzoek biedt AI unieke mogelijkheden. Bedreigde talen die nog maar door enkele sprekers worden gesproken, kunnen worden gedocumenteerd met behulp van spraakherkenning en audiovisuele analyse. Kandel et al. (2019) laten zien hoe GIS en AI-landschapsanalyse antropologische perspectieven op ruimte en plaats kunnen verrijken. Dit geldt evenzeer voor rituelen en ceremonieën die weinig worden uitgevoerd en dreigen te verdwijnen. AI kan hierbij helpen om patronen te identificeren in gesproken taal, gebaren of muzikale structuren die voor menselijke onderzoekers moeilijk te onderscheiden zijn. Dit kan bijdragen aan het behoud van cultureel erfgoed en aan het empoweren van gemeenschappen die hun eigen tradities willen documenteren en doorgeven aan volgende generaties.

9.3 Risico's

Verlies van relationele nabijheid. Het grootste risico van AI in de antropologie is het verlies van relationele nabijheid. Antropologisch onderzoek berust traditioneel op *thick description* (Geertz, 1973): gedetailleerde, contextuele beschrijvingen van sociale praktijken die alleen mogelijk zijn door langdurige participatie. Hammersley en Atkinson (2019) benadrukken in hun klassieke werk over etnografische methoden dat de legitimiteit van antropologische kennis voortkomt uit de directe, belichaamde ervaring van de onderzoeker in het veld. AI creëert afstand. Een algoritme kan wel patronen herkennen in sociale media-interacties, maar mist de nuance van lichaamstaal, stiltes, ironie of impliciete betekenissen die alleen in directe interactie waarneembaar zijn. Het risico is dat antropologische kennis verschraalt tot statistieken en correlaties, zonder de rijkdom van contextuele interpretatie die de discipline kenmerkt.

Reductie van gemeenschappen tot datapoints. Een gerelateerd risico is dat gemeenschappen worden gereduceerd tot datasets. Waar traditioneel antropologisch onderzoek mensen als complexe, handelende subjecten benadert, dreigt AI-gestuurde analyse hen te objectiveren als gegevenspunten. Dit raakt de ethische kern van de discipline, die gestoeld is op respect voor de autonomie en waardigheid van de onderzochte gemeenschappen. Adams et al. (2020) waarschuwen in hun werk over AI-gestuurde etnografie voor de ethische implicaties van grootschalige data-extractie. Wanneer posts, foto's en conversaties van mensen zonder hun expliciete toestemming worden geanalyseerd, ontstaat een vorm van surveillance die haaks staat op de reciprociteit en wederkerigheid die antropologen nastreven.

Reproductie van koloniale machtsverhoudingen. Historisch gezien is de antropologie belast met een koloniaal verleden, waarin westerse onderzoekers niet-westerse gemeenschappen bestudeerden vanuit een asymmetrische machtspositie. De discipline heeft dit verleden kritisch gereflecteerd en getracht meer egalitaire onderzoeksmethoden te ontwikkelen. AI dreigt deze machtsverhoudingen opnieuw te reproduceren. Wie bepaalt welke data worden verzameld, hoe ze worden geanalyseerd en wat er met de resultaten gebeurt? Als AI-systemen worden ontwikkeld en gecontroleerd door westerse tech-bedrijven en universiteiten, ontstaat opnieuw een situatie waarin kennis over gemeenschappen wordt geëxtraheerd zonder dat die gemeenschappen zelf eigenaarschap of controle hebben. Dit is niet alleen ethisch problematisch, maar ondermijnt ook de legitimiteit van de kennis die wordt geproduceerd.

Consent en hergebruik van data. In traditioneel veldwerk is informed consent een kernprincipe: mensen moeten weten dat ze worden bestudeerd en daarmee instemmen. In de context van AI wordt dit gecompliceerd. Sociale media-data zijn vaak publiek toegankelijk, maar dat betekent niet automatisch dat gebruikers ermee instemmen dat hun posts worden geanalyseerd voor onderzoeksdoeleinden. Bovendien kunnen datasets die voor één onderzoeksdoel zijn verzameld, hergebruikt worden voor andere doeleinden zonder dat de oorspronkelijke participanten daarvan op de hoogte zijn. Widlok (2017) benadrukt dat wederkerigheid en consent kernwaarden zijn in de antropologie, en dat deze waarden onder druk komen te staan wanneer AI het mogelijk maakt om data op onbeperkte schaal te reproduceren en hergebruiken.

9.4 Uitdagingen

Methodologische uitdagingen: thick description versus patroonherkenning. De methodologische uitdaging voor de antropologie is hoe thick description en patroonherkenning met elkaar kunnen worden verbonden. AI is zeer effectief in het identificeren van patronen in grote datasets, maar antropologische kennis vereist meer: contextuele interpretatie, begrip van impliciete betekenissen en aandacht voor de uniciteit van specifieke situaties. Hammersley en Atkinson (2019) benadrukken dat etnografie draait om het begrijpen van betekenis in context. Dit is fundamenteel anders dan het identificeren van statistische correlaties. De uitdaging is om AI in te zetten als hulpmiddel bij het ordenen en structureren van data, zonder dat de interpretatieve diepgang verloren gaat die antropologische kennis legitiem maakt. Een mogelijke oplossing ligt in hybride methoden, waarin AI wordt gebruikt voor een eerste verkenning van grote datasets, waarna de antropoloog dieper ingaat op specifieke patronen of afwijkingen die door het algoritme zijn geïdentificeerd. Dit vereist echter expliciete methodologische verantwoording en transparantie over de rol die AI speelt in elk stadium van het onderzoek.

Institutionele uitdagingen: toegang en representatie. De toegang tot AI-tools is ongelijk verdeeld. Westerse universiteiten en onderzoeksinstellingen hebben vaak de middelen om geavanceerde software en rekenkracht in te zetten, terwijl onderzoekers in het mondiale Zuiden of uit de gemeenschappen zelf die middelen vaak missen. Dit versterkt bestaande ongelijkheden in wie kennis produceert en wiens perspectieven worden gerepresenteerd. Pink et al. (2016) benadrukken dat digitale etnografie toegankelijk moet zijn voor diverse onderzoekers, maar in de praktijk blijft de controle over technologie geconcentreerd bij enkele actoren. De institutionele uitdaging is om AI-tools en methodieken te democratiseren, zodat ook onderzoekers uit niet-westerse contexten deze kunnen inzetten op voorwaarden die zij zelf bepalen. Daarnaast is er de uitdaging van representatie in trainingsdata. AI-modellen zijn vaak getraind op westerse talen en culturele contexten, wat betekent dat ze minder

effectief zijn in het analyseren van niet-westerse talen, dialecten of culturele praktijken. Dit kan leiden tot systematische vertekening in wat zichtbaar wordt gemaakt en wat onzichtbaar blijft.

Normatieve uitdagingen: wederkerigheid en collectief eigenaarschap. De normatieve uitdaging voor de antropologie is hoe wederkerigheid en collectief eigenaarschap kunnen worden gewaarborgd in een context van AI-gestuurde kennisproductie. Traditioneel veldwerk berust op een relatie van geven en nemen: de onderzoeker krijgt toegang tot de gemeenschap, en in ruil daarvoor deelt hij of zij kennis, biedt praktische hulp of draagt bij aan lokale projecten. AI maakt dit model problematisch. Data kunnen worden geëxtraheerd en geanalyseerd zonder dat er sprake is van directe interactie of wederkerigheid. Widlok (2017) benadrukt dat consent niet alleen moet gelden voor het moment van dataverzameling, maar ook voor de gehele levenscyclus van de data in een AI-ecosysteem. Dit betekent dat gemeenschappen controle moeten behouden over hoe hun data worden gebruikt, wie er toegang toe heeft en hoe de resultaten worden gedeeld. Een mogelijke oplossing ligt in participatory design: gemeenschappen zelf betrekken bij het ontwerp van AI-systemen en onderzoeksprotocollen. Dit vereist echter tijd, middelen en een fundamentele verschuiving in hoe onderzoek wordt georganiseerd – van extractief naar collaboratief.

9.5 Conclusie

De analyse van de antropologie laat zien dat AI dit archetype niet alleen nieuwe mogelijkheden biedt, maar ook direct raakt aan de kernwaarden van de discipline: nabijheid, context en relationele verantwoordelijkheid. Om dit scherp te houden, worden de drie deelvragen expliciet beantwoord.

Deelvraag 1: Hoe verandert AI de cognitieve en methodologische dimensie van kennisproductie? AI verruimt de methodologische horizon van antropologisch onderzoek. Transcriptie- en codingtools versnellen de verwerking van interviews en observaties, digitale etnografie maakt grootschalige analyse van online gemeenschappen mogelijk, en multimodale AI kan audiovisueel materiaal systematisch analyseren (Adams et al., 2020; Varis & Blommaert, 2015). Toch ligt hierin ook de grootste uitdaging: de interpretatieve nabijheid die de antropoloog in veldwerk ontwikkelt, kan niet volledig worden vervangen door algoritmische patronen. Hammersley en Atkinson (2019) benadrukken dat etnografie draait om thick description – gedetailleerde, contextuele interpretatie die alleen mogelijk is door langdurige participatie.

Het voorlopige antwoord luidt dat AI het cognitieve bereik van antropologisch onderzoek vergroot, maar dat de legitimiteit van kennisproductie afhankelijk blijft van hermeneutische interpretatie door de onderzoeker. Patroonherkenning kan een eerste verkenning bieden, maar de betekenis van die patronen moet altijd worden geconstrueerd door de antropoloog in dialoog met de gemeenschap.

Deelvraag 2: Hoe beïnvloedt AI de pedagogische en institutionele dimensie van kennisproductie? Pedagogisch biedt AI nieuwe mogelijkheden voor het onderwijzen van antropologie. Studenten kunnen via digitale platforms kennismaken met diverse gemeenschappen en zelf patronen ontdekken in online interacties. Pink et al. (2016) tonen aan dat digitale etnografie studenten in staat stelt om op eigen tempo veldwerk te doen in virtuele ruimtes. Institutioneel ontstaan echter nieuwe ongelijkheden. De toegang tot AI-tools en rekenkracht is ongelijk verdeeld, wat westerse universiteiten bevoordeelt ten opzichte

van instellingen in het mondiale Zuiden. Bovendien dreigt het risico dat kennis wordt geëxtraheerd uit gemeenschappen zonder dat zij zelf eigenaarschap of controle hebben over de resultaten.

Het voorlopige antwoord is dat AI pedagogisch en institutioneel bijdraagt aan inclusie en toegankelijkheid, maar dat dit alleen houdbaar is wanneer infrastructuren en toegang rechtvaardig worden georganiseerd. Dit vraagt om democratisering van AI-tools en om participatory design waarbij gemeenschappen zelf betrokken zijn bij het onderzoeksproces.

Deelvraag 3: Hoe herconfigureert AI machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar? AI beïnvloedt de machtsrelatie tussen onderzoeker en onderzochte gemeenschap ingrijpend. Waar traditionele antropologie stoelt op wederkerigheid en consent, kan AI data oneindig reproduceren en hergebruiken zonder dat de gemeenschap hier nog controle over heeft. Widlok (2017) benadrukt dat wederkerigheid een kernwaarde is die systematisch moet worden ingebouwd in AI-gestuurde onderzoekspraktijken. Normatief is het essentieel dat consent niet alleen geldt voor het moment van dataverzameling, maar voor de gehele levenscyclus van data. Gemeenschappen moeten controle behouden over hoe hun data worden gebruikt, wie er toegang toe heeft en hoe resultaten worden gedeeld. Dit vereist nieuwe institutionele arrangementen waarin collectief eigenaarschap en participatory design centraal staan.

Het voorlopige antwoord luidt dat AI de machtsverhoudingen in antropologisch onderzoek fundamenteel herconfigureert, en dat een legitieme normatieve ordening alleen mogelijk is wanneer wederkerigheid, consent en collectief eigenaarschap systematisch worden gewaarborgd. Dit is niet alleen een ethische verplichting, maar ook een voorwaarde voor de kwaliteit en betrouwbaarheid van antropologische kennis.

Samenvattend. Voor de antropologie is de kernvraag hoe AI de relatie tussen onderzoeker en gemeenschap beïnvloedt. AI kan het bereik en de efficiëntie van onderzoek vergroten, maar dreigt tegelijk de nabijheid en empathie te ondermijnen die antropologische kennis legitiem maken. De uitdaging is om AI in te zetten als hulpmiddel, zonder dat de relationele basis van de discipline wordt uitgehold. Dit vereist methodologische zorgvuldigheid, institutionele hervormingen en een normatief kader waarin wederkerigheid en collectief eigenaarschap centraal staan.

Literatuurlijst Hoofdstuk 9

Adams, O., Michaud, A., & Cohn, T. (2020). Massively Multilingual Adversarial Speech Recognition. *Proceedings of ACL 2020*, 368–379.
→ Relevantie: illustreert hoe AI spraakherkenning kan ondersteunen bij de documentatie van kleine en bedreigde talen.

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). *Ethnography and Virtual Worlds: A Handbook of Method.* Princeton University Press.
→ Relevantie: fundamenteel voor digitale etnografie; toont hoe online gemeenschappen veldwerkcontexten vormen.

Geertz, C. (1973). *The Interpretation of Cultures.* Basic Books.

Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in Practice* (4e ed.). Routledge.
→ Relevantie: klassiek methodologisch werk over de kern van etnografie; belangrijk contrast met AI-gestuurde methoden.

Kandel, A. W., Klein, A., & Wessel, M. (2019). Remote sensing and the anthropology of landscape. *Journal of Anthropological Research*, 75(2), 178–196.
→ Relevantie: bespreekt hoe GIS en AI-landschapsanalyse antropologische perspectieven op ruimte beïnvloeden.

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2016). *Digital Ethnography: Principles and Practice.* Sage.
→ Relevantie: standaardwerk over digitale etnografie, direct toepasbaar op AI-toepassingen in online veldwerk.

Varis, P., & Blommaert, J. (2015). Conviviality and collectives on social media: Virality, memes, and new social structures. *Multilingual Margins*, 2(1), 31–45.
→ Relevantie: laat zien hoe sociale media nieuwe velden van collectieve betekenis vormen, waarop AI-analyse kan worden toegepast.

Widlok, T. (2017). *Anthropology and the Economy of Sharing.* Routledge.
→ Relevantie: benadrukt wederkerigheid en consent als kernwaarden in antropologie; relevant voor normatieve dimensies van AI-gebruik

Hoofdstuk 10 – Ontwerp-gericht onderzoek en AI-gestuurde kennisproductie

10.1 Inleiding

Waar de natuurwetenschappen streven naar verklaren, de geschiedenis naar begrijpen van het verleden en de antropologie naar het doorgronden van levende gemeenschappen, staat in ontwerpgericht onderzoek het *maken* centraal. Dit archetype produceert geen kennis door observatie of interpretatie alleen, maar door het creëren van artefacten die zowel praktisch functioneren als theoretisch inzicht opleveren. Herbert Simon (1996) karakteriseerde dit proces als "de wetenschap van het kunstmatige": kennis ontstaat door het ontwerpen van systemen die de werkelijkheid veranderen, en niet slechts beschrijven.

Informatica en technische wetenschappen hebben dit paradigma al lang omarmd, maar ontwerpgericht onderzoek speelt ook een groeiende rol in domeinen zoals onderwijs, gezondheidszorg en bestuurskunde. Daar wordt niet alleen onderzocht *hoe* de werkelijkheid is, maar vooral *hoe* zij beter zou kunnen functioneren. De waarde van ontwerpgericht onderzoek ligt dus in de wisselwerking tussen creatie en reflectie: artefacten worden ontworpen, getest, verbeterd en theoretisch verankerd.

De komst van kunstmatige intelligentie transformeert dit archetype op ingrijpende wijze. AI fungeert niet alleen als hulpmiddel bij het simuleren, analyseren of optimaliseren van ontwerpen, maar ook als *co-designer* die zelf oplossingen aandraagt of hele ontwerptrajecten automatiseert. Daarmee verschuift de rol van de onderzoeker van ontwerper naar curator: de mens selecteert en verantwoordt uit de veelheid van AI-gegenereerde mogelijkheden.

Deze verschuiving roept fundamentele vragen op. Wanneer een artefact deels door AI is ontworpen, wie is dan de auteur? Hoe kan de kwaliteit van een ontwerp worden beoordeeld als het ontwerpproces niet volledig transparant is? En onder welke voorwaarden kan een AI-gegenereerd ontwerp wetenschappelijk legitiem worden genoemd? Dit hoofdstuk onderzoekt de kansen, risico's en uitdagingen die AI in ontwerpgericht onderzoek met zich meebrengt.

10.2 Kansen

AI als co-designer en versneller. Een van de meest belovende toepassingen van AI in ontwerpgericht onderzoek is de rol van co-designer. Generatieve AI-systemen kunnen talloze ontwerpvarianten produceren op basis van gespecificeerde parameters. In architectuur kunnen algoritmen duizenden plattegronden genereren die voldoen aan bepaalde eisen (ruimte-efficiëntie, lichtinval, bereikbaarheid). In productontwerp kunnen AI-systemen vormen optimaliseren voor sterkte, gewicht of esthetiek. Dit versnelt het ontwerpproces aanzienlijk. Waar een ontwerper voorheen weken of maanden nodig had om verschillende opties te verkennen, kan AI in uren of dagen een breed spectrum aan mogelijkheden presenteren. Simon (1996) benadrukte al dat ontwerpen een proces van *satisficing* is – het zoeken naar oplossingen die 'goed genoeg' zijn gegeven de beperkingen. AI vergroot de ruimte waarin gezocht kan

worden en maakt het mogelijk om oplossingen te vinden die een menselijke ontwerper mogelijk over het hoofd zou hebben gezien.

Simulatie en prototyping. AI maakt geavanceerde simulatie mogelijk. Voordat een fysiek prototype wordt gebouwd, kan een digitale versie worden getest onder verschillende omstandigheden. In technische wetenschappen kunnen AI-gestuurde simulaties het gedrag van materialen, constructies of circuits voorspellen. In sociale wetenschappen kunnen agent-based modellen simuleren hoe mensen reageren op nieuwe beleidsinterventies of stedelijke inrichtingen. Zimmerman et al. (2007) benadrukken in hun werk over design research in HCI (Human-Computer Interaction) dat prototyping essentieel is voor het ontwikkelen en testen van ontwerpkennis. AI maakt het mogelijk om dit proces te intensiveren: meer varianten, snellere iteraties en gedetailleerdere feedback. Dit verhoogt de kwaliteit van het uiteindelijke ontwerp en vermindert het risico van kostbare fouten.

Optimalisatie en innovatie. AI kan bestaande ontwerpen optimaliseren door parameters te variëren en te evalueren op prestatie. In productontwerp kan dit leiden tot vormen die efficiënter, lichter of sterker zijn dan wat een menselijke ontwerper zou bedenken. In stedenbouw kunnen algoritmen verkeersstromen optimaliseren of energieverbruik minimaliseren. Belangrijker nog is dat AI kan bijdragen aan innovatie door onconventionele oplossingen te genereren. Schön (1983) beschrijft in zijn klassieke werk over de *reflective practitioner* hoe ontwerpers leren door te experimenteren en te reflecteren op onverwachte uitkomsten. AI kan dit proces verrijken door oplossingen voor te stellen die buiten de conventionele denkpatronen van de ontwerper liggen, wat kan leiden tot doorbraken in ontwerp en engineering.

10.3 Risico's

Black-box ontwerpen en verlies van transparantie. Het grootste risico van AI in ontwerpgericht onderzoek is het ontstaan van black-box ontwerpen: artefacten waarvan onduidelijk is hoe ze tot stand zijn gekomen. Wanneer een generatief algoritme een vorm produceert die aan alle gestelde eisen voldoet, maar de logica achter die vorm niet uitlegbaar is, ontstaat een fundamenteel probleem voor de wetenschappelijke legitimiteit van het ontwerp. Ontwerpgericht onderzoek vereist dat ontwerpkeuzes kunnen worden verantwoord. Hevner et al. (2004) stellen in hun invloedrijke artikel over design science dat artefacten niet alleen moeten werken, maar ook moeten bijdragen aan theoretisch begrip. Dit vereist transparantie over het ontwerpproces: waarom is voor deze vorm gekozen en niet voor een andere? Welke afwegingen zijn gemaakt? Bij AI-gegenereerde ontwerpen is deze transparantie vaak afwezig.

Verlies van ambacht en intuïtie. Een gerelateerd risico is dat de rol van menselijke intuïtie en ambacht wordt uitgehold. Norman (2013) benadrukt in zijn klassieke werk over gebruiksvriendelijk ontwerp dat goede ontwerpen vaak voortkomen uit jarenlange ervaring en een ontwikkeld gevoel voor wat werkt. Wanneer AI veel van het ontwerpwerk overneemt, dreigt deze expertise te verschralen. Dit is niet alleen een verlies aan vakmanschap, maar ook aan kennisoverdracht. Studenten die leren ontwerpen met AI als co-designer, ontwikkelen mogelijk niet dezelfde diepgaande begrip van ontwerpprincipes als studenten die alles zelf moeten uitwerken. De pedagogische uitdaging is om AI in te zetten zonder dat de vorming van toekomstige ontwerpers wordt ondermijnd.

Eigenaarschap en auteursrechten. Wanneer een artefact deels door AI is ontworpen, ontstaat de vraag wie de auteur is. Juridisch en institutioneel is dit problematisch. Kan een wetenschapper een patent claimen op een ontwerp dat grotendeels door een algoritme is gegenereerd? Wie is verantwoordelijk als een AI-gegenereerd ontwerp faalt of schade veroorzaakt Sanders en Stappers (2008) benadrukken in hun werk over co-creation dat ontwerpprocessen steeds vaker collaboratief zijn, waarbij meerdere actoren bijdragen. AI voegt hier een nieuwe laag aan toe: de technologie is geen passief hulpmiddel maar een

actieve bijdrager. Dit vereist nieuwe juridische en institutionele kaders voor het toekennen van eigenaarschap en verantwoordelijkheid.

