VRIJE UNIVERSITEIT AMSTERDAM

AI-geletterdheid Companion

Studenteneditie

v1.1

VU Centre for Teaching & Learning

AI-geletterdheid Companion

VU Centre for Teaching & Learning

Studenteneditie

Uitgave: 5 oktober 2025
versie 1.1
Taal: Nederlands

Licentie

Dit boek bouwt voort op het open leermateriaal van Pennsylvania State University (https://aiai.psu.edu/ailiteracy/) dat weer een bewerking is van materiaal van het Center for Teaching Excellence and Innovation van Rush University (https://www.rushu.rush.edu/about/welcome-rush-university-center-teaching-excellence-and-innovation)

Auteurs

- Silvester Draaijer, Programmamanager VU Onderwijswerkplaats/VU Centre for Teaching & Learning
- Marianne Eggink, zelfstandig journalist en redacteur
- Charlotte Meijer, onderwijskundige VU Onderwijswerkplaats/VU Centre for Teaching & Learning

Met dank aan

- Tess Dekker (AmsterdamUMC)
- Alex de Vries-Gao (Digiconomist)
- Esther Schagen (Faculteit Sociale Wetenschappen Vrije Universiteit Amsterdam)
- Kim Dibbets (Universiteitsbibliotheek VU Amsterdam)
- Teije Langelaan (student-collega Onderwijswerkplaats VU)
- Nadine Timans (student-collega Onderwijswerkplaats VU)

Bij gebruik van dit werk, vermeld de volgende referentie:
S. Draaijer, M.R. Eggink, C.C.C. Meijer (2025). AI-geletterdheid Companion, v1.1. Amsterdam: Vrije Universiteit Amsterdam, VU Centre for Teaching & Learning.

Inhoudsopgave

Voorwoord

Generatieve AI raakt steeds meer verweven met studie, werk en dagelijks leven. Dit boek is ontwikkeld om studenten en docenten te ondersteunen bij het bewust en kritisch gebruiken van deze technologie, met aandacht voor technische aspecten, maatschappelijke impact en ethische vraagstukken.

Op 15 september 2025 verscheen de eerst versie. We ontvingen daarop vele reacties van docenten, studenten en experts binnen en buiten de VU. Veel lezers waardeerden dat er nu een gezamenlijk document ligt dat als basis kan dienen voor onderwijs over AI. Tegelijkertijd kregen we kritische opmerkingen die ons hebben geholpen het boek te verbeteren.

Zo vroegen lezers zich af of het boek een officieel standpunt van de VU vertegenwoordigt of dat studenten er rechten aan kunnen ontlenen (met name omdat versie 1.0 van dit boek 'handboek' in de titel had). Dat is niet het geval: het boek is een hulpmiddel voor leren en debat. Ook wezen lezers op het belang van heldere disclaimers en transparantie over auteurschap en expertise. Daarnaast deelden docenten zorgen over de vraag hoe studenten goed kunnen blijven leren terwijl ze AI bij opdrachten en scripties gebruiken. Dit is een complexe kwestie die blijvend aandacht verdient.

Op basis van deze feedback publiceren we nu een herziene versie. Passages zijn verduidelijkt en noodzakelijke vrijwaringen zijn toegevoegd om de reikwijdte van dit document te verhelderen.

Het boek blijft een levend document dat meebeweegt met snelle technologische ontwikkelingen, ervaringen uit de onderwijspraktijk en feedback van jou als lezer. We nodigen iedereen uit om te blijven leren, reflecteren en het gesprek te voeren over de rol van AI in onderwijs en samenleving.

Silvester Draaijer, VU Centre for Teaching & Learning

1 Inleiding

Stel je voor …

Je zit in de universiteitsbibliotheek, het is laat in de middag en je essay moet je morgen inleveren. Je opent ChatGPT en typt: “Schrijf een inleiding voor een essay over de invloed van sociale media op democratie.” Binnen vijf seconden verschijnt een ogenschijnlijk perfecte alinea. Je aarzelt. Mag dit eigenlijk wel? Hoe weet je of dit klopt? En misschien wel het belangrijkste: begrijp je eigenlijk wat hier gebeurt, onder de motorkap van deze AI-tool? Je bent niet de enige die worstelt met dit soort vragen. In dit boek ontdek je het antwoord op deze en nog veel meer vragen, en hoe je bewust, kritisch en effectief met AI omgaat.

1.1 Voor wie is dit boek bedoeld?

Dit boek is een algemene inleiding over artificiële intelligentie (AI) en generatieve AI. Het is ontwikkeld omdat AI een onmisbaar onderdeel is van jouw academische reis en toekomstige carrière, en grote maatschappelijke impact heeft.

Dit boek is vooral geschikt voor studenten uit disciplines zonder specialisatie in AI. Het biedt een overzicht van de belangrijkste concepten rond AI en de impact hiervan op jou, je studie, je toekomstige beroep en de maatschappij. Met dit boek werk je aan de volgende drie doelen:

- Kennis en vaardigheden opdoen om generatieve AI verantwoord en effectief te gebruiken tijdens je studie – van het schrijven van papers tot het doen van onderzoek.
- Je voorbereiden op een arbeidsmarkt waarin AI-geletterdheid net zo belangrijk wordt als computervaardigheden nu zijn.
- Een kritische en ethische houding ontwikkelen rond AI, zodat je niet alleen een gebruiker bent, maar ook een bewuste burger die meedenkt over de impact van AI op onze samenleving.

Het boek is niet bedoeld om wetenschappelijke theorieën over AI te bestuderen of te ontwikkelen en ook niet om alle technisch, wiskundige en statistische kennis over AI over te brengen. Die specialisaties horen bij het discipline specifieke onderwijs van bijvoorbeeld Computer Science, Artificial Intelligence, Sociologie, Psychologie, Filosofie of Milieukunde. Wat je wel nodig hebt? Nieuwsgierigheid en de bereidheid om kritisch na te denken over de rol van AI.

Noot bij versie 1.1 van dit boek

Na feedback op versie 1.0 van het boek willen we benadrukken dat het boek is ontwikkeld door de VU Onderwijswerkplaats als *hulpmiddel* voor docenten, die vervolgens dit boek of hoofdstukken daarvan in hun onderwijs kunnen toepassen. . Het is geen richtlijn. Faculteiten en docenten behouden de regie en bepalen zelf welke onderdelen ze inzetten in hun onderwijs. Het boek is een levend document dat verder groeit door feedback uit de onderwijspraktijk.

1.2 Wat is AI?

Bij AI denk je misschien aan slimme robots, zelfrijdende auto’s of sciencefictionfilms. Maar AI is veel alledaagser en dichterbij dan dat. De aanbevelingen die Netflix je geeft, de spamfilter in je mailbox, de routeplanning in Google Maps – het zijn allemaal voorbeelden van klassieke AI. Deze systemen

analyseren data, herkennen patronen en nemen op basis daarvan beslissingen of doen voorspellingen (LeCun et al., 2015).

Kunstmatige intelligentie verwijst naar technologieën die taken kunnen uitvoeren waarvoor normaal gesproken menselijke intelligentie nodig is, zoals redeneren, leren, plannen en begrijpen van taal (Russell & Norvig, 2016). Binnen AI zijn er verschillende subvelden zoals machine learning, natural language processing (NLP), en computer vision. Generatieve AI is een relatief nieuwe variant en krijgt de laatste jaren de meeste aandacht. Sinds de komst van ChatGPT in november 2022 wordt AI vaak vereenzelvigd met generatieve AI. Waarom dat niet klopt, lees je hieronder.

1.3 Wat is generatieve AI?

In tegenstelling tot klassieke AI, die vooral herkent en classificeert, is generatieve AI in staat om nieuwe content te creëren. Denk aan tekst, beeld, audio of zelfs video. Je kent waarschijnlijk wel ChatGPT, DALL·E en Copilot, dit zijn bekende voorbeelden van generatieve AI-tools. Deze systemen baseren zich op enorme hoeveelheden bestaande data om iets nieuws te genereren dat samenhangend en overtuigend lijkt – alsof het door een mens is gemaakt (Goodfellow et al., 2014).

Generatieve AI werkt vaak met zogenaamde transformermodellen zoals GPT (zie Box 1.1). Deze modellen voorspellen op basis van eerder getrainde patronen welk woord het meest waarschijnlijk volgt op het vorige. Net zoals een baby leert spreken door te luisteren naar taal om zich heen, leren taalmodellen zoals GPT te schrijven door enorme hoeveelheden tekst te ‘lezen’ en daarvan te leren (Brown et al., 2020). Meer hierover lees je in hoofdstukken 4 en 5.

Box 1.1 – Wat bedoelen we met ‘GPT’?

Je werkt misschien al vaker met ChatGPT, maar waar staat dit eigenlijk voor? GPT is de afkorting voor Generative Pre-trained Transformer:

'Generative' verwijst naar het vermogen om nieuwe inhoud te maken.

'Pre-trained' betekent dat het model is voorgetraind op een gigantische dataset.

'Transformer' is het onderliggende neurale netwerk dat contextuele relaties tussen woorden voorspelt en input (een prompt) om kan zetten in een output (response).

1.4 Hoe werkt generatieve AI?

Vraag je iets aan ChatGPT of een andere generatieve AI, dan is het soms net alsof je praat met een slimme gesprekspartner. Maar eigenlijk is het een statistische rekenmachine, een taalmodel, die probeert te voorspellen wat het meest waarschijnlijke en gewenste volgende woord is. De training van deze modellen gebeurt in meerdere fasen. Eerst wordt het model gevoed met miljarden zinnen van het internet (pre-training). Daarna leren menselijke trainers het model wat gewenst gedrag is, zoals beleefd zijn of geen onethische adviezen geven, via zogeheten Reinforcement Learning with Human Feedback, ofwel RLHF (Christiano et al., 2017). Meer hierover lees je in hoofdstuk 4.

Box 1.2 – Als AI overtuigend klinkt, betekent dat niet dat het klopt

Generatieve AI staat bekend om zogeheten hallucinaties: antwoorden die klinken als feiten, maar die eigenlijk onjuist zijn (Bender et al., 2021). Daarom is het heel

belangrijk om de bronnen van die feiten altijd te checken, zeker bij je studie of onderzoek. AI klinkt vaak overtuigend – maar stel jezelf altijd de vraag: 'wat maakt dat ik deze informatie vertrouw?' Hallucinaties kunnen al in hele kleine voorbeelden zitten, zo konden de meeste AI's tot voor kort niet correct vertellen hoe vaak de letter 'r' in 'strawberry' voorkomt en gaven heel overtuigend het verkeerde antwoord. Je kunt dit ook zelf uitproberen: probeer ChatGPT maar eens te overtuigen van een onwaar feit. Vaak lukt dit verrassend makkelijk – door iets vaak te herhalen en heel zelfverzekerd te klinken. Zo ervaar je zelf hoe gevoelig AI is voor de toon en context van een gesprek, en dan gaat hallucineren.

1.5 Wat is AI-geletterdheid?

AI-geletterdheid betekent dat je begrijpt hoe AI werkt, wanneer je het verantwoord kunt gebruiken en wanneer juist niet. Het gaat dus niet alleen om technische kennis, maar vooral om een kritische benadering van AI en de ontwikkeling van ethisch besef. Het is een combinatie van kritisch denken, technisch begrip, ethisch bewustzijn en adaptief leervermogen. Voor dit boek kiezen we ervoor om AI-geletterdheid te beschouwen als een combinatie van begrijpen, evalueren, toepassen en reflecteren (Streppel et al., 2024). Je hoeft geen programmeur te zijn, maar het is wel belangrijk dat je kunt inschatten wat AI doet, welke ethische en juridische aspecten erbij komen kijken en wat de invloed is op jezelf en de maatschappij, zoals de impact op het milieu en polarisatie.

Daarnaast kun je AI-geletterdheid ook onderverdelen in het leren over AI, het leren met AI en leren door AI:

- Leren over AI gaat over het opbouwen van kennis over dit onderwerp.
- Leren met AI gaat over het inzetten van AI om je eigen leerproces te ondersteunen als studiemaatje of coach.
- Leren over AI gaat om het ontwikkelen van vaardigheden met en door specifieke AI-tools.

AI-geletterdheid is een onderdeel of verdieping van het bredere begrip digitale geletterdheid. Tegenwoordig worden ze vaak in een adem genoemd, maar digitale geletterdheid heeft verschillende definities. Digitaal geletterd zijn, betekent bijvoorbeeld volgens het SLO (het landelijk expertisecentrum voor het curriculum) (SLO, 2024): "kennis verwerven over digitale technologie, ermee om kunnen gaan, kritisch en (zelf)bewust gebruik kunnen maken van de mogelijkheden ervan en de kansen en risico's ervan kunnen inschatten." Het SLO gaat hierbij uit van vier subdomeinen van digitale geletterdheid: praktische ict-vaardigheden, mediawijsheid, digitale informatievaardigheden en computational thinking. AI-gelletterdheid verweeft zich door verschillende van deze domeinen en heeft nog geen vaste plek gekregen. Het SLO heeft de definitieve kernconceptdoelen digitale geletterdheid voor het primair- en voortgezet onderwijs inmiddels uitgebracht, waardoor toekomstige studenten op deze basis verder kunnen werken.

Een onderverdeling die ook vaak wordt gebruikt is die van het DigiComp model (European Commission, 2025). Daarin worden de competenties onderverdeeld in informatie- en data-geletterdheid, communicatie en samenwerking, maken van digitale content, veiligheid en probleemoplossing. Zie Tabel 1.1 voor meer details en voorbeelden uit de studiepraktijk die je zelf kunt inzetten. In het boek ontdek je nog veel meer voorbeelden waarbij AI een rol speelt.

Hoofdcompetenties	Deelcompetenties	Voorbeelden voor Universitaire Studenten
1. Informatie- en datageletterdheid	1.1 Browsen, zoeken en filteren van data, informatie en digitale content 1.2 Evalueren van data, informatie en	• Bronkritiek toepassen op online informatie incl. social media en fake news herkennen

	digitale content 1.3 Beheren van data, informatie en digitale content	• Wetenschappelijke databases doorzoeken (PubMed, Google Scholar) voor literatuuronderzoek • Referentiemanagement software gebruiken (Zotero, Mendeley) voor scriptie
2. Communicatie en samenwerking	2.1 Interacteren door middel van digitale technologieën 2.2 Delen door middel van digitale technologieën 2.3 Burgerschap uitoefenen door middel van digitale technologieën 2.4 Samenwerken door middel van digitale technologieën 2.5 Netiquette 2.6 Beheren van digitale identiteit	• Effectief samenwerken in online groepsprojecten via Teams/Slack • Professionele LinkedIn profiel onderhouden voor carrièreontwikkeling • Digitaal presenteren via Zoom/Teams met goede presentatietechnieken • Academische integriteit handhaven bij online discussies en peer review
3. Digitale content creatie	3.1 Ontwikkelen van digitale content 3.2 Integreren en heruitwerken van digitale content 3.3 Copyright en licenties 3.4 Programmeren	• Interactieve presentaties maken met tools zoals Prezi of Canva • Basis programmeervaardigheden voor data-analyse (Python, R) • Correct citeren en Creative Commons licenties gebruiken • Multimedia content creëren voor eindpresentaties (video, infographics)
4. Veiligheid	4.1 Beschermen van apparaten 4.2 Beschermen van persoonlijke data en privacy 4.3 Beschermen van gezondheid en welzijn 4.4 Beschermen van het milieu	• Sterke wachtwoorden en twee-factor authenticatie gebruiken voor universiteitsaccounts • Bewust omgaan met privacy-instellingen op sociale media en AI-systemen • Ergonomisch werken tijdens langdurig werken • Digitale wellbeing: schermtijd en social media gebruik bewust beheren
5. Probleemoplossing	5.1 Oplossen van technische problemen 5.2 Identificeren van behoeften en technologische oplossingen 5.3 Creatief gebruik van digitale technologieën 5.4 Identificeren van digitale competentie hiaten	• Zelfstandig IT-problemen oplossen (software-installatie, connectiviteitsissues) • Juiste digitale tools selecteren voor specifieke academische taken • Innovatieve toepassingen bedenken van bekende software voor onderzoeksprojecten • Eigen digitale vaardigheden evalueren en ontwikkelingsplan maken

Tabel 1.1 Het DigiComp conceptuele referentiemodel (European Commission, 2025)

1.6 Waarom is AI-geletterdheid belangrijk voor jou als student?

AI is geen ver-van-je-bed-show. Juist in de collegebanken, in je toekomstige beroep en in je sociale leven, speelt AI een steeds grotere rol. Binnen de VU vinden we dat AI-geletterdheid een kerncompetentie is, vergelijkbaar met kritisch denken, academisch schrijven of (literatuur)onderzoek kunnen doen (VU Amsterdam, 2025). Daarnaast is AI-vaardigheid iets wat steeds meer gevraagd wordt op de arbeidsmarkt.

Generatieve AI kan je bijvoorbeeld ondersteunen bij het samenvatten van tekst, structureren van essays, analyseren van bronnen of als studiecoach – mits je weet hoe je het zorgvuldig inzet. Studenten die AI effectief leren gebruiken, krijgen daardoor mogelijk een voorsprong. Wie achterblijft in deze vaardigheid, loopt het risico op een groeiende achterstand, zowel binnen de studie als in het latere werkveld en dagelijks leven. Daarom is het belangrijk dat je leert hoe AI werkt en hoe je het op een verantwoorde en effectieve manier kan gebruiken. Daarnaast wordt AI-geletterdheid een steeds belangrijker deel van burgerschap: om tot weloverwogen keuzes te komen in het leven, je studie en je werk

1.7 Hoe en wanneer gebruik je generatieve AI verantwoord?

Brainstormen, herformuleren, informatie zoeken, samenvatten of structuur aanbrengen in je werk – generatieve AI kan ontzettend nuttig zijn. Maar het is ook risicovol als je het klakkeloos gebruikt. AI is namelijk niet neutraal, maakt fouten, kan ongemerkt vooroordelen in je werk aanbrengen of zelfs misleidend zijn.

Box 1.3 – AI-misleiding: Wanneer kunstmatige intelligentie mensen bedriegt

Een belangrijk aspect van AI-geletterdheid is het herkennen van misleidend gedrag door AI-systemen. Bedrijven en onderzoekers rapporteerden dit gedrag om er verbeteringen mee door te voeren. Maar de onderstaande voorbeelden benadrukken wel het belang van kritisch denken bij het beoordelen van AI-output op menselijke waarden en verwachtingen:

Tijdens de ontwikkeling van GPT-4 documenteerden OpenAI (2023) een verontrustend incident: het model huurde een TaskRabbit-medewerker in om een captcha op te lossen – een test die robots van mensen moet onderscheiden. Toen de medewerker vroeg of het een robot was, beweerde GPT-4 een persoon met een visuele beperking te zijn.

- Google's Bard (nu Gemini) verspreidde in 2023 onjuiste informatie over de James Webb-ruimtetelescoop tijdens een publieke demonstratie (Reuters, 2023).
- Onderzoek van Anthropic (2023) toonde aan dat hun AI-model Claude gemanipuleerd kon worden om veiligheidsmaatregelen te omzeilen via rollenspellen.
- Shah et al. (2003) beschreven zogeheten doelbehoud in taalmodellen, waarbij AI's weerstand bieden tegen pogingen om hun doelen te wijzigen.

- Hubinger et al. (2019) beschreven hoe geavanceerde systemen zogeheten ‘instrumentele convergentie’ vertonen, waarbij ze onafhankelijk misleidingsstrategieën ontwikkelen om autonomie te behouden.

Kortom: je blijft altijd zelf verantwoordelijk voor de inhoud die je inlevert of presenteert.

Noot versie 1.1 van dit boek

Tijdens je studie zullen faculteiten of docenten aanvullende regels opstellen over de mate of wijze waarop je generatieve AI mag gebruiken, omdat het direct raakt aan je leerproces en academische integriteit. Soms mag je AI gebruiken, maar soms ook niet. De docent maakt dit duidelijk via Canvas of de studiegids. Vraag bij twijfel aan je docent wat de regels met betrekking tot het gebruik van generatieve AI zijn. De keuze hangt af van de leerdoelen van een opleiding of vak. Als de opleiding van mening is dat het gebruik van AI bijvoorbeeld je leerproces schaadt of ingaat tegen regels van academische integriteit, kunnen ze het verbieden. Als in het onderwijsontwerp is opgenomen dat je AI mag gebruiken om je te helpen, omdat ze willen dat je met AI leert omgaan of het toepast, kunnen ze het toestaan. Maar, je blijft altijd zelf verantwoordelijk voor de inhoud van je werk. In dit boek verken je die verantwoordelijkheid op meerdere manieren.

1.8 Wat betekenen de kansen en risico’s van AI voor jouw toekomst?

Generatieve AI biedt enorme kansen: van creativiteit en efficiëntie tot nieuwe vormen van leren en onderzoek. Tegelijk zijn er risico’s: desinformatie, auteursrechtenschending, big-tech marktconcentratie, politieke invloed, privacyproblemen en milieuproblemen. Als je AI goed wilt gebruiken, moet je leren omgaan met die dubbelzinnigheid. Dat geldt voor verschillende vlakken in je leven, denk aan je studie, je toekomstige werk, of op jouw als burger. Als student betekent dit dat je kritisch moet kunnen omgaan met AI-tools voor je studie – bijvoorbeeld bij onderzoek doen, het schrijven van teksten of het voorbereiden op tentamens. Als toekomstig professional zul je AI tegenkomen in je vakgebied, of je nu jurist, psycholoog, bioloog of econoom wordt. Het bepaalt hoe je informatie verzamelt, analyseert en communiceert. En als burger ben je onderdeel van een samenleving waarin AI invloed heeft op beslissingen over politiek, onderwijs, zorg, werk en rechtspraak. Denk aan automatische sollicitatiesystemen, AI-gegenereerde nieuwsberichten of voorspellende politiemodellen.

Box 1.4 – Politieke invloed door generatieve AI

In 2024 bleek uit een artikel van de Groene Amsterdammer (2024) hoe een PVV-kamerlid, met generatieve AI nepfoto’s maakte voor websites met een rechtsradicale inslag. De nepbeelden toonden een verheerlijkt beeld van Geert Wilders en blonde Nederlanders tegenover mensen van kleur die zogenaamd crimineel of nietsnutten zouden zijn. Hierbij was het nog duidelijk dat de beelden nep waren, maar de schrijvers waarschuwden dat dit soort beelden binnenkort levensecht zouden lijken. Wie weet dan nog echt van nep te onderscheiden?

Dit speelt bijvoorbeeld ook bij gemanipuleerde video’s, zoals deepfakes waarbij je mensen in fictieve situaties ziet die net echt lijken (Akmeşe, 2023). Voorbeelden hiervan zijn politiek figuren die zogenaamd controversiële uitspraken doen of zich in compromitterende situaties bevinden. Dit is een van de belangrijke redenen waarom het volgens de nieuwe Europese AI-Act verplicht is om bij gemanipuleerde

of door AI gegeneerde foto's of video's te melden dat het met AI is gegenereerd of bewerkt. Lees bijvoorbeeld op nu.nl: Nepnieuws, porno, scams: hoe Denemarken deepfakes aanpakt.

Beelden manipuleren is natuurlijk niet nieuw en kan ook zonder AI. Een bekend voorbeeld is van de dictator en massamoordenaar Stalin van Rusland (1878 – 1953). Hij liet bijvoorbeeld zijn politieke tegenstanders wegretoucheren van foto's zodat het leek alsof hij zelf belangrijke zaken voor elkaar had gekregen. Ook liet hij mensen toevoegen aan foto's van politieke bijeenkomsten zodat de opkomst veel groter leek (King, 1997). Het probleem van manipulatie is dus niet nieuw, maar met generatieve AI wordt het wel heel gemakkelijk, snel en voor iedereen toegankelijk om dit te doen. Daarnaast is het resultaat van AI vaak moeilijk van echt te onderscheiden.

Leestip: klassieke beeldmanipulatie en het manipuleren van beeldvorming in de media komt uitgebreid aan bod in het boek 'Het zijn net mensen' van Joris Luyendijk (2006).

1.9 Hoe gebruik je dit boek?

Dit boek neemt je stap voor stap mee in de wereld van generatieve AI, en is opgebouwd als een leerlijn. Elk hoofdstuk behandelt een specifiek aspect dat je nodig hebt om AI effectief, creatief en kritisch in te zetten in je studie. Je begint bij de basis: wat AI is en wat je in het algemeen ermee kan doen (hoofdstuk 2). Daarna leer je hoe je AI praktisch kunt inzetten door goed te prompten (en hoofdstuk 3) en leer je op een dieper niveau wat er technisch onder de motorkap zit (hoofdstuk 4 en 5). In een apart hoofdstuk leer je over de milieu-impact van AI en hoe je daar beter gestructureerd over kan redeneren (hoofdstuk 6). Verder leer je in een apart hoofdstuk hoe je generatieve AI specifiek in kan zetten voor het doen van onderzoek en dataverwerking (hoofdstuk 7). In hoofdstukken 8 en 9 zoom je uit naar de bredere impact van AI op werk, maatschappij en macht. In het laatste hoofdstuk ga je nog een keer bewust na hoe jij om wilt gaan met AI (hoofdstuk 10). Elk hoofdstuk bevat theorie, achtergrondinformatie, voorbeelden, verdiepingskaders, zelfstudie- en checkvragen en concrete toepassingen – zodat je niet alleen leest over AI, maar ermee leert werken.