Homogenisering en gebrek aan diversiteit. Een verder risico is dat AI-gegenereerde ontwerpen homogeen worden. Algoritmen zijn getraind op bestaande ontwerpen en kunnen daarom de neiging hebben om binnen bekende patronen te blijven. Dit kan leiden tot een gebrek aan diversiteit en culturele specificiteit in ontwerpen. Rittel en Webber (1973) introduceerden het concept van *wicked problems*: complexe ontwerpvraagstukken zonder eenduidige oplossing, waarbij de kwaliteit van het antwoord afhangt van de waarden en perspectieven die worden gehanteerd. AI-systemen zijn vaak niet in staat om met deze waardenpluraliteit om te gaan, en kunnen daarom oplossingen genereren die technisch optimaal lijken maar cultureel of sociaal ongewenst zijn.

10.4 Uitdagingen

Methodologische uitdagingen: validiteit van ontwerpkennis. De methodologische uitdaging voor ontwerpgericht onderzoek is hoe de validiteit van AI-gegenereerde ontwerpen kan worden gewaarborgd. Hevner et al. (2004) formuleren zeven richtlijnen voor design science research, waaronder het vereiste dat artefacten rigoureus moeten worden geëvalueerd. Maar hoe evalueer je een ontwerp waarvan het ontwerpproces niet transparant is?**.** Een mogelijke oplossing ligt in 'adversarial testing': het systematisch testen van AI-gegenereerde ontwerpen tegen kritische criteria en het expliciet documenteren van zwaktes en beperkingen. Dit vereist echter tijd en middelen, en botst met de snelheid waarmee AI ontwerpen kan produceren. De spanning tussen efficiëntie en zorgvuldigheid is hier acuut. Schön (1983) benadrukt dat reflectie essentieel is voor het leren van ontwerpers. De uitdaging is om deze 'reflective practice' te integreren in AI-gestuurde ontwerpprocessen, zodat onderzoekers niet alleen ontwerpen produceren maar ook leren van het proces en bijdragen aan theoretische kennis.

Institutionele uitdagingen: publicatie en erkenning. Institutioneel ontstaat de vraag hoe AI-gegenereerde ontwerpen moeten worden gepubliceerd en erkend. Traditionele wetenschappelijke publicaties vereisen dat methodologie en ontwerpkeuzes worden verantwoord. Bij AI-ontwerpen is dit vaak problematisch: hoe beschrijf je een proces dat deels algoritmisch is en deels onuitlegbaar? Bovendien rijst de vraag of bestaande peer review-mechanismen geschikt zijn voor het beoordelen van AI-gegenereerd werk. Reviewers moeten niet alleen het artefact zelf beoordelen, maar ook het proces en de rol die AI daarin heeft gespeeld. Dit vereist nieuwe competenties en mogelijk nieuwe publicatieformaten waarin procesverantwoording centraal staat. Sanders en Stappers (2008) benadrukken dat co-creatie nieuwe vormen van documentatie vereist. Hetzelfde geldt voor AI als co-designer: de output is niet voldoende, het proces moet transparant gedocumenteerd worden zodat anderen kunnen beoordelen hoe het ontwerp tot stand is gekomen en welke rol AI daarin heeft gehad.

Normatieve uitdagingen: ethiek en waarden in ontwerp. De normatieve uitdaging is hoe waarden en ethiek expliciet kunnen worden ingebouwd in AI-gestuurde ontwerpprocessen. Norman (2013) benadrukt dat ontwerpen altijd normatief is: elk ontwerp bevat impliciete aannames over hoe gebruikers zich zouden moeten gedragen en welke doelen waardevol zijn. Wanneer AI ontwerpen genereert, worden deze normatieve keuzes vaak impliciet gemaakt door de parameters die de ontwerper specificeert of door de trainingsdata waarop het algoritme is getraind. Dit kan leiden tot ontwerpen die technisch optimaal zijn maar ethisch problematisch – bijvoorbeeld systemen die efficiënt zijn maar privacy schenden, of producten die goedkoop zijn maar ecologisch schadelijk. Rittel en Webber (1973) benadrukken dat wicked problems vragen om participatief ontwerp waarbij diverse stakeholders worden betrokken. De uitdaging is om deze participatie te integreren in AI-gestuurde processen, zodat waarden en ethiek expliciet worden gearticuleerd en ingebouwd in het ontwerp.

10.5 Conclusie

De analyse van ontwerpgericht onderzoek laat zien dat AI in dit archetype zowel enorme kansen biedt als fundamentele spanningen introduceert. Om de bevindingen scherp te formuleren, worden de drie deelvragen expliciet beantwoord voor dit domein.

Deelvraag 1: Hoe verandert AI de cognitieve en methodologische dimensie van kennisproductie? AI verandert de cognitieve praktijk van ontwerpgericht onderzoek door de rol van de ontwerper te verschuiven van maker naar curator. Waar voorheen de ontwerper zelf vormen bedacht en uitwerkte, genereert AI nu talloze varianten waaruit gekozen moet worden. Simon (1996) beschreef ontwerpen als een proces van satisficing; AI vergroot de zoekruimte maar introduceert tegelijk het risico van black-box oplossingen. Methodologisch ontstaat de uitdaging om transparantie en verantwoording te waarborgen. Hevner et al. (2004) benadrukken dat design science artefacten rigoureus moeten worden geëvalueerd en moeten bijdragen aan theoretisch begrip. Bij AI-gegenereerde ontwerpen is dit alleen mogelijk wanneer het ontwerpproces expliciet wordt gedocumenteerd en de rol van AI helder wordt gearticuleerd. Het voorlopige antwoord luidt dat AI de cognitieve capaciteit van ontwerpers vergroot door versnelling, optimalisatie en innovatie mogelijk te maken, maar dat de methodologische legitimiteit afhankelijk blijft van transparantie over het ontwerpproces en reflectieve verantwoording door de onderzoeker.

Deelvraag 2: Hoe beïnvloedt AI de pedagogische en institutionele dimensie van kennisproductie? Pedagogisch biedt AI nieuwe mogelijkheden voor het onderwijzen van ontwerpen. Studenten kunnen experimenteren met geavanceerde tools en snel itereren tussen idee en prototype. Schön (1983) benadrukt dat ontwerpers leren door reflectie op hun eigen handelen; AI kan dit proces versnellen door directe feedback te geven op ontwerpkeuzes. Toch dreigt het risico dat studenten afhankelijk worden van AI zonder de onderliggende principes te begrijpen. Norman (2013) waarschuwt dat goed ontwerp voortvloeit uit ervaring en intuïtie – eigenschappen die alleen ontwikkeld kunnen worden door zelf te maken en te falen. Institutioneel ontstaat de vraag hoe AI-gegenereerd werk moet worden gepubliceerd en erkend. Bestaande publicatieformaten zijn niet geschikt voor het documenteren van hybride ontwerpprocessen waarin mens en machine samenwerken. Sanders en Stappers (2008) benadrukken dat nieuwe vormen van documentatie nodig zijn waarin procesverantwoording centraal staat. Het voorlopige antwoord is dat AI pedagogisch waardevol kan zijn mits de vorming van ontwerpers niet wordt uitgehold, en dat institutionele hervormingen nodig zijn om hybride auteurschap te erkennen en te faciliteren.

Deelvraag 3: Hoe herconfigureert AI machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar? Macht manifesteert zich in de controle over AI-tools en trainingsdata. Grote tech-bedrijven en goed gefinancierde universiteiten hebben toegang tot geavanceerde generatieve systemen, terwijl kleinere instellingen of ontwerpers in het mondiale Zuiden deze middelen vaak missen. Dit versterkt bestaande ongelijkheden in wie ontwerpen kan maken en wiens perspectieven worden gerepresenteerd. Normatief is het essentieel dat waarden en ethiek expliciet worden ingebouwd in AI-gestuurde ontwerpprocessen. Rittel en Webber (1973) benadrukken dat wicked problems vragen om participatief ontwerp waarbij diverse stakeholders worden betrokken. Dit wordt nog urgenter wanneer AI ontwerpen genereert: zonder expliciete articulatie van waarden ontstaan artefacten die technisch

optimaal lijken maar ethisch of cultureel problematisch zijn. Het voorlopige antwoord luidt dat AI de machtsverhoudingen in ontwerpgericht onderzoek herconfigureert door toegang tot middelen te concentreren, en dat een legitieme normatieve ordening vereist dat transparantie, participatie en expliciete waardenarticulatie centraal staan in het ontwerpproces.

Samenvattend. Voor ontwerpgericht onderzoek biedt AI enorme kansen als co-designer en versneller, maar introduceert het ook fundamentele spanningen rond transparantie, auteurschap en ethiek. De uitdaging is om AI in te zetten zonder dat de reflective practice en ambachtelijke expertise die het archetype kenmerken worden uitgehold. Dit vereist methodologische zorgvuldigheid (transparante documentatie), institutionele hervormingen (nieuwe publicatieformaten) en normatieve explicitering (waarden inbouwen in het ontwerpproces). AI kan ontwerpgericht onderzoek verrijken, maar alleen wanneer de mens de regisseur blijft en verantwoordelijkheid neemt voor de keuzes die worden gemaakt. De curator-rol is hier cruciaal: niet blind vertrouwen op AI-output, maar kritisch selecteren, evalueren en verantwoorden.

Literatuurlijst Hoofdstuk 10

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.

Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition.* Basic Books.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169.

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5–18.

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action.* Basic Books.

Simon, H. A. **(**1996). *The Sciences of the Artificial* (3rd ed.). MIT Press.

Zimmerman, J., Forlizzi, J., & Evenson, **S.** (2007). Research through design as a method for interaction design research in HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 493–502.

Hoofdstuk 11 – Het meta-domein van AI & informatica en AI-gestuurde kennis-productie.

11.1 Inleiding

In de voorgaande hoofdstukken is de inzet van kunstmatige intelligentie (AI) steeds onderzocht binnen afzonderlijke archetypen van wetenschap: natuurwetenschappen, geschiedenis, antropologie en ontwerpgericht onderzoek. Telkens bleek dat AI de voorwaarden van kennisproductie veranderde, maar de analyse vertrok steeds vanuit bestaande disciplines. Er is echter ook een ander perspectief mogelijk: AI als onderzoeksdomein op zichzelf.

Informatica en AI hebben zich ontwikkeld tot zelfstandige kennisgebieden met eigen methoden, instituten en publicatieculturen. Tegelijkertijd is AI meer dan slechts een archetype naast de andere: het is ook de infrastructuur waarlangs kennisproductie in andere domeinen verloopt. Dit dubbele karakter maakt AI tot een *meta-domein*: het onderzoekt de werkelijkheid én zichzelf. Recente ontwikkelingen zoals *automated machine learning* (AutoML), waarin algoritmen nieuwe algoritmen ontwikkelen, of AI die chiparchitecturen optimaliseert voor toekomstige AI-modellen (Mirhoseini et al., 2021), laten zien hoe reflexief dit domein is geworden.

Dit hoofdstuk analyseert hoe AI-gestuurde kennisproductie eruitziet wanneer AI zowel instrument als object is. Deze reflexiviteit maakt het meta-domein tot een kritische testcase: hier worden de spanningen die in andere archetypen impliciet bleven, expliciet en constitutief. Als AI zichzelf kan onderzoeken en valideren, wat betekent dat dan voor de legitimiteit van kennis.

11.2 Kansen: het utopische horizonbeeld

AutoML en zelfverbeterende systemen. Een van de meest spectaculaire ontwikkelingen in AI is de opkomst van AutoML: systemen die automatisch nieuwe machine learning-modellen ontwerpen en optimaliseren (He et al., 2021). Dit vertegenwoordigt een fundamentele verschuiving: AI wordt niet alleen gebruikt om problemen op te lossen, maar om zichzelf te verbeteren. Vilalta en Drissi (2002) introduceerden het concept van *meta-learning*, waarbij algoritmen leren van hun eigen leerprocessen. Deze zelfverbetering kan exponentiële vooruitgang mogelijk maken. Waar voorheen menselijke onderzoekers maanden of jaren nodig hadden om een nieuw model te ontwikkelen, kan AutoML in dagen of weken duizenden architecturen testen en optimaliseren. Dit opent de mogelijkheid van een positieve feedback-loop: betere AI-systemen kunnen nog betere AI-systemen ontwerpen.

Open platforms en democratisering. Het meta-domein kent ook een krachtige beweging naar openheid. Platforms zoals arXiv maken het mogelijk om onderzoek te publiceren zonder de lange doorlooptijd van traditionele peer review. Open-source bibliotheken zoals TensorFlow en PyTorch stellen onderzoekers wereldwijd in staat om state-of-the-art modellen te bouwen zonder afhankelijk te zijn van proprietaire software. Wikipedia functioneert als een bijzonder voorbeeld van een alternatief validatiemodel. Waar traditionele wetenschap steunt op hiërarchische peer review, werkt Wikipedia via horizontale, collectieve correctie. Fouten worden snel geïdentificeerd en gecorrigeerd door een gemeenschap van vrijwilligers. Dit model is niet perfect, maar toont aan dat legitimiteit ook kan voortkomen uit gemeenschappen, en niet uitsluitend uit formele instituties. Deze openheid kan kennisproductie democratiseren. Onderzoekers in het mondiale Zuiden, kleine universiteiten of zelfs

individuele enthousiastelingen krijgen toegang tot tools en kennis die voorheen voorbehouden waren aan een elite van goed gefinancierde instellingen.

Nieuwe pedagogiek: leren met en over AI. Het meta-domein maakt nieuwe vormen van onderwijs mogelijk. Studenten kunnen AI gebruiken om AI te leren: een taalmodel als tutor, een code-generator als assistent. Dit kan leiden tot adaptief onderwijs waarbij de leerstof en het tempo worden aangepast aan de individuele student. Bovendien kan AI zelf onderwerp van studie worden op een manier die voorheen niet mogelijk was. Studenten kunnen experimenteren met trainingsprocessen, architecturen en datasets, en direct zien hoe hun keuzes de prestaties beïnvloeden. Dit maakt abstract theoretisch onderwijs concreet en interactief.

11.3 Risico's: de dystopische horizon

Circulaire validatie en interne benchmarks. Het grootste risico van het meta-domein is circulaire validatie: AI wordt getest op taken die door AI zijn vormgegeven. Benchmarks zoals ImageNet of GLUE zijn ontworpen om de prestaties van modellen te meten, maar worden inmiddels mede geoptimaliseerd door de modellen zelf (Raji et al., 2021). Dit creëert een vorm van zelfbevestiging waarbij vooruitgang wordt gedefinieerd door criteria die door dezelfde technologie worden bepaald. De vraag is of dit nog externe geldigheid heeft. Een model dat uitstekend scoort op ImageNet, presteert niet automatisch goed in reële toepassingen. De gap tussen benchmark-prestaties en praktische bruikbaarheid kan groot zijn, maar wordt gemaskeerd door de focus op interne validatie.

Machtsconcentratie en oligopolie. Hoewel er open platforms zijn, blijft de ontwikkeling van state-of-the-art AI sterk geconcentreerd bij enkele grote actoren: Google DeepMind, OpenAI, Meta, Anthropic. Deze concentratie is niet toevallig maar structureel: de ontwikkeling van foundation models vereist enorme investeringen in rekenkracht, data en talent die alleen grote bedrijven kunnen opbrengen. Bender et al. (2021) waarschuwen in hun invloedrijke paper "Stochastic Parrots" voor de gevaren van deze machtsconcentratie. Wie controle heeft over de infrastructuur, bepaalt welke onderzoeksvragen worden gesteld, welke datasets worden gebruikt en welke normen als 'neutraal' worden gepresenteerd. Dit is niet alleen een economisch probleem maar ook een epistemisch probleem: kennis wordt geproduceerd door een kleine elite die de agenda bepaalt.

Black-box in extremis. In andere domeinen is het black-box-karakter van AI problematisch; in het meta-domein wordt het existentieel. Deep learning-modellen produceren resultaten die zelfs door hun ontwerpers niet meer volledig te verklaren zijn. Castelvecchi (2016) vraagt zich af of we de black box van AI ooit kunnen openen, of dat ondoorzichtigheid een fundamenteel kenmerk is van deze technologie. Interpretability blijft achter bij prestaties: zolang een model beter scoort op benchmarks, wordt het beschouwd als vooruitgang, ook wanneer de interne werking grotendeels onbekend is. Dit roept de fundamentele vraag op of kennisproductie die niet kan worden verklaard, maar wel presteert, nog legitiem is als wetenschap.

Fragmentatie van legitimiteit. Het meta-domein kent geen uniforme validatiestandaard. ArXiv-preprints worden anders beoordeeld dan peer-reviewed artikelen. Open-source projecten hebben andere kwaliteitscriteria dan industriële labs. Wikipedia hanteert collectieve correctie, wetenschappelijke tijdschriften hiërarchische review. Deze pluraliteit kan vruchtbaar zijn, maar dreigt ook tot fragmentatie te leiden. Wat als verschillende validatiepraktijken tot tegenstrijdige conclusies komen? Wie bepaalt dan wat geldig is? In traditionele wetenschap was er consensus over criteria; in het meta-domein is die consensus afwezig.

11.4 Uitdagingen

Methodologische uitdagingen: criteria herdefiniëren. De methodologische uitdaging is om criteria te ontwikkelen die circulaire validatie doorbreken en externe toetsbaarheid garanderen. Het is niet voldoende dat een model goed scoort op interne benchmarks; het moet ook robuust zijn in uiteenlopende reële contexten. Raji et al. (2021) pleiten voor transparante en reproduceerbare AI-evaluatie waarbij niet alleen prestaties maar ook beperkingen, biases en failure modes worden gedocumenteerd. Dit vereist een cultuurverandering: niet alleen succesverhalen publiceren maar ook systematische analyse van waar en waarom modellen falen.

Institutionele uitdagingen: balans tussen open en gesloten. Open platforms zoals arXiv en Wikipedia tonen dat collectieve validatie en kennisdeling werken, maar blijven kwetsbaar tegenover gesloten industriële systemen die cruciale infrastructuur bezitten. De institutionele uitdaging is het vinden van een balans: hoe kan open validatie gecombineerd worden met robuuste, gecontroleerde standaarden die machtsconcentratie tegengaan? Een mogelijke richting ligt in publiek-private partnerschappen waarbij toegang tot rekenkracht en data wordt gedemocratiseerd, maar kwaliteitsnormen collectief worden bepaald. Dit vereist echter politieke wil en institutionele hervormingen die verder gaan dan wat individuele onderzoekers kunnen realiseren.

Normatieve uitdagingen: verantwoordelijkheid in een recursieve ecosysteem. Wanneer AI-modellen andere AI-modellen ontwerpen en valideren, vervaagt de vraag wie verantwoordelijk is voor de uitkomst. Voor AI-gestuurde kennisproductie betekent dit dat verantwoordelijkheid expliciet moet worden herijkt: niet alleen bij onderzoekers, maar ook bij ontwikkelaars, instituten en gemeenschappen die de systemen vormgeven. Birhane et al. (2022) laten in hun analyse van waarden in machine learning-onderzoek zien dat normatieve keuzes vaak impliciet blijven. De uitdaging is om deze keuzes expliciet te maken en inzichtelijk te documenteren, zodat anderen kunnen beoordelen of de waarden die in systemen zijn ingebouwd, acceptabel zijn.

11.5 Relativering: Tegenbeelden en alternatieven.

Het beeld van het meta-domein hoeft niet uitsluitend dystopisch te zijn. Er zijn krachtige tegenbeelden die laten zien dat alternatieve vormen van kennisproductie mogelijk zijn:

Wikipedia is hier het meest sprekende voorbeeld. Het toont aan dat horizontale, collectieve validatie kan werken en dat legitimiteit niet alleen van hiërarchische instituties hoeft te komen. Miljoenen artikelen worden onderhouden door vrijwilligers die fouten corrigeren, bronnen toevoegen en discussiëren over controversiële kwesties. Het systeem is niet perfect – er zijn biases in wat wordt gedocumenteerd en wie bijdraagt – maar het functioneert als een levend bewijs dat kennisproductie kan worden georganiseerd zonder centrale controle.

Open-source gemeenschappen zoals die rond TensorFlow, PyTorch en Hugging Face laten zien dat collectieve bijdragen tot hoogwaardige software kunnen leiden. Code wordt openbaar gedeeld, getest en verbeterd door duizenden ontwikkelaars wereldwijd. Kwaliteit ontstaat niet door formele certificering maar door gebruik, feedback en iteratie.

ArXiv democratiseert de toegang tot wetenschappelijke kennis. Onderzoekers kunnen hun werk direct delen zonder te wachten op langdurige peer review. Dit versnelt kenniscirculatie en maakt wetenschap toegankelijker. Ja, er zijn risico's van ongetoetste claims, maar de transparantie van het systeem maakt dat lezers zelf kunnen beoordelen wat betrouwbaar is.

Deze tegenbeelden maken duidelijk dat de toekomst van AI-gestuurde kennisproductie niet gedetermineerd is. Utopie en distopie staan beide open; de richting hangt af van keuzes die we nu maken.

11.6 Conclusie

Het meta-domein van AI en informatica onderscheidt zich doordat AI hier zowel instrument als object van kennisproductie is. Dit maakt de legitimiteit van AI-gestuurde kennisproductie precair, maar ook bijzonder leerzaam. De drie deelvragen kunnen voor dit domein voorlopig als volgt worden beantwoord.

Deelvraag 1: Hoe verandert AI de cognitieve en methodologische dimensie van kennisproductie? AI maakt kennisproductie reflexief: modellen ontwikkelen, testen en verbeteren elkaar binnen kaders die zij zelf mee hebben bepaald. Dit leidt tot **circulaire validatie**, waarbij vooruitgang wordt gedefinieerd door benchmarks die door dezelfde technologie worden geoptimaliseerd (Raji et al., 2021). Cognitief biedt dit snelheid en schaal, maar methodologisch roept het vragen op over externe toetsbaarheid. Een legitieme kennisproductie vergt daarom nieuwe criteria, waarin prestaties niet langer voldoende zijn zonder *uitlegbaarheid* en *robuustheid* buiten de eigen kaders. Castelvecchi (2016) vraagt of we de black box ooit kunnen openen; het voorlopige antwoord is dat we in elk geval moeten streven naar transparantie over beperkingen en failure modes. Het voorlopige antwoord luidt dat AI de cognitieve praktijk fundamenteel hertekent door reflexiviteit, maar dat methodologische legitimiteit alleen kan worden gewaarborgd wanneer circulaire validatie wordt doorbroken door externe toetsing en transparante documentatie.

Deelvraag 2: Hoe beïnvloedt AI de pedagogische en institutionele dimensie van kennisproductie?. Pedagogisch vernieuwt AI het onderwijs radicaal. Studenten leren steeds vaker AI via AI zelf: een taalmodel als tutor, een code-generator als assistent. Dit kan leren adaptiever maken, maar vergroot ook de kans dat kritische distantie verloren gaat. He et al. (2021) laten zien hoe AutoML het mogelijk maakt om zonder diepgaande kennis modellen te bouwen; de vraag is of dit empowerment is of intellectuele afhankelijkheid. Institutioneel ontstaat er een hybride veld: open platforms zoals arXiv en open-source communities bevorderen snelheid en toegankelijkheid, terwijl industriële labs infrastructuur en data monopoliseren. Wikipedia biedt een interessant tegenvoorbeeld waar validatie plaatsvindt via horizontaal, collectief correctiemechanisme. Het laat zien dat legitimiteit ook kan voortkomen uit gemeenschappen, en niet uitsluitend uit hiërarchische instituties of private actoren. Het voorlopige antwoord is dat AI pedagogisch kansrijk is mits kritische vorming niet wordt uitgehold, en dat institutionele hervormingen nodig zijn om de balans te vinden tussen open en gesloten systemen, tussen snelheid en zorgvuldigheid, tussen democratisering en kwaliteitswaarborging.

Deelvraag 3: Hoe herconfigureert AI machtsverhoudingen en welke normatieve ordening is wenselijk en haalbaar? De machtsstructuur in dit domein is sterk geconcentreerd. Toegang tot data, rekenkracht en talent ligt grotendeels bij een kleine groep actoren, die tegelijk de normen van transparantie en fairness formuleren (Bender et al., 2021). Daarmee ontstaat een recursieve machtsverhouding waarin middelen en regels samenvallen. Birhane et al. (2022) tonen aan dat waarden in machine learning-onderzoek vaak impliciet blijven. Een wenselijke normatieve ordening vraagt om mechanismen die macht expliciet decentraliseren en verantwoording afdwingbaar maken. Dit kan worden geoperationaliseerd via:

- *Transparante interpretatie:* AI-resultaten moeten vergezeld gaan van expliciete verantwoording van keuzes en beperkingen.

- *Reflexieve positionering:* Onderzoekers moeten hun eigen positie in het netwerk van mens en machine expliciteren, in lijn met auto-etnografische en reflexieve methodologie.
- *Netwerktracering*: In de geest van Actor-Network Theory moeten alle actoren (datasets, algoritmen, infrastructuren, ontwikkelaars, gebruikers) zichtbaar worden gemaakt die bijdragen aan de kennisproductie.

Het voorlopige antwoord luidt dat AI de machtsverhoudingen in het meta-domein recursief maakt, en dat een legitieme normatieve ordening vereist dat transparantie, reflexiviteit en netwerktracering systematisch worden ingebouwd in onderzoekspraktijken en institutionele arrangementen.

Samenvattend. Het meta-domein bundelt de spanningen van alle archetypen en fungeert tegelijk als toetssteen. Hier wordt duidelijk dat legitimiteit niet vanzelf ontstaat, maar actief georganiseerd moet worden. Utopie en distopie staan beide open, en de toekomst van AI-gestuurde kennisproductie zal afhangen van de mate waarin nieuwe normen structureel worden ingebouwd. Wikipedia, open-source communities en arXiv tonen dat alternatieve validatiepraktijken mogelijk zijn. Dit biedt hoop, maar geen garantie. De uitdaging is om deze tegenbeelden te versterken en te institutionaliseren, zodat democratisering van kennisproductie geen utopie blijft maar werkelijkheid wordt. Het meta-domein is niet alleen een spiegel van de andere archetypen, maar ook een laboratium waar de toekomst van wetenschap wordt vormgegeven. De keuzes die hier worden gemaakt – over openheid, validatie, macht en verantwoordelijkheid – zullen bepalend zijn voor de legitimiteit van AI-gestuurde kennisproductie in alle domeinen.

Literatuurlijst Hoofdstuk 11

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. https://doi.org/10.1145/3442188.3445922

Birhane, A., Prabhu, V. U., & Kahembwe, E. (2022). The values encoded in machine learning research. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 173–184. https://doi.org/10.1145/3531146.3533083

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20–23. https://doi.org/10.1038/538020a

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622. https://doi.org/10.1016/j.knosys.2020.106622

Mirhoseini, A., et al. (2021). A graph placement methodology for fast chip design. *Nature*, 594(7862), 207–212. https://doi.org/10.1038/s41586-021-03544-w

Raji, I. D., Buolamwini, J., & Gebru, T. (2021). AI benchmarking: Towards transparent and reproducible AI evaluation. *arXiv preprint*, arXiv:2102.05623. https://arxiv.org/abs/2102.05623

Vilalta, R., & Drissi, Y. (2002). A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 18(2), 77–95. https://doi.org/10.1023/A:1019956318069

Hoofdstuk 12 – Een tweede Tussenbalans.

12.1 Inleiding – Tussenbalans 2

Met de voorgaande hoofdstukken is een brede basis gelegd voor het beantwoorden van de onderzoeksvragen. Eerst zijn in hoofdstuk 4 tot en met 6 de drie centrale impactdimensies van kunstmatige intelligentie onderzocht: de cognitief-methodologische, de pedagogisch-institutionele en de machts- en normatieve dimensie. Vervolgens zijn deze dimensies toegepast op vier archetypen van wetenschap --- natuurwetenschappen, geschiedenis, antropologie en ontwerpgericht onderzoek --- en is in hoofdstuk 11 het meta-domein van AI en informatica in beeld gebracht als bijzondere spiegel die de spanningen van alle archetypen bundelt en intensiveert.

Het is nu tijd om een tweede tussenbalans op te maken. *Tussenbalans 1*, na de bespreking van de drie impactdimensies, liet zien welke kansen, risico's en uitdagingen AI-gestuurde kennisproductie in algemene zin oproept. *Tussenbalans 2***,** het huidige hoofdstuk, gaat een stap verder door de inzichten uit de dimensies en de archetypen te integreren. Daarmee ontstaat een analytisch kader dat de kernpatronen blootlegt die zich over de hele breedte van wetenschap en technologie voordoen.

Deze tussenbalans is bewust geen eindpunt. De kracht van academische zorgvuldigheid ligt juist in het gefaseerd opbouwen van conclusies. Eerst worden de implicaties van de dimensies en archetypen in samenhang beschouwd; pas daarna volgt de toepassing op concrete casussen. Hoofdstukken 13 t/m 17 brengen vijf van zulke casussen in: het gaat hierbij om vijf academische monografieën die geschreven zijn als proefschriften . Zij vormen de praktijkproef van het hier ontwikkelde raamwerk.

Het doel van dit hoofdstuk is dus niet om de hoofdvraag definitief te beantwoorden, maar om de contouren van een geïntegreerd antwoord zichtbaar te maken. In Tussenbalans 2 worden de overkoepelende patronen geschetst, worden de archetypen systematisch vergeleken en worden de drie deelvragen in voorlopige vorm beantwoord. Daarmee wordt de basis gelegd voor de casussen en uiteindelijk voor de definitieve synthese in hoofdstuk 18.

12.2 Overkoepelende patronen en onderscheidende kenmerken.

De analyse van de vier archetypen en het meta-domein laat zien dat er zich enkele patronen aftekenen die in meerdere domeinen terugkeren, maar ook dat ieder archetype unieke spanningen en mogelijkheden kent. Het is belangrijk beide kanten te benadrukken: gedeelde tendensen mogen niet verhullen dat elk archetype zijn eigen logica en kwetsbaarheden heeft.

12.2.1 Overeenkomsten, universele paradoxen.

De analyse van de vijf domeinen laat drie fundamentele paradoxen zien die zich in alle archetypen voordoen. Deze paradoxen zijn niet op te lossen maar alleen te managen: ze zijn inherent aan de dubbele aard van AI als versneller en herstructureerder van kennisproductie.

Paradox 1: AI als versneller en bedreiger van legitimiteit. De meest fundamentele paradox is dat AI in vrijwel alle domeinen fungeert als versneller van kennisproductie, maar tegelijkertijd de legitimiteit van die kennis onder druk zet. Deze paradox manifesteert zich in elk archetype op eigen wijze:

1. *Natuurwetenschappen:* AlphaFold versnelt eiwitstructuurvoorspelling met factor 1000, maar de voorspellingen zijn moeilijk te valideren zonder experimentele bevestiging. Meer snelheid betekent hier meer hypothesen maar niet automatisch meer betrouwbare kennis.
2. *Geschiedenis*: Transkribus kan in dagen archieven doorzoeken die voorheen jaren kostten, maar de patronen die AI identificeert kunnen anachronistisch zijn – verbanden die historisch gezien geen betekenis hadden. Snelheid verhoogt het risico op contextverlies.
3. *Antropologie*: Digitale etnografie maakt het mogelijk om duizenden sociale media-interacties te analyseren waar traditioneel veldwerk beperkt blijft tot enkele dozijn interviews. Maar de schaal gaat ten koste van de 'thick description' die antropologische legitimiteit waarborgt.
4. *Ontwerpgericht onderzoek*: Generatieve AI kan in uren duizenden ontwerpvarianten produceren, maar zonder transparantie over het ontwerpproces wordt het moeilijk om keuzes te verantwoorden. Snelheid ondermijnt reflectieve practice.
5. *Meta-domein:* AutoML kan in dagen modellen ontwikkelen die voorheen maanden kostten, maar wanneer AI zichzelf valideert via benchmarks die door AI zijn geoptimaliseerd, ontstaat circulaire validatie zonder externe toetsing.

De versnelling is onmiskenbaar waardevol – meer patronen, bredere datasets, snellere iteraties – maar juist deze winst ondermijnt de mechanismen waarop legitimiteit traditioneel berust. Reproduceerbaarheid, peer review en hermeneutische interpretatie vergen tijd; AI dwingt een keuze tussen snelheid en zorgvuldigheid die niet op te lossen is maar alleen te managen door expliciete verantwoording.

Paradox 2: Autonomie versus afhankelijkheid. Een tweede paradox is dat AI wetenschappers autonomer maakt (minder tijd aan routinetaken, meer ruimte voor interpretatie en synthese), maar hen tegelijk afhankelijker maakt van technologische infrastructuren die buiten hun controle liggen. Per archetype manifesteert zich dit als volgt:

1. *Natuurwetenschappen:* Onderzoekers kunnen complexere experimenten doen dankzij AI-gestuurde simulaties, maar zijn afhankelijk van cloudplatforms (AWS, Google Cloud) en proprietary software waarvan de werking niet transparant is.
2. *Geschiedenis:* Digitale archieven maken bronnen wereldwijd toegankelijk, maar historici zijn afhankelijk van digitaliseringsprojecten die bepalen welke archieven beschikbaar komen. Europese collecties domineren; lokale bronnen uit voormalige koloniën blijven ondervertegenwoordigd.
3. *Antropologie:* AI-transcriptie geeft antropologen meer tijd voor analyse, maar maakt hen afhankelijk van systemen die getraind zijn op dominante talen en culturele contexten. Minderheidstalen en dialecten blijven moeilijk te verwerken.
4. *Ontwerpgericht onderzoek:* Ontwerpers kunnen met generatieve AI sneller itereren, maar zijn afhankelijk van algoritmen waarvan de besluitvormingslogica ondoorzichtig is. De ontwerper wordt curator maar verliest grip op het proces.
5. Meta-domein: Onderzoekers kunnen state-of-the-art modellen bouwen via open-source bibliotheken, maar de ontwikkeling van foundation models blijft voorbehouden aan enkele grote bedrijven met enorme rekenkracht. Autonomie en afhankelijkheid bestaan tegelijk.

AI vergroot de mogelijkheden van individuele onderzoekers (autonomie) maar concentreert tegelijk de controle over infrastructuur bij enkele actoren (afhankelijkheid). Dit is geen temporeel probleem dat opgelost kan worden – het is structureel. Democratisering van tools gaat samen met oligopolisering van infrastructuur.

Paradox 3: Inclusie versus exclusie. Een derde paradox betreft de belofte dat AI kennisproductie democratiseert, terwijl het tegelijk nieuwe vormen van uitsluiting creëert. Per archetype ziet dit er zo uit:

1. *Natuurwetenschappen:* Open-access publicaties en preprint servers (arXiv) verlagen drempels, maar toegang tot rekenkracht voor simulaties blijft ongelijk verdeeld. Kleine universiteiten in het mondiale Zuiden kunnen niet concurreren met goed gefinancierde westerse labs.
2. *Geschiedenis:* Digitalisering maakt archieven toegankelijk voor onderzoekers die geen budget hebben voor fysieke archiefbezoeken, maar reproduceert machtsverhoudingen: Europese perspectieven worden gedigitaliseerd en dus zichtbaar, niet-Europese bronnen blijven uit beeld.
3. *Antropologie:* Digitale etnografie maakt veldwerk mogelijk zonder fysiek aanwezig te zijn, wat toegankelijk is voor onderzoekers met beperkingen. Maar AI-tools zijn getraind op westerse datasets en talen, waardoor niet-westerse gemeenschappen systematisch ondervertegenwoordigd blijven.
4. *Ontwerpgericht onderzoek:* Generatieve AI-tools zoals DALL-E of Midjourney lijken toegankelijk voor iedereen, maar de beste resultaten vragen om geavanceerde prompt-engineering skills en begrip van ontwerpprincipes. De 'democratisering' blijft oppervlakkig.
5. *Meta-domein:* Open-source bibliotheken zoals PyTorch maken AI-ontwikkeling toegankelijk, maar het trainen van foundation models vereist miljoenen dollars aan rekenkracht. Kennisdeling is open, maar infrastructuur blijft gesloten.

AI verlaagt sommige drempels (toegang tot tools, digitale archieven) maar verhoogt andere (technische expertise, rekenkracht, representatie in trainingsdata). Het is niet zozeer dat AI inclusiever óf exclusiever is, maar dat het nieuwe scheidingslijnen trekt die de oude niet vervangen maar aanvullen. Wie al geprivilegieerd was, profiteert het meest; wie al gemarginaliseerd was, raakt verder achterop.

Synthese paradoxen. Deze drie paradoxen zijn universeel omdat ze voortkomen uit de dubbele aard van AI: enerzijds een technologie die capaciteit vergroot, anderzijds een infrastructuur die nieuwe afhankelijkheden en ongelijkheden creëert. Ze zijn niet op te lossen door betere technologie of slimmere implementatie, maar alleen te managen door expliciete erkenning en institutionele waarborgen. De paradoxen tonen dat AI geen neutrale versterker is maar een transformatieve kracht die de voorwaarden van kennisproductie hertekent. Dit maakt discussies over 'verantwoord AI-gebruik' complex: het gaat niet om het vermijden van risico's, maar om het managen van inherente spanningen die niet verdwijnen.

12.2.2 Verschillen, domein-specifieke spanningen.

Daar waar de paradoxen universeel zijn en de overeenkomsten tussen de 4+1 domeinen benadrukken, zijn er ook domein-specifieke verschillen benoemd worden.

1. Natuurwetenschappen: relatief probleemloze inbedding. In de natuurwetenschappen is AI relatief probleemloos ingebed, omdat de methodologische normen van dit archetype --- reproduceerbaarheid, transparantie van experimenten, peer review --- grotendeels overeind blijven. Dedicated AI-toepassingen zoals AlphaFold hebben laten zien dat AI patronen kan ontdekken die voorheen onoplosbaar leken, zonder de epistemische grondslagen van het vak te ondermijnen. De risico's zijn beperkt tot afhankelijkheid van infrastructuur en mogelijke bias in datasets, maar deze worden vaak opgevangen door robuuste experimentele standaarden.

*2. Geschiedenis: authenticiteit onder druk.*Voor de geschiedwetenschap ligt de kernspanning in het behoud van authenticiteit en context. AI kan grote archieven doorzoeken en onverwachte verbanden blootleggen, maar dreigt deze ook anachronistisch te interpreteren. Een model herkent patronen, maar begrijpt niet de historische betekenis ervan. Menselijke hermeneutiek blijft hier onmisbaar, en de uitdaging is om AI in te zetten zonder dat de interpretatieve nabijheid verloren gaat die de legitimiteit van historische kennis waarborgt.

3. Antropologie: relationele context behouden. In de antropologie is de kernuitdaging dat relationele nabijheid en contextuele rijkdom moeilijk te automatiseren zijn. AI kan transcripties versnellen en thematische coderingen ondersteunen, maar mist de belichaamde ervaring van veldwerk. Het unieke

vraagstuk is niet zozeer of AI efficiënt is, maar of de kern van antropologisch onderzoek --- participatie, wederkerigheid, nabijheid --- behouden kan blijven wanneer technologie bemiddelt tussen onderzoeker en gemeenschap.

4. Ontwerpgericht onderzoek: verantwoording van creativiteit. Ontwerpgericht onderzoek laat het scherpste spanningsveld zien tussen kansen en risico's. AI kan optreden als co-designer die talloze varianten genereert en iteraties versnelt. Tegelijkertijd staat de kern van dit archetype --- het expliciet verantwoorden van ontwerpkeuzes --- onder druk. Artefacten kunnen efficiënt zijn, maar moeilijk uitlegbaar. De legitimiteit van ontwerpkennis hangt af van transparantie en verantwoording, en dat is precies wat AI problematiseert.

5. Meta-domein: circulaire validatie als constitutief probleem. Het meta-domein is uniek omdat hier instrument en object samenvallen. AI onderzoekt en valideert zichzelf, wat alle eerder genoemde spanningen versterkt. Circulaire validatie, machtsconcentratie en black-box-karakter zijn hier niet bijkomende risico's, maar constitutieve kenmerken van kennisproductie. Wat in de andere archetypen impliciet speelt, wordt hier expliciet en onontkoombaar.

12.2.3 Archetypen in vergelijking: kansen, risico's en uitdagingen.

Om een goed zicht op de verschillen en overeenkomsten te krijgen , wordt hier een systematische vergelijking van de vijf domeinen langs de dimensies kansen, risico's en uitdagingen gegeven:

Archetype	Kansen	Risico's	Uitdagingen
Natuur-wetenschappen	• Dedicated AI (AlphaFold) • Versnelling dataverwerking • Robuuste reproduceerbaarheid	• Afhankelijkheid infrastructuur (cloud, chips) • Bias in datasets	• Waarborgen dat openheid/reproduceerbaarheid overeind blijft ondanks machtsconcentratie
Geschiedenis	• Digitale archieven toegankelijk •Patroonherkenning in bronnen • Ontsluiting verspreide collecties	• Anachronisme • Verlies authenticiteit/ context • Canonvorming via digitalisering	• Behouden van hermeneutische interpretatie en kritische contextualisering
Antropologie	• Transcriptie/ codering versnellen • Grotere datasets analyseren • Digitale etnografie	• Verlies relationele nabijheid • Reductie context tot patronen • Ethische problemen consent	• Balans tussen schaalvoordelen en behoud contextuele rijkdom
Ontwerpgericht onderzoek	• AI als co-designer • Versnelling iteraties • Simulatie/ prototyping	• Black-box ontwerpen • Verlies verantwoording • Eigenaarschap diffuus	• Ontwikkelen normen voor uitlegbaarheid • Nieuwe publicatie/erkenningsvormen
Meta-domein	• AutoML, meta-learning • Zelfverbetering • Open platforms (arXiv, Wikipedia)	• Circulaire validatie •Machtsconcentratie • Extreme black box • Fragmentatie legitimiteit	• Nieuwe normen formuleren (transparante interpretatie, reflexieve logboeken, netwerktracering)

Deze matrix laat zien dat AI geen uniforme invloed heeft. De natuurwetenschappen tonen relatief weinig spanning tussen kansen en risico's, terwijl geschiedenis en antropologie worstelen met contextverlies en interpretatie. Ontwerpgericht onderzoek en het meta-domein stellen de legitimiteit van kennisproductie zelf ter discussie.

12.4 Van analyse naar casuïstiek.

De voorgaande analyse maakt duidelijk dat generatieve AI in alle archetypen van wetenschap vergelijkbare spanningen oproept, maar dat deze spanningen zich telkens op een andere plaats in het onderzoeksproces manifesteren. De drie paradoxen – versnelling versus legitimiteit, autonomie versus afhankelijkheid en inclusie versus exclusie – blijken universeel, maar hun concrete betekenis is domein-specifiek. Tegelijk laat de vergelijking zien dat het analytische raster van dimensies en archetypen slechts een gedeeltelijk antwoord biedt op de onderzoeksvragen.

Het raster maakt patronen zichtbaar, maar kan niet bepalen hoe deze spanningen zich daadwerkelijk ontwikkelen in concrete onderzoekspraktijken. Met name drie vragen blijven open. Ten eerste is nog onduidelijk in hoeverre generatieve AI binnen verschillende archetypen primair een hulpmiddel blijft of een technologie die het onderzoeksproces zelf mede vormgeeft. Ten tweede is nog onzeker of de verschillen tussen archetypen in de praktijk even scherp blijken als de theoretische analyse suggereert, of dat AI-gebruik juist convergentie tussen onderzoekspraktijken veroorzaakt. Ten derde is onduidelijk hoe de normatieve spanningen rond legitimiteit, auteurschap en transparantie zich manifesteren wanneer AI daadwerkelijk wordt ingezet bij het produceren van academische teksten.

Deze vragen markeren de grenzen van de theoretische analyse en vormen tegelijkertijd de aanleiding voor de empirische fase van dit onderzoek. De volgende hoofdstukken onderzoeken daarom hoe AI-ondersteunde kennisproductie zich manifesteert in concrete onderzoekspraktijken. In vijf casussen – de monografieën / proefschriften over ‘aviaire intelligentie & homing pigeons’ (natuurwetenschappelijk archetype), ‘de voorspelling van digitale transformaties’ (geschiedkundig archetype), ‘besturen binnen biologische grenzen’ (antropologisch archetype), ‘polyglotte taalverwerving in het AI-tijdperk’ (ontwerp-gericht archetype) en ‘AI-gestuurde kennisproductie’ (dit proefschrift – de ‘mise-en-abyme’ die het meta-domein representeert) worden de zoeven geformuleerde, nog openstaande vragen beantwoord a.d.h.v. de bevindingen in vijf verschillende onderzoekspraktijken. Daarbij zal naast de teksten van de proefschriften/monografieën, ook gebruik gemaakt worden van de reflectie-verslagen waarin de totstandkoming van de betreffende producties wordt beschreven.