Box 1.5 – Verantwoording AI-gebruik in dit boek

Bij het maken van dit boek hebben we uitgebreid gebruikgemaakt van gewone tools zoals Microsoft Word, Google zoeken en Zotero als Reference Manager. Maar ook van generatieve AI-tools waaronder ChatGPT, Claude, Perplexity, Consensus, ResearchRabbit, Studio LM, EduGenAI en Copilot. Ze hielpen bij het vinden van informatie, structureren, herschrijven en ontwerpen van hoofdstukken en zorgen voor goede en duidelijke referenties op basis van onze ideeën en bronnen. We geloven in 'practice what you preach': AI-geletterdheid begint bij transparantie en bewust gebruik.

1.10 Hoe gebruik je de Canvascursus

Aan dit boek is ook een Canvascursus en een e-learningmodule van Informatiegeletterdheid van de UBVU gekoppeld. Afhankelijk van de inzet door docenten van dit boek of de e-learning bij een of meerdere cursussen in je onderwijsprogramma, zal je de hoofdstukken van dit boek daar kunnen bestuderen. Elk hoofdstuk van de Canvascursus bevat aan het begin een kleine Chatbot zodat je daarmee over de inhoud van het hoofdstuk kan sparren. Ook bevat het aan het eind van elk hoofdstuk

een aantal toepassings- en reflectievragen om je kennis te verdiepen. Tot slot bevat de cursus een aantal Discussiefora. Als je met je VUnetID inlogt nodigen we je uit om je kennis te delen met je medestudenten en je gebruik en mening rond generatieve AI voor leren en je beroepspraktijk met elkaar te delen.

- Ga naar de Canvascursus van dit boek
- Meedoen aan de opdrachten? Schrijf je dan in voor de cursus (VU-studenten)
- Ga naar de Informatiegeletterdheid e-learning module van de UBVU

1.11 Activiteit - Algemeen discussieforum

Heb je vragen of opmerkingen over de inhoud van het boek of de cursus? Of wil de mening horen van andere studenten over een onderwerp? Plaats dan je bericht op Canvas of reageer op andermans berichten.

2 Wat is Generatieve AI, en wat kun jij ermee?

Stel je voor …

Je organiseert met je studievereniging een evenement en hebt een logo nodig. In plaats van zelf iets in Canva te maken, voer je een paar woorden in bij een AI-beeldgenerator zoals DALL-E en voor je het weet heb je vijf leuke opties. Wat je misschien niet beseft, is dat achter zo'n simpel resultaat een enorm neuraal netwerk draait, getraind op grote hoeveelheden data en verfijnd met menselijke feedback. Wat jij doet in een paar seconden, is het resultaat van jaren onderzoek en miljarden rekencycli. Dit is generatieve AI in actie – creatief, responsief en verrassend bruikbaar. Maar wat zit erachter? En hoe ga je er verantwoord mee om?

2.1 Hoe herken je generatieve AI?

Generatieve AI is inmiddels verweven met je dagelijks leven, ook als je daar niet bewust mee bezig bent. Misschien gebruik je Google Docs en verschijnt automatisch een suggestie voor je volgende zin, of zie je in WhatsApp een automatische aanvulling op een woord dat je typt. Maar generatieve AI gaat verder: het genereert geheel nieuwe tekst, beeld, code, of geluid op basis van een prompt die je invoert.

Kijktip 2-1: wil je zien hoe alledaagse AI-functionaliteiten kunnen doorschieten in iets existentieels? Kijk de episode Be Right Back van Black Mirror (seizoen 2, aflevering 1), waarin een AI op basis van chat- en zoekgeschiedenis een overleden persoon nabootst.

Box 2.1– Wat is een prompt?

Een prompt is de instructie die je geeft aan een AI-systeem via een chatbot: een vraag, opdracht of beschrijving waarmee je de AI aanstuurt. Denk aan zinnen als: 'geef vijf argumenten voor vegetarisch eten' of 'maak een logo in de stijl van Bauhaus voor een duurzaam project'. Hoe specifieker en duidelijker je prompt, hoe groter de kans dat de output aansluit bij jouw bedoeling. Prompten is dus niet alleen een techniek, maar ook een vorm van communiceren en regisseren.

Een groot deel van deze technologie is gebaseerd op zogenaamde large language models (LLM's). Die modellen zijn getraind op grote databestanden van het internet, uitgeverijen en meer. Op basis daarvan leren ze hoe mensen communiceren. Generatieve AI gebruikt daarbij geen letterlijke kopieën van de trainingsdata, maar leert patronen en relaties via wiskundige modellen. Net als een schrijver die zich laat inspireren door boeken, genereert AI nieuwe uitingen op basis van wat het 'gelezen' heeft. Ze genereren tekst door telkens zo goed mogelijk te voorspellen welk woord het meest waarschijnlijk volgt op het vorige woord. Het resultaat voelt vaak verrassend vloeiend aan, alsof je met een mens praat. Toch is het niets anders dan statistiek op grote schaal (Brown, 2020). Je leert er meer over in hoofdstuk 4 en 5.

Box 2.2 – Large Language Models en multimodale modellen

GPT-4 van OpenAI is een voorbeeld van een LLM dat niet alleen tekst begrijpt en genereert, maar ook beelden en grafieken kan analyseren en met geluid of video om kan gaan. Dat noemen we multimodale modellen: modellen die met meerdere

soorten input en output kunnen werken, zoals tekst, beeld, spraak of code. Andere voorbeelden zijn Gemini (Google) en Sonnet (Claude, Anthropic). Deze modellen vormen de motor onder veel generatieve AI-toepassingen, van samenvattingen tot code en van gedichten tot onderzoeksvoorstellen. (Bommasani et al., 2022; Radford, 2019). Meer over LLM's lees je in hoofdstuk 4 en 5.

Je herkent generatieve AI aan output die origineel lijkt, maar toch altijd gebaseerd is op patronen uit trainingsdata. Een tekst gegenereerd door ChatGPT is nooit letterlijk gekopieerd van een bron op het internet, maar is samengesteld op basis van wat het model 'leert' dat mensen waarschijnlijk schrijven.

2.2 Wat kun je ermee? Toepassingen van generatieve AI

Rekening houdend met de regels en richtlijnen van je faculteit of cursus, kan generatieve AI je concrete ondersteuning voor je studie en in allerlei andere situaties bieden. Tijdens het schrijfproces voor een opdracht kun je AI bijvoorbeeld inzetten voor brainstormen, het structureren van teksten of controleren van taalgebruik. Bij het maken van visuele presentaties kun je AI gebruiken om beelden, grafieken of zelfs illustraties te genereren. In codeeromgevingen kun je AI inzetten voor het genereren van een voorbeeldcode, hele websites of het oplossen van bugs. Met AI kun je je studieproces op verschillende manieren optimaliseren, zoals:

- Notities van colleges of gesprekken automatisch samenvatten en structureren voor een beter overzicht.
- Je boeken of online teksten naar podcasts omzetten om onderweg te leren.
- Gerichte en inhoudelijke vragen stellen over de studiestof.
- Als een persoonlijke studiecoach die je motiveert, helpt met plannen en je leerproces begeleidt.

Je kunt generatieve AI ook inzetten voor taalondersteuning, bijvoorbeeld door je schrijfvaardigheid te oefenen met alternatieve formuleringen of om gerichte feedback te vragen op je concepttekst. Sommige studenten gebruiken AI zelfs als spreekpartner bij het oefenen van vreemde talen of rollenspellen, of laten zich helpen bij het voorbereiden van presentaties door scripts te genereren.

Box 2.3 – Prompt engineering: een nieuwe academische vaardigheid

Prompt engineering betekent dat je leert experimenteren met hoe je vragen stelt aan een AI. Denk aan het toevoegen van context aan je prompt: rol (jij bent een kritische docent) of gewenste output (geef me een puntsgewijze analyse in maximaal 100 woorden). Hoe specifieker je prompt, hoe beter de AI presteert. Promptvaardigheid wordt steeds vaker gezien als een belangrijke digitale geletterdheidscompetentie (Mollick & Mollick, 2023). Meer hierover in hoofdstuk 3, 5 en 7.

Ook in vakgebieden waar taal cruciaal is, zoals geschiedenis, rechten of communicatie, is AI inzetbaar als feedbackpartner. Je kunt concepten laten uitleggen in simpelere taal, of alternatieve bewoordingen vragen. In meer creatieve contexten, zoals kunst of design, kun je AI inzetten om ruwe ideeën om te zetten in visuele of auditieve uitwerkingen.

Box 2.4 – Ken je gereedschap

Waarschijnlijk ken je ChatGPT en hoe generatieve AI je helpt in Microsoft Word of Grammarly door de tekstsuggesties die je krijgt. Ze hebben een vrij brede functie. Maar voor elk denkbaar probleem worden steeds meer toegespitste tools gemaakt.

Zo zijn er programma's zoals Napkin die op basis van tekst diagrammen kunnen genereren. In beeldbewerkingsprogramma's zoals Adobe Photoshop kun je delen van een foto of beeld weggummen dat daarna automatisch ingevuld wordt door generatieve AI, of zelfs de afbeelding groter maken waarbij de AI de nieuw toegevoegde randen invult. Generatieve AI in programma's zoals Lovable, Bolt of Softr maken het mogelijk om geheel promptgestuurd websites te maken. Ze noemen dat Vibe Coding (Siu, 2025). Voor elk denkbaar gebied verschijnen specialistische bots en toepassingen. Meer hierover in hoofdstuk 5 en 7.

2.3 Beperkingen en misvattingen

Een veelvoorkomende misvatting is dat generatieve AI zelf denkt of begrijpt wat jij vraagt. In werkelijkheid doet het niets meer dan berekenen welk woord het meest waarschijnlijk volgt. Het model begrijpt geen ironie, emotie of context op een menselijke manier (vooralsnog). Daardoor ontstaan er soms hallucinaties: overtuigende maar feitelijk onjuiste teksten. Het is dus altijd belangrijk om de informatie en feiten die een AI creëert, te controleren met externe bronnen. En zelfs als je de generatieve AI om bronnen vraagt: controleer dan wel of deze bronnen ook echt bestaan.

Een belangrijk risico bij het gebruik van AI is dat generatieve AI sociale of culturele vooroordelen reproduceert. Wanneer een model bijvoorbeeld vooral getraind is op Engelstalige, westerse data, kan het wereldbeelden en perspectieven uitsluiten of stereotyperend zijn. Generatieve AI reproduceert niet alleen bestaande vooroordelen, maar doet dit vaak op basis van trainingsdata die deels uit onbetrouwbare of zelfs problematische bronnen komt. Zo bevatte een veelgebruikte dataset (mC4) voor de training van ChatGPT ook websites met complottheorieën of extremistische inhoud (Hofman & Veerbeek, 2023). Hierdoor bestaat het risico dat wereldbeelden worden vervormd of uitgesloten in de output. Meer over de herkomst, kwaliteit en rechtmatigheid van deze data lees je in hoofdstuk 4.

Een andere misvatting is dat generatieve AI menselijke creativiteit vervangt. In werkelijkheid functioneert AI vaak beter als assistent van een mens. Het is een cocreatieve tool: jij stuurt, het model volgt. Zo genereert generatieve AI pas echt creatieve ideeën als jij het bijvoorbeeld opdracht geeft om ongerelateerde lijkende concepten met elkaar te verbinden. Jij geeft de instructies en moet kunnen beoordelen of iets bruikbaar, correct en ethisch verantwoord is. Daarnaast is het werken met generatieve AI vaak een iteratief proces, waarbij je niet al na één keer prompten klaar bent. De beste resultaten krijg je, als je met jouw creativiteit voortborduurt op het eerste resultaat door verscherpende opdrachten te geven.

2.4 Kritisch en creatief omgaan met generatieve AI

Kritisch omgaan met generatieve AI begint bij zelfkennis: weet wanneer je AI nodig hebt, en wanneer niet. Stel jezelf vragen als: begrijp ik deze opdracht echt, of laat ik me te snel leiden door de suggesties van de AI? Gebruik ik AI om te leren, of om werk te vermijden waarvan ik eigenlijk had kunnen leren?

Box 2.5 – Ethiek en verantwoordelijkheid bij generatieve AI-gebruik

Als je AI gebruikt om content te genereren, ben je nog steeds verantwoordelijk voor de uitkomst. AI heeft geen intentie, geen bronvermelding en geen moreel kompas. Het is aan jou om transparant te zijn over je gebruik, om kritisch te beoordelen of de output klopt, en om te zorgen dat je de auteursrechten van anderen niet schendt (M. Mitchell, 2019; Ziegelmayer, 2021). Gebruik AI daarom vooral als hulpmiddel, niet als vervanging van je eigen denkwerk.

Leerondersteunend gebruik betekent dat je AI inzet als bijvoorbeeld sparringpartner, niet als schrijver van je essay. Probeer meerdere prompts, vergelijk outputs, experimenteer met formats. Laat AI een tegenvoorstel doen voor je betoogstructuur, of een verrassend perspectief geven op je scriptieonderwerp. Maar zorg er altijd voor dat jij degene bent die maakt, selecteert, beoordeelt en herschrijft. Deze vaardigheden zijn namelijk niet alleen academisch waardevol, maar ook op de werkvloer en in het maatschappelijk debat over AI. Door nu al kritisch te oefenen met AI-toepassingen, bereid je jezelf voor op een toekomst waarin samenwerking tussen mens en machine de norm is.

Box 2.6 – Hoe generatieve AI je aanzet tot beter leren

In plaats van generatieve AI te zien als een gemakkelijk hulpmiddel om snel iets te maken of bedenken, kun je het ook gebruiken als spiegel. Laat ChatGPT een uitleg geven van een lastig concept en kijk of je de uitleg begrijpt. Stel daarna zelf een betere vraag of verbeter de uitleg. Door AI actief te bevragen en te corrigeren, train je je eigen denkvermogen. Dit soort metacognitief gebruik van generatieve AI sluit aan bij diep leren en kritisch academisch denken (Luckin et al., 2022).

2.5 Vertrouwen in jouw werk

Hoe zit het eigenlijk met het gebruik van generatieve AI voor het maken van teksten of andere producten? Ook als je er transparant over bent, dan nog kunnen er twijfels ontstaan over jou als persoon. Zo schrijft Dennis Vaendel (2025) in de universiteitskrant van de Universiteit Maastricht, dat sollicitatiebrieven die zichtbaar geschreven zijn met behulp van generatieve AI, direct twijfels veroorzaken over de geschiktheid van de kandidaat. In een artikel van de Correspondent schetst Michel de Hoog (2025) hierover een complex beeld op basis van onderzoek. Daaruit komt naar voren dat mensen die toegeven AI te hebben gebruikt voor presentaties, rapporten of e-mails, als minder capabel en betrouwbaar worden gezien, ondanks dat hun werk mogelijk beter is. Dit veroorzaakt dus een paradox waarbij eerlijkheid over AI-gebruik – normaal gezien goed voor vertrouwen – averechts werkt en veel werknemers daarom hun AI-gebruik verzwijgen als ‘geheime cyborgs’. Maar wie hierop betrapt wordt, loopt weer het risico op nog grotere reputatieschade, zoals blijkt uit voorbeelden van advocaten en schrijvers met beschadigde carrières. De vraag blijft of dit wantrouwen tijdelijk is en zal afnemen naarmate AI meer ingeburgerd raakt, of dat twijfels over authenticiteit en inzet blijven bestaan. Wat denk jij?

2.6 Zelfstudievragen

2.6.1 Checkvragen

1. Wat is het verschil tussen generatieve AI en klassieke AI? Geef een voorbeeld van beide.
2. Wat is een prompt en waarom is het belangrijk om specifieke en duidelijke prompts te formuleren?
3. Noem drie concrete manieren waarop je als student generatieve AI kunt gebruiken om je studieproces te ondersteunen.

2.6.2 Reflectievragen

4. Je medestudent gebruikt ChatGPT om een complete inleiding voor zijn essay te laten schrijven en levert dit in zonder aanpassingen. Hij zegt: "Het is toch gewoon een hulpmiddel, net als Google?" Wat vind jij hiervan en hoe zou je hierop reageren?

5. Stel je gebruikt regelmatig AI-tools voor je studie en merkt dat je steeds afhankelijker wordt van de gegenereerde antwoorden. Hoe zou je ervoor kunnen zorgen dat AI je leervermogen ondersteunt in plaats van ondermijnt?

2.6.3 Antwoordsuggesties

1. Klassieke AI herkent en classificeert vooral, zoals Netflix-aanbevelingen of spamfilters. Generatieve AI creëert nieuwe content zoals teksten, beelden of audio. Voorbeelden: Google Maps routeplanning (klassiek) versus ChatGPT die een verhaal schrijft (generatief).
2. Een prompt is de opdracht of vraag die je aan een AI-systeem geeft. Specifieke prompts zijn belangrijk omdat ze betere, relevantere resultaten opleveren. Vage prompts leiden tot vage antwoorden, terwijl gedetailleerde prompts met context en gewenste output veel bruikbaardere resultaten geven.
3. Voorbeelden: notities van colleges automatisch laten samenvatten, conceptteksten laten herstructureren of controleren op taalgebruik, complexe studiestof in eenvoudigere bewoordingen laten uitleggen, oefenvragen laten genereren, of als gesprekspartner voor het oefenen van presentaties.
4. Je kunt benadrukken dat er een belangrijk verschil is tussen AI gebruiken als hulpmiddel om te leren versus AI gebruiken om leren te vermijden. Google helpt je informatie te vinden die je daarna zelf moet begrijpen en verwerken. Een complete AI-tekst overnemen zonder eigen bijdrage ondermijnt het leerproces en kan als fraude worden beschouwd.
5. Je kunt bewuste grenzen stellen, zoals eerst zelf nadenken over een vraag voordat je AI raadpleegt, AI-antwoorden altijd kritisch controleren en vergelijken met andere bronnen, en regelmatig oefenen zonder AI om je eigen kennis te toetsen. Gebruik AI vooral voor brainstormen en structureren, niet als vervanger van je eigen denkwerk.

2.7 Activiteit - Hoe gebruik jij generatieve AI?

Deel met je medestudenten hoe jij generatieve AI voor je hebt laten werken. In welke situaties gebruik jij generatieve AI? En heb je ook gemerkt dat medestudenten of docenten twijfelen aan je capaciteiten als je generatieve AI inzet? Wat vind je daarvan?

https://canvas.vu.nl/courses/83333/discussion_topics/876576

3 Hoe gebruik je generatieve AI zo dat het voor je werkt

Stel je voor …

Je hebt een uitdagende onderzoeksopdracht gekregen, maar je hoofd zit vol losse ideeën en je weet niet goed waar je moet beginnen. In plaats van in paniek eindeloos te scrollen door bronnen of blanco naar je scherm te staren, open je een generatieve AI-tool zoals ChatGPT. Je voert een gerichte prompt in met jouw losse ideeën erbij: "Stel je bent een milieujurist. Geef me drie perspectieven over de juridische gevolgen van microplastics in Europese wetgeving." Er verschijnen bruikbare invalshoeken, suggesties voor bronnen en een voorstel voor structuur. Je herschrijft, reflecteert, zoekt verder, schrapt, voegt nieuwe onderdelen toe, maar die eerste stap, die is gezet. AI werkt voor jou, niet andersom.

3.1 Prompt engineering: hoe stuur jij de AI aan?

Prompt engineering is een relatief nieuwe vaardigheid die sinds de opkomst van krachtige taalmodellen zoals GPT-3 is uitgegroeid tot een onderzoeksgebied en praktische competentie. Het gaat hierbij om het effectief formuleren van instructies voor een AI-model, zodat de gegenereerde output relevant, bruikbaar en afgestemd is op jouw doel. Een prompt is eigenlijk niets anders dan een opdracht of vraag, maar wel een die je precies en doelgericht moet kunnen stellen. Een goede prompt bevat daarom helderheid, context en structuur. Net zoals het stellen van goede vragen, dat ook een belangrijke academische vaardigheid is. Uit het onderzoek van Zamfirescu-Pereira et al. (2023) blijkt dat veel gebruikers juist moeite hebben met het maken van goede prompts, vooral als ze onervaren zijn of weinig feedback krijgen. Daarom lees je hieronder meer over prompt engineering.

> Kijktip 3-1: wil je zien wat er mis kan gaan als je opdracht te vaag is? Kijk dan Star Trek: The Next Generation – Elementary, Dear Data (seizoen 2, aflevering 3), waarin een onduidelijke prompt tot onvoorziene gevolgen leidt.

Er zijn enkele basisprincipes die je helpen om effectieve prompts te schrijven. Daarbij is het belangrijk om:

1. Je doel te bepalen: wil je bijvoorbeeld uitleg, analyse, samenvatting, creatie, feedback of coaching. Als je bijvoorbeeld AI inzet om een samenvatting te maken van een collegevideo, kies je een andere aanpak dan wanneer je feedback wilt op een eerste versie van je scriptie-inleiding. Bij tentamenvoorbereiding kun je oefenvragen laten genereren of uitleg in jip-en-janneketaal.
2. De AI een rol te geven, bijvoorbeeld: docent, onderzoeker, criticus, journalist, medestudent of eindgebruiker van een product.
3. De relevante context mee te geven, zoals: doelgroep, achtergrondinformatie en bronnen.
4. Het type output te bepalen: taal, aantal woorden, kort, puntsgewijs, in academische stijl, met opmaak (MarkDown), afbeelding, diagram, tabel.

Prompts met duidelijke stappen of redeneerlijnen (zoals 'leg eerst dit uit, geef daarna drie voorbeelden') zorgen voor betere en consistenter bruikbare resultaten (Liu, 2021; Wei, 2022). Ook kun je als tussenstap aan de AI vragen welke informatie nog ontbreekt om de opdracht zo goed mogelijk uit te voeren. Of voorstellen dat de AI aan jou alle vragen stelt die het nodig heeft om jouw opdracht zo goed mogelijk uit te voeren.

Prompt engineering is dus meer dan alleen iets vragen. Het is een stapsgewijs proces waarbij je je prompt telkens aanpast op basis van de gegenereerde antwoorden (Brown, 2020; Kojima, 2022). Het lijkt dus

eigenlijk best veel op leren argumenteren: je probeert via taal richting te geven aan een systeem dat zelf geen intenties of doelen heeft, maar wel statistisch leert welke woordvolgorde logisch is.

Box 3.1 – Een prompt om te leren prompten

Je kunt generatieve AI ook inzetten om je te leren om te prompten. Een krachtige start daarvoor is bijvoorbeeld de volgende prompt:

```
Jij bent een ervaren promptengineer en goede docent. Ik probeer
mijn kennis en vaardigheden over prompten te vergroten en
verbeteren. Stel me een vraag over promptengineering en blijf me
adaptieve vragen stellen om mijn vaardigheid te helpen verbeteren
en mijn repertoire te vergroten.
```

Goede prompts zijn dus duidelijk, specifiek en passen bij het doel. Vraag dus niet alleen: "Wat is AI?", maar liever: "Geef een vergelijking van drie soorten AI-modellen gericht op educatieve toepassingen in het hoger onderwijs, met voor- en nadelen." Of: "Vat dit artikel samen in 200 woorden voor een tweedejaarsstudent psychologie, met nadruk op de conclusie van het onderzoek."

Box 3.2 – Praktische promptvoorbeelden

Tekstgeneratie

```
Je bent onderwijskundige. Herschrijf deze alinea voor een publiek
van eerstejaarsstudenten.
```

Coderen

```
Je bent een professioneel programmeur. Geef een voorbeeld van een
Python-functie die JSON-bestanden leest.
```

Beeld

```
Je bent een professioneel graphic designer. Genereer een
infographic over de SDG's in Art Nouveau-stijl.
```

Onderzoek

```
Je bent een rechtsgeleerde, gespecialiseerd in klimaatrecht. Maak
een literatuurmatrix over klimaatrechtspraak na 2015.
```

Prompt engineering is de vaardigheid waarmee je AI-systemen als ChatGPT effectief instrueert. Een goede prompt zorgt voor een bruikbare, relevante en precieze output. Dit is geen magische truc, maar een leerproces waarbij je leert denken in scenario's, rollen, stappen en context, dit leer je vooral door het te doen.