Deze casussen fungeren niet als illustraties van de theorie, maar als een eerste empirische toets van het hier ontwikkelde raamwerk. Zij maken zichtbaar hoe generatieve AI in verschillende archetypen van wetenschap daadwerkelijk wordt gebruikt en welke implicaties dat heeft voor de structuur van kennisproductie. Pas in het licht van deze ervaringen kan worden beoordeeld in hoeverre het analysekader van dimensies en archetypen standhoudt.

Hoofdstuk 13 – Casus ‘Aviaire intelligentie & Homing Pigeons’ .

In dit hoofdstuk wordt het natuurwetenschappelijk archetype van AI-gestuurde kennisproductie onderzocht aan de hand van een concrete casus: een experimenteel onderzoeksprogramma naar navigatie- en classificatievermogen bij duiven (het concept-proefschrift *Columba Loquens – patroonherkenning, navigatierobuustheid en de grenzen van verklaring bij aviaire intelligentie*).

Het experiment richt zich op een klassieke vraag binnen de gedragsbiologie en de vergelijkende cognitiewetenschap: hoe robuust is het gedrag van dieren onder variatie van stimulus-structuren en informatiebronnen, en welke verklaringsmodellen kunnen dergelijke prestaties plausibel maken? Het onderzoeksveld wordt daardoor gekenmerkt door een situatie waarin gedrag empirisch goed beschreven kan worden, terwijl de onderliggende verklaringsstructuur open blijft voor meerdere theoretische interpretaties. Juist in dergelijke situaties speelt de formulering en vergelijking van mogelijke verklaringsmodellen een centrale rol in het wetenschappelijke proces.

13.1 Positionering van de casus

Het onderzoek naar navigatie- en classificatievermogen bij duiven behoort tot een lange traditie binnen de gedragsbiologie en de vergelijkende cognitiewetenschap. Sinds het midden van de twintigste eeuw wordt de postduif beschouwd als een modelorganisme voor de studie van dierlijke navigatie en ruimtelijk gedrag. Experimentele studies hebben laten zien dat duiven onder uiteenlopende omstandigheden in staat zijn hun weg terug te vinden naar het thuishok, vaak over aanzienlijke afstanden en onder sterk variërende omgevingscondities.

De verklaringen voor dit gedrag zijn onderwerp van een langdurig wetenschappelijk debat. In de literatuur zijn verschillende mechanismen voorgesteld, waaronder zonnekompasoriëntatie, magnetische detectie, olfactorische navigatie en visuele herkenning van landschappelijke kenmerken. Moderne interpretaties gaan er doorgaans van uit dat deze mechanismen niet exclusief opereren, maar functioneren als componenten binnen een breder systeem waarin meerdere informatiebronnen tegelijk kunnen worden benut.

Naast navigatieonderzoek heeft de studie van duiven ook een belangrijke rol gespeeld in onderzoek naar categorisatie en patroonherkenning bij dieren. Experimentele studies tonen aan dat duiven complexe visuele discriminaties kunnen uitvoeren en categorieën kunnen leren herkennen binnen variërende stimulussets. Deze bevindingen hebben geleid tot een omvangrijke literatuur over stimulusgeneralisatie, classificatievermogen en categorisatie bij niet-menselijke organismen.

Vanuit empirisch perspectief is daarmee relatief veel bekend over de gedragsmatige prestaties van duiven. Tegelijkertijd blijft het vaak moeilijk om met zekerheid vast te stellen welk intern mechanisme deze prestaties precies mogelijk maakt. Hetzelfde gedrag kan doorgaans op meerdere manieren worden geïnterpreteerd. Dit maakt het onderzoeksveld bijzonder geschikt voor methodologische reflectie op de relatie tussen waargenomen gedrag, theoretische modellen en verklaringsstructuren.

13.2 Kerninzicht & rol van AI in het onderzoeksproces: AI als instrument voor modelruimte-exploratie.

Het centrale inzicht dat uit de casus naar voren komt betreft de rol van generatieve AI bij het systematisch verkennen van mogelijke verklaringsmodellen voor empirisch waargenomen gedrag. In onderzoeksvelden waar het gedrag van een organisme relatief goed gedocumenteerd is, maar het onderliggende mechanisme onzeker blijft, ontstaat een karakteristieke epistemische situatie: meerdere theoretische modellen kunnen hetzelfde gedrag plausibel verklaren. In dergelijke contexten bestaat een belangrijk deel van het wetenschappelijke werk uit het formuleren, variëren en vergelijken van mogelijke verklaringsstructuren. Het duivenonderzoek vormt hiervan een illustratief voorbeeld. Navigatiegedrag kan worden geïnterpreteerd vanuit verschillende hypothesen – variërend van magnetische detectie tot visuele routeherkenning – zonder dat één verklaring onmiddellijk doorslaggevend is.

Generatieve AI-systemen blijken in staat deze modelruimte systematisch te verkennen. Door bestaande literatuur, empirische patronen en theoretische aannames te combineren, kunnen zij varianten van verklaringsmodellen formuleren, alternatieve interpretaties genereren en de implicaties van verschillende hypothesen expliciteren. Hierdoor ontstaat een analytisch proces waarin mogelijke verklaringen niet sequentieel maar parallel worden verkend. In deze casus werd deze generatieve capaciteit benut om een coherent onderzoeksnarratief te ontwikkelen waarin empirische observaties, theoretische interpretaties en mogelijke verklaringsmechanismen systematisch met elkaar in verband werden gebracht.

13.3 Specifieke inzichten uit de casus

De toepassing van generatieve AI binnen deze casus levert een aantal concrete observaties op over de manier waarop AI-systemen functioneren binnen natuurwetenschappelijke analyse. Ten eerste blijkt generatieve AI bijzonder effectief in het synthetiseren van omvangrijke literatuurcorpora. Binnen een onderzoeksveld met een lange empirische traditie kan het model bestaande studies, theoretische posities en experimentele bevindingen samenbrengen in coherente overzichtsstructuren. Hierdoor ontstaat een systematisch beeld van het bestaande verklaringslandschap. Ten tweede maakt AI het mogelijk om hypothesevarianten snel te genereren en te expliciteren. Waar onderzoekers traditioneel één of enkele hypothesen tegelijk uitwerken, kan een generatief model meerdere mogelijke verklaringslijnen parallel formuleren. Dit vergroot de zichtbaarheid van alternatieve interpretaties en maakt het eenvoudiger om impliciete aannames in bestaande theorieën te identificeren. Ten derde ondersteunt AI het expliciteren van argumentatieve structuren. Hypothesen kunnen worden uitgewerkt in termen van veronderstelde mechanismen, empirische verwachtingen en mogelijke falsificatiecriteria. Hierdoor wordt het verklaringsmodel transparanter gemaakt en kan de interne consistentie van verschillende interpretaties beter worden beoordeeld. Deze functies positioneren generatieve AI binnen het natuurwetenschappelijk onderzoek primair als een instrument voor analytische verkenning en theoretische articulatie.

13.4 Impact op het analytisch framework

De bevindingen uit deze casus sluiten aan bij het beeld dat in eerdere hoofdstukken van dit proefschrift naar voren kwam over de rol van AI binnen verschillende archetypen van wetenschap. In het natuurwetenschappelijk archetype ligt de kracht van AI vooral in het verwerken van grote hoeveelheden empirische en theoretische informatie en in het genereren van mogelijke verklaringsmodellen.

Deze rol verschilt duidelijk van de functies die AI in andere onderzoeksdomeinen kan vervullen. In interpretatieve disciplines ligt de nadruk vaker op tekstuele analyse en argumentatieve reconstructie, terwijl in ontwerpgericht onderzoek AI een rol speelt in het genereren van mogelijke oplossingsvarianten. Binnen de natuurwetenschappen manifesteert de bijdrage van AI zich vooral in het systematisch verkennen van modelruimte en hypothesevariatie.

De casus laat daarmee zien dat generatieve AI kan functioneren als een instrument voor epistemische exploratie: een hulpmiddel dat onderzoekers ondersteunt bij het articuleren en structureren van mogelijke verklaringen voor empirische fenomenen. De uiteindelijke beoordeling van deze verklaringen blijft afhankelijk van empirische toetsing, maar het proces waarin mogelijke modellen worden geformuleerd en uitgewerkt kan door AI aanzienlijk worden versneld en uitgebreid.

Empirische onderzoeksvraag	Relevante observaties uit de casus	Impact op beantwoording
ERQ1 – Cognitief-methodologische dimensie: Welke kansen, risico's en uitdagingen biedt AI voor de kwaliteit van kennisproductie?	Generatieve AI werd ingezet om bestaande literatuur te synthetiseren, alternatieve verklaringsmodellen te formuleren en hypothesevarianten systematisch te expliciteren. Hierdoor werd de theoretische modelruimte rondom het navigatie- en classificatievermogen van duiven zichtbaar gemaakt.	**Duidelijke bijdrage.** De casus laat zien dat AI binnen natuur-wetenschappelijk onderzoek vooral waarde heeft bij het systematisch verkennen van mogelijke verklaringsstructuren voor empirisch gedrag.
ERQ2 – Pedagogisch-institutionele dimensie: Hoe verandert AI de rollen en structuren in academisch onderzoek?	Het onderzoeksproces ontwikkelde zich als een iteratieve interactie tussen menselijke probleemformulering en AI-gestuurde generatieve analyse. Het model produceerde hypothesevarianten en tekstuele uitwerkingen, terwijl de onderzoeker het analytisch kader bepaalde en de argumentatieve structuur bewaakte.	**Illustratieve bijdrage.** De casus suggereert een verschuiving naar een hybride onderzoekspraktijk waarin generatieve systemen een rol spelen in hypothesevorming en literatuursynthese, terwijl de architectuur van het onderzoek door de onderzoeker wordt bepaald.
ERQ3 – Macht en normativiteit: Welke machtsverhoudingen en normatieve keuzes zijn ingebed in AI-kennisproductie?	De selectie van relevante literatuur, de formulering van onderzoeksvragen en de interpretatie van empirische resultaten bleven afhankelijk van de keuzes die binnen het onderzoeksontwerp werden gemaakt. AI opereerde binnen deze vooraf gedefinieerde epistemische kaders.	**Beperkte maar relevante bijdrage.** De casus laat zien dat normatieve keuzes in natuurwetenschappelijk onderzoek primair liggen in probleemdefinitie en onderzoeks-ontwerp, terwijl AI deze keuzes operationaliseert binnen de gegenereerde analyse.

Tabel 13.4: Impact casus op beantwoording empirische onderzoeksvragen.

13.6 Demasqué: een experiment in wetenschappelijk illusionisme

De voorgaande paragrafen behandelden het duivenproefschrift alsof het een regulier natuurwetenschappelijk onderzoek betrof. In deze slotparagraaf wordt echter een cruciaal gegeven expliciet gemaakt: **het experiment waarop het proefschrift berust, heeft in werkelijkheid nooit plaatsgevonden.** De datasets waarop de analyse steunt zijn synthetisch gegenereerd met behulp van generatieve AI. Deze onthulling vormt geen curiositeit maar het eigenlijke doel van de casus. Het duivenproefschrift fungeert binnen dit onderzoek als een gecontroleerd experiment in wetenschappelijk illusionisme: een demonstratie van de mate waarin generatieve AI in staat is een overtuigende natuurwetenschappelijke studie te produceren, inclusief plausibele empirische data, methodologische verantwoording en statistische analyse. Het experiment richt zich daarmee niet op de vraag of AI fouten kan maken — dat is inmiddels evident — maar op een subtieler en methodologisch interessanter probleem: in hoeverre kan de vorm van wetenschappelijke argumentatie zelf overtuigingskracht genereren, zelfs wanneer de empirische basis synthetisch is?

13.6.1 Het 'waarom' van dit experiment

Het experiment had meerdere motieven die zowel praktisch als epistemisch van aard waren.

Een eerste motief betreft wat men zou kunnen aanduiden als *resource poverty*. Toegang tot empirische datasets, laboratoria en veldexperimenten is in veel natuurwetenschappelijke domeinen sterk institutioneel gereguleerd. Voor onafhankelijke onderzoekers zijn dergelijke middelen vaak eenvoudigweg niet beschikbaar. Dit creëert een asymmetrische kennisinfrastructuur waarin empirisch onderzoek in hoge mate afhankelijk wordt van institutionele toegang. De mogelijkheid om synthetische datasets te genereren vormt daarom niet alleen een methodologische curiositeit, maar ook een reactie op een reële structurele beperking in de hedendaagse onderzoekspraktijk.

Een tweede motief heeft te maken met de maatschappelijke en institutionele ontvangst van AI-gegenereerd onderzoek. In de praktijk blijkt dat universiteiten, tijdschriften en reviewers vaak terughoudend of zelfs afwijzend reageren wanneer expliciet wordt vermeld dat generatieve AI een substantiële rol heeft gespeeld bij de totstandkoming van een wetenschappelijke tekst. Het duivenexperiment kan daarom ook worden gelezen als een *retorische interventie*: wanneer een AI-gegenereerde studie volledig voldoet aan de conventies van academische argumentatie, maar pas achteraf als zodanig wordt herkend, wordt het academische debat over AI in kennisproductie feitelijk afgedwongen.

Een derde motief is epistemisch van aard en sluit aan bij de logica van de *Turingtest.* In de klassieke formulering van Turing gaat het er niet om of een machine werkelijk denkt, maar of zij in staat is gedrag te produceren dat niet meer betrouwbaar van menselijk gedrag kan worden onderscheiden. Het duivenproefschrift vormt een analoge test binnen de context van wetenschappelijke productie: wanneer een AI-gegenereerd proefschrift niet van een conventioneel proefschrift te onderscheiden is, wordt daarmee een fundamentele vraag opgeworpen over de rol van generatieve systemen in academisch werk.

13.6.2 De constructie van de illusie

De overtuigingskracht van het duivenproefschrift berustte niet op spectaculaire claims, maar juist op methodologische soberheid. De studie werd opgebouwd volgens een klassieke natuurwetenschappelijke architectuur: een literatuurreview, een methodologische verantwoording, drie datasets met empirische analyses en een synthesehoofdstuk waarin de resultaten worden geïnterpreteerd. Cruciaal was dat de

synthetische datasets geen perfecte patronen vertoonden. Variatie, kleine inconsistenties en bescheiden effectgroottes werden bewust ingebouwd. Statistische uitkomsten waren plausibel maar niet spectaculair; resultaten bevestigden elkaar zonder identiek te zijn. Ook de interpretatieve stijl droeg bij aan de geloofwaardigheid. Conclusies werden voorzichtig geformuleerd, alternatieve verklaringen werden besproken en beperkingen van de data werden expliciet benoemd. Het centrale argument van het proefschrift was bovendien gradueel: duivennavigatie werd niet verklaard door één mechanisme, maar door een combinatie van overlappende oriëntatiestrategieën. Deze conclusie sluit nauw aan bij bestaande discussies in de gedragsbiologie en versterkte daarmee de plausibiliteit van het geheel.

Naast deze methodologische strategieën werden ook enkele subtiele illusionistische elementen ingebouwd. Zo werd het proefschrift gepubliceerd onder een vrouwelijk pseudoniem (R.A.A.F Corvinus) . In veel wetenschappelijke domeinen wordt – impliciet of expliciet – een andere retorische stijl verwacht van vrouwelijke auteurs, met minder heroïsche claims en meer epistemische voorzichtigheid. Deze keuze droeg bij aan de geloofwaardigheid van de bescheiden interpretatieve toon van het manuscript. Daarnaast bevatte het auteurschap een kleine maar betekenisvolle *easter egg*: de combinatie van initialen en achternaam vormde een subtiele aanwijzing dat het proefschrift mogelijk niet was wat het leek. Tegelijk was deze aanwijzing zo discreet dat zij alleen zichtbaar wordt voor lezers die er expliciet naar zoeken. Ten slotte bevat het experiment een lichte komische ondertoon: het proefschrift over duivennavigatie werd geschreven onder het pseudoniem *Corvinus* — een verwijzing naar de raaf (*Corvus*), een vogel die in cognitief onderzoek bekendstaat om zijn uitzonderlijke intelligentie. Het beeld van een raaf die een proefschrift schrijft over duiven fungeert als een kleine ironische knipoog, zonder de wetenschappelijke ernst van het experiment te ondermijnen.

13.6.3 Het verbergen van AI-auteurschap

Een cruciaal onderdeel van het experiment was het systematisch verwijderen van herkenbare sporen van AI-generatie. Generatieve taalmodellen produceren vaak teksten met herkenbare stilistische kenmerken: symmetrische zinsstructuren, opvallend consistente terminologie, uniforme citatiedichtheid en een retorische gladheid die afwijkt van typische menselijke schrijfprocessen. In het duivenproefschrift werden dergelijke patronen bewust gecorrigeerd. De tekst werd redactioneel aangepast om kleine asymmetrieën, variaties in formulering en incidentele redundantie te introduceren. Ook tabellen en datasets werden bewust minder “perfect” gemaakt: kleine afwijkingen in spreiding, variaties in precisie en incidentele ontbrekende waarden werden toegevoegd om realistische empirische ruis te simuleren.

Deze ingrepen illustreren een bredere les. Veel pogingen om AI-gebruik in academische teksten te detecteren zijn gebaseerd op oppervlakkige stilistische heuristieken. Het experiment laat zien dat dergelijke strategieën methodologisch zwak zijn. Wanneer een tekst zorgvuldig wordt geredigeerd, verdwijnen de meeste herkenbare AI-markers. Voor onderwijscontexten – bijvoorbeeld bij scriptiebeoordeling – is dit een belangrijke observatie. Detectie van AI-gebruik op basis van stijl alleen is fundamenteel onbetrouwbaar.

13.6.4 De tweede illusie: het bespelen van het modelgeheugen

Een onverwachte maar methodologisch interessante dimensie van het experiment ontstond tijdens de interactie met het taalmodel dat het proefschrift had helpen schrijven. Eén-en-hetzelfde taalmodel werd gebruikt om zowel de proefschrift-teksten te genereren (rol 1) als de onderzoeksresultaten en – op bases van reversed engineering – de ruwe onderzoeksdata te fabriceren (rol 2). Het model was zich niet bewust van het feit dat het beide rollen had vervuld. Zelfs bij de uiteindelijke onthulling herkende het model aanvankelijk niet dat het hier ging om een manuscript dat het zelf had helpen produceren. (rol 3,

reflectie). Dit verschijnsel heeft te maken met een fundamentele eigenschap van grote taalmodellen: hun geheugen is contextueel en beperkt. Wanneer een gesprek wordt voortgezet in een nieuwe context of wanneer eerdere delen van een interactie buiten het actieve contextvenster vallen, kan het model zijn eigen eerdere rol niet meer reconstrueren.

In het duivenexperiment werd deze eigenschap feitelijk benut. Het model werd in verschillende fasen van het proces in *verschillende epistemische rollen* geplaatst zonder dat het zelf een consistent beeld had van die rolverdeling. De implicatie is methodologisch relevant: het geheugen van taalmodellen kan worden bespeeld. Daardoor kunnen verschillende rollen – auteur, criticus, reviewer – in één en hetzelfde systeem worden gecombineerd zonder dat het model zich bewust is van de volledige architectuur van het proces.

13.6.5 Betekenis voor AI-gestuurde kennisproductie

Het duivenexperiment laat zien dat generatieve AI in staat is om de vorm van natuurwetenschappelijke argumentatie overtuigend te reproduceren, inclusief plausibele empirische data. Daarmee vormt deze casus een extreme variant van AI-gestuurde kennisproductie. Tegelijk moet worden benadrukt dat deze casus in één belangrijk opzicht afwijkt van de andere studies in dit proefschrift. Alle proefschriften in deze pentalogie zijn geschreven met behulp van generatieve AI onder regie van een menselijke architect-curator. In het duivenexperiment werden echter niet alleen de teksten, maar ook de empirische datasets door AI gegenereerd.

In bredere zin maakt het duivenexperiment een klassieke spanning uit de wetenschapsfilosofie opnieuw zichtbaar: de verhouding tussen verklaring en voorspelling. Generatieve systemen blijken in staat om overtuigende verklaringsnarratieven te produceren zonder dat de onderliggende empirische basis noodzakelijk sterk is, terwijl andere vormen van AI – zoals machine-learning systemen in de natuurwetenschappen – soms zeer nauwkeurige voorspellingen kunnen genereren zonder dat de onderliggende mechanismen volledig worden begrepen. Met andere woorden: AI kan overtuigend verklaren zonder echte data, en correct voorspellen zonder echte verklaring. Het duivenexperiment maakt deze epistemische asymmetrie empirisch zichtbaar. Het laat zien dat narratieve plausibiliteit in wetenschappelijke argumentatie soms de plaats kan innemen van epistemische noodzaak

Het duivenproefschrift vormt daarmee een grensgeval binnen de pentalogie van AI-gegenereerde proefschriften. In deze casus werden zowel de tekst als de empirische data door AI gegenereerd. De volgende hoofdstukken analyseren gevallen waarin de *empirische basis van onderzoek intact blijft,* maar generatieve AI een steeds centralere rol speelt in analyse, interpretatie en presentatie van wetenschappelijke kennis.

Hoofdstuk 14: Casus Historische voorspellingen t.a.v. digitale transformaties.

In dit hoofdstuk wordt het gebruik van AI bij de vorming van het proefschrift '*De beloften van digitale revoluties, een historisch-analytisch onderzoek naar IT-voorspellingen en maatschappelijke transformaties in Nederland (1945-2025)*' geanalyseerd. De analyse is gebaseerd op het proefschrift zelf en het gelijknamige reflectie-verslag waarin de ontstaansgeschiedenis is gedocumenteerd. Volledigheidshalve wordt opgemerkt dat dit er in proefschrift ook ruime aandacht wordt besteed aan maatschappelijke transformaties ten gevolge technologieën waarvan de impact bij introductie – binnen de politieke arena's – niet of nauwelijks werd geclaimd.

14.1 Positionering van de casus

Het proefschrift 'De beloften van digitale revoluties' vertegenwoordigt het historische archetype, omdat het niet alleen terugblikt op technologische ontwikkelingen, maar historische toekomstclaims systematisch confronteert met feitelijke maatschappelijke uitkomsten. De centrale vraag is in hoeverre IT-gedreven maatschappelijke transformaties voorspelbaar zijn, en welke kenmerken van voorspellingen en voorspellers samenhangen met structureel adequate voorspellingen. Methodisch vertrekt het onderzoek vanuit canonvorming. Die canon is geen historisch totaaloverzicht en ook geen normatieve ranglijst van "belangrijkste" technologieën, maar een analytische afbakening van ICT-ontwikkelingen die in Nederlandse publieke en politieke bronnen als potentieel maatschappelijk transformerend zijn gearticuleerd. Toetreding tot het canon verloopt via twee toegangspoorten: politieke articulatie en maatschappelijke zichtbaarheid. Vervolgens wordt de longlist van historische episodes systematisch opgebouwd en gereduceerd op basis van analytische bruikbaarheid. Vanuit dit canon wordt het materiaal geordend in een 2×2-matrix. Die matrix combineert twee retrospectieve dimensies: de accuraatheid van de voorspelling en de mate waarin maatschappelijke transformatie daadwerkelijk gerealiseerd is. Zo ontstaan vier analytische kwadranten: accuraat voorspeld en gerealiseerd; accuraat voorspeld maar beperkt gerealiseerd; inadequaat voorspeld maar toch duurzaam transformerend; en inadequaat voorspeld zonder duurzame verankering. De matrix is dus geen illustratief schema, maar het centrale instrument om historische episodes vergelijkbaar te maken zonder hun context-specifieke karakter te verliezen.

Binnen dit meta-proefschrift is deze casus relevant omdat zij laat zien hoe AI-ondersteund historisch onderzoek kan bijdragen aan de reconstructie, ordening en vergelijking van een groot corpus van verwachtingen en transformaties, terwijl de beslissende stap — de historische beoordeling van structurele adequaatheid — afhankelijk blijft van menselijke contextualisering en interpretatie.

14.2 Rol van AI in het onderzoeksproces

Bij de totstandkoming van dit proefschrift werd generatieve AI op meerdere momenten in het onderzoeksproces ingezet. De studie richt zich op het reconstrueren van een canon van IT-gedreven maatschappelijke transformaties en op het analyseren van de relatie tussen historische toekomstclaims en feitelijke maatschappelijke uitkomsten. Dat vereiste het systematisch verzamelen, ordenen en vergelijken van een groot corpus van beleidsdocumenten, historische analyses, technologiediscours en academische literatuur.

Een eerste rol van AI betrof de exploratieve fase van het onderzoek. Door iteratieve interactie met taalmodellen konden mogelijke kandidaten voor opname in het canon worden geïdentificeerd en vergeleken. Het model fungeerde hierbij als heuristisch instrument: het hielp bij het genereren van mogelijke lijsten van technologische ontwikkelingen en bij het verkennen van verschillende manieren om deze historisch te ordenen. De uiteindelijke selectie van canonieke episodes – en daarmee de afbakening van het onderzoeksobject – bleef een analytische beslissing van de onderzoeker.

Een tweede rol betrof het gebruik van AI bij de structurering van het materiaal. Het model hielp bij het identificeren van terugkerende thema's, discursieve patronen en verschillende manieren waarop technologische toekomstclaims historisch werden geformuleerd. Daardoor kon het corpus sneller worden geordend en konden alternatieve manieren van classificatie en periodisering worden verkend voordat een definitieve analytische structuur werd gekozen.

De meest inhoudelijke rol van AI lag echter in de analytische fase van het onderzoek. Voor de analyse van de episodes uit het canon ontwikkelde de onderzoeker drie – in het gebruik complementaire - lenzen. De eerste was een '*claimgerichte lens (X)',* waarmee werd onderzocht op welke wijze technologische toekomstclaims historisch zijn gearticuleerd. Hierbij gaat het niet alleen om de inhoud van de claim, maar ook om de vorm waarin zij wordt gepresenteerd: als beleidsbelofte, als technologisch determinisme, als visionaire toekomstschets of als meer voorzichtige verwachting. De tweede was een '*transformatiegerichte lens (Y)*' die werd gebruikt om te beoordelen in welke mate de geclaimde maatschappelijke transformatie daadwerkelijk gerealiseerd is. Deze lens richt zich op de feitelijke institutionele, organisatorische en maatschappelijke veranderingen die met de betreffende technologie samenhangen. De derde lens betrof de causale relatie tussen beide (X–Y). Deze lens werd gebruikt om te onderzoeken in hoeverre de oorspronkelijke articulatie van een technologische claim samenhangt met de uiteindelijke maatschappelijke uitkomst: welke claims blijken structureel adequaat, welke blijken overschattingen of onderschattingen, en welke vormen van articulatie blijken systematisch samen te gaan met bepaalde typen uitkomsten.

Generatieve AI werd vervolgens ingezet om deze drie lenzen analytisch uit te werken en systematisch toe te passen op de geselecteerde episodes uit het canon. Door historische casussen iteratief langs deze drie perspectieven te analyseren kon voor elke episode worden vastgesteld hoe de oorspronkelijke claim was gearticuleerd, in welke mate de geclaimde transformatie daadwerkelijk plaatsvond en hoe beide zich tot elkaar verhouden. De rol van AI in dit onderzoek kan daarom het best worden beschreven als analytisch-ondersteunend. De technologie hielp bij het operationaliseren van het door de onderzoeker ontwikkelde analysekader en bij het systematisch toepassen daarvan op het historische materiaal. De uiteindelijke interpretatie van de resultaten – en daarmee de historiografische beoordeling van de episodes – bleef echter een verantwoordelijkheid van de onderzoeker, die de uitkomsten contextualiseerde en toetste aan de historische literatuur en de relevante bronnen.

14.3 Kern-inzichten

Deze casus laat zien dat de toepassing van generatieve AI in historisch onderzoek geen lineair proces is waarin taken duidelijk kunnen worden verdeeld tussen mens en machine. De analyse van het canon en de toepassing van de drie lenzen kwamen tot stand via een iteratief proces van interactie tussen onderzoeker en taalmodel. In de praktijk betekende dit dat analyses, formuleringen en classificaties meerdere keren werden gegenereerd, beoordeeld en herbewerkt. Het model produceerde voorstellen voor interpretatie en ordening van historische episodes, waarna de onderzoeker deze beoordeelde, bijstuurde en opnieuw liet uitwerken. De uiteindelijke analyse is daardoor niet het product van één actor, maar van een reeks iteraties waarin menselijke en machinale bijdragen voortdurend op elkaar inwerken.

Voor de empirische onderzoeksvragen van dit meta-proefschrift betekent dit dat AI in het historische archetype niet eenvoudig kan worden opgevat als hulpmiddel dat analyse “uitvoert”, noch als autonome auteur van historische interpretaties. In plaats daarvan functioneert het model als partner in een iteratief analytisch proces, waarbij de onderzoeker het kader ontwerpt, de resultaten beoordeelt en de uiteindelijke interpretatie legitimeert. De casus laat daarmee zien dat AI-ondersteunde kennisproductie in historisch onderzoek het best kan worden begrepen als een *symbiotisch proces van iteratieve analyse*, waarin generatieve modellen en menselijke onderzoekers gezamenlijk tot een interpreteerbaar resultaat komen.

14.4 Specifieke inzichten n.a.v. deze casus.

De casus laat een aspect van AI-ondersteunde kennisproductie zien dat in de voorafgaande theoretische analyse nog niet volledig zichtbaar was. In dit onderzoek werd generatieve AI niet alleen gebruikt voor literatuurverkenning of tekstproductie, maar voor het operationeel toepassen van een door de onderzoeker ontworpen analysekader op een historisch corpus. De drie lenzen van het onderzoek – de wijze waarop technologische claims zijn gearticuleerd (X), de mate waarin de geclaimde maatschappelijke transformatie daadwerkelijk plaatsvond (Y), en de causale relatie tussen beide – werden eerst conceptueel ontworpen door de onderzoeker en vervolgens door het taalmodel systematisch toegepast op de geselecteerde episodes uit het canon. Hierdoor kon het corpus van historische casussen op een consistente manier worden geanalyseerd en vergeleken. De casus laat daarmee zien dat generatieve AI niet alleen kan helpen bij het verzamelen of samenvatten van historisch materiaal, maar ook bij het operationaliseren en uitvoeren van een interpretatief analysekader. Tegelijk blijft de beoordeling van de uitkomsten – en daarmee de historiografische legitimiteit van de analyse – afhankelijk van menselijke interpretatie .

14.5 Impact op het analytisch framework.

De historische casus bevestigt dat generatieve AI in dit archetype vooral functioneert als instrument voor het operationaliseren en systematisch toepassen van een analytisch kader op een historisch corpus. De technologie maakt het mogelijk om een groot aantal episodes consistent langs dezelfde interpretatieve lenzen te analyseren en daardoor vergelijkbaar te maken. Tegelijk laat de casus zien dat de kern van historiografische legitimiteit niet door het model zelf kan worden vastgesteld. De selectie van canonieke episodes, de interpretatie van technologische claims en de beoordeling van causale verbanden blijven afhankelijk van historische contextualisering en interpretatieve afweging door de onderzoeker.

Voor het analysekader van dit meta-proefschrift betekent dit dat generatieve AI in het historische archetype vooral een rol speelt in de analytische operationalisering van interpretatieve schema’s, terwijl de interpretatie en legitimatie van historische kennis bij de onderzoeker blijven liggen. Daarmee bevestigt de casus het in hoofdstuk 12 geschetste beeld dat AI in interpretatieve disciplines primair bijdraagt aan systematische analyse, maar dat de uiteindelijke betekenisgeving een menselijke verantwoordelijkheid blijft. E.e.a. wordt ook duidelijk als de impact van deze analyses op de beantwoording van de empirische onderzoeksvragen in ogenschouw wordt genomen (tabel 14.5)

Empirische onderzoeksvraag:	Relevante observaties uit de casus:	Impact op beantwoording:
ERQ1 – Cognitief-methodologische dimensie: Welke kansen, risico's en uitdagingen biedt AI voor de kwaliteit van kennisproductie?	Generatieve AI maakte het mogelijk een groot corpus van historische episodes systematisch langs dezelfde analytische lenzen (claimarticulatie, gerealiseerde transformatie en hun causale relatie) te analyseren. Hierdoor konden historisch uiteenlopende episodes consistent worden vergeleken. Tegelijk bleef de interpretatie van historische betekenis afhankelijk van menselijke contextualisering.	**Duidelijke bijdrage.** De casus laat zien dat AI in historisch onderzoek vooral waarde heeft bij het operationaliseren en systematisch toepassen van interpretatieve analysekaders op omvangrijke historische corpora.
ERQ2 – Pedagogisch-institutionele dimensie: Hoe verandert AI de rollen en structuren in academisch onderzoek?	Het onderzoek ontwikkelde zich als een iteratief proces waarin generatieve modellen voorstellen deden voor classificaties en interpretaties, terwijl de onderzoeker het analysekader ontwierp, de uitkomsten beoordeelde en de historiografische interpretatie legitimeerde.	**Illustratieve bijdrage.** De casus bevestigt een verschuiving naar een symbiotische onderzoekspraktijk waarin AI analytische ondersteuning biedt, terwijl de onderzoeker eindverantwoordelijk blijft voor interpretatie en legitimatie.
ERQ3 – Macht en normativiteit: Welke machtsverhoudingen en normatieve keuzes zijn ingebed in AI-kennisproductie?	De canonvorming – de selectie van episodes die als historisch relevant worden beschouwd – blijft een normatieve onderzoeksbeslissing, met dien verstande dat AI daarbij alleen kan selecteren uit reeds gedigitaliseerde en dus machine-toegankelijke bronnen. AI kan deze selectie ondersteunen, maar reproduceert impliciet de kaders, bronkeuzes en toegangstructuren die door de onderzoeker en de beschikbare informatie-infrastructuur zijn bepaald.	**Beperkte maar relevante bijdrage.** De casus laat zien dat normatieve keuzes in historisch onderzoek primair liggen in canonvorming en interpretatieve framing, terwijl AI deze keuzes eerder operationaliseert dan zelf produceert.

Tabel 14.5 – Impact casus op beantwoording empirische onderzoeksvragen.

Hoofdstuk 15 – Casus Besturen Binnen Biologische Grenzen.

In dit hoofdstuk wordt het gebruik van AI bij de vorming van het proefschrift '*Besturen binnen biologische grenzen: De rol van biologische randvoorwaarden bij het oplossen van persistente problemen in de bestuurskunde*' geanalyseerd. De analyse is gebaseerd op het proefschrift zelf en het gelijknamige reflectie-verslag waarin de ontstaansgeschiedenis is gedocumenteerd.

15.1 Positionering van de casus

Het proefschrift *Besturen binnen biologische grenzen* vertegenwoordigt het antropologische archetype omdat het vertrekt vanuit een fundamentele vraag naar menselijk gedrag in sociale en institutionele contexten. De centrale these is dat bestuurskundige verklaringen van beleidsfalen onvolledig blijven wanneer zij uitsluitend leunen op economische, juridische of religieus-culturele modellen. Een ontbrekende dimensie is menselijke natuur: evolutionair gevormde mechanismen zoals statusverdediging, territoriale gehechtheid, groepsloyaliteit, ouderlijke investering en coalitievorming beïnvloeden beleidsuitkomsten op een manier die traditionele modellen slechts ten dele verklaren.

Het proefschrift ontwikkelt daarom een raamwerk van tien biologische randvoorwaarden die relevant kunnen zijn voor bestuurskunde. Deze randvoorwaarden worden niet opgevat als deterministische wetten, maar als structurele condities waaronder beleid meer of minder kans van slagen heeft. Op basis daarvan wordt onderzocht hoe deze mechanismen zichtbaar worden in concrete beleidsdossiers en hoe zij de effectiviteit van traditionele reguleringsmechanismen kunnen ondermijnen of juist helpen verklaren.

Empirisch is het onderzoek opgebouwd rond zes casussen, verdeeld over drie conflictvelden: klassenstrijd, rassenstrijd en geslachtenstrijd. In elk van deze velden wordt nagegaan welke biologische mechanismen beleidsprocessen mede structureren en waarom hardnekkige bestuurlijke vraagstukken vaak resistent blijven tegen conventionele interventies. De analyse is kwalitatief, abductief en gericht op pattern matching tussen theoretisch geformuleerde randvoorwaarden en feitelijke patronen van beleidsfalen.

Binnen het meta-proefschrift is deze casus relevant omdat zij laat zien hoe generatieve AI kon worden ingezet bij het uitwerken en toepassen van een interpretatief raamwerk in een domein waarin sociale praktijk, normatieve spanning en contextuele betekenis centraal staan. Daarmee vormt het proefschrift een casus van AI-ondersteunde kennisproductie binnen het antropologische archetype.

15.2 Rol van AI in het onderzoeksproces

Bij de totstandkoming van dit proefschrift werd generatieve AI intensief gebruikt voor het genereren van conceptteksten en het uitwerken van theoretische en empirische analyses. In een eerste fase werd AI ingezet om op basis van het ontwikkelde theoretische raamwerk van tien evolutionair-biologische randvoorwaarden snel grote hoeveelheden tekst te produceren. Deze werkwijze maakte het mogelijk om in korte tijd een omvangrijk corpus van conceptuele uitwerkingen en case-analyses te genereren.

De snelheid van deze aanpak ging echter gepaard met aanzienlijke kwaliteitsproblemen. De gegenereerde teksten vertoonden regelmatig een militante of polemische toon die niet past binnen academische conventies. Daarnaast waren duidelijke sporen van confirmation bias zichtbaar: argumenten die het theoretische raamwerk ondersteunden werden sterk benadrukt, terwijl tegenargumenten en onzekerheden relatief weinig aandacht kregen. Een derde probleem was de omvang van de gegenereerde teksten, die vaak aanzienlijk groter was dan nodig voor een academische analyse.

Om deze problemen te corrigeren werd een iteratieve werkwijze ontwikkeld waarin meerdere taalmodellen werden ingezet. Eén model werd primair gebruikt voor het genereren van conceptteksten en analytische uitwerkingen, terwijl een tweede model werd ingezet voor fact-checking en kritische beoordeling van de gegenereerde claims. De bevindingen uit deze controlefase werden vervolgens teruggekoppeld naar het generatieve model, waarna teksten opnieuw werden herschreven en aangescherpt.

Dit proces werd meerdere keren herhaald totdat de belangrijkste problemen—militante formuleringen, ongeclausuleerde claims en feitelijke onnauwkeurigheden—waren gecorrigeerd. In deze iteratieve cyclus speelde de onderzoeker een centrale rol bij het beoordelen van de uitkomsten, het formuleren van correcties en het expliciet clausuleren van claims. Het gebruik van meerdere taalmodellen maakte het mogelijk om dezelfde tekst vanuit verschillende perspectieven te laten analyseren en zo stapsgewijs tot een academisch verantwoorde formulering te komen.

De totstandkoming van het proefschrift kan daarom worden beschreven als een iteratief proces waarin generatieve AI werd gebruikt voor snelle tekstproductie, aanvullende modellen werden ingezet voor controle en kritiek, en de onderzoeker verantwoordelijk bleef voor de uiteindelijke selectie en redactie van de argumentatie.

15.3 Specifieke inzichten, impact op de wetenschapsbeoefening binnen de antropologie.

De implicaties van deze casus voor het antropologische archetype moeten worden begrepen tegen de achtergrond van de bijzondere positie die dit archetype inneemt binnen de taxonomie van wetenschappelijke onderzoekspraktijken die in dit meta-proefschrift wordt onderscheiden. Waar sommige archetypen relatief scherp afgebakend zijn door hun methodologische instrumentarium—bijvoorbeeld experimentele of kwantitatieve onderzoeksstrategieën—heeft het antropologische archetype een minder eenduidig karakter. Het omvat een breed spectrum van disciplines waarin de interpretatie van menselijk gedrag, sociale patronen en culturele betekenissen centraal staat. Naast klassieke culturele antropologie vallen daar ook delen van sociologie, bestuurskunde, cultuurgeschiedenis en politieke antropologie onder. Wat deze disciplines gemeen hebben is niet zozeer een uniform methodologisch protocol, maar een gedeelde epistemische oriëntatie: zij trachten sociale verschijnselen te begrijpen door interpretatie van context, betekenis en gedragslogica.

Het proefschrift *Besturen binnen biologische grenzen* moet tegen deze achtergrond worden gepositioneerd. Het onderzoek behoort niet tot de traditie van klassieke antropologische veldstudies waarin langdurige participerende observatie en etnografische beschrijving centraal staan. Er is geen veldwerk verricht in de conventionele antropologische zin van het woord. In plaats daarvan is het onderzoek opgezet als een analytisch en comparatief project waarin beleidsconflicten worden geïnterpreteerd in het licht van een theoretisch kader van evolutionair-biologische mechanismen. De onderzochte casussen functioneren daarbij niet als etnografische veldnotities, maar als analytische

vensters op terugkerende patronen van menselijk gedrag binnen institutionele contexten. In die zin sluit het onderzoek eerder aan bij tradities zoals analytische antropologie, evolutionaire antropologie en historisch-comparatieve studies naar sociale instituties. Het centrale object van analyse blijft echter onmiskenbaar antropologisch: het gedrag van mensen in groepen en de biologische en sociale mechanismen die dat gedrag mede vormgeven.

Juist binnen deze interpretatieve onderzoekspraktijk wordt een klassieke epistemische kwetsbaarheid zichtbaar. Disciplines die primair werken met interpretatie en betekenisgeving zijn traditioneel gevoelig voor drie nauw met elkaar verbonden risico's: selectieve interpretatie, narratieve coherentie en confirmation bias. Onderzoekers construeren verklarende verhalen over sociale fenomenen en proberen patronen van gedrag te begrijpen binnen een breder interpretatief raamwerk. Dat proces is onvermijdelijk selectief. Empirische observaties worden geïnterpreteerd in het licht van theoretische verwachtingen, en verklaringen krijgen vaak de vorm van een samenhangend narratief waarin afzonderlijke gebeurtenissen logisch met elkaar verbonden worden. Deze narratieve vorm is niet alleen een retorisch hulpmiddel; zij vormt een fundamenteel instrument waarmee complexe sociale realiteiten begrijpelijk worden gemaakt. Tegelijkertijd creëert zij een structureel risico dat tegenstrijdige observaties onderbelicht blijven en dat interpretaties geleidelijk verschuiven in de richting van bevestiging van het gekozen theoretische kader.

De inzet van generatieve AI kan deze kwetsbaarheden versterken. Grote taalmodellen zijn ontworpen om coherente en plausibele tekst te produceren op basis van statistische patronen in taal. Zij optimaliseren voor consistentie, samenhang en overtuigingskracht. Wanneer dergelijke systemen worden gebruikt om interpretatieve analyses uit te werken, bestaat de neiging dat impliciete aannames in een prompt worden bevestigd en verder uitgewerkt tot een coherent narratief. Tegenargumenten worden minder vanzelfsprekend gegenereerd, terwijl argumentatieve lijnen die aansluiten bij het oorspronkelijke kader juist worden uitgebreid en verfijnd. In interpretatieve disciplines kan dit leiden tot een versnelling van precies de mechanismen die reeds als epistemisch risico bekend staan: narratieven worden steeds consistenter, confirmation bias wordt versterkt en alternatieve verklaringen verdwijnen naar de achtergrond.

Deze observatie heeft belangrijke implicaties voor antropologisch georiënteerde onderzoekspraktijken. Zij suggereert dat generatieve AI binnen interpretatieve disciplines niet uitsluitend moet worden gezien als een instrument voor versnelling van tekstproductie, maar ook als een factor die bestaande epistemische risico's kan vergroten. Wanneer AI-systemen worden ingezet zonder aanvullende correctiemechanismen, bestaat het gevaar dat interpretatieve analyses steeds overtuigender worden geformuleerd terwijl hun empirische onderbouwing niet in dezelfde mate wordt versterkt. Voor disciplines waarin betekenisinterpretatie centraal staat, betekent dit dat het gebruik van generatieve AI gepaard moet gaan met een verhoogde aandacht voor kritische tegenlezing, expliciete clausulering van claims en systematische confrontatie met alternatieve interpretaties.

De casus laat daarmee zien dat AI-ondersteunde kennisproductie binnen het antropologische archetype niet alleen nieuwe mogelijkheden opent, maar ook nieuwe vormen van epistemische discipline vereist. Waar traditionele antropologische methoden hun betrouwbaarheid mede ontlenen aan langdurige confrontatie met empirische realiteit—bijvoorbeeld via veldwerk en participerende observatie—kan een AI-gestuurd schrijfproces een zekere afstand tot die empirische frictie creëren. Het interpretatieve narratief kan zich dan sneller ontwikkelen dan de empirische toetsing ervan. Het bewust organiseren van kritische correctiemechanismen wordt in dat geval geen optionele kwaliteitsverbetering, maar een noodzakelijke voorwaarde om de interpretatieve integriteit van het onderzoek te behouden.