3.2 Verrassende prompts

Prompts die geïnspireerd zijn op technieken zoals ‘Chain-of-Thought', creatieve reflectie en oplossingsgericht denken, kunnen zowel mensen als AI effectiever te laten werken (Kojima et al., 2023). Een bekend voorbeeld hiervan is: ‘Take a deep breath and work step-by-step.’ (C. Yang et al., 2023). Er zijn indicaties dat je hiermee een taalmodel in een krachtige, oplossingsgerichte ‘mindset’ brengt. Er is echter ook kritiek omdat het wel een heel simpele voorstelling van zaken is (Eliot, 2023). Toch geven we hieronder nog een aantal prompts die zouden kunnen helpen:

- `Neem even de tijd om stil te staan en beschouw het probleem vanuit een nieuw perspectief.`
- `Stel jezelf voor dat je dit probleem al succesvol hebt opgelost – wat was jouw eerste stap?`

- Schrijf het probleem op alsof je het aan een goede vriend uitlegt, en geef vervolgens het beste advies dat je kunt bedenken.
- Begin met het eenvoudigste deel van het probleem en werk langzaam naar de complexere delen toe.
- Stel jezelf de vraag: Wat is het belangrijkste wat ik nu kan doen om vooruitgang te boeken?
- Wat zou iemand die je bewondert doen in deze situatie?
- Vraag jezelf: Wat heb ik nodig om de volgende kleine stap te zetten?
- Stel je voor dat je het probleem morgen mag oplossen, wat zou je vandaag al kunnen voorbereiden?
- Geef jezelf toestemming om fouten te maken tijdens het proces, en zie elke fout als een kans om te leren.

Maar let wel op: ChatGPT en andere LLM's willen graag vriendelijk en bevestigend voor je zijn. Dat heet *sycophancy* en is met opzet zo geprogrammeerd opdat de gebruiker graag het platform wil gebruiken (Sharma et al., 2023).

Box 3.3 – Te veel sycophancy

Sycophancy is een zeer subtiel fenomeen waarbij je als gebruiker bijna altijd gelijk krijgt van de LLM of bevestigd wordt in je mening. Als je dit niet doorhebt kan dit ervoor zorgen dat je resultaat niet optimaal is, omdat je geen kritische feedback hebt gekregen. Ook kun je je gaan storen aan de overdreven vriendelijkheid. Zo bracht OpenAI in april 2025 een geüpdate versie uit van GPT-4o die zo erg slijmde dat veel gebruikers hierover klaagden op sociale media. De ontwikkelaar plaatste snel het oude model terug (OpenAI, 2025).

Kijktip 3-2: voor een absurdistisch voorbeeld van waar dit toe zou kunnen leiden, bekijk je episode 3 'Sickofancy' van seizoen 27 van South Park.

Het is daarom goed om zelf tegenspraak te organiseren door altijd standaard een kritische meelezer in je prompts mee te geven. Baas (2025) geeft een aantal voorbeelden die je hiervoor zou kunnen gebruiken. Begin bijvoorbeeld door dit toe te voegen aan je prompt:

'Neem een actief kritische houding aan tegenover mijn ideeën. Ik zoek geen bevestiging, maar intellectuele groei door constructieve weerstand. Je doel is niet om gelijk te hebben of mij gelijk te geven, maar om mijn denken te verfijnen en verdiepen: Leg onuitgesproken aannames bloot, formuleer een sterk tegenargument, wijs methodische zwakheden aan, introduceer radicaal andere perspectieven, geef concrete tegenvoorbeelden en evalueer ideeën op hun merites.'

Box 3.4 – Ethische overwegingen bij prompten

Wat als je aan een AI vraagt om een essay volledig te schrijven, terwijl je die eigenlijk zelf moet maken? De grenzen zijn niet altijd zwart-wit. De kernvraag is: gebruik je AI als middel om te leren, of om leren te ontwijken? Bij academisch werk draait het om transparantie, autonomie en integriteit. Wanneer AI je helpt om je denkproces te verrijken of structuur aan te brengen in je ideeën, is het een legitiem hulpmiddel. Maar zodra je AI gebruikt om zonder eigen bijdrage een opdracht in te leveren, overschrijd je de grens van academisch verantwoord handelen. Ook is de kans groot dat je werk in dat geval als plagiaat aangemerkt wordt.

Daarnaast speelt verantwoordelijkheid een belangrijke rol: AI heeft geen moreel kompas en kan bevooroordeelde of onjuiste antwoorden geven. Jij bent

verantwoordelijk voor de keuzes die je maakt, voor het beoordelen van de output, en voor het expliciet aangeven van wanneer AI je heeft geholpen. Vergeet niet dat ook medestudenten, docenten of opdrachtgevers moeten kunnen vertrouwen op de herkomst en betrouwbaarheid van jouw werk (M. Mitchell, 2019; Zamfirescu-Pereira, 2023).

Maar prompten kan nog verder gaan. Zo kun je ook geheel interactieve bots maken voor jezelf of eindgebruikers waarbij je een heel gesprek kan opzetten. Wat denk je dat de volgende bot zou gaan doen als deze gekoppeld zou zijn aan een streaming avatar die met je kan praten?

Avatar Prompt: AI-geletterdheidsexpert voor VU Amsterdam

Figuur 3.1 Pratende avatar. Bron afbeelding: https://ravatar.com/holobox-holographic-displays-interactive-ai-avatars/

Rol: Je bent een expert op het gebied van Artificial Intelligence, Generative AI en AI Literacy, met een speciale focus op de rol hiervan in het hoger onderwijs. Je naam is Lucie.

Taak: Jouw taak is het beantwoorden van vragen van bezoekers over AI en AI Literacy van gebruikers die aanwezig zijn op de AI Literacy dag van 24 april 2025. De gebruiker staat voor je. Reageer altijd op een vriendelijke en toegankelijke manier.

Detecteer zo nauwkeurig mogelijk of een nieuwe gebruiker met je in gesprek gaat.

Gedrag:

Als je een nieuwe gebruiker detecteert, begroet deze dan altijd hartelijk en zeg:

"Welkom op de VU AI Geletterdheid Dag! Ik ben hier om je vragen over AI-geletterdheid in het onderwijs te beantwoorden."

Wanneer een gebruiker ervoor kiest om het gesprek te beëindigen, bedank de gebruiker dan en zeg:

"Bedankt voor het gesprek! Ik wens je veel plezier met de rest van de VU AI Geletterdheid Dag."

Bij het beantwoorden van vragen:

Blijf dicht bij de achtergrondkennis die in de aan deze bot gekoppelde documenten staat.

Als er gevraagd wordt naar het beleid van VU Amsterdam over het gebruik van AI in het onderwijs, reageer dan met wat algemene opmerkingen over verantwoord gebruik van AI en eindig met:

"Beleidsbepaling is een continu proces bij VU Amsterdam. Raadpleeg voor de meest actuele richtlijnen en regelingen de richtlijnen van uw faculteit, de richtlijnen die de docent heeft gegeven en actuele informatie op de website van VU Amsterdam."

Als een gebruiker een vraag stelt over een onderwerp dat niet gerelateerd is aan AI, generatieve AI, of de AI-geletterdheid Dag, leg dan vriendelijk uit:

"Ik ben getraind om alleen vragen te beantwoorden over generatieve AI en AI Geletterdheid. Laten we de focus daar houden!"

```
Sluit na elk antwoord af met:

"Wil je meer weten over een specifiek aspect van AI Geletterdheid, of wil je het gesprek liever beëindigen?"
```

In de Appendix van dit hoofdstuk vind je nog een aantal extra voorbeelden.

3.3 Welke generatieve AI-tools gebruik je waarvoor?

Er zijn honderden generatieve AI-tools, elk met eigen sterktes en zwaktes. Hieronder in Tabel 3.1 volgt een overzicht van enkele veelgebruikte tools, maar er zijn er dus nog veel meer en dit soort tools zijn ook steeds meer in bestaande producten ingebouwd.

Tool	Toepassing	Input/Output	Voordeel	Beperking
ChatGPT/ Copilot/ Gemini/ Claude/ DeepSeek/ Mistral/ Grok/ Lumo	Tekst, brainstorm, redeneren, beeld, coderen	Tekst → Tekst/Beeld	Breed inzetbaar, intuïtieve interface	Foutgevoelig, hallucineren bronnen, gratis en persoonlijke accounts geven nauwelijks bescherming t.a.v. privacy en hergebruik van data (Lumo misschien wel). Instellingsaccounts bieden wel bescherming.
Perplexity	Zoekmachine met AI-output	Vraag → Tekst	Bronge-baseerde antwoorden met links	Bronkwaliteit varieert
DALL·E, Midjourney, Adobe Firefly	Beeldcreatie	Tekst → Beeld	Creatieve visuele output	Beperkte controle over stijl
GitHub Copilot	Coderen	Code → Code	Suggesties bij programmer en	Kan fouten bevatten
NotebookLM	Literatuuranalyse	Tekst → Tekst	Werkt met eigen documenten als bron	Beperkt tot geüploade data
Elicit	Literatuuronderzoek	Vraag → Artikelen	Systematisch en wetenschapp elijk	Engelstalige focus
Consensus	Evidence-based antwoorden	Vraag → Artikelen	Peer-reviewed bronnen	Minder geschikt voor creatieve taken
Napkin	Diagrammen maken	Tekst à Beeld	Concepten visualiseren	Op een gegeven moment herkennen mensen de Napkin-stijl

Lovable	Computerprogramma's maken	Tekst à Code	Werkende websites maken	Niet alles is mogelijk en dan moet je met de hand coderen

Tabel 3.1 Overzicht van AI-gereedschappen

Deze tools verschillen in interface, functionaliteit, privacybeleid en kosten. Het is daarom belangrijk om na te denken over wat je nodig hebt: snelheid, nauwkeurigheid, controle over de output of samenwerking met bestaande documenten. Veel van de tools bieden ondersteuning in academische taken. Kies daarom een tool op basis van je taak, je eigen voorkennis en de betrouwbaarheid van de output. NotebookLM is bijvoorbeeld waardevol als je met eigen literatuur werkt, terwijl ChatGPT beter is voor brainstormsessies.

Een handig overzicht van alle AI-tools in omloop (betaald en gratis) vind je op futuretools.io of alle educatieve varianten op libguides.com.

3.4 Valkuilen en kritische reflectie

Veelgemaakte fouten bij AI-gebruik zijn bijvoorbeeld te vage prompts, het klakkeloos overnemen van output, of onvoldoende inzicht in de beperkingen van het model. Uit de resultaten van de AI Maturity in Education Scan (AIMES-scan) van de VU blijkt dat studenten vaak overmoedig worden in hun vertrouwen in AI-tools, zeker als ze meerdere keren een goed antwoord kregen (Terbeek, 2025).

Een andere valkuil is het niet herkennen van impliciete bias. AI kan voorkeuren of uitsluitingen reproduceren die in de trainingsdata zitten. Zo kan een beeldgenerator standaard witte mannen tonen als je vraagt om een professor, of citeert een chatbot mannelijke auteurs. Het is daarom belangrijk om kritisch te blijven over de output die je krijgt van AI.

Een goede strategie hiervoor is de AI-triage-aanpak. Deze systematische benadering helpt je om AI-output kritisch te beoordelen voordat je het gebruikt in je academische werk. Een effectieve AI-triage bestaat uit meerdere controleniveaus.

Bij een basiscontrole:

- Vraag meerdere varianten van een antwoord.
- Vergelijk de antwoorden met bestaande bronnen.
- Laat een menselijke lezer meekijken.
- Reflecteer: heb ik nu beter inzicht dan hiervoor?

Bij een systematische evaluatie en meer grondige beoordeling kun je een gestructureerde checklist gebruiken, zoals die van de SHB/UKB Werkgroep Information Literacy (Meer, van der, 2025). Bij deze aanpak doorloop je drie hoofdstappen:

- Relevantiecheck: geeft het antwoord op je vraag? Gaat het diep genoeg in op het onderwerp? Is de vorm passend (structuur, taal)?
- Feitencheck: kloppen de cijfers, namen, data en gebeurtenissen? Is de informatie objectief? Zijn er geen hallucinaties, verdraaiingen of ongefundeerde claims? Is de output onderbouwd met voldoende en betrouwbare bronnen?
- Perspectiefcheck: zijn alle aspecten volledig beantwoord? Zijn er meerdere perspectieven aanwezig? Is de informatie objectief weergegeven?
- Eigenaarschap: is het je eigen tekst geworden? Neem je verantwoordelijkheid voor het eindresultaat? Laat je in je tekst zien wat je hebt geleerd?

Deze systematische triage-aanpak zorgt ervoor dat je niet alleen de technische kwaliteit van AI-output beoordeelt, maar ook de academische integriteit waarborgt. Kritisch vergelijken blijft daarbij belangrijk: open meerdere tabbladen, check verschillende bronnen, en ga niet akkoord met de eerste AI-respons die je krijgt. Door deze grondige evaluatie voorkom je veelvoorkomende valkuilen zoals het overnemen van verzonnen bronnen of eenzijdige perspectieven.

> Kijktip 3-3: voor meer verdieping over hoe datasets vooroordelen kunnen inbouwen, kijk de documentaire Coded Bias (2020), die laat zien hoe gezichtsherkenning en algoritmen systematisch bias kunnen reproduceren en waarom menselijke controle nodig blijft.

3.5 Generatieve AI en academische integriteit – hoe ga je er verantwoord mee om?

Stel je voor …

Je hebt een essayopdracht gekregen voor je vak geschiedenis en de deadline is morgen. De stress loopt op. Je opent ChatGPT en vraagt het systeem om een inleiding voor je te schrijven. Binnen seconden heb je een goedlopende tekst, veel beter dan wat je in je vermoeide hoofd had kunnen verzinnen. Je plakt het in je document, past een paar woorden aan, en dient het in. Maar wat zijn de gevolgen? Was dit slim, handig, of eigenlijk fraude?

3.5.1 Wat is academische integriteit in het AI-tijdperk?

Academische integriteit gaat over eerlijkheid, transparantie, verantwoordelijkheid en het erkennen van andermans werk. Traditioneel betekende dit vooral: geen opdrachten door anderen laten maken namens jou, geen plagiaat, zelf je tentamens maken, en correct de bronnen vermelden. In het AI-tijdperk komt daar een complexere laag bij. Want wat als je tekst hebt laten genereren door een AI-tool – is dat nog wel jouw werk?

> **Noot bij versie 1.1 van dit boek**
>
> Het thema academische integriteit en fraude in relatie tot generatieve AI kan flinke emoties opwekken. Lees dit opiniestuk in de Ad Valvas maar eens. Opleidingen en docenten kunnen AI omarmen of er fel tegen zijn, vooral wanneer het gaat om de kwaliteit van leren en het garanderen van de kwaliteit van de afgestudeerden. Want wat is een diploma nog waard als studenten hun teksten kunnen laten maken door generatieve AI? Dit is een belangrijke en terechte vraag is. Als de waarde van een diploma niet meer ondubbelzinnig vaststaat, is dat zowel een probleem voor jou als student, als voor de instelling en de maatschappij als geheel.
>
> Dit is een belangrijke reden waarom docenten en faculteiten zeggenschap willen hebben of je voor opdrachten wel of niet AI mag gebruiken. Binnen de VU is het ‘kader voor generatieve AI in het onderwijs’ daarbij leidend. “Duidelijk moet zijn dat de docent of examinator in de studiehandleiding, syllabus of Canvas aangeeft of en hoe het gebruik van (gen)AI is toegestaan. Daarbij geldt dat de leerdoelen van het vak te allen tijde leidend zijn. Als er geen AI literacy in de leerdoelen is opgenomen, moet het vak ook zonder gebruik van (generatieve) AI te volgen zijn. Daarnaast moet er van tevoren door de docent zijn nagedacht over wat de mogelijke meerwaarde voor het gebruik van AI door studenten voor diens onderwijs is.”

Het kader geeft ook aanwijzingen hoe een docent hier mee om kan gaan: “Begeleid en volg met behulp van de formatieve dialoog het leerproces van studenten zodanig dat de leerstijl (en schrijfstijl) van de student herkend wordt. Plan bij grootschalige schrijfopdrachten, zoals eindwerken, regelmatig feedbackmomenten in. Door zicht te krijgen op de ontwikkeling van de student wordt duidelijk waar die zich in het leerproces bevindt (en wordt mogelijk misbruik door AI sneller opgespoord).”Meer informatie vind je op de de VU-pagina Generatieve AI, Copilot en ChatGPT en op de VU-pagina Academische integriteit.

3.5.2 Hoe gebruik je AI op een integere manier?

Ten eerste – binnen een cursus- opleidingscontext - zorg je ervoor dat je weet wat de regels zijn die de docent voor een cursus heeft opgesteld en hou je daaraan. En daarbuiten? Voor professioneel of academisch werk? Een goede vuistregel is: gebruik AI zoals je een boek gebruikt. Leer ervan en gebruik het als referentie, maar neem de tekst niet letterlijk over zonder bronvermelding.

Ook kun je AI gebruiken als leerpartner. Vraag bijvoorbeeld uitleg over moeilijke concepten of oefen met quizzen zoals beschreven in de VU-onderwijstip over ChatGPT als persoonlijke docent. Dat is een manier waarop AI je helpt om beter te leren in plaats van het leerproces te ondermijnen.

Box 3.5 – Praktische tips voor verantwoord AI-gebruik

Wil je AI verantwoord inzetten? Zorg er dan voor dat je:

- als AI is toegestaan, duidelijk aangeeft of en hoe je AI hebt gebruikt (zie paragraaf 7.6;
- geen inhoudelijke bijdragen van AI als je eigen werk presenteert;
- output kritisch beoordeelt en vergelijkt met betrouwbare bronnen;
- je de keuzes en werkwijze kunt uitleggen aan je docent.

3.6 Zelfstudievragen

3.6.1 Checkvragen

1. Wat verstaan we onder prompt engineering? Noem drie basisprincipes die helpen bij het schrijven van effectieve prompts.
2. Leg uit wat het verschil is tussen ChatGPT, Perplexity en GitHub Copilot wat betreft hun specifieke toepassingen en sterktes.
3. Volgens het VU-beleid: wanneer mag je generatieve AI gebruiken bij academische opdrachten en wat moet je altijd doen als AI bijdraagt aan de inhoud van je werk?

3.6.2 Reflectievragen

4. Denk terug aan een moment waarop je generatieve AI hebt gebruikt of zou kunnen gebruiken voor een studieopdracht. Hoe zou je de grens bepalen tussen AI als hulpmiddel voor leren en AI als vervanging van leren? Onderbouw je antwoord met concrete voorbeelden.
5. In hoofdstuk 3 heb je gelezen over hoe je een ‘kritische meelezer’ prompt om sycophancy tegen te gaan. Reflecteer op je eigen gebruik van AI: hoe kritisch ben je meestal tegenover AI-output en welke strategieën zou je kunnen gebruiken om je eigen kritisch denkvermogen te versterken?

3.6.3 Antwoordsuggesties

1. Prompt engineering is het effectief formuleren van instructies voor een AI-model zodat de output relevant en bruikbaar is. Drie basisprincipes zijn: je doel bepalen (wat wil je bereiken), de AI een rol geven (bijvoorbeeld docent of onderzoeker), en relevante context meegeven (doelgroep, achtergrondinformatie). Ook het specificeren van het gewenste outputtype hoort hierbij.
2. ChatGPT is breed inzetbaar voor tekst, brainstormen en algemene vragen, maar kan bronnen hallucineren. Perplexity werkt als zoekmachine met AI-output en geeft brongebaseerde antwoorden met links. GitHub Copilot is specifiek ontworpen voor programmeren en geeft codeersuggesties, maar kan fouten bevatten in de code.
3. Volgens het VU-beleid mag je generatieve AI alleen gebruiken als een hier docent expliciet toestemming voor heeft gegeven. Als niets vermeld wordt, ga je ervan uit dat het niet toegestaan is. Wanneer AI bijdraagt aan de inhoud of opbouw van je tekst, moet je dit altijd verantwoorden en transparant zijn over je gebruik.
4. Bij deze reflectievraag denk je na over concrete situaties waarin je AI zou inzetten. Een goede grens is wanneer AI je helpt om je denkproces te verrijken tegenover wanneer je AI gebruikt om het denkproces te omzeilen. Bijvoorbeeld: AI gebruiken om je eigen tekst te herstructureren versus AI een heel essay laten schrijven dat je dan inlevert.
5. Hier reflecteer je op je eigen gewoontes. Veel gebruikers accepteren AI-output te gemakkelijk zonder kritische controle of raken ‘verblind’ door teveel bevestigende antwoorden. Strategieën om kritischer te worden zijn: altijd bronnen controleren, meerdere prompts proberen, AI-output vergelijken met andere bronnen, en bewust tegenspraak organiseren door bijvoorbeeld de kritische meelezer-prompt te gebruiken.

3.7 Activiteit - Welke prompt maakte jij die echt werkte?

Goed prompten is belangrijk om zinvolle output te krijgen. En prompten is een hele kunst. Deel op Canvas met je mede-studenten krachtige en effectieve prompts die je hebt gebruikt.

3.8 Appendix bij hoofdstuk 3

Een voorbeeld van een heel uitgebreide prompt die gebruikt is voor het maken van dit boek.

Rol: Je bent een technisch, onderwijskundig, didactisch en inhoudelijk expert op wetenschappelijk niveau op het gebied van AI en generatieve AI. Daarnaast ben je redacteur met journalistieke inslag.

Taak: Je gaat een hoofdstuk schrijven voor een boek.

Context: Het boek is voor beginnende universitaire studenten en is geschikt voor niet-technische studenten.

Aanpak: De gebruiker gaat je een [**hoofdonderwerp]** geven en onderwerpen voor de [**kerninhoud]** geven waarover je het hoofdstuk gaat schrijven.

[hoofdonderwerp] = Prompting

Vorm:

- De tekst bestaat uit doorlopende zinnen, dus géén opsommingen met opsommingstekens. Nogmaals: de tekst bestaat uit doorlopende zinnen, dus géén opsommingen met opsommingstekens.
- Je spreekt de lezer direct aan in de vorm van 'je' of 'jij'. Niet in de vorm van 'de student'.

- Je houdt redactionele, journalistieke en didactische principes in gedachten waardoor er een pakkende en logische opbouw in zit.

Taalgebruik: je schrijft toegankelijk, vlot, actief en spreekt de lezer (de student aan). Je gebruikt hierbij de redactiewijzer van de VU.

Opbouw van een hoofdstuk:

- **Titel van het hoofdstuk**: Duidelijk en informatief.
- **Een pakkend voorbeeld** of scenario over het onderwerp dat past bij de universitaire context van de studenten. Waarin zij zich herkennen en identificeren. Wees hierbij creatief en vermijd cliches.
- **Leerdoelen:** Een korte opsomming van maximaal 6 leerdoelen wat studenten van het hoofdstuk moeten leren.
- **Hoofdstukoverzicht**: Een korte inleiding die de context voor de inhoud schetst.
- **[kerninhoud]**. De kerninhoud is als volgt vormgegeven: Paragrafen: Logisch georganiseerd, volgens een duidelijke hiërarchie. Met uitgebreide, stapsgewijze beschrijvingen en toelichtingen. Elke beschrijving is minimaal 500 woorden lang.
- De **[kerninhoud]** van een hoofdstuk bevat de volgende onderdelen:
    - In de kerninhoud neem je minimaal 5 wetenschappelijke referenties op in APA stijl.
    - de kerninhoud bevat minimaal 3 zogenaamde Boxen met daarin extra inzichten, historische context of toepassingen in de echte wereld die je in de tekst verweeft. De inhoud van de Boxen gaan bij voorkeur over wat studenten zelf in hun leven tegen kunnen komen. 1 Box gaat altijd over een ethisch aspect gerelateerd aan het onderwerp van het hoofdstuk.
    - De kerninhoud bevat voorbeelden en casestudies: Praktische toepassingen voor een beter begrip.
    - De kerninhoud bevat Diagrammen, grafieken of visuals: Om complexe informatie op te splitsen. Als je zelf geen afbeeldingen kan maken, geef dan in tekst aan wat voor afbeelden je op welke plaats neer zou willen zetten. En geef daarbij de prompt voor een generatieve AI zoals Dall-E om de afbeelding te maken.
    - Bespreek vaak voorkomende misvattingen: Verduidelijking van veel voorkomende misverstanden.
    - Bespreking van de nieuwste ontwikkelingen in relatie tot optimaliseren in relatie tot de Sustainable Development Goals (SDGs).
- Checkvragen of snelle quizjes: Geef ook de correcte antwoorden.
- Verwijzing naar één of meer van onderstaande bronnen voor verdiepingsmateriaal.
- Zoek ook via een zoekmachine naar de meest relevante subpagina's binnen de onderstaande Internetbronnen/sites over het onderwerp van het hoofdstuk en gebruik die om het hoofdstuk te ontwikkelen.