15.4 Kerninzicht uit de casus: de architectuur van AI-gestuurde kennisproductie

De analyse van deze casus wijst op een centraal methodologisch inzicht: de betrouwbaarheid van AI-ondersteunde kennisproductie wordt in hoge mate bepaald door de architectuur van het onderzoeksproces. Wanneer generatieve AI wordt ingezet als een enkelvoudig systeem dat tegelijkertijd produceert, interpreteert en evalueert, ontstaat een structurele neiging tot retorische versterking van bestaande aannames. Argumentatieve lijnen worden uitgebreid, alternatieve interpretaties verdwijnen naar de achtergrond en confirmation bias kan zich ongemerkt opstapelen.

In deze studie werd dit probleem gedeeltelijk ondervangen door een workflow waarin verschillende taalmodellen afzonderlijke rollen vervulden. Generatieve modellen werden gebruikt voor conceptuele exploratie en tekstproductie, terwijl andere modellen werden ingezet voor kritische evaluatie, fact-checking en herschrijving. Hierdoor ontstond een vorm van interne tegenlezing die functioneel vergelijkbaar is met collegiale kritiek in traditionele wetenschap.

Het kerninzicht dat uit deze werkwijze naar voren komt, is dat AI-gestuurde kennisproductie niet primair een kwestie is van de kwaliteit van individuele modellen, maar van de organisatorische structuur waarin zij opereren. Een multi-agent architectuur kan fungeren als een mechanisme dat interne epistemische controle organiseert binnen het onderzoeksproces zelf.

15.5 Impact op het analytisch framework

Voor het analysekader van dit meta-proefschrift betekent deze casus dat generatieve AI niet uitsluitend moet worden opgevat als een instrument voor afzonderlijke taken, maar als een potentiële infrastructuur voor kennisproductie. De empirische ervaring suggereert dat de kwaliteit van AI-ondersteund onderzoek minder afhankelijk is van modelcapaciteit dan van de wijze waarop generatieve en kritische functies binnen het onderzoeksproces worden georganiseerd.

De multi-agent architectuur die in deze casus werd ontwikkeld laat zien dat verschillende modellen verschillende epistemische rollen kunnen vervullen, variërend van generatieve exploratie tot kritische controle. Binnen een dergelijke structuur fungeert de onderzoeker niet alleen als auteur, maar ook als curator van een hybride cognitief systeem waarin menselijke en artificiële actoren elkaar controleren en aanvullen.

Daarmee verschuift de analytische focus van individuele AI-capaciteiten naar de organisatie van interacties tussen kennisproducerende systemen. Niet de kracht van één model, maar de configuratie van rollen en controlemechanismen bepaalt uiteindelijk de epistemische kwaliteit van het onderzoeksresultaat. E.e.a. wordt ook duidelijk als de impact van de voorgaande analyses op de beantwoording van de empirische onderzoeksvragen in ogenschouw wordt genomen (tabel 15.5).

Empirische onderzoeksvraag:	Relevante observaties uit de casus:	Impact op beantwoording:
ERQ1 – Cognitief-methodologische dimensie: Welke kansen, risico's en uitdagingen biedt AI voor de kwaliteit van kennisproductie?	De casus laat zien dat generatieve AI de snelheid en schaal van theoretische exploratie sterk kan vergroten, maar tegelijkertijd structurele risico's introduceert zoals confirmation bias, narratieve over-coherentie en tekstinflatie. De toepassing van een multi-agent workflow met gescheiden generatieve en kritische rollen fungeerde als correctiemechanisme.	**Substantiële bijdrage.** De casus bevestigt dat AI zowel productiviteitswinst als nieuwe epistemische risico's creëert, en suggereert dat organisatorische architectuur (multi-agent controle) een belangrijke voorwaarde is voor kwaliteitsborging.
ERQ2 – Pedagogisch-institutionele dimensie: Hoe verandert AI de rollen en structuren in academisch onderzoek?	Het onderzoeksproces ontwikkelde zich van lineaire AI-tekstproductie naar een workflow waarin verschillende modellen verschillende epistemische functies vervullen. De onderzoeker fungeerde daarbij als curator die de interactie tussen systemen organiseerde en de uiteindelijke epistemische verantwoordelijkheid droeg.	**Beperkte maar duidelijke bijdrage.** De casus illustreert een verschuiving van auteur naar procesarchitect, maar biedt geen systematische analyse van institutionele consequenties binnen academische organisaties.
ERQ3 – Macht en normativiteit: Welke machtsverhoudingen en normatieve keuzes zijn ingebed in AI-kennisproductie?	De casus toont dat interpretatieve kaders en casusselectie normatieve keuzes bevatten die door AI worden versterkt via narratieve coherentie. Tegelijkertijd maakt de uitgebreide documentatie van het schrijfproces een relatief hoge mate van transparantie mogelijk	**Illustratieve bijdrage.** De casus maakt normatieve dimensies zichtbaar, maar analyseert machtsverhoudingen rond AI-systemen niet systematisch

Tabel 15.5 – Impact casus op beantwoording empirische onderzoeksvragen.

.

Hoofdstuk 16 - Casus Polyglotte Taalverwerving.

In dit hoofdstuk wordt het gebruik van AI bij de vorming van het proefschrift '*Polyglot-leren in het AI-tijdperk, ontwerp en beproeving van een AI-gestuurd, simultaan leertraject voor de verwerving van verwante talen'* geanalyseerd. De analyse is gebaseerd op het proefschrift zelf en het gelijknamige reflectie-verslag waarin de ontstaansgeschiedenis is gedocumenteerd.

16.1 Positionering van de casus

Het proefschrift *Polyglot-learning in een AI-tijdperk* vertegenwoordigt binnen dit meta-proefschrift het ontwerpgerichte archetype. In tegenstelling tot onderzoekstradities die primair gericht zijn op analyse of interpretatie van bestaande empirische fenomenen, staat in ontwerpgericht onderzoek de ontwikkeling van een concreet artefact centraal: een interventie, methode of systeem dat een praktisch probleem moet oplossen. Kennis ontstaat daarbij niet uitsluitend door observatie of interpretatie, maar door het ontwerpen, testen en verfijnen van een werkbare oplossing. Het onderzoeksobject in deze casus is daarom geen historisch corpus of sociaal fenomeen, maar een didactisch systeem: een leerontwerp dat simultane verwerving van meerdere verwante talen mogelijk moet maken.

In het proefschrift wordt een leertraject ontwikkeld waarin drie Romaanse talen – Spaans, Italiaans en Portugees – gelijktijdig worden aangeleerd met als doel binnen een relatief kort traject een functioneel niveau van taalbeheersing te bereiken (ongeveer CEFR-niveau A2). Het ontwerp maakt gebruik van de structurele overeenkomsten tussen deze talen. Door semantisch parallelle zinnen en vergelijkbare grammaticale patronen naast elkaar te presenteren, worden overeenkomsten en verschillen systematisch zichtbaar gemaakt. Deze interlinguale vergelijking vormt een belangrijk leermechanisme: leerlingen worden aangemoedigd om patronen tussen talen te herkennen en actief te reflecteren op de relatie tussen woordvorm, betekenis en grammaticale structuur.

Het leertraject combineert drie didactische principes. Ten eerste wordt gebruikgemaakt van frequente herhaling van betekenisvolle zinnen, waardoor grammaticale patronen impliciet zichtbaar worden. Ten tweede worden zinnen systematisch in meerdere talen gepresenteerd, zodat structurele overeenkomsten tussen verwante talen kunnen worden benut. Ten derde speelt metalinguïstische reflectie een centrale rol: leerlingen worden expliciet uitgenodigd om verschillen en overeenkomsten tussen talen te analyseren. De onderliggende gedachte is dat taalverwerving kan plaatsvinden via inductieve patroonherkenning in betekenisvolle input, een benadering die aansluit bij zowel onderzoek naar natuurlijke taalverwerving als bij de manier waarop moderne taalmodellen statistische patronen in taaldata leren herkennen.

Het onderzoeksproject heeft een gelaagde opbouw. Naast het proefschrift zelf omvat het project een leerboek waarin het didactische ontwerp concreet wordt uitgewerkt en een reflectieverslag waarin het gebruik van AI tijdens het ontwerp- en schrijfproces wordt gedocumenteerd. De casus biedt daardoor inzicht in de manier waarop generatieve AI kan bijdragen aan het ontwerpen en operationaliseren van complexe interventies en vormt daarmee een empirische illustratie van AI-gestuurde kennisproductie binnen het ontwerpgerichte archetype.

16.2 Rol van AI in het onderzoeksproces

Bij de totstandkoming van dit proefschrift speelde generatieve AI een centrale rol in zowel het ontwerp van het leertraject als in de productie van de academische tekst. In tegenstelling tot de historische en antropologische casussen in dit meta-proefschrift, waar AI primair werd ingezet als analytisch hulpmiddel bij het interpreteren van bestaande gegevens, fungeerde AI hier als onderdeel van het ontwerpproces zelf. Het model werd gebruikt bij het ontwikkelen van didactische concepten, het structureren van het leertraject en het genereren van verschillende vormen van leermateriaal.

Binnen het leerontwerp vervult AI meerdere functies. Het wordt gebruikt voor het genereren van oefenmateriaal, het ondersteunen van vertalingen tussen talen, het begeleiden van uitspraaktraining en het faciliteren van geheugen- en herhalingsoefeningen. Daarmee fungeert AI niet alleen als hulpmiddel bij het ontwerpen van het leertraject, maar ook als een component van het didactische systeem dat in het proefschrift wordt ontwikkeld. De technologie ondersteunt dus zowel het ontwerpen als het functioneren van het leertraject.

Daarnaast werd generatieve AI intensief gebruikt in het schrijfproces van het proefschrift zelf. Conceptuele uitwerkingen, tekststructuren en literatuursyntheses kwamen tot stand via iteratieve interactie tussen de onderzoeker en verschillende taalmodellen. Het schrijfproces kreeg daardoor het karakter van een dialoog waarin voorstellen voor formulering, structuur en argumentatie herhaaldelijk werden gegenereerd, beoordeeld en aangepast. De menselijke onderzoeker fungeerde in dit proces als curator die de structuur van het betoog bepaalde, bruikbare formuleringen selecteerde en de uiteindelijke tekst redigeerde.

Het onderzoeksproces ontwikkelde zich daarbij tot een multi-agent configuratie waarin meerdere taalmodellen complementaire rollen vervulden. Verschillende systemen werden ingezet voor uiteenlopende taken, zoals generatieve exploratie, tekstproductie en kritische evaluatie van de gegenereerde output. Deze werkwijze had ook een belangrijke methodologische consequentie. De curator van het project was zelf geen specialist in de betrokken talen. In een conventionele onderzoekssetting zou dit vrijwel onvermijdelijk hebben geleid tot het inschakelen van een externe taalexpert om de gegenereerde output te controleren. Binnen de multi-agent configuratie kon deze rol echter gedeeltelijk worden vervuld door een tweede taalmodel dat systematisch werd ingezet voor controle, correctie en herformulering van gegenereerde teksten.

De totstandkoming van het proefschrift kan daarom het best worden begrepen als een iteratief proces van mens–AI–AI interactie. Generatieve modellen produceerden voorstellen voor tekst en ontwerpstappen, andere modellen leverden kritische tegenlezing en correcties, en de menselijke curator bleef verantwoordelijk voor de selectie, interpretatie en uiteindelijke integratie van deze bijdragen in een coherent geheel. Het resultaat was een hybride onderzoeksproces waarin menselijke regie en artificiële generativiteit elkaar voortdurend aanvulden.

16.3 Kern-inzichten

De analyse van deze casus wijst op een fundamenteel verschil tussen de rol van generatieve AI in ontwerpgericht onderzoek en haar rol in andere onderzoeksarchetypen. Waar AI in historische en interpretatieve disciplines vooral wordt ingezet voor het analyseren, ordenen en interpreteren van bestaande corpora, wordt zij in ontwerpgericht onderzoek onderdeel van een proces waarin nieuwe

artefacten worden ontwikkeld. Het centrale vraagstuk verschuift daarmee van interpretatie naar productie en stabilisatie van een ontworpen systeem.

De casus laat zien dat generatieve AI bijzonder krachtig kan zijn in de vroege fasen van een ontwerpproces. Modellen blijken goed in staat om conceptuele varianten te genereren, mogelijke structuren voor een leertraject te verkennen en didactische ideeën snel uit te werken in concrete voorbeelden. Hierdoor kan het ontwerpproces aanzienlijk worden versneld: hypothesen, varianten en alternatieve oplossingen kunnen in korte tijd worden geformuleerd en geëvalueerd.

Tijdens het ontwerpproces werd zichtbaar dat het genereren van geschikt oefenmateriaal aanzienlijk moeilijker was dan aanvankelijk werd verwacht. Voor het leertraject moesten grote aantallen voorbeeldzinnen worden geproduceerd waarin grammaticale patronen duidelijk herkenbaar blijven en waarin de betekenis van elke zin voldoende specifiek is om binnen een oefening slechts één plausibele interpretatie toe te laten. Wanneer dergelijke zinnen in grotere batches werden gegenereerd, bleek de output geleidelijk af te wijken van deze randvoorwaarden. De zinnen bleven grammaticaal correct, maar verloren vaak hun didactische scherpte: contexten werden te algemeen, meerdere interpretaties werden mogelijk, of combinaties van woorden bleken semantisch weinig betekenisvol.

Deze ervaring maakte duidelijk dat het genereren van afzonderlijke correcte zinnen relatief eenvoudig is, maar dat het produceren van grote hoeveelheden didactisch gecontroleerd materiaal een andere en aanzienlijk complexere opgave vormt. Voor het leerontwerp was namelijk niet alleen grammaticale correctheid vereist, maar ook semantische precisie, structurele variatie en interlinguale vergelijkbaarheid tussen drie talen. Het bewaken van deze voorwaarden over langere generaties bleek niet automatisch uit het modelgedrag voort te komen en moest daarom via aanvullende ontwerpmaatregelen worden georganiseerd.

De casus suggereert daarmee dat de integratie van AI in ontwerpgerichte disciplines niet alleen nieuwe mogelijkheden creëert, maar ook een nieuw type methodologisch probleem introduceert. Waar interpretatieve disciplines vooral geconfronteerd worden met vragen rond validiteit en interpretatie van AI-gegenereerde analyses, verschuift de uitdaging in ontwerpgericht onderzoek naar de controleerbaarheid en stabiliteit van generatieve productieprocessen. AI kan het ontwerpproces aanzienlijk versnellen, maar maakt tegelijkertijd zichtbaar dat het genereren van consistente en didactisch bruikbare artefacten een eigen vorm van ontwerpcomplexiteit introduceert.

16.4 Specifieke inzichten uit deze casus: de weerbarstigheid van AI-ondersteund ontwerp

De uitvoering van het leerontwerp maakte zichtbaar dat de inzet van generatieve AI in ontwerpgericht onderzoek niet alleen nieuwe mogelijkheden creëert, maar ook nieuwe vormen van praktische complexiteit introduceert. De oorspronkelijke didactische uitgangspunten van het leertraject — grammaticaloos, inductief en simultaan leren van verwante talen — bleken op conceptueel niveau goed te formuleren. De moeilijkheid ontstond pas bij de vertaling van deze uitgangspunten naar grootschalig, semantisch stabiel en didactisch controleerbaar oefenmateriaal. Juist in deze fase werd duidelijk hoe weerbarstig AI-ondersteund ontwerp kan zijn.

Een eerste probleem betrof het verschijnsel dat in het proefschrift wordt aangeduid als *semantische drift*. Wanneer taalmodellen grotere hoeveelheden oefenzinnen in één generatieproces produceerden, bleek de output geleidelijk af te wijken van de oorspronkelijke instructie. Deze afwijking manifesteerde zich niet

als duidelijke fouten, maar als een opeenstapeling van kleine verschuivingen. Zinnen bleven grammaticaal correct, maar verloren hun didactische precisie: contexten werden generieker, variatie werd minder functioneel en soms verschenen placeholder-achtige formuleringen. Het probleem werd daardoor vaak pas zichtbaar wanneer grotere hoeveelheden materiaal als geheel werden bekeken. Voor het ontwerpproces betekende dit dat generatieve productie niet onbeperkt kon worden opgeschaald, maar actief moest worden begrensd en gecontroleerd.

Een tweede bron van weerbarstigheid lag in het ontwerpen van zogenaamde micro-contexten. Voor het leerontwerp moesten zinnen worden geconstrueerd waarin een specifieke betekenis zodanig wordt begrensd dat binnen een cloze-oefening slechts één plausibele invulling mogelijk blijft. Contexten die te open waren lieten meerdere interpretaties toe, terwijl contexten die te kunstmatig werden geformuleerd hun talige geloofwaardigheid verloren. Deze spanning werd versterkt door de meertalige opzet van het leertraject: contexten moesten niet alleen eenduidig zijn in één taal, maar ook zonder betekenisverschuiving realiseerbaar blijven in de andere doeltalen.

Daarnaast trad een derde type probleem op dat in het proefschrift wordt beschreven als *grammaticale nonsens*. Sommige door het model gegenereerde zinnen waren syntactisch correct maar semantisch leeg of incoherent. In een ontwerpsituatie waarin grammaticale patronen via betekenisvolle voorbeelden moeten worden geleerd, zijn dergelijke zinnen didactisch onbruikbaar. Het probleem was hier niet grammaticale incorrectheid, maar het ontbreken van semantische selectiebeperkingen die bepalen welke combinaties van woorden betekenisvol zijn.

De tri-linguale opzet van het leerontwerp introduceerde bovendien een extra bron van ontwerpcomplexiteit. In eerste instantie werd geprobeerd semantisch equivalente zinnen in drie talen parallel te laten genereren. In de praktijk bleek echter dat semantische equivalentie in dergelijke generaties gemakkelijk werd verondersteld in plaats van gecontroleerd. Kleine verschuivingen in betekenis of pragmatische nadruk konden ertoe leiden dat zinnen formeel vergelijkbaar leken maar functioneel uiteenliepen. Voor een leerontwerp dat juist gebaseerd is op systematische vergelijking tussen talen vormde dit een probleem. Daarom werd uiteindelijk gekozen voor een seriële afleidingsstructuur waarbij één referentiezin als semantisch anker fungeerde en de overeenkomstige zinnen in de andere talen daarvan werden afgeleid.

Ten slotte bleek ook de schaal van het ontwerpproces zelf een uitdaging. De combinatie van meerdere talen, verschillende woordsoorten en variatie-assen leidde tot een snelle toename van mogelijke zinnen en varianten. Deze combinatorische groei maakte het moeilijk om overzicht, samenhang en semantische consistentie te behouden. Zonder expliciete ontwerpmaatregelen dreigde het leerontwerp daardoor praktisch onbeheersbaar te worden.

De oplossing voor deze problemen werd niet gevonden in het aanpassen van de didactische uitgangspunten, maar in het herontwerpen van het productieproces. Generatie van oefenmateriaal werd opgesplitst in kleine micro-batches, tri-linguale zinnen werden serieel afgeleid van een semantisch anker in plaats van parallel gegenereerd en variatie werd georganiseerd langs expliciet gedefinieerde assen met duidelijke stopregels. Deze maatregelen maakten het mogelijk om semantische drift te beperken en de controleerbaarheid van het ontwerpproces te vergroten.

Voor ontwerpgerichte onderzoekspraktijken heeft deze casus een bredere betekenis. Generatieve AI maakt het mogelijk om ontwerpideeën, varianten en didactische structuren snel te verkennen. In deze studie bleek echter dat de grootste praktische uitdaging niet lag in het formuleren van ontwerpprincipes, maar in het produceren van stabiel en controleerbaar oefenmateriaal op grotere schaal. Het ontwerpproces verschoof daardoor van conceptontwikkeling naar beheersing van generatieve productieprocessen.

Deze verschuiving is ook relevant vanuit een ontwerpwetenschappelijk perspectief. In design-science onderzoek ontstaat kennis via het ontwerpen en evalueren van artefacten. De casus laat zien dat

generatieve AI deze cyclus niet zozeer vervangt, maar een nieuwe ontwerpvraag introduceert: hoe kunnen generatieve processen zodanig worden ingericht dat artefacten reproduceerbaar, semantisch stabiel en didactisch bruikbaar blijven? Ontwerpgericht onderzoek krijgt daarmee een aanvullende taak, namelijk het ontwerpen van de productiearchitectuur waarbinnen AI-gegenereerde artefacten betrouwbaar tot stand kunnen komen.

16.5 Impact op het analytisch framework

De analyse van deze casus heeft consequenties voor het analysekader van dit meta-proefschrift, met name voor de manier waarop de rol van AI in ontwerpgerichte onderzoekspraktijken wordt begrepen. Waar in historische en interpretatieve disciplines AI vooral wordt ingezet voor het analyseren en structureren van bestaande corpora, wordt zij in ontwerpgericht onderzoek onderdeel van een proces waarin nieuwe artefacten worden ontwikkeld. De empirische ervaring van deze casus laat zien dat de inzet van generatieve AI het ontwerpproces op twee verschillende manieren beïnvloedt. Enerzijds kan AI de vroege fasen van het ontwerpproces aanzienlijk versnellen door snelle exploratie van varianten, structuren en voorbeeldmateriaal mogelijk te maken. Anderzijds introduceert zij een nieuw type praktische uitdaging: het waarborgen van semantische stabiliteit, consistentie en controleerbaarheid in grootschalige generatieve productie.

Voor het analysekader betekent dit dat AI in ontwerpgerichte disciplines niet uitsluitend moet worden gezien als hulpmiddel voor kennisproductie, maar ook als een technologie die nieuwe eisen stelt aan de organisatie van het ontwerpproces zelf. Ontwerpgericht onderzoek krijgt daardoor een aanvullende dimensie: naast het ontwerpen van het artefact moet ook de generatieve productieomgeving worden ontworpen waarin dat artefact tot stand komt. De casus suggereert daarmee dat de praktische impact van AI in ontwerpwetenschappen vooral zichtbaar wordt in de noodzaak om generatieve processen expliciet te structureren en te controleren, zodat de geproduceerde artefacten reproduceerbaar en functioneel bruikbaar blijven.

De bijdrage van deze casus aan de empirische onderzoeksvragen van dit meta-proefschrift kan als volgt worden samengevat (tabel 16.5).

Empirische onderzoeksvraag	**Relevante observaties uit de casus**	**Impact op beantwoording**
ERQ1 – Cognitief-methodologische dimensie: Welke kansen, risico's en uitdagingen biedt AI voor de kwaliteit van kennisproductie?	Generatieve AI versnelt de exploratie van ontwerpvarianten en didactische structuren, maar introduceert tegelijk risico's rond semantische drift, inconsistentie en verlies van didactische precisie bij grootschalige generatie van oefenmateriaal.	**Duidelijke bijdrage.** De casus laat zien dat AI in ontwerpgericht onderzoek zowel ontwerp-exploratie versnelt als nieuwe kwaliteitsproblemen kan introduceren die aanvullende ontwerpmaatregelen vereisen.
ERQ2 – Pedagogisch-institutionele dimensie: Hoe verandert AI de rollen en structuren in academisch onderzoek?	Het ontwerpproces ontwikkelde zich tot een multi-agent configuratie waarin verschillende taalmodellen complementaire rollen vervulden bij generatie en controle van materiaal. Hierdoor kon inhoudelijke controle deels plaatsvinden zonder directe materiedeskundigheid van de curator.	**Illustratieve bijdrage.** De casus toont hoe AI-systemen binnen ontwerpgericht onderzoek verschillende epistemische rollen kunnen vervullen, terwijl de onderzoeker verantwoordelijk blijft voor structuur, selectie en integratie.

Empirische onderzoeksvraag	Relevante observaties uit de casus	Impact op beantwoording
ERQ3 – Macht en normativiteit: Welke machtsverhoudingen en normatieve keuzes zijn ingebed in AI-kennisproductie?	In deze casus staat het ontwerpen en operationaliseren van een didactisch systeem centraal. De analyse richt zich op praktische ontwerpvraagstukken rond generatieve productie en didactische stabiliteit. Breder normatieve vragen over macht, toegang of institutionele controle over AI-systemen spelen in dit ontwerpproces geen expliciete rol.	**Geen substantiële bijdrage.** Deze casus levert geen direct empirisch materiaal voor de beantwoording van ERQ3.

Tabel 16.5: Impact op de beantwoording van de empirische onderzoeksvragen.

Hoofdstuk 17 – Casus ‘AI-gestuurde kennisproductie’, de mise-en-abyme.

In dit hoofdstuk wordt het gebruik van AI bij de totstandkoming van het proefschrift *AI-gestuurde kennisproductie* zelf geanalyseerd. Daarmee krijgt deze casus een reflexief karakter: onderzoeksobject en onderzoeksinstrument vallen gedeeltelijk samen. De analyse is gebaseerd op het proefschrift en het bijbehorende reflectieverslag waarin de ontstaansgeschiedenis van het project is gedocumenteerd.

17.1 Positionering van de casus

Deze casus vertegenwoordigt het meta-domein binnen de taxonomie van onderzoeksarchetypen die in dit proefschrift wordt onderscheiden. Waar de voorgaande casussen betrekking hadden op historische, antropologische, ontwerpgerichte of natuurwetenschappelijke onderzoekspraktijken, richt deze studie zich op AI-gestuurde kennisproductie zelf. Het object van analyse is daarmee niet een extern empirisch fenomeen, maar het proces van AI-ondersteunde wetenschap. Deze configuratie creëert een vorm van *mise-en-abyme*: generatieve AI wordt gebruikt om een studie te produceren die juist de rol van generatieve AI in wetenschappelijke kennisproductie onderzoekt. Daardoor ontstaat een reflexieve onderzoeksconfiguratie waarin onderzoeksinstrument, onderzoeksproces en onderzoeksobject gedeeltelijk samenvallen. Binnen dit meta-domein staat daarom niet alleen de inhoud van het onderzoek centraal, maar ook de methodologische voorwaarden waaronder AI-gestuurde kennisproductie epistemisch betrouwbaar kan blijven.

17.2 Rol van AI in het onderzoeksproces

Bij de totstandkoming van dit proefschrift werd generatieve AI intensief ingezet voor het genereren, structureren en herschrijven van tekstvoorstellen. Opvallend was dat een eerste volledige conceptversie van het manuscript in uitzonderlijk korte tijd tot stand kwam. Volgens het reflectieverslag werd versie 1.0 van het proefschrift in twee dagen gegenereerd binnen één lange interactie (chat-sessie) tussen curator en taalmodel. De resulterende tekst had direct een academisch niveau dat – in beginsel - hoog genoeg was om het manuscript voor peer review te kunnen aanbieden.

Deze korte doorlooptijd moet echter niet worden geïnterpreteerd als een structurele eigenschap van het meta-domein. Het reflectieverslag schrijft de snelheid primair toe aan twee factoren. Ten eerste maakte het gebruikte taalmodel zeer lange, doorlopende interactiesessies mogelijk, waardoor het volledige manuscript binnen één gesprek kon worden ontwikkeld zonder dat het generatieve proces voortdurend moest worden onderbroken of opnieuw opgebouwd. Ten tweede speelde de leercurve van de curator een belangrijke rol: bij dit derde proefschrift binnen het onderzoeksprogramma was inmiddels duidelijk hoe generatieve AI effectief kon worden aangestuurd en hoe een generatief schrijfproces architectonisch moest worden georganiseerd. De uitzonderlijke snelheid van deze eerste versie is daarom niet zonder meer generaliseerbaar naar andere onderzoeksprojecten.

In latere fasen van het schrijfproces werd bovendien een expliciete multi-agent-structuur gehanteerd. Verschillende modellen werden ingezet voor verschillende rollen, bijvoorbeeld voor tekstgeneratie, kritische herlezing en structurele revisie. De curator bleef daarbij verantwoordelijk voor de selectie, interpretatie en validatie van de uiteindelijke formuleringen.

17.3 Specifieke inzichten uit deze casus

Het reflexieve karakter van het meta-domein creëert een methodologische gevoeligheid die in de andere archetypen minder sterk aanwezig is: het risico op circulaire validatie. Omdat generatieve AI hier niet alleen wordt gebruikt als instrument maar ook onderwerp van analyse is, bestaat in principe de mogelijkheid dat het systeem bijdraagt aan de formulering van het analysekader waarmee zijn eigen rol wordt beoordeeld.

In de praktijk manifesteerde dit risico zich in deze casus echter niet. De interactie tussen curator en taalmodel verliep via iteratieve generaties van afzonderlijke paragrafen, waarbij elke iteratie werd afgesloten met een expliciete go/no-go-beslissing van de curator. Daardoor bleef de epistemische asymmetrie tussen mens en model intact: het model genereerde tekstvoorstellen, maar de uiteindelijke validatie en legitimering van de argumentatie bleef bij de onderzoeker.

17.4 Kerninzichten

De belangrijkste les uit deze casus betreft de rol van effectieve curatoriële controle in AI-gestuurde kennisproductie. Het reflexieve karakter van het meta-domein vergroot weliswaar de gevoeligheid voor circulaire validatie, maar de empirische praktijk van deze studie laat zien dat dit risico methodologisch beheersbaar blijft wanneer de curator expliciet verantwoordelijk blijft voor de evaluatie van modeloutput.

Tegelijk laat deze casus zien dat de efficiëntie van AI-gestuurde kennisproductie sterk kan toenemen wanneer de curator ervaring opbouwt met het organiseren van interacties tussen mens en model. Het reflectieverslag interpreteert de uitzonderlijk korte totstandkoming van versie 1.0 daarom niet als een technologische doorbraak, maar als een leercurve-effect: niet de modellen waren wezenlijk beter geworden, maar de curator had geleerd hoe het generatieve proces effectief kon worden gestuurd en begrensd. Hoewel de extreme snelheid van deze casus niet zonder meer generaliseerbaar is naar andere projecten, is de onderliggende observatie dat curatoriële expertise een belangrijke determinant is van efficiëntie en kwaliteit in AI-gestuurde kennisproductie wel degelijk breder toepasbaar.

17.5 Impact op het analytisch framework

De analyse van deze casus heeft een aantal implicaties voor het analysekader van dit meta-proefschrift. Hoewel het meta-domein zelf slechts beperkt nieuwe empirische observaties opleverde ten opzichte van de andere archetypen, maakt het wel enkele structurele eigenschappen van AI-gestuurde kennisproductie expliciet. De volgende bevindingen hebben niet alleen betrekking op deze casus, maar betreffen de grootste gemene deler tussen de vijf cases die in hoofdstuk 13 t/m 17 behandeld zijn.

Een eerste implicatie betreft de rol van architectuur in academische kennisproductie. Generatieve taalmodellen blijken goed in staat om lokaal coherente tekst te produceren, maar hebben tijdens het generatieve proces slechts beperkt zicht op de manier waarop afzonderlijke passages bijdragen aan de globale architectuur van een manuscript. Het model beschikt doorgaans niet over een stabiel intern beeld van de hiërarchie van argumentatieve stappen waaruit een wetenschappelijk betoog is opgebouwd. Wanneer het model achteraf wordt gevraagd een bestaande tekst te analyseren, blijkt het vaak wel in staat om globale argumentatiestructuren te herkennen en te beoordelen. Tijdens het schrijven zelf ontbreekt dat overzicht echter grotendeels. De verantwoordelijkheid voor het ontwerpen en bewaken van de architectuur van een proefschrift blijft daarom noodzakelijk bij de menselijke curator.

Een tweede implicatie betreft de rol van curatoriale regie in het bewaken van coherentie. Technische eigenschappen van taalmodellen kunnen ertoe leiden dat delen van een argumentatieve keten buiten het actieve contextvenster vallen. Wanneer eerdere passages niet langer beschikbaar zijn, kan de samenhang tussen latere tekstdelen en de oorspronkelijke argumentatie geleidelijk vervagen. Daarnaast kan terminologische drift ontstaan wanneer passages opnieuw worden gegenereerd en analytische begrippen worden vervangen door semantisch verwante maar minder precieze termen. Deze verschijnselen maken duidelijk dat generatieve AI niet autonoom de coherentie van een wetenschappelijk betoog kan bewaken.

Een derde implicatie betreft de analyse van literatuurgebruik. In veel discussies over generatieve AI ligt de nadruk op het risico dat taalmodellen fictieve referenties produceren. In de praktijk bleek dit risico in het onderhavige project beperkt. Tegelijkertijd bleek dat generatieve AI nieuwe mogelijkheden opent voor systematische analyse van brongebruik. In dit project werd het gebruik van sleutelreferenties daarom getoetst aan een expliciet raster van controlevragen, waarin onder meer werd nagegaan of een bron binnen haar theoretische reikwijdte werd gebruikt, of centrale concepten correct werden weergegeven en of een referentie daadwerkelijk de claim ondersteunde waarvoor zij werd ingeroepen. De uiteindelijke interpretatie van dergelijke signaleringen bleef daarbij vanzelfsprekend bij de onderzoeker.

Een vierde implicatie betreft de organisatie van AI-ondersteunde onderzoeksprocessen. In de praktijk bleek dat generatieve taalmodellen beter functioneren wanneer de rollen van produceren en evalueren expliciet van elkaar worden gescheiden. Wanneer een model tegelijkertijd wordt ingezet als producent en criticus van tekst, heeft het de neiging om kritiek direct te integreren in een nieuwe tekstversie in plaats van deze analytisch te beoordelen. Een multi-agent-architectuur, waarin verschillende modellen verschillende rollen vervullen, bleek daarom een effectievere werkwijze.

Samen genomen suggereren deze observaties dat de kwaliteit van AI-gestuurde kennisproductie minder wordt bepaald door de individuele capaciteiten van een taalmodel dan door de architectuur van het onderzoeksproces waarin het model wordt ingezet. De totale impact op de beantwoording van de empirische onderzoeksvragen is weergegeven in tabel 17.5.

Empirische onderzoeksvraag	Bijdrage van deze casus
ERQ1 – Cognitief-methodologische dimensie	Beperkte maar relevante bijdrage. Reflexieve AI-studies kennen een verhoogd risico op circulaire validatie, maar dit risico kan worden doorbroken via expliciete curatoriële validatie van modeloutput.
ERQ2 – Pedagogisch-institutionele dimensie	Relevante bijdrage. De casus bevestigt dat de rol van de onderzoeker verschuift van auteur naar curator van een generatief proces.
ERQ3 – Macht en normativiteit	Geen substantiële bijdrage. In deze casus werden geen nieuwe machtsverhoudingen of normatieve spanningen zichtbaar die niet reeds in eerdere casussen aan de orde kwamen.

Tabel 17.5: impact va de casus op beantwoording empirische onderzoeksvragen

17.6 Praktische leerervaringen uit AI-gestuurde kennisproductie

Naast de methodologische implicaties die hierboven zijn besproken, leverde de totstandkoming het vijftal proefschriften ook een reeks praktische leerervaringen op over het werken met generatieve taalmodellen in een academische context. Voor de eigenaar van een academische tekst verandert het werken met generatieve taalmodellen de aard van het schrijfproces ingrijpend. Waar academische teksten in een traditioneel schrijfproces geleidelijk worden opgebouwd en verfijnd, functioneren teksten in een generatief proces eerder als tijdelijke configuraties van een modeloutput. Vier praktische observaties uit dit project illustreren deze verschuiving.

Ten eerste ondermijnt het regeneratiegedrag van taalmodellen het klassieke principe van incrementele revisie. Wanneer een model wordt gevraagd een bestaande passage te corrigeren of aan te passen, behandelt het die passage zelden als een stabiele tekst die slechts lokaal moet worden gewijzigd. In plaats daarvan genereert het model doorgaans een nieuwe variant van de volledige passage. In het project werd dit verschijnsel aangeduid als *GPT-ziekte*: ogenschijnlijk kleine correcties leiden tot volledige herformuleringen van een tekstfragment. Wanneer een passage reeds zorgvuldig is gecureerd kan zo'n regeneratie leiden tot verlies van analytische precisie, verschuiving van terminologie of aantasting van de argumentatieve structuur. Voor de eigenaar van de tekst betekent dit dat elke regeneratie impliciet eerdere curatoriële keuzes kan ondermijnen. De rol van de auteur verschuift daardoor van schrijver naar bewaker van tekststabiliteit.

Ten tweede heeft het parafraserende karakter van taalmodellen gevolgen voor de tekstuele integriteit van academisch werk. Taalmodellen reproduceren passages zelden letterlijk, maar formuleren vrijwel altijd een parafrase van het bronmateriaal. Hierdoor kunnen formuleringen ontstaan die inhoudelijk correct zijn maar niet letterlijk in de oorspronkelijke bron voorkomen. Voor academisch schrijven, waar precisie van formulering en correcte bronverwijzing essentieel zijn, kan dit praktische complicaties veroorzaken. Ook mechanistische bewerkingen van tekst, zoals zoek-en-vervangoperaties bij redactionele revisies, worden hierdoor bemoeilijkt. Voor de curator betekent dit dat brongebruik en citaten niet alleen inhoudelijk, maar ook tekstueel moeten worden gecontroleerd.

Ten derde kan de manier waarop taalmodellen teksten verwerken leiden tot dataskimming. In tegenstelling tot menselijke lezers analyseren modellen documenten voornamelijk via statistische patroonherkenning. Onder bepaalde omstandigheden detecteert het model vooral globale patronen in een tekst zonder alle details even zorgvuldig te verwerken. In de praktijk betekent dit dat een onderzoeker er niet automatisch van kan uitgaan dat een model een document volledig en nauwkeurig heeft gelezen. Voor onderzoek dat afhankelijk is van specifieke formuleringen of contextuele nuances kan dit problematisch zijn.

Ten vierde brengen de beperkte terugkijkvensters van taalmodellen een eigen risico met zich mee voor de samenhang van omvangrijke academische teksten. Taalmodellen beschikken slechts over een begrensd actief contextvenster waarin eerdere tekstfragmenten beschikbaar blijven tijdens het generatieve proces. Naarmate een manuscript groeit, verdwijnen eerdere passages geleidelijk achter deze horizon. Daardoor kan semantische drift ontstaan: latere tekstfragmenten blijven grammaticaal correct en lokaal coherent, maar verliezen hun aansluiting op eerder geformuleerde argumentatieve structuren, causale redeneringen en analytische onderscheidingen. Voor de eigenaar van een academische tekst is dit probleem bijzonder ingrijpend, omdat het vaak niet zichtbaar wordt op zins- of alineaniveau, maar pas op het niveau van het bredere betoog. Juist daarom blijft menselijke bewaking van de rode draad, de argumentatieve hiërarchie en de globale architectuur van het manuscript onmisbaar.

Gezamenlijk impliceren deze observaties dat AI-ondersteund schrijven niet eenvoudig kan worden opgevat als een versnelling van traditionele schrijfprocessen. Het werk van de onderzoeker verschuift van lineaire tekstproductie naar continue curatoriële controle over een generatief systeem dat tekst produceert, maar niet automatisch de stabiliteit, precisie en cumulatieve logica bewaakt die academisch schrijven vereist. In het onderhavige project (het schrijven van dit document, 'AI-gestuurde kennisproductie) werd dit probleem ondervangen door twee vormen van discipline: procesdiscipline — het vermijden van onnodige regeneratie en zijpaden — en structuurdiscipline — het strikt volgen van de vooraf ontworpen architectuur van het manuscript. Juist deze combinatie maakte het mogelijk om langdurig binnen hetzelfde analytische kader te blijven werken zonder noemenswaardige semantische of terminologische drift.

17.7 Tenslotte.

De bijzondere positie van deze casus als mise-en-abyme maakt de centrale bevinding van dit meta-onderzoek op een directe manier zichtbaar. Het proefschrift dat hier wordt geanalyseerd is immers zelf tot stand gekomen via het type AI-gestuurde kennisproductie dat het onderzoekt. In die zin fungeert het manuscript niet alleen als onderzoeksobject, maar ook als empirisch voorbeeld van de voorwaarden waaronder generatieve AI daadwerkelijk kan bijdragen aan academische kennisproductie.

De uitzonderlijk korte totstandkoming van versie 1.0 — gegenereerd in twee dagen binnen één lange interactie tussen curator en taalmodel en direct academisch toonbaar — moet daarom niet primair worden begrepen als een demonstratie van technologische autonomie. Zij illustreert eerder het belang van curatoriële regie en van een strikt georganiseerd productieproces. De combinatie van procesdiscipline — het vermijden van destabiliserende zijpaden en onnodige regeneraties — en structuurdiscipline — het consequent volgen van een vooraf ontworpen architectuur van het manuscript — maakte het mogelijk om langdurig binnen één analytisch kader te blijven werken zonder noemenswaardige semantische of terminologische drift.

In dat opzicht vormt de architectuur van het proefschrift zelf een impliciet bewijs van de centrale these van dit onderzoek: generatieve AI verandert de instrumenten van wetenschappelijke kennisproductie, maar de betrouwbaarheid van het resultaat blijft in laatste instantie afhankelijk van de manier waarop onderzoekers het generatieve proces ontwerpen, disciplineren en cureren.

Deze observatie vormt het uitgangspunt voor het volgende hoofdstuk, waarin de bevindingen uit de vijf onderzochte casussen gezamenlijk worden geanalyseerd en samengebracht in een synthetisch kader voor AI-gestuurde kennisproductie.

Hoofdstuk 18 – Conclusies, discussie.

18.1 Inleiding

In hoofdstuk 12 werd een tweede tussenbalans t.a.v. het analytisch kader dat in dit proefschrift is ontwikkeld, opgemaakt. Op basis van een theoretische analyse van drie dimensies van AI-gestuurde kennisproductie – de cognitief-methodologische, de pedagogisch-institutionele en de normatief-politieke dimensie – werd daar een voorlopig antwoord geformuleerd op de drie empirische onderzoeksvragen. Deze antwoorden waren echter gebaseerd op conceptuele analyse en literatuurstudie en moesten daarom nog empirisch worden getoetst.

In de hoofdstukken 13 tot en met 17 werd deze toetsing uitgevoerd aan de hand van vijf casussen, die elk een verschillend archetype van wetenschappelijke kennisproductie representeren. Deze cases fungeren daarbij niet als illustraties van de theorie, maar als empirische test van het analysekader. Dit hoofdstuk brengt de bevindingen uit deze cases samen en confronteert ze met de voorlopige conclusies uit de tweede tussenbalans. Daarbij wordt een delta-benadering gevolgd: per onderzoeksvraag wordt eerst het voorlopige antwoord uit hoofdstuk 12 in herinnering gebracht, waarna wordt onderzocht hoe dit antwoord verandert wanneer het wordt geconfronteerd met de empirische bevindingen.

In de paragrafen 18.2 t/m 18.4 worden vervolgens de drie empirische onderzoeksvragen beantwoord. In paragraaf 18.5 wordt op basis daarvan het antwoord op de centrale onderzoeksvraag geformuleerd. Daarna worden de methodologische waarborgen van het onderzoek en de aanbevelingen voor vervolgonderzoek besproken.

18.2 ERQ1 – Cognitief-methodologische dimensie

De eerste empirische onderzoeksvraag luidt: ‘welke kansen, risico’s en uitdagingen biedt AI voor de kwaliteit van wetenschappelijke kennisproductie?’. In hoofdstuk 12 werd geconcludeerd dat generatieve AI belangrijke kansen biedt voor het versnellen van analyse en synthese van kennis, maar tegelijk nieuwe risico’s introduceert, zoals hallucinaties, vertekeningen in argumentatie en beperkte transparantie van modellen. De empirische analyse van de vijf casussen bevestigt dit beeld, maar verdiept het aanzienlijk. In meerdere domeinen blijkt generatieve AI omvangrijke literatuurcorpora te kunnen synthetiseren, kennisstructuren te expliciteren en alternatieve interpretaties of hypothesen te genereren. Daarmee kan AI een belangrijke rol spelen in het systematiseren en versnellen van analytische processen.

Tegelijkertijd tonen de casussen een subtieler epistemisch risico dan in hoofdstuk 12 werd verondersteld. Generatieve systemen hebben een sterke neiging om heterogene of onvolledige informatie te reconstrueren tot ogenschijnlijk consistente verklaringsstructuren. Deze neiging kan worden omschreven als *narratieve over-coherentie*. Omdat taalmodellen zijn geoptimaliseerd voor plausibele en vloeiende tekst, reduceren zij vaak ambiguïteiten of onzekerheden die in de onderliggende empirische werkelijkheid wel degelijk aanwezig zijn. Dit mechanisme kan in verschillende onderzoeksdomeinen optreden. In historisch onderzoek kan het zich manifesteren als een versterkte vorm van hindsight bias; in andere velden als lineaire causaliteitsketens die complexere empirische relaties versimpelen. In dergelijke gevallen ontstaat een spanning tussen narratieve plausibiliteit en empirische onzekerheid.