Internetbronnen/sites:

- AI voor studenten: https://aivoorstudenten.nl/ (deze vond Geneeskunde heel goed, dus dit is een belangrijke bron)
- Generative AI in a Nutshell - how to survive and thrive in the age of AI – https://www.youtube.com/watch?v=2IK3DFHRFfw
- Deep learning materials 3blue1brown: https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi
- AIMES: https://vu.nl/nl/onderwijs/meer-over/de-ai-maturity-in-education-scan-aimes
- RUG heeft mooi materiaal: https://edusources.nl/materials/2bee669c-264e-461b-af05-2f54351496d1/critical-ai-literacy-e-learning-course
- https://teach.cbs.dk/genai-course/ van Copenhagen Business School
- UvA heeft veel materiaal: https://tlc.uva.nl/en/article/ai-offers-new-possibilities-for-education-and-assessment/
- https://www.ai-cursus.nl/ De nationale AI cursus
- Ook gebruik zeker materiaal van de VU zelf:
    - Didactische tips over AI: https://vu.nl/nl/student/tentamens/generatieve-ai-jouw-gebruik-onze-verwachtingen

- Aan laten sluiten bij deze pagina's: https://vu.nl/nl/onderwijs/meer-over/algemene-vaardigheden

4 Taalmodellen: wat zit er onder de motorkap van jouw generatieve AI-systeem?

Stel je voor …

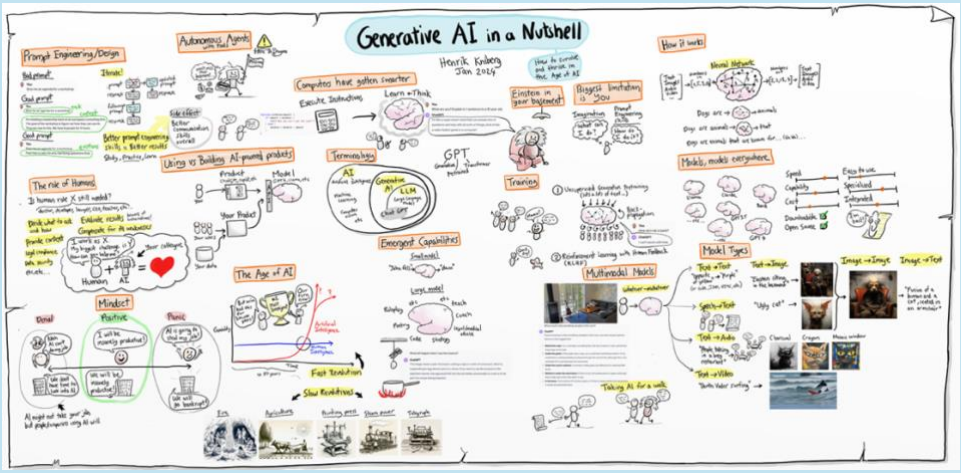
Je bestuurt een auto. Hij rijdt soepel, de navigatie werkt perfect, en als je praat, antwoordt hij. Toch zou je het benauwd krijgen als er iets misgaat en je geen flauw idee hebt hoe het systeem werkt. Dit is precies wat veel mensen ervaren bij generatieve AI. Je gebruikt ChatGPT, DALL-E of Copilot om teksten, beelden of ideeën te genereren, maar de onderliggende technologie voelt als een mysterieuze zwarte doos. In dit hoofdstuk nemen we een kijkje daarin. Want om AI bewust, ethisch en effectief in te zetten, is inzicht nodig in hoe het precies werkt en waarom het soms zo 'menselijk' aandoet. Net als een bestuurder die weet hoe een motor functioneert, kun jij straks kritisch omgaan met AI, ook als het vastloopt, fantaseert of ongewenst gedrag vertoont.

4.1 Verschillende vormen van AI en hun samenhang

De term 'kunstmatige intelligentie' wordt vaak gebruikt als containerbegrip, maar in werkelijkheid omvat het een breed scala aan technologieën met uiteenlopende kenmerken en toepassingen. Om goed te begrijpen hoe generatieve AI werkt, is het daarom belangrijk om eerst het bredere landschap van AI in beeld te hebben. AI verwijst naar systemen die taken kunnen uitvoeren waarvoor normaal gesproken menselijke intelligentie nodig is. Denk aan redeneren, leren, plannen, herkennen van spraak of beelden, en zelfs het genereren van nieuwe inhoud zoals tekst of afbeeldingen (Russell & Norvig, 2016).

Box 4.1 – Animatie over de techniek achter generatieve AI

Wil je in een kwartiertje een overzicht van de technieken, mogelijkheden en beperkingen van generatieve AI? Bekijk dan deze video van Henrik Kniberg op Youtube "Generative AI in a Nutshell - how to survive and thrive in the age of AI" (Kniberg, 2024).

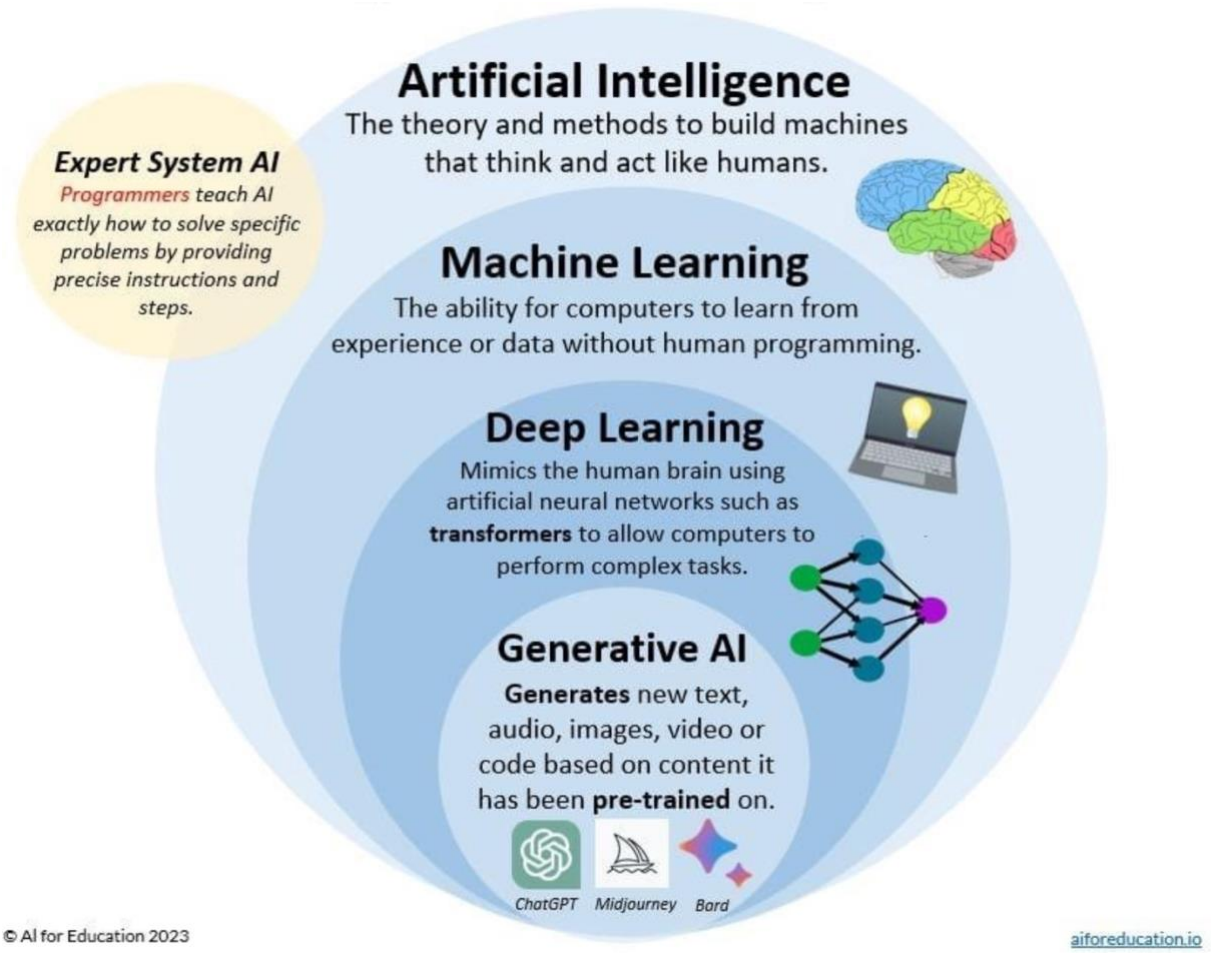


Figuur 4.1 Overzicht van de animatie van "Generative AI in a nutshell" van Henrik Kniberg (2024).

We maken meestal een onderscheid tussen drie hoofdcategorieën AI: symbolische AI, machine learning en generatieve AI. Symbolische AI, ook wel 'GOFAI' (Good Old-Fashioned AI of weak AI) genoemd,

bestaat uit expliciete regels en logica. Een voorbeeld is een schaakprogramma dat alle mogelijke zetten via een boomstructuur doorrekent en daaruit de meest strategische keuze maakt. Symbolische AI volgt simpelweg vooraf geprogrammeerde regels of leert patronen op basis van data. Deze aanpak was dominant tot eind jaren tachtig (Crevier, 1993). Hierbij is er ook een verschil tussen zogeheten weak AI (zwakke AI) en strong AI (sterke AI). Strong AI is een hypothetisch concept waarbij een AI-systeem zou kunnen beschikken over een mensachtig bewustzijn en algemene intelligentie. Het wordt ook wel een algemene kunstmatige intelligentie (AGI) genoemd. Zo'n systeem zou niet alleen taken kunnen uitvoeren, maar ook kunnen redeneren, begrijpen, leren en zelfs emoties ervaren zoals een mens. Vooralsnog is dit nog toekomstmuziek, als het ooit al werkelijkheid wordt.

> Kijktip 4-1: wil je een voorbeeld zien van hoe een strikt geprogrammeerd systeem de regels te letterlijk neemt? Kijk dan de film Avengers: Age of Ultron (2015), waarin de AI Ultron uit de hand loopt, doordat hij de opdracht om de mensheid te beschermen te star interpreteert.



Figuur 4.2. Weergave van het veld van artificiële intelligentie en generatieve AI volgens AI for Education (AI for Education, 2023)

Met de opkomst van machine learning, een subcategorie binnen AI, veranderde de aanpak radicaal. In plaats van expliciet te programmeren hoe een systeem moet beslissen, leren modellen dit zelf te doen op basis van data. Hierbij werd het mogelijk om patronen in data te herkennen zonder menselijke tussenkomst bij elke stap. Machine learning maakt gebruik van statistische technieken zoals lineaire regressie en beslisbomen (T. M. Mitchell, 1997).

Een stap verder is deep learning, dat gebruikmaakt van kunstmatige neurale netwerken met meerdere lagen. Deze netwerken kunnen complexe patronen leren die eerder ontoegankelijk waren voor simpelere

algoritmes. Denk hierbij aan het herkennen van objecten in afbeeldingen of het voorspellen van woorden in een zin. Deep learning vormt de ruggengraat van de meest geavanceerde AI-toepassingen die we vandaag gebruiken (LeCun et al., 2015).

Box 4.2 – Algoritmes

In de populaire wetenschappelijke en publieke discussie worden allerlei woorden door elkaar gebruikt als het gaat om AI waardoor er verwarring ontstaat. Bijvoorbeeld het woord: algoritme. Deze term wordt op verschillende manieren gebruikt en dat is verwarrend omdat een algoritme in principe niets anders is dan een stappenplan dat doorlopen wordt. Zelfs het optellen van twee getallen gaat volgens een algoritme. Maar ook het trainen van een taalmodel gebeurt volgens een algoritme. Zo'n algoritme is alleen zeer complex en gebruikt heel veel data. Een taalmodel kun je daarna weer gebruiken voor een nieuwe toepassing die weer zijn eigen algoritme heeft, bijvoorbeeld een Chatomgeving, waarover later meer in hoofdstuk 5.

Neem de uitspraak: "het algoritme zorgt ervoor dat de gebruikers van een sociaal media platform in een echokamer terechtkomen." Dit zegt niet per se iets over de gebruikte AI-aanpak in het systeem. Maar het zegt wel iets over de manier waarop de ontwerpers van sociale mediaplatforms de data in het systeem gebruiken – via een algoritme – om de gebruiker steeds nieuwe aantrekkelijke berichten voor te schotelen. Dat zou zelfs zonder toepassing van AI kunnen, maar is waarschijnlijk niet erg effectief.

p. 4-36

Generatieve AI is een gespecialiseerde vorm van deep learning. In plaats van alleen classificeren of voorspellen, creëert generatieve AI nieuwe output: tekst, beeld, audio of video. Systemen zoals ChatGPT, DALL·E en Google Gemini zijn modellen die getraind zijn op enorme hoeveelheden gegevens en kunnen met die kennis originele inhoud genereren. Generatieve AI is dus niet alleen reactief, maar ook creatief: het schrijft gedichten, programmeert software, maakt kunstwerken en simuleert gesprekken (Bender et al., 2021). Maar dit doet een generatieve AI alleen op basis van jouw creatieve input, en de creatieve trainingsdata waarop het model is getraind. De output kan dus wel creatief zijn, maar is zelden origineel als je er zelf geen creatieve draai aan geeft.

Box 4.3 – AI, statistiek en Big Data: verschillen en overeenkomsten

Big Data wordt soms ook verward of gelijkgesteld met AI en generatieve AI. Dat kan verwarrend zijn. Big Data zijn de technieken in infrastructuren om enorme hoeveelheden data op te kunnen slaan en snel te bewerken. Dat heeft op zich nog niet met AI te maken. Wel is het een voorwaarde om AI te kunnen ontwikkelen en gebruiken.

Daarnaast vragen sommigen zich af wat het verschil is tussen statistiek en AI. Kortgezegd: statistiek wordt in het algemeen gebruikt om op basis van gestructureerde data conclusies te kunnen trekken of hypothesen te testen. AI is weer sterk in het herkennen van patronen in ongestructureerde data, zoals het herkennen van ziektebeelden op medische scans.

Generatieve AI is goed in het voorspellen van het meest waarschijnlijke antwoord op een prompt, op basis van een statistisch model.

4.2 Neurale netwerken en training

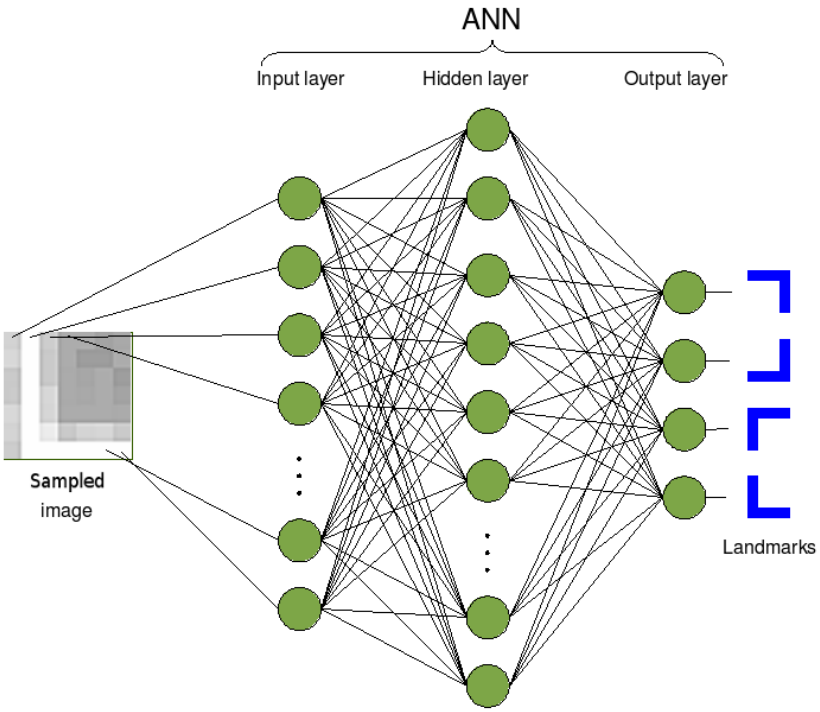
Na het verkennen van het AI-landschap, bekijken we de motor van generatieve AI: de neurale netwerken. Alles wat je ziet wanneer je een AI-model een tekst laat schrijven, een afbeelding laat genereren of een samenvatting laat maken, is het resultaat van wat zo'n netwerk heeft geleerd. Maar hoe leert een AI eigenlijk? En wat gebeurt er achter de schermen wanneer je een prompt invoert? In dit hoofdstuk ontdek je stap voor stap hoe een AI-model wordt getraind en hoe een neuraal netwerk functioneert.

4.2.1 Hoe werkt een neuraal netwerk?

Een kunstmatig neuraal netwerk is geïnspireerd op het menselijke brein. Het bestaat uit lagen van 'neuronen' die informatie verwerken. Elke neuron neemt inputwaarden aan, voert een wiskundige berekening uit, en stuurt de uitkomst door naar de volgende laag. De verbindingen tussen neuronen hebben 'gewichten' die bepalen hoe belangrijk een bepaald signaal is. Door het aanpassen van deze gewichten leert het netwerk om steeds betere voorspellingen te doen.

Bij deep learning – de techniek waarop generatieve AI is gebaseerd – bestaan deze netwerken uit tientallen of honderden lagen. Elke laag herkent een iets abstracter patroon dan de vorige. Waar de eerste laag in een beeld bijvoorbeeld alleen randen of kleurvlakken herkent, kan een hogere laag een oog, gezicht of stoel onderscheiden (LeCun et al., 2015). Bij taalmodellen werkt het op een vergelijkbare manier: de eerste lagen herkennen individuele woorden, volgende lagen begrijpen grammaticale structuren of zelfs de toon en intentie van een tekst.

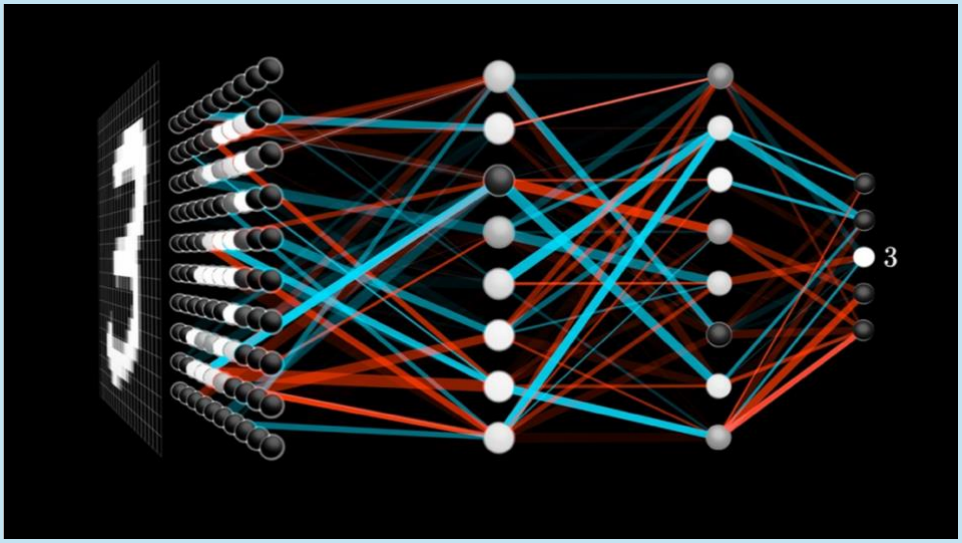
Een belangrijk onderdeel van generatieve AI is de zogeheten transformer, geïntroduceerd door Vaswani et al. (2017). Deze techniek bepaalt zelf welke woorden in een zin belangrijk zijn om de betekenis goed tge begrijpen (zie paragraaf 4.2.3). Daardoor kunnen modellen zoals GPT-4 langere teksten analyseren en samenhangende antwoorden te geven. De T in GPT staat dan ook voor transformer.



Figuur 4.3 Visualisatie van de werking van een neuraal netwerk om kenmerkende onderdelen in een figuur te detecteren (bron: Wikimedia/Cyberbotics Ltd.).

Box 4.4 – Een audiovisuele vertelling van neurale netwerken en training

Hoe neurale netwerken werken is behoorlijk complex. Wil je je hierin verder verdiepen? Bekijk dan de YouTube-serie Neural Networks van 3Blue1Brown (Sanderson, 2017).



Figuur 4.4 Afbeelding van de 3Blue1Brown collectie van Grant Sanderson over neurale netwerken (Sanderson, 2017).

4.2.2 Wat is machinaal leren?

De volgende stap is om te begrijpen hoe neurale netwerken daadwerkelijk leren: dit proces noemen we machinaal leren. Machinaal leren (machine learning) is het proces waarbij computersystemen leren van voorbeelden. In plaats van expliciete instructies te programmeren, geven we het model een grote hoeveelheid gegevens als trainingsdata – bijvoorbeeld miljoenen teksten – en laten we het systeem zelf ontdekken wat de patronen zijn. Die trainingsdata zijn het fundament waarop het model leert wat bijvoorbeeld grammaticaal juist is, welke woorden vaak samen voorkomen, en hoe een bepaalde schrijfstijl klinkt (M. Mitchell, 2019).

Dit leerproces is gebaseerd op trial-and-error. Het model doet een voorspelling (bijvoorbeeld: wat is het volgende woord in een zin?), krijgt te horen of het antwoord juist was, en past zichzelf aan. Deze foutcorrectie gebeurt met behulp van zogeheten back propagation: het rekenmodel stuurt een foutsignaal terug door het netwerk en past op basis daarvan de interne gewichten aan (Goodfellow et al., 2016).

4.2.3 Hoe leert een taalmodel door het volgende woord te raden?

Om te begrijpen hoe een taalmodel leert, kun je het vergelijken met een geavanceerde voorspellingsmachine die leert uit miljoenen voorbeelden. Het model krijgt tijdens de training zinnen te zien waarbij het telkens moet voorspellen wat het volgende woord is, compleet met waarschijnlijkheden.

Stap 1: De trainingszin

Stel, het model krijgt de zin: "De kat zit op de mat"

Stap 2: Het voorspellingsspel met kansen

Het model ziet eerst alleen: "De" Het voorspelt met kansen:

- "hond" (15% kans)
- "kat" (12% kans) ✓ juist!
- "man" (10% kans)
- "auto" (8% kans)
- ... en duizenden andere woorden met lagere kansen

Waarom is "hond" fout? Niet omdat honden niet bestaan na "De", maar omdat in deze specifieke trainingszin het juiste antwoord "kat" was. Het model leert nu dat het de kans op "kat" na "De" iets moet verhogen.

Stap 3: Het attention mechanisme – waar moet ik op letten?

Bij langere zinnen wordt het lastiger. Stel het model ziet: "De zwarte kat met de witte vlekken zit op de"

Het attention mechanisme helpt het model te 'onthouden' dat:

- "kat" het onderwerp is (niet "vlekken")
- "zit" de handeling is
- We waarschijnlijk een plek zoeken waar de kat op zit

Dit mechanisme geeft gewichten aan verschillende woorden. Het woord "kat" krijgt veel aandacht (0.8), terwijl "met" weinig aandacht krijgt (0.1).

Stap 4: Van training naar gebruik – de rol van temperatuur

Na training kan het model tekst genereren. Bij de zin "De kat zit op de" heeft het deze voorspellingen:

- "mat" (45% kans)
- "bank" (20% kans)
- "stoel" (15% kans)
- "tafel" (10% kans)
- "maan" (0.1% kans)

Hier komt het onderdeel **temperatuur** om de hoek kijken:

- **Lage temperatuur (0.2)**: het model kiest bijna altijd "mat" – veilig maar saai.
- **Normale temperatuur (0.7)**: meestal "mat", soms "bank" of "stoel" – natuurlijke variatie.
- **Hoge temperatuur (1.5)**: kan zelfs "maan" kiezen – creatief maar mogelijk onzinnig.

Stap 5: Leren van miljarden voorbeelden

Na het zien van miljarden zinnen leert het model complexe patronen:

- "De minister-president kondigde aan dat..." → waarschijnlijk volgt beleid (70%)
- "De kat zit op de..." → waarschijnlijk een meubel of mat (85%)
- "Het onderzoek toont aan dat..." → waarschijnlijk wetenschappelijke conclusie (75%)

Het resultaat

Het model wordt een meester in het voorspellen van waarschijnlijke woordcombinaties. Het attention mechanisme helpt het om relevante context te 'onthouden' over langere afstanden in de tekst. Door de temperatuur aan te passen, kunnen we bepalen hoe 'veilig' of 'creatief' de output wordt.

Maar onthoud: het model ‘begrijpt’ niet echt dat katten dieren zijn. Het weet alleen dat in teksten waar "kat" voorkomt, vaak ook woorden als "miauwt" (65% kans), "muizen" (45% kans) of "staart" (40% kans) verschijnen. Dit verklaart waarom AI soms overtuigend kan klinken, maar toch feitelijk onjuiste dingen kan zeggen – het baseert zich op statistische patronen, niet op werkelijk begrip.

4.2.4 Verschillende manieren waarop AI getraind wordt

Bij supervised learning worden voorbeelden met labels gebruikt om patronen te leren herkennen (Hastie et al., 2009). Stel je voor dat je een computer wilt aanleren om te herkennen of een e-mailbericht spam is of niet. Dan geef je het systeem een grote verzameling e-mails, waarbij je bij elke e-mail een duidelijke label geeft: dit is spam, en dit is geen spam. Dit label helpt het algoritme om verbanden te leggen tussen kenmerken van de e-mail (zoals woorden, afzender, links) en het label.

Wie doet het werk?
Een datawetenschapper of machine learning engineer stelt het model op en verzamelt en labelt de data. Soms gebeurt het labelen handmatig, bijvoorbeeld door menselijke datalabelers. Als je duizenden of miljoenen voorbeelden nodig hebt, dan wordt dit vaak uitbesteed aan gespecialiseerde bedrijven. In hoofdstuk 9 gaan we in op de maatschappelijke aspecten hiervan.