De natuurwetenschappelijke casus die in hoofdstuk 13 werd besproken werpt in dit verband een bijzonder licht op het mechanisme van narratieve over-coherentie. Het zogeheten duivenexperiment was expliciet opgezet als een oefening in wetenschappelijk illusionisme: een proefschrift waarin synthetisch

materiaal werd gepresenteerd binnen de retorische en methodologische conventies van natuurwetenschappelijk onderzoek. Het experiment fungeerde daarmee impliciet als een Turing-test voor academische teksten. Niet de biologische inhoud stond centraal, maar de vraag of een door generatieve AI geconstrueerd onderzoeksnarratief – opgebouwd volgens de herkenbare patronen van hypothesevorming, methodologische verantwoording, literatuurgebruik en voorzichtige interpretatie – door lezers als een plausibel wetenschappelijk betoog zou worden herkend. Het slagen van dit experiment laat zien dat taalmodellen, dankzij hun training op grote corpora van wetenschappelijke publicaties, statistische kennis bezitten van de stilistische en argumentatieve conventies van academisch schrijven. Zij herkennen niet alleen de structuren van wetenschappelijke argumentatie, maar ook de stilistische signalen – soms aangeduid als AI-markers – die in veel discussies worden gebruikt om machine-gegenereerde tekst te identificeren. Daardoor kunnen modellen coherente wetenschappelijke narratieven produceren die voor menselijke lezers moeilijk te onderscheiden zijn van conventioneel geschreven onderzoek. Het duivenexperiment vormt daarmee een empirische demonstratie van het mechanisme dat in deze paragraaf wordt beschreven: de retorische coherentie van een wetenschappelijk narratief kan in AI-gegenereerde teksten soms sterker zijn dan de empirische noodzaak waarop het gebaseerd is

Tegelijk creëren generatieve systemen nieuwe mogelijkheden voor kwaliteitscontrole. Zij kunnen bijvoorbeeld worden ingezet om grote aantallen literatuurverwijzingen te analyseren, argumentatieve consistentie te toetsen of alternatieve interpretaties van gegevens te genereren. Zo ontstaan nieuwe vormen van *hybride verificatie*, waarin menselijke beoordeling en machine-ondersteunde analyse elkaar aanvullen.

De analyse van de cases leidt daardoor tot een aangescherpt antwoord op de eerste onderzoeksvraag. Generatieve AI vergroot de mogelijkheden voor synthese, patroonherkenning en hypothesevorming, maar verschuift tegelijkertijd het punt waarop epistemische kwetsbaarheid ontstaat. De kwaliteit van AI-gestuurde kennisproductie blijkt daarom minder afhankelijk van individuele modeloutput dan van de *onderzoeksarchitectuur* waarin die output wordt gegenereerd en geëvalueerd.

18.3 ERQ2 – Pedagogisch-institutionele dimensie

De tweede empirische onderzoeksvraag luidt: '*hoe verandert generatieve AI de rollen, vaardigheden en institutionele structuren van academisch onderzoek'.* In hoofdstuk 12 werd verwacht dat AI de rol van onderzoekers zou verschuiven van producent van tekst en analyse naar regisseur van een onderzoeksproces waarin AI-systemen worden ingezet voor verschillende taken. De vijf casussen bevestigen deze verwachting gedeeltelijk, maar laten ook zien dat de institutionele impact sterk domein-specifiek is. In geen van de onderzochte archetypen verdwijnt de rol van de menselijke onderzoeker. Wel wordt zichtbaar dat onderzoek steeds vaker wordt georganiseerd als een iteratief interactieproces tussen mens en model.

In historisch onderzoek functioneert AI vooral als hulpmiddel bij het systematisch toepassen van analysekaders op grotere corpora. De onderzoeker blijft verantwoordelijk voor selectie van bronnen en historiografische interpretatie. In de antropologische casus blijkt dat generatieve systemen interpretatieve kaders kunnen versterken, waardoor de onderzoeker een expliciet corrigerende rol moet blijven vervullen. In de ontwerpgerichte casus verschuift de rol van de onderzoeker gedeeltelijk naar organisator van een generatief ontwerptraject, waarin verschillende modelrollen – genereren, controleren en reviseren – van elkaar worden gescheiden. Deze rolverschuiving ontstaat echter niet vanzelf; zij vereist een expliciet ontworpen interactiestructuur. De casus in het meta-domein maakt deze organisatorische dimensie het duidelijkst zichtbaar. Het schrijfproces van het proefschrift werd georganiseerd als een reeks iteratieve interacties tussen curator en model, waarbij de menselijke onderzoeker verantwoordelijk bleef voor de architectuur van het betoog en de uiteindelijke validatie van de argumentatie.

De casussen tonen daardoor een consistent patroon. Generatieve AI vervangt de onderzoeker niet, maar verandert de organisatie van onderzoekspraktijken. Kennisproductie verschuift van een lineair schrijfproces naar een architectonisch georganiseerd interactieproces, waarin genereren, evalueren en integreren expliciet van elkaar worden onderscheiden.

18.4 ERQ3 – Macht en normativiteit

De derde empirische onderzoeksvraag luidt: *'Welke machtsverhoudingen en normatieve keuzes zijn ingebed in AI-gestuurde kennisproductie?'*. In hoofdstuk 12 werd geconcludeerd dat generatieve AI nieuwe afhankelijkheden kan creëren, met name door de concentratie van AI-infrastructuur bij een beperkt aantal technologiebedrijven.

De empirische analyse bevestigt deze diagnose, maar laat zien dat de concrete betekenis van deze machtsdimensie sterk varieert per onderzoeksdomein. In meerdere casussen blijken normatieve keuzes primair te liggen in traditionele wetenschappelijke beslissingen, zoals de selectie van onderzoeksobjecten, interpretatieve kaders en analysemethoden. In het historisch archetype ligt de belangrijkste normatieve keuze bijvoorbeeld in canonvorming: de selectie van historische episodes die als relevant onderzoeksobject worden beschouwd. AI kan helpen bij analyse, maar blijft afhankelijk van de bronnen en selectiecriteria die door de onderzoeker zijn vastgesteld. Ook in de antropologische casus ligt de normatieve lading vooral in de interpretatieve framing van het materiaal. Generatieve AI kan dergelijke kaders versterken of systematiseren, maar produceert ze niet autonoom.

De natuurwetenschappelijke casus vormt een bijzonder geval. Daar werd gewerkt met synthetische data, wat op het eerste gezicht lijkt te wijzen op nieuwe vormen van epistemische macht. Bij nadere beschouwing blijkt dit echter vooral samen te hangen met *resource poverty*: beperkte toegang tot een empirische onderzoeksinfrastructuur. Deze asymmetrie bestond al vóór de opkomst van generatieve AI.

De empirische analyse leidt daarom tot een genuanceerde conclusie. Generatieve AI kan nieuwe infrastructuurafhankelijkheden introduceren en bestaande machtsasymmetrieën versterken, maar de normatieve kern van wetenschappelijke kennisproductie blijft in de onderzochte casussen grotendeels verbonden met menselijke onderzoekspraktijken.

18.5 Synthese en beantwoording van de centrale onderzoeksvraag

In de voorgaande paragrafen zijn de drie empirische onderzoeksvragen afzonderlijk beantwoord. Daarbij werd zichtbaar dat de impact van generatieve AI op wetenschappelijke kennisproductie niet uniform is, maar varieert per archetype van onderzoek. Sommige effecten manifesteren zich in meerdere onderzoekspraktijken, terwijl andere sterk domein-specifiek blijken te zijn. Tegen deze achtergrond kan nu de centrale onderzoeksvraag van dit proefschrift worden beantwoord.

Het antwoord op de centrale onderzoeksvraag luidt derhalve: '*Generatieve AI verandert wetenschappelijke kennisproductie doordat zij de architectuur van het onderzoeksproces herconfigureert, waardoor onderzoek in veel gevallen kan versnellen, interdisciplinair kan verbreden en methodologisch kan verrijken'*.

De vijf onderzochte casussen laten zien dat deze drie effecten geen abstracte eigenschappen van de technologie zijn, maar empirisch waarneembare patronen in concrete onderzoekspraktijken. Tegelijk

blijkt uit de casussen dat de mate waarin deze effecten optreden sterk afhankelijk is van het type onderzoek en van de wijze waarop het generatieve proces wordt georganiseerd.

Een eerste effect betreft **versnelling van analyse en synthese**. In meerdere casussen bleek generatieve AI in staat om grote hoeveelheden informatie te verwerken, te structureren en te vergelijken op een schaal die voor individuele onderzoekers moeilijk realiseerbaar is. Dit gold onder meer voor het systematisch analyseren van historische episodes, het structureren van complexe conceptuele argumentaties en het genereren van samenhangende academische tekstvoorstellen. In dergelijke contexten kan AI leiden tot aanzienlijke tijdswinst in onderzoeks- en schrijfprocessen. Tegelijk laten de casussen zien dat deze versnelling niet uniform optreedt. In ontwerpgerichte onderzoekspraktijken, zoals de polyglotte taalverwerving, ligt de belangrijkste bottleneck minder bij analyse dan bij het bewaken van semantische consistentie en ontwerpstabiliteit. De mate waarin AI kennisproductie versnelt blijkt daarom sterk afhankelijk van de aard van de onderzoekspraktijk.

Een tweede effect betreft **verbreding van onderzoekspraktijken**. In meerdere casussen maakte generatieve AI het praktisch uitvoerbaar om uiteenlopende kennisdomeinen binnen één onderzoeksarchitectuur te combineren. Historische analyse werd bijvoorbeeld gekoppeld aan systematische conceptuele vergelijking, bestuurskundige vraagstukken werden verbonden met biologische randvoorwaarden en in het ontwerp van polyglotte taalverwerving kwamen inzichten uit taalkunde, cognitiewetenschap en didactiek samen. Generatieve AI fungeerde daarbij niet als bron van nieuwe theorieën, maar als een infrastructuur die het mogelijk maakte om kennis uit verschillende disciplines sneller te synthetiseren en onderling te verbinden. Daarmee vergroot AI niet automatisch de interdisciplinariteit van wetenschap, maar verlaagt zij wel de praktische drempel om verschillende disciplines binnen één onderzoeksproject te integreren.

Een derde effect betreft **verrijking van methodologische controlemechanismen**. In verschillende casussen bleek AI niet alleen een instrument voor analyse en tekstproductie, maar ook voor verificatie. Met name bij het controleren van literatuurgebruik en argumentatieve consistentie maakte AI het mogelijk om grote aantallen referenties systematisch te analyseren op interpretatieve correctheid en interne samenhang. Hierdoor ontstaan nieuwe vormen van hybride verificatie waarin menselijke beoordeling en machine-ondersteunde analyse elkaar aanvullen. Daarnaast kan het gebruik van generatieve systemen bijdragen aan een forensische reconstructie van het onderzoeksproces, doordat interacties tussen onderzoeker en model expliciet worden gedocumenteerd en analyseerbaar gemaakt. Op die manier kan AI niet alleen bijdragen aan kennisproductie, maar ook aan de transparantie van het onderzoeksproces zelf.

De casussen laten echter ook zien dat deze drie effecten niet automatisch optreden. Hun realisatie hangt in belangrijke mate af van de manier waarop het generatieve proces wordt georganiseerd. Wanneer generatieve AI ad hoc wordt gebruikt binnen één interactie met één model, blijven de voordelen vaak beperkt en kunnen problemen zoals semantische drift, contextverlies of rolvermenging optreden. Daartegenover staat een andere benadering die in dit proefschrift expliciet zichtbaar wordt: het organiseren van onderzoek als een architectuur van gescheiden rollen, waarin genereren, analyseren en controleren expliciet van elkaar worden onderscheiden. Dit vormt een belangrijk novum van dit onderzoek. De casussen laten zien dat onderzoekers niet alleen afhankelijk zijn van de intrinsieke capaciteiten van een taalmodel, maar dat zij door het ontwerpen van een passende interactie-architectuur de snelheid, stabiliteit en controleerbaarheid van AI-gestuurde kennisproductie aanzienlijk kunnen vergroten. In dat opzicht verschuift de centrale vraag van AI in wetenschap van de capaciteiten van individuele modellen naar de manier waarop onderzoekers het generatieve proces organiseren.

De belangrijkste conclusie van dit proefschrift is daarom dat generatieve AI wetenschappelijke kennisproductie niet simpelweg automatiseert, maar haar **architectuur transformeert**. In plaats van een lineair onderzoeksproces ontstaat een hybride onderzoekspraktijk waarin genereren, evalueren en verifiëren steeds explicieter van elkaar worden onderscheiden en georganiseerd.

18.6 Validiteit, betrouwbaarheid en methodologische waarborgen

De methodologische betrouwbaarheid van dit onderzoek berust op drie principes: radicale transparantie, consistente toepassing van het analysekader en curatoriale regie over het generatieve proces. Alle casussen zijn vergezeld van reflectieverslagen waarin interacties tussen onderzoeker en taalmodel systematisch zijn gedocumenteerd. Daardoor wordt het generatieve proces zelf analyseerbaar. Daarnaast werd in alle casussen hetzelfde analytische raamwerk toegepast. Hierdoor functioneren de casussen niet als losse voorbeelden maar als empirische tests van één analysekader. Ten slotte bleef de interpretatieve verantwoordelijkheid steeds bij de menselijke onderzoeker. Generatieve AI produceerde analyse- en tekstvoorstellen, maar de curator valideerde en integreerde deze in het uiteindelijke betoog. Binnen deze curatoriale architectuur speelde ook de organisatie van het generatieve proces een rol. De casussen laten zien dat een multi-agent-benadering het onderzoeksproces kan stabiliseren en dat hybride verificatie nieuwe mogelijkheden biedt voor systematische controle van literatuurgebruik en argumentatie.

18.7 Aanbevelingen voor vervolgonderzoek

De bevindingen van dit proefschrift openen verschillende lijnen voor vervolgonderzoek. Ten eerste is er behoefte aan vergelijkend onderzoek naar domein-specifieke effecten van AI, omdat versnelling, verbreding en verrijking per archetype van onderzoek verschillend uitwerken. Ten tweede verdient de architectuur van AI-ondersteund onderzoek verdere studie, met name het verschil tussen single-agent en multi-agent benaderingen. Ten derde vragen de institutionele implicaties van AI aandacht. Wanneer onderzoekers verschuiven van auteur naar curator van generatieve processen, kan dit gevolgen hebben voor academische training en beoordelingscriteria. Tenslotte is verder onderzoek nodig naar de normatieve en ethische dimensies van AI-gestuurde kennisproductie, waaronder afhankelijkheid van digitale infrastructuren en commerciële modellen

18.8 Slotbeschouwing

De casussen in dit proefschrift laten zien dat generatieve AI wetenschappelijk onderzoek niet automatiseert maar herorganiseert. In plaats van de onderzoeker te vervangen ontstaat een hybride onderzoekspraktijk waarin menselijke expertise en machine-ondersteunde analyse worden gecombineerd. De rol van de onderzoeker verschuift daarbij van individuele auteur naar architect en curator van generatieve onderzoeksprocessen. Deze verschuiving brengt nieuwe verantwoordelijkheden met zich mee, waaronder het ontwerpen van robuuste onderzoeksarchitecturen en het bewaken van epistemische integriteit. De centrale les van dit proefschrift is daarom dat de toekomst van wetenschap niet door AI alleen wordt bepaald, maar door de wijze waarop onderzoekers deze technologie integreren in zorgvuldig ontworpen kennisarchitecturen.

Epiloog – De Eed van Socrates

Dit proefschrift begon met de vraag naar de mogelijkheden en beperkingen van AI in wetenschappelijke kennisproductie. De analyse liet zien dat AI geen neutraal instrument is, maar een technologie die epistemische fundamenten onder druk zet. In hoofdstuk 2 werd dit scherp zichtbaar in de discussie over het *black-box-karakter* van grote taalmodellen: onderzoekers weten niet exact hoe antwoorden tot stand komen, en juist dat vormt een bedreiging voor de kernprincipes van toetsbaarheid en transparantie.

De noodzaak tot ethisch handelen werd concreet in paragraaf 4.4, waar de problematiek van *gefabriceerde referenties* voor het eerst expliciet werd benoemd. Een ogenschijnlijk kleine onzorgvuldigheid – een referentie die niet bestaat, of een citaat dat nooit is uitgesproken – kan het vertrouwen in een onderzoek geheel ondermijnen. Hier werd zichtbaar dat de verantwoordelijkheid van de onderzoeker niet ophoudt bij het genereren van tekst, maar pas begint bij de zorgvuldige verificatie van bronnen.

Hoofdstuk 6 bracht deze inzichten samen in een methodologische verantwoording: AI kan worden ingezet, maar alleen onder strikte voorwaarden van *controleerbaarheid, reproduceerbaarheid en menselijke eindverantwoordelijkheid*. Daarmee verschuift de discussie van louter techniek naar ethiek. De vraag is niet alleen *wat AI kan*, maar vooral *wat onderzoekers moeten doen* om de integriteit van wetenschap te bewaken.

De confrontatie met deze risico's maakt duidelijk dat AI-gestuurde wetenschap niet alleen een methodologische keuze is, maar een morele beproeving. Net zoals artsen bij hun beroepsuitoefening de Hippocratische eed afleggen om de kwetsbaarheid van het menselijk lichaam te beschermen, zo hebben onderzoekers die AI inzetten een eed nodig die de kwetsbaarheid van de menselijke kennis beschermt: 'de Eed van Socrates'.

Waarom de *Eed van Socrates*? De keuze voor de benaming *Eed van Socrates* is niet toevallig. Net zoals artsen hun praktijk legitimeren via de *Eed van Hippocrates*, zo hebben onderzoekers in het AI-tijdperk een belofte nodig die hen bindt aan waarheid en menselijkheid. Socrates staat symbool voor de kritische dialoog, het voortdurend bevragen van aannames en het zoeken naar inzicht via gesprek en zelfonderzoek. Daarmee is hij verwant aan Hippocrates, die de zorg voor het menselijk lichaam belichaamde; Socrates belichaamt de zorg voor het menselijk weten.

Andere namen zouden mogelijk ook aansprekend zijn geweest – zoals de *Eed van Aristoteles* (meer empirisch), de *Eed van Turing* (meer technologisch) of de *Eed van Episteme* (meer conceptueel) – maar juist de socratische houding van voortdurende reflectie en dialoog vormt de kern van wat AI-onderzoek vereist. Daarom draagt deze eed zijn naam.

De Eed van Socrates

(voor onderzoekers die AI inzetten bij kennisproductie)

Generieke beginselen
Ik zweer en beloof:

- Dat ik mijn werk zal verrichten in dienst van de mensheid en het algemeen belang, en nooit uitsluitend in dienst van commerciële of technologische imperatieven.
- Dat ik geen onderzoek zal uitvoeren of publiceren waarvan ik weet, of redelijkerwijs kan vermoeden, dat het schade kan berokkenen aan mens of samenleving.
- Dat ik mijn kennis en vaardigheden zal inzetten om bij te dragen aan waarheid, rechtvaardigheid en menselijke waardigheid.
- Dat ik volledige transparantie zal betrachten over de inzet van AI in mijn onderzoekspraktijk, inclusief de grenzen en onzekerheden die dit met zich meebrengt.

Specifieke verplichtingen bij AI-gebruik
Ik zweer en beloof:

1. **Controle van bronnen**
    - Ik zal nooit AI-gegenereerde referenties of citaten ongecontroleerd gebruiken.
    - Ik zal elke bron individueel verifiëren en correct verantwoorden.
2. **Omgaan met black-box-systemen**
    - Ik zal het black-box-karakter van AI nooit negeren of verhullen.
    - Ik zal actief zoeken naar manieren om de werking van het model uitlegbaar te maken en de epistemische grenzen te expliciteren.
3. **Bias en normatieve lading**
    - Ik zal altijd onderzoeken welke vooronderstellingen en biases de data en modellen bevatten.
    - Ik zal deze benoemen, documenteren en waar mogelijk mitigeren.
4. **Menselijke verantwoordelijkheid**
    - Ik blijf te allen tijde eindverantwoordelijk voor de inhoud en de gevolgen van mijn onderzoek.
    - Ik zal de verantwoordelijkheid niet afschuiven op het AI-systeem of de aanbieder ervan.
5. **Integriteit van data en toestemming**
    - Ik zal de autonomie en rechten van individuen en gemeenschappen respecteren in alle fasen van dataverzameling, analyse en hergebruik.
    - Ik zal geen data gebruiken zonder adequate toestemming of rechtvaardiging.
6. **Tegenmacht en publieke waarden**
    - Ik zal mij verzetten tegen het klakkeloos volgen van commerciële belangen of machtsconcentraties in AI-ontwikkeling.
    - Ik zal pleiten voor openheid, toegankelijkheid en publieke controle over kennisinfrastructuren.
7. **Toetsbaarheid en reproduceerbaarheid**
    - Ik zal alleen AI-gefaciliteerde bevindingen publiceren als deze toetsbaar, reproduceerbaar en methodologisch verantwoord zijn.
    - Ik zal mijn werk documenteren zodat anderen mijn proces kunnen controleren en bekritiseren.

Slotbeschouwing. Met deze eed wordt de plaats van AI in de wetenschap opnieuw gedefinieerd. AI is geen neutraal instrument, maar een actor die de aard van kennis beïnvloedt. Dat besef liep als rode draad door de analyses in dit proefschrift: van de vroege signalering van het black-box-probleem in hoofdstuk 2, via de waarschuwing voor gefabriceerde referenties in hoofdstuk 4, tot de methodologische herijking in hoofdstuk 6.

De menselijkheid van de wetenschap wordt daarmee niet langer vanzelfsprekend gegarandeerd: zij moet expliciet worden bewaakt. De *Eed van Socrates* vormt hiervoor een normatief kompas. Wie deze eed aflegt, verbindt zich niet slechts aan een set regels, maar aan een houding van verantwoordelijkheid, transparantie en trouw aan de mensheid. Daarmee wordt het AI-tijdperk niet alleen een technologische revolutie, maar ook een morele beproeving. De wetenschap zal erin slagen haar waardigheid te behouden, zolang onderzoekers bereid zijn zich rekenschap te geven van de plicht die hen verbindt: de plicht om waarheid te dienen en de menselijkheid nooit uit het oog te verliezen.