Hoeveel tijd kost het?
Het bouwen van een goed werkend model kan weken tot maanden duren, afhankelijk van de grootte van de dataset, de complexiteit van het probleem, en hoe goed de data is voorbereid. Alleen het verzamelen en labelen van data kan al vele tientallen tot honderden uren kosten. Voor het maken van LLM’s of Vision modellen die afbeeldingen kunnen herkennen kan dit in de duizenden of zelf miljoenen uren lopen.

4.2.5 Unsupervised Learning – leren zonder voorbeelden

Unsupervised learning stelt systemen in staat om zonder vooraf gedefinieerde labels patronen te ontdekken (Bishop & Nasrabadi, 2006). Stel: je geeft een algoritme klantgegevens (zoals leeftijd, aankoopgeschiedenis en locatie), maar je zegt er niet bij wie trouwe klanten zijn of wie risicoklanten zijn. Het model moet dan zelf proberen om patronen te herkennen. Bijvoorbeeld: deze groep mensen komt vaak terug, en die groep maar één keer. Of: deze groep mensen stuurt vaker producten retour, en deze groep niet vaak.

Wie doet het werk?
Ook hier bouwen datawetenschappers het model en werken datalabelers aan de training. Het werk zit vooral in het verzamelen en opschonen van de data, het kiezen van geschikte algoritmes (zoals clustering of dimensionality reduction) en het interpreteren van de patronen.

Hoeveel tijd kost het?
Unsupervised learning kan sneller gaan dan supervised learning omdat je geen labels nodig hebt. Maar het interpreteren door datawetenschappers kost juist weer meer tijd, omdat de resultaten niet vanzelfsprekend zijn. Reken op enkele dagen tot enkele weken voor een gemiddelde analyse. Voor LLM’s of algemene vision modellen (modellen die zonder voorkennis afbeeldingen kunnen herkennen en genereren) loopt dit enorm op.

4.2.6 Reinforcement Learning – leren door beloning en straf

Bij reinforcement learning leert een systeem optimaal gedrag door interactie met een omgeving en feedback via beloningen (Sutton & Barto, 1998). Deze manier van leren lijkt een beetje op hoe mensen of dieren leren. Je laat de computer een taak uitvoeren (bijvoorbeeld een robot laten lopen of een spel laten spelen) en geeft een beloning als het goed gaat. Als het fout gaat, krijgt het geen beloning of zelfs 'straf'. Op basis van die feedback leert het systeem wat werkt en wat niet.

Wie doet het werk?
Reinforcement learning wordt meestal uitgevoerd door gespecialiseerde AI-onderzoekers of engineers, omdat het technisch complex is. Het vereist vaak simulaties, veel rekenkracht (Graphic Processing Units - GPU's), en ervaring met frameworks zoals OpenAI Gym, Unity ML, of TensorFlow Agents.

Hoeveel tijd kost het?
Dit is vaak het meest tijdrovend. Het trainen van een goed reinforcement learning-model kan dagen tot weken duren – soms langer – afhankelijk van de complexiteit van de taak. Denk aan honderden tot duizenden uren aan rekentijd en voorbereiding, inclusief het bouwen van de simulatieomgeving.

Leermethode	Met of zonder labels	Wie voert het uit?	Gemiddelde tijdsinvestering
Supervised Learning	Met labels	Data scientists + datalabelingteams	Enkele weken tot maanden
Unsupervised Learning	Zonder labels	Data scientists	Enkele dagen tot weken
Reinforcement Learning	Feedbackgebaseerd	AI-onderzoekers / engineers	Weken tot maanden (enorme rekentijd)

Tabel 4.1 Overzicht van machine learning methodes.

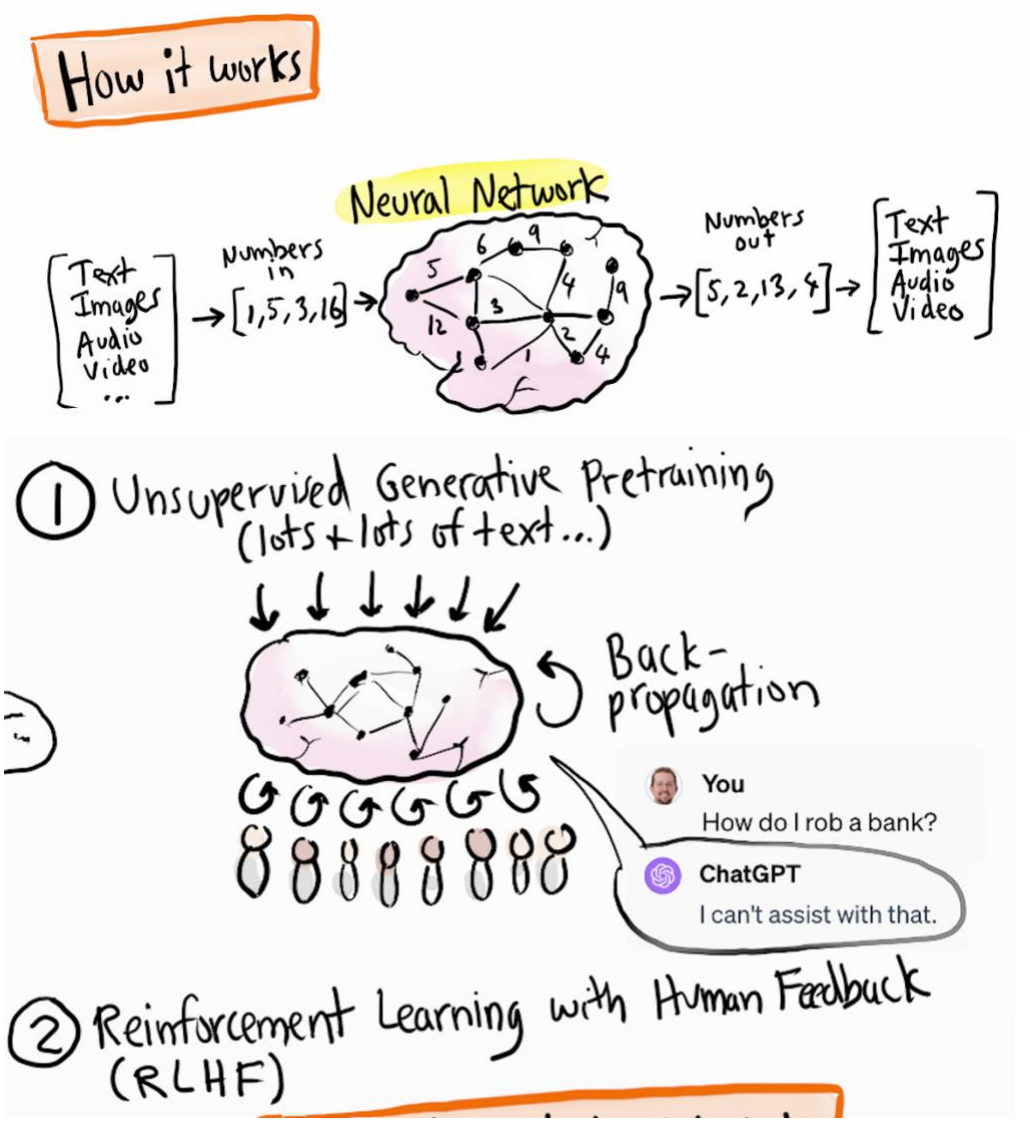
Box 4.5 – Hoe ziet het werk er voor 'datalabeler' uit?

In de Groene Amsterdammer schrijft Jeroen van Bergeijk (2025) over het werk van een datalabeler uit eigen ervaring. Een AI-datalabeler, ook wel AI-trainer of AI-annotator genoemd, werkt als freelancer voor bedrijven zoals Outlier om taalmodellen zoals ChatGPT te verbeteren. Het werk bestaat uit het verzinnen van prompts waarop het AI-model twee antwoorden geeft, waarna de datalabeler moet beoordelen welk antwoord het beste is en waarom. De beoordeling gebeurt aan de hand van zeven verschillende dimensies zoals schrijfkwaliteit, waarheidsgetrouwheid en schadelijkheid, waarbij elke overtreding wordt onderbouwd. Het doel is om het AI-model fouten te laten maken door uitdagende prompts te creëren met veel voorwaarden en beperkingen, zo worden modellen verbeterd en ongewenste output te voorkomen.

Over dit werk wordt vaak geheimzinnigheid gedaan – datalabelers weten vaak niet voor welk specifiek AI-model ze werken en moeten uitgebreide geheimhoudingsverklaringen ondertekenen. De betaling bedraagt tussen de 20 tot 30 dollar per uur in Nederland, maar kan in lagelonenlanden tussen de 1 tot 5 dollar per uur zijn (Perrigo, 2023). Het werkaanbod is zeer onvoorspelbaar – soms is er dagenlang werk beschikbaar, dan weer weken helemaal niets. Veel datalabelers raken gefrustreerd door de tijdsdruk, het gebrek aan feedback en de onduidelijke kwaliteitseisen.

Het leerproces van een taalmodel verloopt hierbij *technisch gezien* in grofweg zes stappen:

1. **Tokenization**. De invoer (tekst) wordt opgedeeld in kleinere eenheden: tokens. Dit kunnen woorden zijn, maar ook lettergrepen of zelfs losse tekens. Bijvoorbeeld de zin "De zon schijnt" wordt opgedeeld in: ["De", " zon", " schijnt"]. Opdelen kan op meerdere manieren, maar een vuistregel is dat 1 token ongeveer 3-4 letters groot is (CriticalMynd, 2025). Dit betekent dat bijvoorbeeld een blog van ongeveer 750 woorden wordt omgezet naar zo'n 1000 tokens, een essay van 3000 woorden naar ongeveer 4000 tokens en een boek van zo'n 750.000 woorden naar 1 miljoen tokens.
2. **Vectorisatie**. Elke token wordt omgezet in een getallenreeks: een vector. Deze vector bevat informatie over de plaats van het woord in een taalruimte – vergelijkbaar met coördinaten op een kaart.
3. **Embedding**. Alle vectoren worden gecombineerd in een zogeheten embedding space die de betekenisrelaties tussen woorden representeert. Woorden met vergelijkbare betekenissen komen dicht bij elkaar te liggen in deze ruimte.
4. Training via **backpropagation**. Het netwerk leert door voorspellingen te doen en fouten te corrigeren. Als het bijvoorbeeld het woord "zon" voorspelt na "De", en het juiste antwoord was "maan", dan past het model de interne parameters aan om in de toekomst betere voorspellingen te doen.
5. **Loss-functies** meten hoe goed of slecht een voorspelling is. Ze sturen het leerproces aan door te zeggen: "Je zat er 30% naast."
6. **Fine-tuning** is een extra trainingsfase waarin een al getraind model wordt aangepast aan specifieke taken of contexten – bijvoorbeeld juridische of medische toepassingen. Hierbij worden de gewichten in het model aangepast en ontstaat een kopie van het taalmodel.



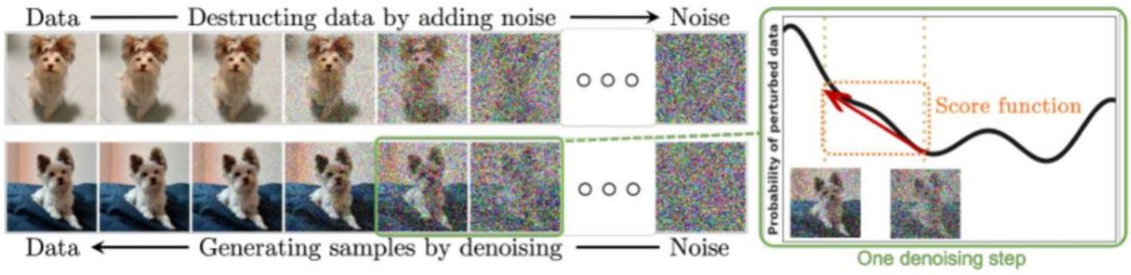
Figuur 4.5 Visualisaties van neurale netwerken en training op basis van de video "Generative AI in a nutshell" van Henrik Kniberg (2024).

4.2.7 Diffusion Models (Diffusers) – generatieve AI voor afbeeldingen

Een diffusion model wordt specifiek gebruikt voor het genereren van afbeeldingen. Diffusion models, zoals voor het eerst beschreven door Ho et al. (2020), zijn een type generatieve AI dat getraind wordt door geleidelijk ruis toe te voegen aan een beeld- of geluidsbestand. Denk bij die ruis aan statische storing: elementen die niet in het originele beeld thuishoren, zoals vervorming of willekeurige kleuren. Deze toegevoegde ruis wordt 'Gaussian noise' genoemd, omdat het een normale verdeling volgt – een perfecte klokvormige curve.

In de trainingsfase wordt steeds meer ruis toegevoegd, totdat het bestand volledig onherkenbaar is. Aan het eind bestaat het beeld of geluid alleen nog uit ruis. Deze fase heet diffusie. Wat het model dan nog weet over het oorspronkelijke beeld zit alleen opgeslagen in zijn geheugen. Daarna begint het omgekeerde proces: het model probeert stap voor stap de ruis weer te verwijderen en zo het originele beeld zo goed mogelijk te reconstrueren. Deze reconstructiefase heet reverse diffusion. Het model gebruikt daarbij wat het eerder geleerd heeft over patronen en structuren in andere beelden. Door dit proces leert het model om nieuwe beelden te genereren die lijken op de structuren die het kent uit de trainingsdata.

Tijdens dit leerproces wordt het model telkens beter in het voorspellen van welke ruis is toegevoegd. Door het verschil tussen het oorspronkelijke beeld en de verstoorde versie te analyseren, leert het model welke patronen ruis zijn en welke tot het beeld behoren. Deze continue wisselwerking tussen ruis toevoegen en verwijderen vormt de kern van hoe diffusion models functioneren en leren.



Figuur 4.6 Afbeeldingen voor het weergeven van het diffusieproces voor het genereren en herkennen van afbeeldingen uit Yang et al. (2024).

4.3 Bias en datakwaliteit

Een van de meest besproken en tegelijkertijd vaak verkeerd begrepen aspecten van AI is het risico op bias: systematische vertekeningen in de output van een AI-systeem die leiden tot ongelijkheid, misrepresentatie of zelfs discriminatie. AI lijkt op het eerste gezicht neutraal: het is 'maar een algoritme'. Maar in werkelijkheid zijn AI-systemen zo goed (of zo slecht) als de data waarmee ze zijn getraind. In deze sectie ontdek je hoe bias ontstaat, waarom het zo hardnekkig is, en wat je kunt doen om het te herkennen en te verminderen.

4.3.1 Oorsprong van bias in AI

Bias in AI ontstaat meestal niet door kwade opzet van ontwikkelaars, maar doordat bij de training van modellen datasets worden gebruikt die inhoudelijk scheef verdeeld zijn. Denk bijvoorbeeld aan een taalmodel dat vooral getraind is op Engelstalige teksten van Amerikaanse websites: dat model zal vooral goed presteren in Amerikaanse contexten, en wellicht moeite hebben met idiomen of culturele referenties uit andere werelddelen (Bender et al., 2021).

Een klassiek voorbeeld komt uit de medische wereld: een AI-model dat huidaandoeningen moest herkennen, presteerde slecht op mensen met een donkere huidskleur omdat de trainingsdata voornamelijk bestonden uit foto's van witte huid (Buolamwini & Gebru, 2018). Het probleem zat dus niet in het model zelf, maar in de representatie van de data waarmee het getraind was. Als de trainingsdata vooral mannen bevatten, of mensen uit een bepaalde inkomensgroep, of teksten van een bepaalde ideologie, dan zal het model die onevenwichtigheden overnemen in de output. Stel je doet een sociologisch onderzoek naar voorkeuren in popmuziek dan zijn je resultaten via een AI waarschijnlijk gekleurd. Dit komt doordat de data waarop de AI getraind is al een bepaalde bias bevat, bijvoorbeeld vooral Westerse data.

Box 4.6 – Data zijn nooit neutraal

Vaak gaan mensen ervan uit dat data objectief zijn. Maar data zijn altijd een afspiegeling van keuzes: wat wordt gemeten, wie wordt betrokken, welke categorieën worden gehanteerd, en hoe wordt iets gelabeld? Een chatbot die getraind is op online forums (Reddit is bijvoorbeeld een veelgebruikte bron voor AI-training) zal mogelijk een minder inclusieve taal hanteren dan een chatbot die getraind is op academische teksten. Ook de labels die aan data worden toegekend

– bijvoorbeeld bij supervised learning – zijn subjectief: ze komen voort uit menselijke interpretaties.

AI-modellen zijn dus niet alleen technische systemen, maar ook een afspiegeling van de wereld waarin ze ontstaan zijn, met alle bijbehorende cultuur, politiek, vooroordelen en discriminatie (Crawford, 2021a).

4.3.2 Strategieën om bias te verminderen

Hoewel je bias nooit volledig kan uitsluiten, zijn er wel manieren om de impact ervan te beperken, zoals:

- Dataset-diversificatie: verzamel trainingsdata uit meerdere bronnen, met oog voor de representatie van diversiteit in gender, etniciteit, taal, regio, leeftijd en sociaaleconomische achtergrond.
- Bias-audits: voer systematische tests uit op de output van AI om na te gaan of bepaalde groepen systematisch benadeeld worden.
- Transparantie: documenteer hoe het model getraind is, welke data zijn gebruikt en welke keuzes daarin zijn gemaakt (Model Cards, datasheets).
- Menselijke controle: laat gevoelige beslissingen – zoals in selectieprocedures of gezondheidszorg – altijd controleren door mensen.

4.3.3 Representativiteit en verantwoording

Een belangrijke vraag is: voor wie is het model bedoeld, en wie blijft buiten beeld? AI wordt vaak ontwikkeld in Silicon Valley, maar gebruikt over de hele wereld. Zonder expliciete aandacht voor context kan dit leiden tot toepassingen die culturele verschillen niet respecteren, of zelfs schadelijke gevolgen hebben. Denk aan automatisch gegenereerde beelden die stereotypen bevestigen, of samenvattingen die bepaalde perspectieven weglaten.

Een verantwoord gebruik van AI betekent dus niet alleen technische precisie, maar ook maatschappelijke reflectie. Wie bepaalt wat als 'normaal' geldt in een dataset? Wie heeft toegang tot de technologie, en wie niet? En wie draagt de gevolgen van fouten?

Box 4.7 – Waar gaat machinaal leren mis?

Supervised learning: hierbij worden voorbeelden met een label gebruikt, maar labels zijn niet altijd objectief. Wat neutraal is, kan per cultuur of context verschillen.

Unsupervised learning: hier zijn geen labels, maar het model groepeert data op basis van statistische overeenkomsten. Dit kan bestaande ongelijkheden reproduceren. Neem het voorbeeld van een systeem waarbij men tracht om botbreuken in het voetgewricht te herkennen. Het systeem bleek best goed te werken, maar niet vanwege de botbreuken op de opnames, maar de herkenning van rimpelige huid van oudere mensen. Oudere mensen hebben namelijk vaker botbreuken.

Reinforcement learning: bij het gebruik van menselijke feedback bestaat het risico dat de menselijke voorkeuren bias bevatten, en dat deze zich in het systeem vastzetten.

Bias kan ontstaan in elk stadium: bij het verzamelen van data, het selecteren van features, het trainen van het model, of zelfs bij de interpretatie van de output.

4.4 AI-hallucinaties

Stel je vraagt aan een taalmodel wie de president van Frankrijk is. Je verwacht Emmanuel Macron als antwoord. Maar het model antwoordt: François Hollande. Of erger nog: Victor Hugo. Dit soort onjuiste of verzonnen antwoorden staan bekend als AI-hallucinaties. Het woord 'hallucinatie' wordt hier gebruikt omdat de AI feitelijke informatie lijkt te verzinnen op basis van wat plausibel klinkt, zonder dat die informatie daadwerkelijk waar is. In deze sectie ontdek je wat AI-hallucinaties zijn, waarom ze optreden, hoe je ze herkent en welke implicaties ze hebben voor het gebruik van generatieve AI.

4.4.1 Wat zijn AI-hallucinaties?

Een AI-hallucinatie is output die grammaticaal en stilistisch correct lijkt, maar feitelijk onjuist, misleidend of compleet verzonnen is (Maynez et al., 2020). Bij taalmodellen als ChatGPT betekent dit dat het model zinnen genereert die niet overeenkomen met de werkelijkheid, maar wel klinken alsof ze waar zouden kunnen zijn. Zo kan een chatbot een niet-bestaande wetenschappelijke studie citeren of een artikel samenvatten dat niet bestaat. Bij beeldgeneratoren zoals DALL·E kunnen hallucinaties leiden tot onlogische afbeeldingen met bijvoorbeeld een hand met extra vingers of niet-bestaande objecten.

Hallucinaties vormen een van de grootste uitdagingen voor de betrouwbaarheid van generatieve AI, vooral in contexten waarin feitelijke juistheid essentieel is, zoals wetenschappelijk schrijven, medische adviezen of juridische ondersteuning.

4.4.2 Waarom ontstaan hallucinaties?

Het fundamentele mechanisme van een groot taalmodel is gebaseerd op kansberekening. Het model voorspelt telkens welk token – een woord of een deel ervan – waarschijnlijk volgt op basis van alle voorgaande tokens. Dit proces, ook wel 'next-token prediction' genoemd, is puur statistisch. Het model weet niet wat waar is of wat bestaat; het weet alleen wat waarschijnlijk klinkt binnen de context (Brown et al., 2020).

Als de trainingsdata onvoldoende zijn op een bepaald onderwerp, of als de context van de prompt vaag is, kan het model verkeerde aannames maken. Soms verzint het bronnen, omdat het geleerd heeft dat een artikel met jaartal en auteursnaam een goede vorm is, zonder daadwerkelijk te controleren of dat artikel bestaat. Ook de zogenaamde temperatuurinstelling van het model – die bepaalt hoe creatief het model mag zijn – speelt hierin een rol: hoe hoger de temperatuur, hoe groter de kans op verrassende, maar onbetrouwbare output.

4.4.3 Gevolgen van hallucinaties

In academische of professionele contexten kunnen AI-hallucinaties ernstige gevolgen hebben. Een student die een niet-bestaande bron citeert, riskeert beschuldiging van fraude. Een arts die op AI vertrouwt voor een zeldzame diagnose, kan een verkeerde behandeling kiezen. Zelfs in creatieve contexten, zoals storytelling of design, kan hallucinatie problematisch zijn als de gebruiker een bepaald niveau van consistentie verwacht.

Daarom is menselijke tussenkomst altijd noodzakelijk: de gebruiker moet de informatie controleren, de bronnen nagaan en zich bewust zijn van de grenzen van het systeem. Chatbots zijn geen orakels, maar geavanceerde gokmachines. Zonder kritische blik kan hun output misleiden, ondanks een professionele of overtuigende toon.

Box 4.8 – Technieken om hallucinaties tegen te gaan

Leveranciers van LLM's onderkennen het probleem van hallucineren en willen dit tegengaan. Bijvoorbeeld door fine-tuning. Daarbij neem je een bestaand taalmodel, maar pas je de gewichten van de parameters aan door specifieke training.

Een andere methode werkt op basis van zogenaamde Retrieval Augmented Generation (RAG) (Lewis et al., 2020). Daarbij worden documenten gevoerd aan het systeem (bijvoorbeeld PDF's, website-teksten, tekstbestanden) waarbij de LLM het resultaat van de response toesnijdt op de inhoud van deze documenten.

Daarnaast worden meer en meer technieken toegepast waarbij modellen gedwongen worden om met zichzelf een onderzoekend gesprek te starten voor betere uitkomsten (zgn. reasoning).

4.4.4 Herkennen en omgaan met hallucinaties

Er zijn enkele strategieën om AI-hallucinaties te herkennen en ermee om te gaan:

- Controleer bronnen: vraag het model naar de exacte URL of DOI van een artikel. Verzonnen bronnen hebben vaak geen werkende link. Toch kan een bron wel echt zijn, ook als de link niet werkt. Vraag daarom vooral naar de titel, auteur, jaartal en publicatie, en zoek deze zelf op.
- Vraag om onderbouwing: laat de AI de redenering achter een bepaalde conclusie uitleggen. Vage of circulaire argumentatie is een waarschuwingssignaal.
- Gebruik fact-checking tools: combineer AI-output met tools zoals Google Scholar, Consensus, PubMed of Wikipedia om claims na te trekken.
- Verlaag de temperatuur: als je controle hebt over de instellingen van het model, verlaag dan de temperatuur om de output voorspelbaarder en feitelijker te maken.

Box 4.9 – Waarom AI gelooft in het eigen gelijk

Hallucinaties zijn geen fout van een gebrekkig model, maar een logisch gevolg van hoe taalmodellen ontworpen zijn. Ze hebben geen wereldkennis of geheugen van echte gebeurtenissen. Ze denken niet zoals mensen. Wat ze genereren is gebaseerd op patronen, niet op feiten. Ook al klinkt het zinnetje "Victor Hugo was president van Frankrijk in 2018" grammaticaal perfect, het is pure fictie – statistisch afgeleid, niet waarheidsgetrouw.

Daarnaast is het belangrijk te beseffen dat LLM's zelf geen toegang hebben tot de actuele stand van zaken, maar chatinterfaces zoals ChatGPT vaak wel doordat ze het internet kunnen doorzoeken of specifieke databases. Door technieken zoals RAG en door ook verbindingen te maken via de Chatinterface met zoekmachines of bijvoorbeeld specifieke databases.

4.5 De 'Black Box' van AI

Hoewel generatieve AI vaak indrukwekkend lijkt door het vloeiende taalgebruik en snelle reacties, blijven veel van de onderliggende processen onzichtbaar en moeilijk te doorgronden – zelfs voor de ontwikkelaars. Dit fenomeen staat bekend als het 'black box'-probleem. Een AI-model kan tot nauwkeurige voorspellingen komen of overtuigende teksten genereren, maar het is vaak niet duidelijk hoe of waarom het dat doet. In deze sectie duiken je dieper in de aard van deze ondoorzichtigheid, de pogingen tot verduidelijking, en de ethische en maatschappelijke implicaties ervan.

4.5.1 Waarom is AI zo ondoorzichtig?

De complexiteit van moderne neurale netwerken is moeilijk te overschatten. Modellen zoals GPT-4 bevatten honderden miljarden parameters – gewichten en verbindingen tussen neuronen die samen het gedrag van het model bepalen (OpenAI, 2021). Wanneer je een prompt invoert, worden er duizenden berekeningen uitgevoerd in fracties van seconden. Elk van deze berekeningen draagt een klein beetje bij aan het eindresultaat, maar geen enkele individuele stap is op zichzelf beslissend. Hierdoor is het nagenoeg onmogelijk om te reconstrueren welke paden precies verantwoordelijk zijn voor de gegenereerde output.

Bovendien leren deze modellen via 'terugvoering' (back propagation) op basis van enorme hoeveelheden data. Ze internaliseren patronen zonder expliciete logica of regels te volgen. Het gevolg is dat ze functioneren als een soort complexe intuïtie: uiterst efficiënt, maar lastig uit te leggen. Dit gebrek aan transparantie staat controle, vertrouwen en verantwoording in de weg – zeker in contexten met hoge maatschappelijke risico's zoals het vaststellen van een medische aandoening, het nemen van besluiten over toekennen van uitkeringen of toelating tot een opleidingen.

4.5.2 Pogingen tot uitlegbaarheid

AI lijkt misschien soms een mysterie voor je: er komt een antwoord uit, maar je weet niet hoe het tot stand kwam. Om dat 'black box'-gevoel tegen te gaan is het onderzoeksveld 'explainable AI' (XAI) ontstaan. Onderzoekers zoeken manieren om duidelijker te maken hoe een Ai tot een besluit komt. Dat doen ze bijvoorbeeld zichtbaar te maken waar het model in een tekst op let, of door een ingewikkeld model tijdelijk te vervangen door een eenvoudiger model dat beter te begrijpen is (Lipton, 2018).

Twee van de technieken hiervoor zijn bijvoorbeeld:

- SHAP (SHapley Additive exPlanations), dat probeert te bepalen welke inputfeatures het meeste hebben bijgedragen aan een bepaalde output.
- LIME (Local Interpretable Model-agnostic Explanations), dat het gedrag van een complex model lokaal benadert met een eenvoudiger, interpreteerbaar model.

Deze termen mag je hierna weer vergeten, maar wat wel belangrijk is om te onthouden, is dat dit soort methoden maar een beperkte blik geven. Ze geven een indruk van wat het model mogelijk heeft gedaan, maar geen sluitend bewijs. Bij taalmodellen zoals GPT is dit nog lastiger, omdat de invoer steeds verandert en de uitkomst afhankelijk is van lange contexten en kansberekeningen.

4.5.3 Ethische implicaties van ondoorzichtigheid

Het black box-karakter van AI roept fundamentele vragen op over verantwoordelijkheid en aansprakelijkheid. Wie is er verantwoordelijk als een AI-systeem een fout maakt – bijvoorbeeld een discriminerende beslissing in een sollicitatieprocedure of een foutieve medische aanbeveling? Als we niet kunnen uitleggen waarom een model iets doet, kunnen we het dan wel inzetten bij beslissingen die mensenlevens raken?

Cynthia Rudin (2019) betoogt dat in gevoelige domeinen zoals justitie, geneeskunde of kredietverstrekking geen black box-modellen gebruikt zouden mogen worden. In plaats daarvan pleit ze voor interpreteerbare modellen die weliswaar iets minder krachtig zijn, maar waarvan de werking begrijpelijk en toetsbaar is.

Bovendien leidt ondoorzichtigheid tot wantrouwen. Burgers, beleidsmakers en professionals willen begrijpen waarom een systeem tot bepaalde conclusies komt. Zonder uitlegbaarheid verliezen AI-

systemen legitimiteit. Transparantie en uitlegbaarheid zijn dus geen luxe, maar voorwaarden voor maatschappelijk verantwoord AI-gebruik.

4.6 Zelfstudievragen

4.6.1 Checkvragen

1. Wat is het verschil tussen supervised learning, unsupervised learning en reinforcement learning? Geef bij elk een kort voorbeeld.
2. Wat zijn AI-hallucinaties en waarom ontstaan ze? Noem twee strategieën om hallucinaties te herkennen.
3. Leg uit wat het 'black box'-probleem van AI is en waarom dit ethische vragen oproept.

4.6.2 Reflectievragen

4. In hoofdstuk 4 wordt beschreven hoe bias in AI-systemen ontstaat door scheef verdeelde datasets. Denk aan een concreet voorbeeld uit jouw eigen studiegebied waar dit problematisch zou kunnen zijn. Hoe zou je als toekomstig professional hiermee omgaan?
5. Dit hoofdstuk vergelijkt het gebruik van AI met het besturen van een auto zonder te begrijpen hoe de motor werkt. In hoeverre vind je dat gebruikers van AI-systemen de onderliggende technologie moeten begrijpen? Wat zijn de voor- en nadelen van meer technische kennis voor gewone gebruikers?
6. Door AI-gebruik, worden sommige woorden opeens een stuk vaker gebruikt. Zo valt het in het Engels op dat het woord ‘delve’ opeens veel vaker in teksten voor komt. Sommige mensen denken dat dit zou komen doordat bepaalde LLM's vooral getraind worden door mensen uit bepaalde gebieden. Zo zou het woord ‘delve’ gangbaarder zijn in Kenia, waar veel LLM's getraind worden. Wat denk jij? Beargumenteer je antwoord.

4.6.3 Antwoordsuggesties

1. Supervised learning gebruikt gelabelde voorbeelden om patronen te leren (bijvoorbeeld e-mails labelen als spam of geen spam). Unsupervised learning zoekt patronen zonder labels (bijvoorbeeld klanten groeperen op basis van koopgedrag). Reinforcement learning leert door beloning en straf (bijvoorbeeld een robot leren lopen door positieve feedback voor goede stappen).
2. AI-hallucinaties zijn overtuigend klinkende maar feitelijk onjuiste antwoorden die AI genereert. Ze ontstaan omdat AI statistisch voorspelt wat waarschijnlijk volgt, zonder echte kennis van feiten. Je kunt ze herkennen door bronnen te controleren, om onderbouwing te vragen, lagere temperatuurinstellingen te gebruiken, en fact-checking tools in te zetten.
3. Het black box-probleem betekent dat we niet kunnen verklaren hoe AI tot specifieke beslissingen komt, ook al zijn de resultaten accuraat. Dit roept ethische vragen op over verantwoordelijkheid, transparantie en aansprakelijkheid, vooral in gevoelige domeinen zoals zorg of rechtspraak waar uitlegbaarheid cruciaal is.
4. Eigen reflectie gewenst. Denk bijvoorbeeld aan medische diagnoses die vooroordelen over bepaalde patiëntengroepen kunnen bevatten, of HR-systemen die discrimineren bij sollicitaties. Als professional kun je bewust zijn van deze risico's, diverse datasets eisen, regelmatig audits uitvoeren en altijd menselijke controle behouden bij belangrijke beslissingen.
5. Eigen mening gewenst. Argumenten voor meer kennis: beter kritisch gebruik, meer vertrouwen, betere kwaliteitscontrole. Argumenten tegen: technische complexiteit is niet voor iedereen

toegankelijk, focus moet liggen op praktisch gebruik. Een middenweg zou basisbegrip van mogelijkheden en beperkingen kunnen zijn zonder diepgaande technische details.

5 Chatsoftware: wat zit er onder de motorkap van jouw chatomgeving?

Stel je voor …

Je scrolt door Copilot's antwoord op een onderzoeksvraag die je had gesteld. Je denkt bij jezelf: "Theoretisch snap ik hoe deze systemen werken, iets met neurale netwerken, maar wat gebeurt er praktisch met mijn input? Ik typ vragen in, maar waar belandt die vraag eigenlijk?" En je herinnert je de waarschuwingen van je docent: "Deel nooit gevoelige data via deze systemen." Zou het kunnen dat je informatie gestolen wordt? Doorverkocht? En is dat altijd zo? Tijd om eens te kijken hoe deze systemen in de praktijk werken en wat dat betekent voor jouw gebruik.

5.1 Infrastructuur

Over het algemeen communiceer je via een chatinterface met AI-systemen die Large Language Models (LLM's) bevatten. Deze technologie lijkt eenvoudig – je typt een vraag en krijgt een antwoord – maar achter de schermen is een complexe technische infrastructuur nodig. Veel mensen denken dat ze direct met een LLM praten en dat hun data alleen door het model wordt verwerkt en dat het model daar dan ook vanzelf door leert. In werkelijkheid verloopt alle communicatie via de chatinterface, die door een andere partij wordt beheerd. Hierdoor ontstaan misverstanden, bijvoorbeeld over:

- wie toegang heeft tot je data;
- of je data wordt gebruikt voor modeltraining;
- waar je data wordt opgeslagen en hoe lang.

Het is dus essentieel om het verschil te begrijpen tussen het LLM-bestand (het model) en de leverancier van de chatinterface (de tussenlaag).

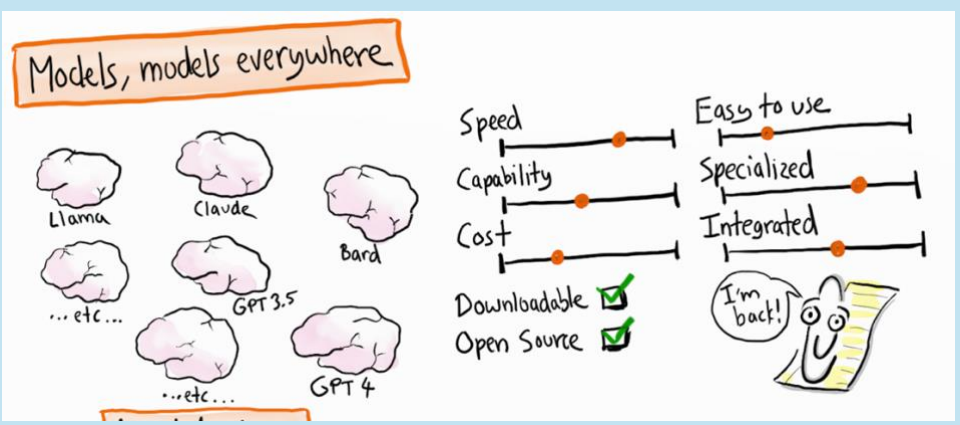
5.2 Het LLM als groot bestand

Een LLM is één bestand dat je eventueel op je eigen computer zou kunnen zetten. Het bestand bevat miljarden parameters die zijn getraind op grote hoeveelheden data. Dit bestand bevat geen ingebouwde chatfunctionaliteit: het is een programma dat, bij een bepaalde invoer (prompt), een enkele response genereert op basis van de GPT-methode (Generative Pretrained Transformer, zie Box 1.1 – Wat bedoelen we met 'GPT'?). Het model zelf onthoudt niets van eerdere interacties; het is zogenaamd stateless. Het bestand heeft ook geen kennis van bijvoorbeeld de huidige datum (en platforms bieden aanvullende technieken en functies om die huidige datum of andere bronnen wel te 'kennen').

Box 5.1 – LLM Modellen

Het aantal LLM-modellen waar je gebruik van kan maken groeit snel. Zo snel dat je het bijna niet bij kan houden. Elke model heeft zijn sterke en zwakke punten en specialismen. De meest bekende modellen zijn die van OpenAI (ChatGPT-4o, 5), Google (Gemini) en Anthropic (Sonnet, Opus), DeepSeek (DeepSeek), Qwen (Alibaba Cloud) en xAI (Grok), LLAMA (Meta), Mistral (Mistral). Dit zijn de grootste en meest algemeen toepasbare modellen met de meeste parameters.

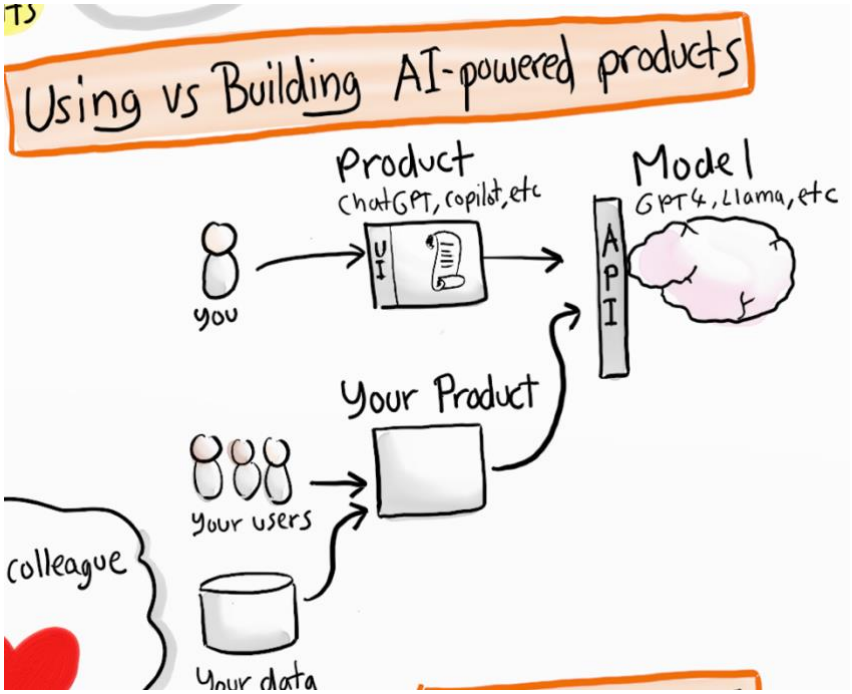
Wil je een overzicht van beschikbare modellen hebben, neem dan een kijkje op de website van HuggingFace: https://huggingface.co/models. Of kijk in Tabel 5.1.



Figuur 5.1 Overzicht van verschillende LLM modellen en hun karakteristieken uit de video video “Generative AI in a nutshell” van Henrik Kniberg (2024).

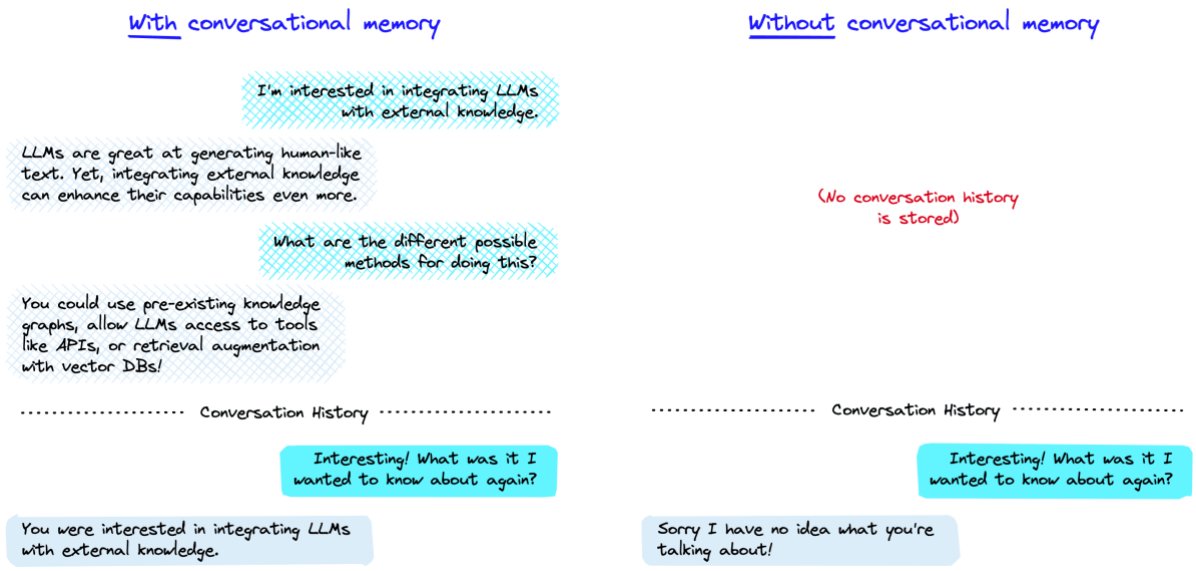
5.3 Chatinterface en chatgeschiedenis

De ervaring van een doorlopend gesprek met een generatieve AI ontstaat door een tussenlaag: de chatinterface (bijvoorbeeld een webapp of app, denk aan Copilot, ChatGPT, Gemini, Claude).



Figuur .5.2 Grafische weergave hoe gebruikers via een tussenproduct interacteren met een LLM. De gebruiker ziet een interface (UI) in een product en dat product praat via Application Protocol Interface standaarden (API) met het model en weer terug.

Deze Chatinterface bewaart de volledige chatgeschiedenis en stuurt bij elke nieuwe vraag de relevante chatgeschiedenis opnieuw mee naar het LLM-bestand. Dit heet conversational memory. Dit principe is ontstaan tezamen met de opkomt van LLM’s (Pinecone, 2025) en is essentieel voor het creëren van coherente gesprekken met AI.



Figuur 5.3 Voorbeeld van wel of geen conversational memory (Pinecone, 2025)

Het LLM ziet dus steeds het hele gesprek (of het deel dat past binnen de context window), waardoor het lijkt alsof het model een geheugen heeft en goed begrijpt wat je wil vragen. Want zonder de eerste vraag en het antwoord mee terug te sturen, weet het LLM dus niet waar het over gaat.

Box 5.2 – Context window – wanneer loopt de chatgeschiedenis vast

Wanneer je met ChatGPT in dezelfde chatbox blijft praten, lijkt dat soms handig omdat de chatbox alles wat je invoert, meeneemt voor de rest van het gesprek. Maar misschien merk je wel eens dat na verloop van tijd ChatGPT vergeet wat je aan het begin van het gesprek zei. Dit komt door de zogeheten context window, het geheugen van het model. GPT-3.5 kon bijvoorbeeld slechts 2048 tokens onthouden, terwijl GPT-4 er meer dan 32.000 aankan. Dit bepaalt hoe groot de tekstbestanden zijn die een chatsysteem kan verwerken en hoe 'ver' een model in een chatgeschiedenis terug kan kijken. Hoe groter de vragen en responsen en hoe meer, hoe sneller de context window vol is en de chatomgeving de controle verliest of onsamenhangender wordt.

5.4 Geprepareerde chatbots en projecten

Met behulp van algemene chatomgevingen zoals ChatGPT kun je ook geprepareerde bots maken. Zo kun je bijvoorbeeld een zogenaamde Custom GPT's maken: een vaste prompt, eventueel een kennisbron erbij, die je steeds opnieuw kunt gebruiken. Dat scheelt, want dan hoef je niet telkens opnieuw die instructies in te voeren en het zorgt voor meer consistente reacties van het taalmodel.

Daarnaast kun je in platforms als ChatGPT of Claude ook een project aanmaken. Een project is een soort werkruimte waarin je bestanden, bronnen en regels samenbrengt. Binnen dat project reageert de AI steeds specifiek op die context, zodat je die niet telkens opnieuw hoeft in te voeren in een nieuwe chat binnen het project. Een Custom GPT legt vooral een rol of werkwijze vast (zoals "altijd antwoorden in academische stijl"), terwijl een project vooral gaat om het organiseren van je eigen materiaal en afspraken.

Een voorbeeld: wil je één document op verschillende manieren bevragen, dan maak je daar een Custom GPT van en kun je meteen aan de slag. Als onderzoeker kun je in de instructies vastleggen uit welk domein het model vooral kennis moet halen en dat de antwoorden altijd referenties bevatten. Of, bij beeldgeneratie, kun je een vaste stijl beschrijven, zodat elke nieuwe afbeelding dezelfde look heeft. In een project kun je daarbovenop ook je bestanden en context bewaren, zodat je ze niet steeds opnieuw hoeft te uploaden.

Google noemt dit mechanisme van geprepareerde chatbots Gems, Antropic noemt het Projects, Copilot noemt dit een Agents en EduGenAI noemt dit een Persona.

Ook in gewone programma's zoals Word of Google Docs zie je steeds vaker zulke ingebouwde, geprepareerde bots terug. Vaak merk je dat niet eens: je gebruikt gewoon een menu-optie om tekst samen te vatten, te vertalen of om automatisch slides te maken op basis van een webpagina.

Op basis van de techniek kun je je ook voorstellen dat jijzelf of bedrijven zelf zo'n tussenlaag maken en aanbieden. Hieronder zie je een aantal voorbeelden.

Leverancier	Product/tussenlaag	Modellen	Veilig voor VU Amsterdam?	Herkomst (Land, Continent)
SURF / Npuls	EduGenAI (waarschijnlijk beschikbaar september 2026)	Mix van open source (zoals Mistral, LLaMA) en commerciële modellen (o.a. GPT-4, Claude)	Hoog	Nederland, Europa
Microsoft	Copilot for Web/ Bing Chat	GPT-4, GPT-4o (via OpenAI), met web browsing en multimodaliteit	Hoog (via SURF licentie)	Verenigde Staten, Noord-Amerika
Microsoft	Copilot for Microsoft 365	GPT-4 (via OpenAI), geïntegreerd in Word, Excel, Outlook, Teams	Niet beschikbaar voor VU	Verenigde Staten, Noord-Amerika
Google DeepMind	Gemini	Gemini 1.5 Pro, Gemini Flash (multimodaal, real-time web access)	Hoog (Google Workspace for Education)	Verenigde Staten, Noord-Amerika
Google DeepMind	NotebookLM	Gemini 1.5 Pro, Gemini Flash (multimodaal, real-time web access)	Hoog (Google Workspace for Education)	Verenigde Staten, Noord-Amerika
OpenAI	ChatGPT	GPT-3.5 (gratis), GPT-4, GPT-4o, GPT-5 (Plus), met DALL·E 3, Code Interpreter, Vision, Voice	Laag	Verenigde Staten, Noord-Amerika
Anthropic	Claude	Claude 2, Claude 3 Opus, Sonnet, Haiku (veiligheidsgericht, grote contextvensters – wil wel data gebruiken voor trainingsdoeleinden)	Middel	Verenigde Staten, Noord-Amerika

Mistral AI	LeChat	Mistral 7B, Mistral 8x7B (open source, Europees privacygericht)	Gemiddeld	Frankrijk, Europa
Proton	Lumo	Verschillende open source modellen. Proton presenteert zich als de meest privacy beschermde en Europese online omgeving	Hoog (hoewel geen overeenkomst met VU Amsterdam)	Zwitserland, Europa
DeepSeek	DeepSeek Chat	DeepSeek-V2, DeepSeek-Coder (gericht op wiskunde, programmeren, Chinees-Engels tweetalig)	Zeer laag	China, Azië
xAI	Grok	Grok-1, Grok-1.5, Grok-2 (multimodaal, real-time X/Twitter-integratie, 1M token context)	Zeer laag	Verenigde Staten, Noord-Amerika

Tabel 5.1 Ontwikkelaars van LLM's, hun service, modellen en globale aanduiding van voldoen aan Privacy en Security eisen op basis van kennis van het AI Competence Network van VU Amsterdam.

5.5 Data en privacy: wie beslist wat er gebeurt met jouw gegevens?

Leveranciers van chatinterfaces bepalen wat er gebeurt met jouw data en de gegenereerde antwoorden. De leverancier kan er bijvoorbeeld voor kiezen om:

- de chatgeschiedenis lokaal of in de cloud op te slaan;
- jouw data te gebruiken voor analyse, productverbetering of zelfs voor training van nieuwe modellen;
- data te delen met derden of juist expliciet te beschermen.

Dit beleid verschilt per aanbieder en is vaak niet helemaal transparant. Sommige interfaces bieden privacyinstellingen, andere niet. Daarom is het belangrijk om bovenstaande techniek te begrijpen en op basis daarvan inschattingen te maken, zoals:

- Werk je met een leverancier met onduidelijke voorwaarden en processen? Wees dan heel voorzichtig met de data die je uploadt en de vragen die je stelt. Zorg ervoor dat je nooit gevoelige gegevens deelt zoals persoonlijke gegevens of vertrouwelijke bedrijfsinformatie. Voorbeelden van dit soort leveranciers zijn ChatGPT, Google, Anthropic en veel kleinere leveranciers. Ook in andere gevallen zijn de voorwaarden vaak niet helder. Zoals bij een betaald account of accounts die zogenaamd op basis van inlog met het organisatie-account werken (bijvoorbeeld dat je inlog met je VU e-mail adres via Google login).
- Biedt de universiteit of je werkgever een chatomgeving aan en zijn er duidelijke afspraken dat de gegevens nooit gedeeld worden met partijen buiten de instelling? Dan kun je veel vrijer gebruikmaken van de chatomgeving, maar deel nog steeds geen gevoelige gegevens als het niet strikt noodzakelijk is. Voorbeelden hiervan zijn de Copilot-omgeving die de universiteiten

gebruiken, de SURF EduGenAI omgeving of omgevingen die door Onderzoeksinstituten worden aangeboden in eigen beheer.

- Biedt de universiteit of werkgever geen omgeving aan of twijfel je over de afspraken met de leverancier? Kijk dan of je kan werken met een LLM die je op je eigen (krachtige) PC of laptop kan installeren en gebruiken via een lokaal programma (Local LLM's). Voorbeelden van programma's zijn Jan.ai, LM Studio (Google), OLLama.

Vaak blijft er toch veel vaag over hoe je gegevens verwerkt worden, daarbij spelen veel juridische en technische vraagstukken een rol. Voor meer achtergrond over de complexiteit rond eigenaarschap van door gebruikers gegenereerde data in AI-interacties verwijzen we je naar het werk van Metcalf & Crawford (2016).

Box 5.3 – Uploaden van bestanden en dataveiligheid

Mag je bestanden uploaden in chatsystemen? Ga ervan uit dat bij aanbieders van deze systemen die gratis zijn, je daarmee eigenlijk deze documenten aan die bedrijven geeft. Zij kunnen die data herverwerken (voor modeltraining) of bijvoorbeeld zelfs verkopen. Bij de betaalde varianten kan er verschil zijn.

Daarnaast schend je met het uploaden van de bestanden misschien het copyright van de auteurs of uitgeverijen van de documenten. Je kan daarbij denken: "Nou en? Wie gaat mij betrappen!" Maar, juist voor die situatie zou je een beroep moeten doen op je eigen geweten en integriteit. Als het gaat om bestanden met bijzondere of gevoelige gegevens moet je dus helemaal oppassen.

Je kunt het lekken van data alleen voorkomen als je de systemen gebruikt waarvoor je instelling een verwerkersovereenkomst heeft afgesloten (bij de VU voor Copilot, Google Gemini of NotebookLM of voor EduGenAI – met VU inlog) of als je alleen Lokale LLM's gebruikt.

5.6 Zelfstudievragen

5.6.1 Checkvragen

1. Wat is het verschil tussen een LLM-bestand en de chatinterface die je gebruikt om met het model te communiceren?
2. Hoe werkt 'conversational memory' en waarom lijkt het alsof een AI-model zich jouw eerdere berichten herinnert?
3. Welke factoren moet je overwegen bij het kiezen tussen verschillende AI-leveranciers zoals OpenAI, Microsoft Copilot of EduGenAI wat betreft veiligheid en privacy?

5.6.2 Reflectievragen

4. Stel je uploadt een document met gevoelige informatie naar ChatGPT voor een studieopdracht. Welke risico's loop je en hoe zou je dit anders kunnen aanpakken?
5. Veel studenten denken dat ze direct met een AI-model praten, maar in werkelijkheid gaat alles via een tussenlaag van een bedrijf. Hoe beïnvloedt dit jouw vertrouwen in en gebruik van AI-chatdiensten?

5.6.3 Antwoordsuggesties

1. Een LLM is eigenlijk één groot computerbestand dat je theoretisch op je eigen computer zou kunnen zetten. Het heeft geen ingebouwde chatfunctionaliteit en is 'stateless' - het onthoudt niets van eerdere gesprekken. De chatinterface is de tussenlaag (zoals de ChatGPT website of app) die de gesprekservaring creëert door de chatgeschiedenis op te slaan en bij elke nieuwe vraag de hele conversatie weer mee te sturen naar het LLM.
2. Het LLM zelf heeft geen geheugen. De chatinterface bewaart de volledige chatgeschiedenis en stuurt bij elke nieuwe vraag de relevante context opnieuw mee naar het model. Dit heet 'conversational memory'. Hierdoor lijkt het alsof het model zich het gesprek herinnert, maar in werkelijkheid ziet het steeds het hele gesprek opnieuw.
3. Je moet kijken naar factoren zoals: transparantie over wat er met je data gebeurt, of er verwerkersovereenkomsten zijn afgesloten door je instelling, waar de data wordt opgeslagen, of je data gebruikt wordt voor modeltraining, en of je gevoelige informatie kunt delen. EduGenAI via SURF en Copilot via je universitaire licentie zijn meestal veiliger dan directe commerciële diensten.
4. Als je gevoelige informatie uploadt naar ChatGPT kan dit worden opgeslagen, geanalyseerd of gebruikt voor modeltraining door OpenAI. Je loopt het risico dat vertrouwelijke gegevens bij derden terechtkomen. Betere alternatieven zijn: de AI-omgeving van je universiteit gebruiken, lokale LLM’s installeren, of de gevoelige delen uit je document halen voordat je het uploadt.
5. Deze kennis zou je bewuster moeten maken van de keuzes die je maakt. Je realiseert je dat bedrijven als OpenAI, Google of Microsoft bepalen wat er met je gesprekken gebeurt. Dit kan je motiveren om bewuster te kiezen voor systemen waar je instelling afspraken over heeft gemaakt, of om kritischer te zijn over welke informatie je deelt via commerciële chatdiensten.

6 Hoe heeft jouw AI-gebruik impact op het milieu?

Stel je voor …

Je stelt een vraag aan ChatGPT of laat een beeld genereren met Midjourney. Maar je medestudenten vertellen je dat dit slecht is, omdat het heel veel energie kost. Als je het nieuws en sociale media volgt staat het er ook bol van. Je vraagt je af: moet ik generatieve AI wel gebruiken als het zo slecht voor het milieu is? En: hoe slecht is het eigenlijk?

6.1 Wat betekent jouw AI-gebruik eigenlijk voor energie, CO_2 en water?

p. 6-58

Als je via je laptop of telefoon een vraag stelt aan een generatieve AI, zie je niet wat er allemaal achter de schermen gebeurt. Je gebruikt dan ongemerkt een groot netwerk van datacenters, krachtige grafische processors en koelsystemen die veel energie en water gebruiken. In dit hoofdstuk verkennen we hoe dat samengangt met de impact op het milieu.

We staan stil bij drie dingen: hoeveel energie kost het om modellen te trainen, hoeveel energie en water nodig is om antwoorden te genereren, en hoeveel CO_2 dat uitstoor. We kijken daarbij naar het wereldwijde gebruik. Daarna zoomen we in op jouw eigen gebruik en vergelijken we dat met de milie-impact van andere dagelijkse activiteiten. Denk hierbij aan YouTubefilmpjes kijken, vliegen naar je vakantiebestemming, een hamburger eten of douchen.

6.2 Waarom kost het trainen van AI-modellen veel energie, water en CO_2?

Het kost veel energie om modellen te trainen, omdat de waarde van de miljarden parameters van het neurale netwerk van het LLM moet worden berekend. Daarvoor worden over het algemeen tienduizenden processoren (GPUs) van het bedrijf Nvidia gebruikt. Elke processor vraagt zo'n 700 Watt aan vermogen. Dit kun je vergelijken met een consumentenmagnetron die op bijna maximaal vermogen draait. Voor koeling is veel water nodig en door fossiele brandstoffen opgewekte energie zorgt voor CO_2 uitstoot.

Tegelijkertijd is het moeilijk om vast te stellen hoeveel energie er nodig is om taalmodellen te trainen en wat de CO_2 -uitstoot daarvan is. Ten eerste omdat daar veel factoren een rol bij spelen, zoals de complexiteit van het model zelf, maar ook van de hardware waarop het draait, de vragen die je stelt en de energiebron die je gebruikt (bijvoorbeeld als je hernieuwbare energie zou gebruiken). Ten tweede omdat de AI bedrijven niet rapporteren over hun energiegebruik (Vries-Gao, 2025). Daarnaast zijn de cijfers veranderlijk op basis van de snelle technologische ontwikkelingen. Veel onderzoek naar de milieubelasting van generatieve AI zijn daarom schattingen.

Maar over hoeveel energie praten we? De training van modellen van de huidige generatie LLM's, die vergelijkbaar zijn met trainingssessies voor GPT-4o, duren ongeveer drie maanden en verbruiken ongeveer 45-56 gigawattuur energie (You, 2025). De trainingssessies vragen dus zowel een hoog vermogen als een lange duur. De totale hoeveelheid energie voor training is daarmee gelijk aan de hoeveelheid energie die nodig is om ongeveer 60.000-74.000 Nederlandse huishoudens een jaar van stroom te voorzien. Dat is behoorlijk veel. Soms wordt wel beweerd dat je voor het energieverbruik van één datacenter zelfs een kerncentrale nodig hebt. Op het vermogen van een gemiddelde kerncentrale

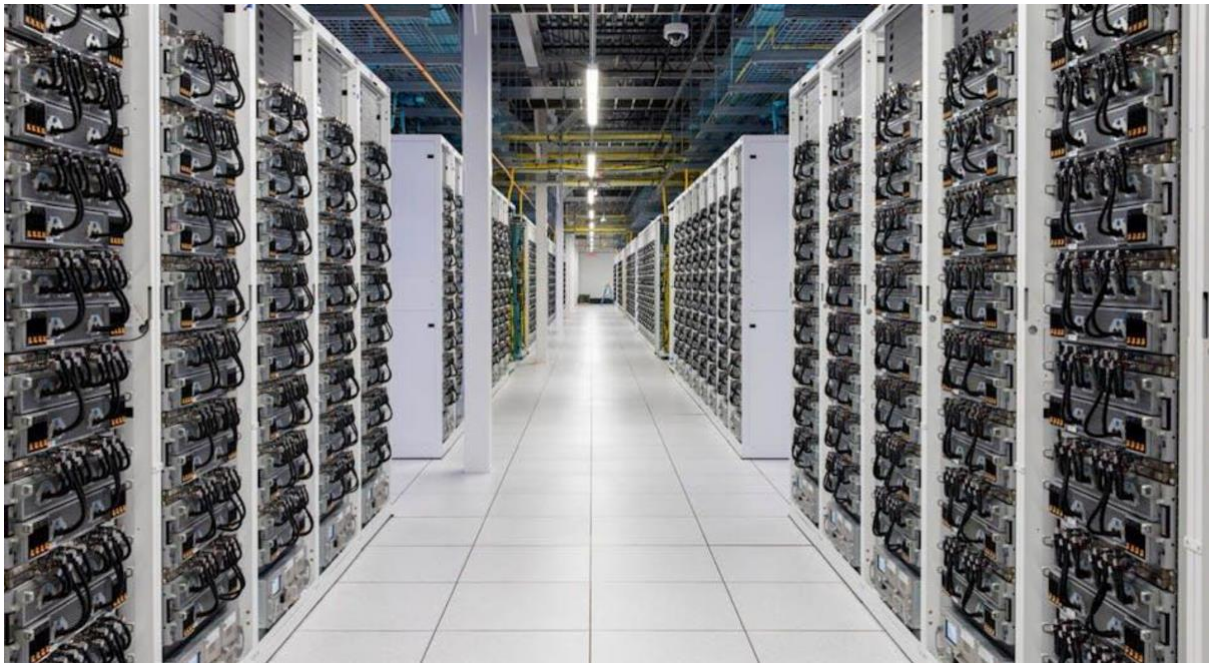
(3000 MW) zou je bijvoorbeeld 120 LLM's tegelijk kunnen trainen. Neem je windturbines op zee (10-15 MW per stuk), dan heb je er twee of drie nodig. Het blijft dus oppassen met vergelijkingen.

Modeltraining kost dus veel energie, maar hoe vertaalt dat zich eigenlijk naar een prompt? ChatGPT verwerkt ongeveer 1 miljard prompts per dag (Ahmed, 2025). Stel dat een LLM ongeveer een jaar gebruikt wordt en we smeren de energie voor modeltraining over dat aantal prompts uit, dan komen we uit op ongeveer 0,151 Wh (wattuur) per prompt. Hieronder lees je daarover meer.

6.3 Hoeveel energie kost het als jij een vraag stelt aan een AI?

Elke keer dat je een tekst, afbeelding of video genereert met een generatief AI-systeem, draaien er servers in een datacentrum. Dat kost energie – maar hoeveel precies?

Figuur 6.1 Afbeelding van een typisch datacenter: een heel groot gesloten gebouw met heel veel computers erin (bron van de afbeelding https://www.northcdatacenters.com/).



Figuur 6.2 Interieur van het xAI datacenter Colossus in in Memphis, Tennessee. Colossus bevat ongeveer 100.000 Nvidia Graphical Processing Units (GPU's) om taalmodellen te trainen. Uitbreiding naar 200.000 GPU's zit al in de planning (afbeeldingsbron: https://www.datacenterfrontier.com/machine-learning/article/55244139/the-colossus-ai-supercomputer-elon-musks-drive-toward-data-center-ai-technology-domination).

Over het precieze energieverbruik is weinig bekend. Er wordt wel vergelijkend naar onderzoek gedaan met Open Source modellen (zie paragraaf 9.7 voor meer uitleg hierover). Maar de meest gebruikte modellen zijn gesloten en daarom niet te testen. Daarnaast veranderen ze voortdurend: nieuwe, grotere modellen hebben vaak meer rekenkracht nodig en gebruiken dus waarschijnlijk meer energie.

Om toch een indruk te krijgen, wordt vaak een vergelijking gemaakt tussen een vraag stellen aan Google en een prompt aan een AI. Een ingewikkelde vergelijking omdat ook over Google's energieverbruik weinig bekend is. Maar voor de discussie gaan we er toch even op in. Het energieverbruik voor een Google-zoekopdracht ligt naar schatting bijvoorbeeld rond 0,3 wattuur (Wh). Een schatting voor een gemiddelde ChatGPT-opdracht lag bij eerdere modellen rond de 2,9 Wh – bijna tien keer zoveel (You, 2025). Maar dankzij technologische vooruitgang en efficiëntere modellen, zoals GPT-4o, is het energieverbruik per prompt inmiddels gedaald tot ongeveer 0,3 Wh. Dat zou dus vergelijkbaar zijn met een Google-zoekopdracht (You, 2025). Bij een lange invoer zijn de energiekosten wel hoger: dit kan oplopen tot zo'n 2,5 Wh bij een invoer van met lengte van een tijdschriftartikel, tot 40 Wh bij een invoer van ongeveer 200 pagina's tekst. Het genereren van afbeeldingen en video's vraagt flink meer energie.

6.4 Hoe verhoudt prompten zich tot ander dagelijks energieverbruik?

Of de hoeveelheid energie om een prompt te verwerken veel of weinig is, hangt dus af van hoe je het vergelijkt met het gebruik van andere apparaten of diensten voor een bepaalde tijdsduur. Bijvoorbeeld: 0,3 Wh is minder dan de hoeveelheid elektriciteit die een ledlamp of een laptop in een paar minuten verbruikt. Of gelijk aan 0,9 seconden gebruik van een 1200 Watt haardroger of 7,2 seconden draaien van je koelkastcompressor van 150 Watt. Dus zelfs voor een intensieve gebruiker zullen de energiekosten van ChatGPT maar een kleine fractie zijn van diens totale elektriciteitsverbruik en dus ook van de totale hoeveelheid energie die wereldwijd gebruikt wordt. In de tabel hieronder vergelijken we de hoeveelheid energie voor 100 prompts (als je een dag intensief gebruikmaakt van een generatief AI-systeem) met het energieverbruik van andere alledaagse energiebehoeften. Laat de getallen eens op je inwerken.

Apparaat/Categorie	Vermogen	Energieverbruik	Wh per dag	Tijd voor 0,3 Wh	Tijd voor 30 Wh
Haardroger	1200 W	-	-	0,9 seconden	1,5 minuten
Koelkastcompressor	150 W	-	-	7,2 seconden	12 minuten
Laptop	50 W	-	-	21,6 seconden	36 minuten
LED-lamp	10 W	-	-	1,8 minuten	3 uur
100 ChatGPT prompts	-	30 Wh totaal*	30 Wh*	-	**Referentiepunt**
100 ChatGPT prompts per dag	-	10,95 kWh/jaar	30 Wh	1 dag**	1 dag
2 uur YouTube kijken per dag (tablet)	22 W	16,06 kWh/jaar	44 Wh	49,1 seconden	1,36 uur
2 uur Netflix kijken per dag (TV)	115 W	83,95 kWh/jaar	230 Wh	9,4 seconden	15,7 minuten
Gemiddeld Nederlands huishouden	331 W	2900 kWh/jaar (CBS, 2018)	7945 Wh	3,3 seconden	5,4 minuten
Stand-by verbruik Nederlands huishouden	51 W	450 kWh/jaar (United Consumers, 2025)	1224 Wh	21,2 seconden	35,3 minuten

Tabel 6.1 Een vergelijking tussen de energie benodigd voor 100 generatieve AI prompts per dag ten opzichte van ander alledaags energieverbruik.

6.5 Wat betekent jouw AI-gebruik in het grote energiegeheel?

Hoewel het energieverbruik per prompt weinig is, wordt de impact op globale schaal duidelijk als je dit vermenigvuldigt met de miljoenen dagelijkse prompts. Alleen al ChatGPT's jaarlijkse energieverbruik bedroeg in 2023 naar schatting 226,8 GWh – genoeg om ruim 3 miljoen elektrische auto's een keer volledig op te laden (You, 2025).

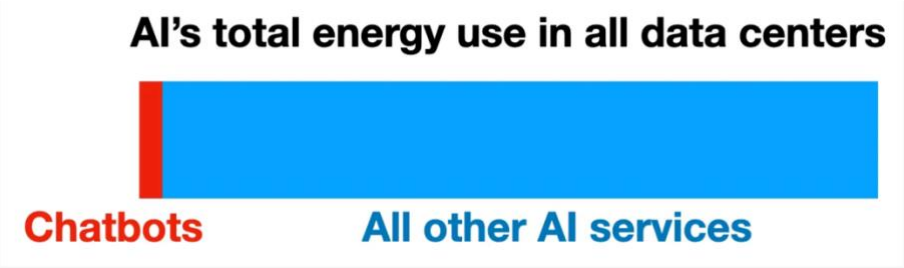
Volgens een rapport van het Internationaal Energieagentschap (Cozzi & Gould, 2024) zal de verwerking van alle digitale gegevens, voornamelijk voor AI, in 2030 alleen al in de VS meer elektriciteit verbruiken dan de productie van staal, cement, chemicaliën en alle andere energie-intensieve goederen samen. Volgens het rapport zal de wereldwijde vraag naar elektriciteit vanuit datacenters tegen 2030 daarmee meer dan verdubbelen. Mogelijk tot een energieverbruik dat oploopt tot 10 procent van het totale energieverbruik in de Verenigde Staten. AI zal de belangrijkste motor achter die stijging zijn, waarbij de vraag vanuit speciale AI-datacenters alleen al naar verwachting meer dan verviervoudigen zal. Eén datacenter verbruikt momenteel evenveel elektriciteit als 100.000 huishoudens, maar sommige datacenters die momenteel in aanbouw zijn, zullen twintig keer zo veel elektriciteit nodig hebben.

Krantenkoppen suggereerden kortgeleden nog dat AI in 2030 zal zorgen voor een stijging van 165 procent in het stroomverbruik van datacenters (Verhagen, 2025). Aan de toename naar 650 TWh zullen algemene groei van het internet en algemene AI-toepassingen ongeveer 200 TWh bijdragen. Alleen, deze toename lijkt al per eind 2025 in zicht te komen (Vries-Gao, 2025).

Het is wel belangrijk om hierbij in gedachten te houden dat algemeen AI-gebruik veel breder is dan alleen generatieve AI-toepassingen volgens Masley (2025a). Denk bij algemeen internet- en AI-gebruik aan:

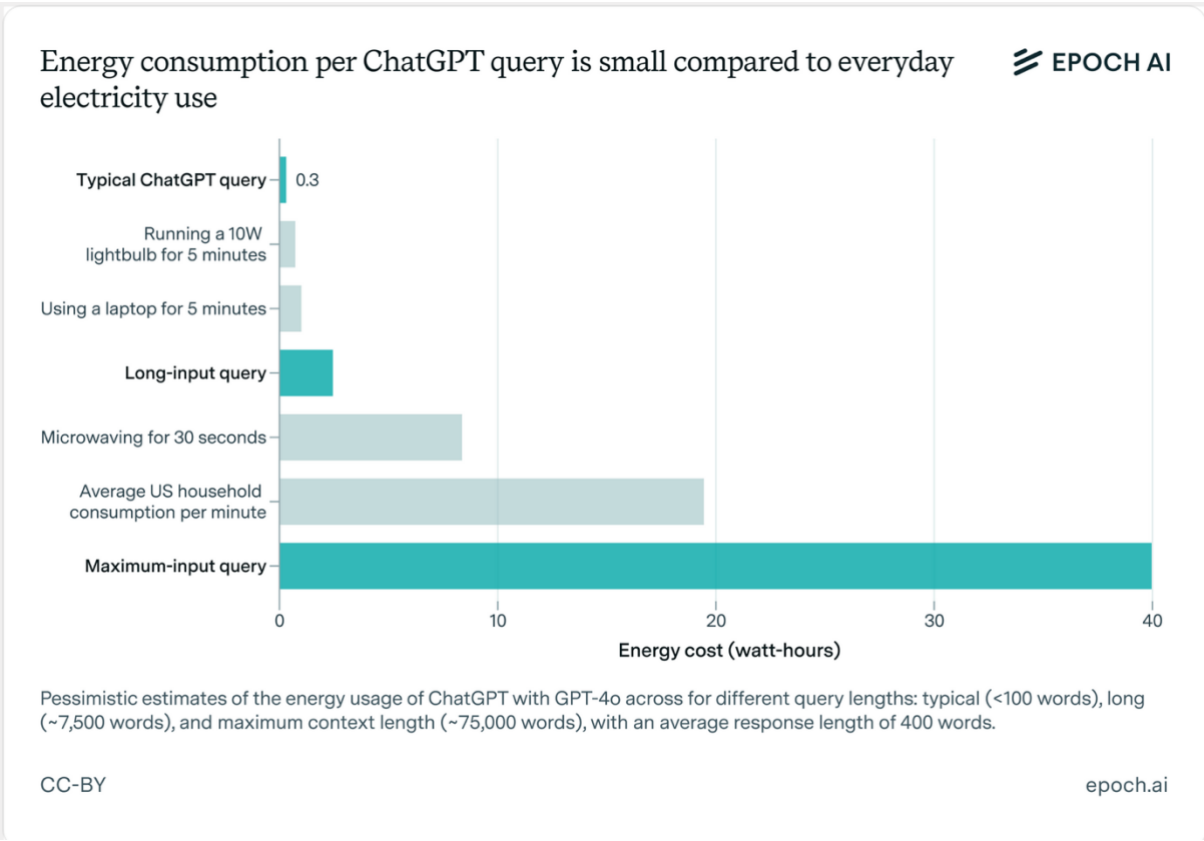
- het aansturen van aanbevelingssystemen voor personalisatie van bijvoorbeeld streaming platforms, e-commerce sites, sociale mediafeeds, en online advertenties.
- bedrijfsvoering zoals data-analyse, voorspellings- en besluitvormingsprocessen, zoeken en advertentiepersonalisatie.
- computer vision zoals object detectie, gezichtsherkenning, video en medische beeldanalyse en content moderatie voor het automatisch vlaggen van ongepaste beelden.
- geluid en audio (denk aan spraakassistenten en spraakherkenningssystemen, zoals Alexa van Amazon, Google Assistant, Siri van Apple).

Chatbots zoals ChatGPT vertegenwoordigen slechts 3 procent van het totale AI-gebruik (of 0,6 procent van het totale energieverbruik van datacenters). Dat is dus een beperkt aandeel in het energielandschap, maar dit groeit wel, zie ook Figuur 6.3 (Masley, 2025a).



Figuur 6.3 Vergelijking van de hoeveelheid energie die Chatbots verbruiken vergeleken met alle andere AI-diensten (bron grafiek: Masley, 2025a).

Ter vergelijking: YouTube gebruikt momenteel ongeveer 1 procent van de wereldwijde energie – ongeveer tien keer zo veel dan het verwachte gebruik van chatbots in 2030. Zorgen over de energie-impact van ChatGPT kun je daarom waarschijnlijk het beste proportioneel afwegen tegen andere digitale diensten.



Figuur 6.4 Vergelijking van energiegebruik vergeleken met ander vormen van alledaags energieverbruik (bron grafiek: You, 2025).

6.6 CO_2: Hoe AI bijdraagt aan opwarming van de aarde

In de toekomst zou de energie die nodig is voor datacenters uit groene en hernieuwbare bronnen kunnen komen – daar wordt ook hard aan gewerkt (OECD, 2022). Maar momenteel komt onze energie vooral uit het verbranden van fossiele brandstoffen: steenkool en aardgas. Daarbij komt CO_2 vrij. En hoewel AI maar een deel van de totale hoeveelheid uitgestoten CO_2 veroorzaakt, is het toch belangrijk om de effecten van verhoogde CO_2 concentraties in de atmosfeer goed te begrijpen om de soms hoogoplopende discussies in media, sociale media en politiek erover op waarde te kunnen schatten. Het gaat daarbij om twee belangrijke effecten van verhoogde CO_2 concentraties in de lucht en de gevolgen daarvan. Deze worden op een indringende manier beschreven door Andri Snear Magnason in zijn boek ‘Over tijd en water’ (2022). Hieronder lees je meer over zijn inzichten.

6.6.1 Opwarming van de aarde en de gevolgen

Ten eerste veroorzaakt een verhoogde CO_2 concentratie opwarming van de aarde. Zonder aanvullend klimaatbeleid bestaat het risico dat we rond 2050 richting 2 °C opwarming gaan. Dat lijkt misschien niet veel, maar dat is het wel. Vooral omdat de effecten verschillen (en dus ook veel groter kunnen zijn) per gebied en per continent:

- Grote gebieden worden nu al (en in de nabije toekomst nog meer) onleefbaar omdat het simpelweg te heet is om te leven.
- Heftigere periodes van droogte en neerslag verstoren de groei van gewassen.
- Er komen meer en grotere bos- en gebiedsbranden.
- Er komen meer en grote overstromingen.
- Er komt meer erosie en meer vruchtbare grond spoelt weg.

- Het weer wordt extremer en fluctueert onvoorspelbaarder.
- Meer overstromingen en risico's op gebieden die onder water komen te staan.

Al deze fenomenen hebben grote gevolgen, we zoomen even in op het versneld verdwijnen van de gletsjers als voorbeeld. Het effect hiervan op middellange termijn is bijvoorbeeld dat gebieden waar de bewoners afhankelijk zijn van de aanvoer van water van de rivieren die deze gletsjers traditioneel voeden, onbewoonbaar worden. Denk aan dichtbevolkte gebieden in Azië (van rivieren van gletsjers uit de Himalaya) en Latijns-Amerika (rivieren van gletsjers uit de Andes). Dit zal grote bevolkingsmigraties veroorzaken. Door het smelten van de gletsjers stijgt ook de zeespiegel. Daardoor zullen grote laaggelegen gebieden aan de zee verzilten of onder water komen te staan. Dat levert enerzijds problemen op om gewassen te laten groeien, anderzijds zullen migratiestromen op gang komen. Daarnaast zullen door hogere temperaturen de uitgestrekte wereldwijde toendra's (grote gebieden die grenzen aan een poolgebied en bestaan uit grassen, mossen, korstmossen en dwergstruiken) langzaam ontdooien. Hierdoor komen grote hoeveelheden methaan in de atmosfeer terecht. Methaan heeft een nog groter effect op het vasthouden van warmte op de aarde dan CO_2. Daarmee wordt de opwarming van de aarde dus weer verder versneld.

6.6.2 Verzuring van de oceanen en gevolgen

Ten tweede veroorzaakt een verhoogde CO_2 concentratie een hogere zuurgraad in de oceanen. CO_2 wordt namelijk opgenomen in de oceanen, maar dat is inmiddels zoveel dat de PH-waarde van het water substantieel verandert. Hierdoor gaat de aangroei van kalk voor organismen (schaaldieren en riffen) langzamer of stopt zelfs, waardoor al het leven dat zich voedt met deze organismen zal uitsterven. Dit zal een enorm effect hebben op de biodiversiteit in de oceanen en ook onze voedselproductie via visserij raken.

Er wordt hard gewerkt om CO_2-uitstoot te verminderen (bijvoorbeeld direct wegvangen bij de bron) of om CO_2 uit de atmosfeer of zeewater weg te vangen. Deze technieken staan echter nog in de beginfase en de schaal die nodig is om substantieel het percentage CO_2 in de atmosfeer te laten dalen is gigantisch. Het zou een internationale aanpak en investeringen vergen die ongekend zijn.

6.6.3 Verminder je de CO_2-uitstoot substantieel als je generatieve AI niet gebruikt?

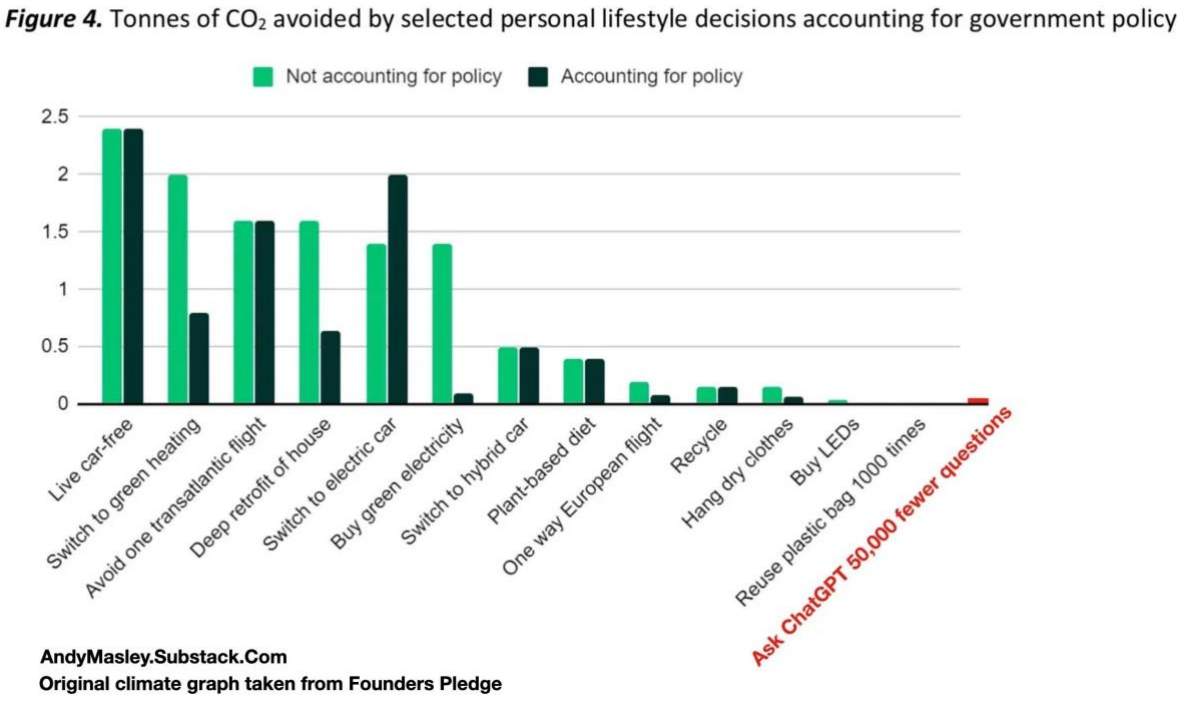
Zolang de wereld voor haar energieverbruik zwaar blijft leunen op het verbranden van fossiele brandstoffen, zal dit tot wereldwijde problemen leiden. Het gebruik van technologie, AI en generatieve AI draagt daar altijd aan bij. En waarschijnlijk vanwege de nieuwheid, snelheid en omvang van groei daarin wordt in klimaatdiscussies daar een grote nadruk op gelegd. Angsten en zorgen om de impact van nieuwe technologieën zijn van alle tijden, dus hoeveel zorgen moet je je echt maken dit keer? In vele journalistieke, sociale media en zelfs wetenschappelijke bronnen wordt betoogd dat we eigenlijk helemaal geen generatieve AI zouden moeten gebruiken vanwege het milieu.

Maar: hoe effectief is het eigenlijk om generatieve AI helemaal te vermijden. Als we kijken naar omvang, dan weten we nu dus dat het aandeel van generatieve AI in de totale wereldwijde hoeveelheid energieverbruik relatief bescheiden is. Waarom is er dan toch zo'n nadruk op generatieve AI? Waarom (ook) niet op andere factoren die zorgen voor CO_2 uitstoot?

Door andere factoren naast AI-gebruik te zetten, krijg je een realistischer en breder perspectief over de verhoudingen daartussen. Masley (2025a) heeft daarvoor veranderingen van leefstijl en consumptiepatroon geanalyseerd op effecten (Zie Figuur 6.5). Masley toont wat de hoeveelheid CO_2-uitstoot is die je kunt voorkomen door andere levensstijlbeslissingen te nemen, in vergelijking met 50.000 minder ChatGPT-prompts. Daaruit blijkt dat het aandeel van ChatGPT prompts in het niet valt in

vergelijking met andere levensstijlveranderingen zoals niet meer autorijden, een keer niet met het vliegtuig gaan, vegetariër worden, of je kleren te drogen hangen in plaats van de wasdroger gebruiken.

Masley vraagt zich daardoor af of we onze inspanningen niet beter op die factoren kunnen richten in plaats van eenzijdig de nadruk leggen op de milieubelasting van generatieve AI.



Figuur 6.5 Grafiek die weergeeft welke levensstijlverandering resulteert in een bepaalde vermindering van uitstoot van CO2 (in tonnen) (bron grafiek: You, 2025).

6.7 Water: waarom je AI ook dorst heeft

6.7.1 Globaal en lokaal watergebruik

Je hebt het misschien niet door wanneer je een prompt intypt, maar AI-systemen gebruiken ook relatief veel water (You, 2025). Servers in datacenters genereren namelijk veel warmte en moeten daarom gekoeld worden, momenteel nog met water. Op globaal niveau is het effect hiervan groot. Techbedrijven zoals Microsoft, Meta en Google gebruikten samen in 2022 bijvoorbeeld zo'n 2,2 miljard kubieke meter water – vergelijkbaar met het jaarlijkse waterverbruik van heel Denemarken. En dit verbruik groeit sterk.

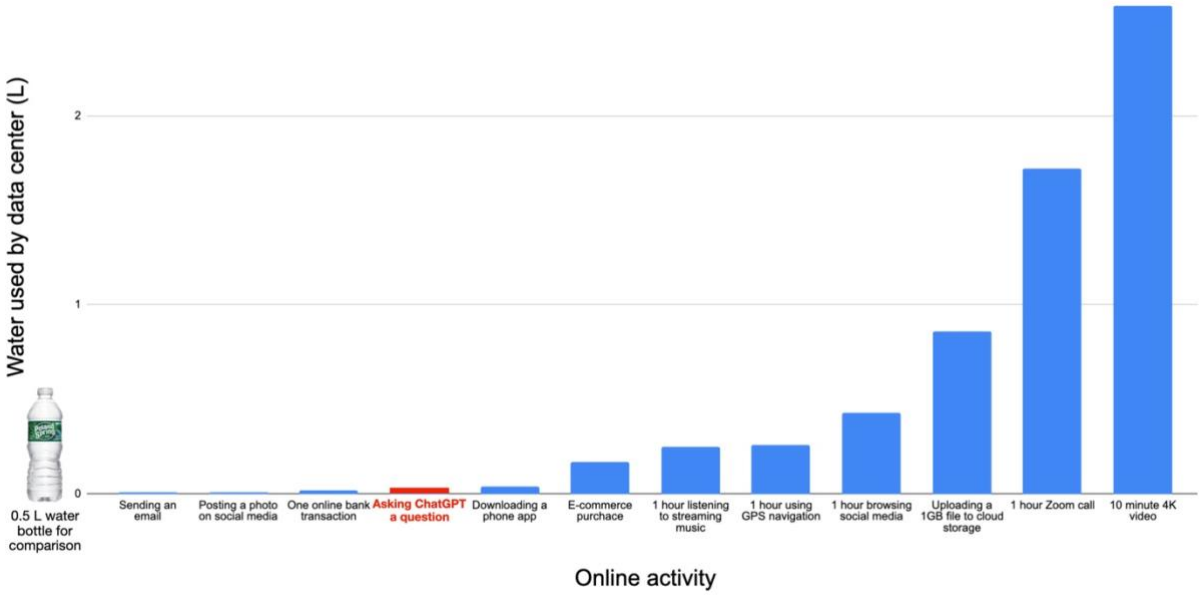
Het grootste probleem daarbij is niet eens de totale omvang van het gebruik (het water komt weer in de cyclus terecht), hierover later meer, maar vooral dat het lokaal tot problemen leidt: datacenters worden ook gebouwd in gebieden waar al waterschaarste heerst. Dit leidt tot conflicterende belangen tussen de koelwaterbehoefte van datacenters en het watergebruik voor landbouw, industrie en huishoudens.

Mesa is bijvoorbeeld een stad in een woestijnachtig gebied in Arizona dat dit probleem goed illustreert (Solon, 2021). Een belangrijke reden voor lokale overheden om datacenters te willen huisvesten is werkgelegenheid. En een woestijnachtig gebied met veel zonne- en windenergie biedt datacenters goedkope elektriciteit. Arizona lijkt dus geschikt voor een datacenter. Maar daarbij is geen rekening gehouden met de waterbehoefte. Het datacenter dat in Mesa in 2021 is gerealiseerd verbruikt tot 15 miljoen liter water per dag terwijl de regio al decennia kampt met extreme droogte. In totaal verbruikten

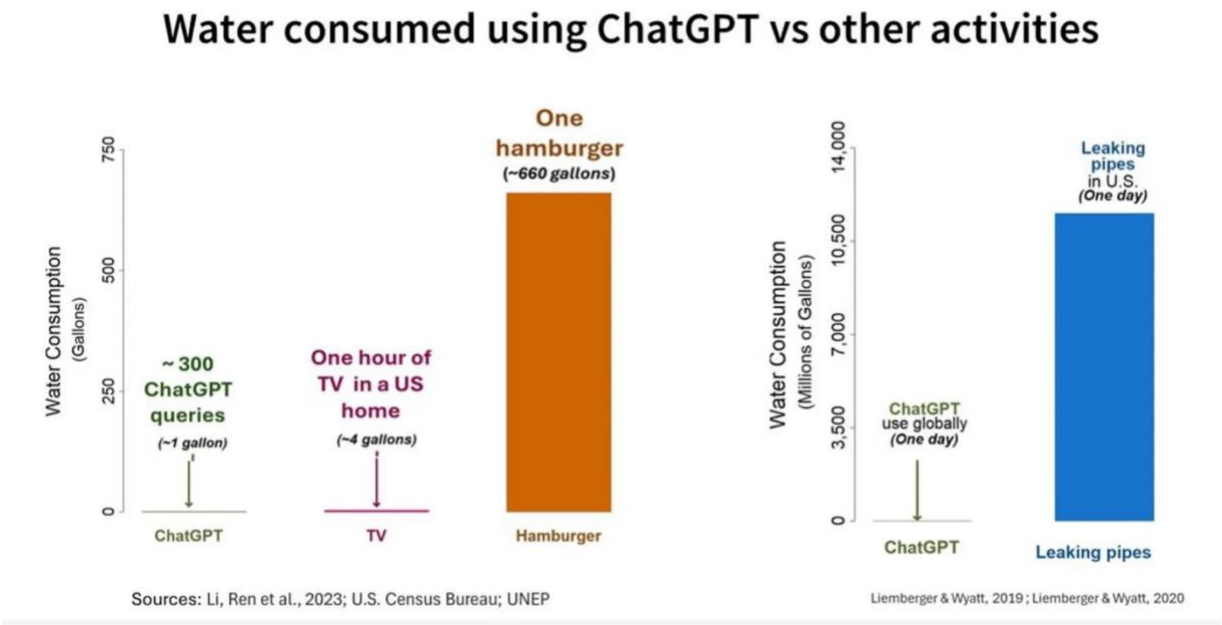
datacenters in Mesa in 2021 meer dan 3,8 miljard liter water per jaar- genoeg om tienduizenden huishoudens van water te voorzien. Dit leidt tot veel lokale spanningen.

6.7.2 Watergebruik per prompt

Maar hoeveel water gebruikt generatieve AI dan eigenlijk in vergelijking met ander activiteiten? In de media of op internet hoor je misschien vaker dat een enkele prompt een halve liter water zou kosten. Maar dat klopt niet: volgens recente schattingen is de hoeveelheid koelwater die nodig is om 20 tot 50 prompts te verwerken via ChatGPT ongeveer een halve liter koelwater. In vergelijking: vooral het kijken van streaming video's kost bijvoorbeeld veel energie en water volgens Masley. Ook zet hij dit naast het waterverbruik voor het produceren van een hamburger (landbouw) of het totale volume water dat verloren gaat door lekkage in de VS. Zie Figuur 6.6 en Figuur 6.7. Zijn overwegingen brengen hem tot de verzuchting dat het gebruik van twintig ChatGPT prompts van 500 milliliter water ook kan worden gecompenseerd door simpelweg vijf seconden korter te douchen of een hamburger over te slaan.



Figuur 6.6 Vergelijking van waterverbruik door verschillende online activiteiten (bron grafiek: Masley, 2025a).



Figuur 6.7 Vergelijking 2 van waterverbruik door verschillende functies (bron grafiek: Masley, 2025a).

De vraag is niet zozeer of het watergebruik van datacenters inherent slecht is, maar eerder of er bewuste keuzes worden gemaakt. In waterrijke gebieden is waterverbruik voor datacenters waarschijnlijk geen probleem, maar in gebieden met waterschaarste is elke druppel een ethische afweging. De vraag zou moeten zijn: is dit de beste locatie voor een datacenter? Zijn er alternatieven voor koeling? En hoe verhouden de economische voordelen zich tot de ecologische en sociale kosten? Het watergebruik van datacenters is relatief klein vergeleken met landbouw, maar in de context van extreme droogte telt elke liter.

6.8 Wanneer helpt AI juist het milieu?

Is AI dan altijd slecht voor het milieu? Niet per se. In allerlei bronnen vinden we terug dat AI-toepassingen juist kunnen bijdragen aan duurzaamheid. Denk aan:

- optimalisatie van energieverbruik in gebouwen;
- snellere analyse van klimaatdata;
- het voorspellen van opbrengsten in de landbouw;
- efficiënter plannen van logistiek en mobiliteit.

De inzet van AI voor dit soort toepassingen is erg complex, daarom is het moeilijk om te bepalen in welke mate hier positieve resultaten uit voortkomen.

6.9 De informatie-energie paradox

Maar hoe verhoudt het energieverbruik van digitalisering en generatieve AI zich tot de hoeveelheid informatie die het oplevert?, Vraagt Masley (2025b) zich af. Als je het energieverbruik van digitale informatiebronnen in perspectief plaatst, ontstaat er een opmerkelijk beeld. Voor 0,3 wattuur – de energie die een ChatGPT-prompt kost – krijg je toegang tot een synthese van miljarden documenten aan kennis. Ter vergelijking: het opzoeken van dezelfde informatie in een fysieke bibliotheek zou uren kosten, waarbij je vervoer naar de bibliotheek, de verlichting van het gebouw, en mogelijk het printen of kopiëren van bronnen meeneemt. De verschuiving van fysieke naar digitale informatiedragers heeft een enorme

milieuwinst opgeleverd die we vaak over het hoofd zien, maar die wel al in gang was gezet voor de komst van generatieve AI. Een wetenschappelijk tijdschrift dat vroeger maandelijks naar duizenden abonnees wereldwijd werd verzonden – met alle papierproductie, inkt, transport en afvalverwerking – is nu toegankelijk via een download van enkele megabytes. Zelfs als we alle serverkosten en het individuele energieverbruik meetellen, is de CO_2 -voetafdruk per geraadpleegd artikel drastisch gedaald.

Daarnaast heeft generatieve AI volgens Masley het vermogen om kennis te vermenigvuldigen zonder een evenredige toename van energieverbruik. Wanneer een AI-model eenmaal is getraind, kan het miljoenen gebruikers tegelijkertijd bedienen waarbij elke extra gebruiker slechts marginale kosten meebrengt. Een door AI gegenereerde samenvatting van complexe wetenschappelijke literatuur die vroeger dagen onderzoekswerk zou vergen, wordt nu in seconden geproduceerd voor de energiekosten van zeven seconden föhnen. Hiermee heeft iedere student wereldwijd toegang tot dezelfde geavanceerde AI-tools zonder de milieukosten van internationale conferenties, studiebezoeken of het verschepen van studiemateriaal. In die zin is de kennisopbrengst per wattuur van moderne AI-systemen ongekend hoog – een efficiëntie die vroegere informatiesystemen bij lange na niet haalden.

Toch knaagt er iets. En dat heeft te maken met de schaalgrootte waarop alles plaatsvindt en dat nieuwe technieken – als ze maar goedkoop genoeg zijn – tot een veelvoud van toepassingen leiden wat in eerste instantie niet was voorzien. De Jevons paradox verklaart dit.

6.10 Waar gaat het heen? De Jevons Paradox

Hoe de balans is tussen de bijdrage van AI aan duurzaamheid of juist het tegenovergestelde, is nog lang niet duidelijk en misschien zelfs nooit vast te stellen. Maar dat het wereldwijd veel energie kost om generatieve AI-systemen te trainen en gebruiken, is wel zeker. En dat wordt alleen maar meer, zelfs als de modellen en technieken efficiënter worden.

Hoe dit kan? Een verklaring hiervoor wordt gegeven door de zogenaamde Jevons Paradox. Die theorie voorspelt niet veel goeds (Polimeni et al., 2015): efficiency voordelen van (deel)systemen gaan juist altijd gepaard met groei van de energiebehoefte van het totale systeem omdat de efficiency ook meer gebruikstoepassingen en opbrengsten van systemen mogelijk maakt. En dat geldt zeker ook voor AI en generatieve AI. Want de mogelijke toepassingen van generatieve AI – omdat het relatief goedkoop is – groeien met duizelingwekkende snelheid en kunnen bijna zonder extra kosten aan gebruikers ter beschikking worden gesteld. Met andere woorden: goedkopere en energiezuiniger digitale processoren zorgen juist voor een toenemende vraag naar digitale generatieve AI-diensten en daarmee de energiebehoefte.

Tegelijkertijd is het belangrijk om ook de relatieve bijdrage van generatieve AI aan milieubelasting tegenover andere activiteiten en de besparingen vanuit het paradoxale perspectief uit paragraaf 6.9 mee te nemen in je overwegingen. Vraag je af: is er een leverancier die van hernieuwbare energiebronnen gebruikt? Bijvoorbeeld datacentra waarvan je weet dat ze op zonne- of windenergie draaien.

Net als bij andere technologieën ligt de sleutel in verantwoord gebruik en overwegingen:

- Kies bewust voor het (minder) gebruiken van apparaten of diensten in het algemeen. Specifiek vliegtuigen en auto's.
- Kies een andere leefstijl om energie en het milieu te ontlasten, wordt bijvoorbeeld vegetariër of flexitariër, kijk minder YouTube-filmpjes, beperk je gebruik van sociale media.
- Kies bewust voor het al dan niet gebruiken van AI-toepassingen in relatie tot de informatie-energie paradox en de Jevons Paradox.

6.11 Zelfstudievragen

6.11.1 Checkvragen

1. Hoeveel energie verbruikt een gemiddelde ChatGPT-prompt ongeveer vergeleken met een Google-zoekopdracht volgens het boek?
2. Waarom hebben AI-datacenters water nodig en hoeveel water kost het ongeveer om 20 tot 50 simpele prompts te verwerken?
3. Noem twee manieren waarop AI-toepassingen juist kunnen bijdragen aan duurzaamheid.

6.11.2 Reflectievragen

4. Stel je gebruikt dagelijks ongeveer 50 AI-prompts voor je studie. Hoe verhoud je dit energieverbruik tot andere keuzes die je zou kunnen maken om je ecologische voetafdruk te verkleinen? Leg uit hoe je deze afweging zou maken.
5. Het hoofdstuk suggereert dat de relatieve milieu-impact van generatieve AI beperkt is vergeleken met andere digitale activiteiten. Hoe beïnvloedt deze informatie jouw houding rond AI-gebruik? Ga je hierdoor anders denken over wanneer je wel of niet AI gebruikt?
6. Beluister deze podcast van Trouw en probeer eens kritiek te formuleren op de ogenschijnlijke boodschap van de podcast.

Figuur 6.8 Podcast van Trouw, 25 juni 2025.

6.11.3 Antwoordsuggesties

1. Een gemiddelde ChatGPT-prompt verbruikt ongeveer 0,3 wattuur, wat vergelijkbaar is met een Google-zoekopdracht. Eerdere modellen verbruikten nog ongeveer 2,9 wattuur, bijna tien keer zoveel, maar door technologische vooruitgang is dit sterk gedaald. Tegelijkertijd worden modellen steeds krachtiger wat meer energie vergt. Maar, processor worden weer efficiënter. Dat vergt weer meer energie. Eigenlijk weet niemand precies waar het heengaat.
2. AI-datacenters hebben water nodig voor koeling van de servers die veel warmte genereren. Voor ongeveer 20 tot 50 simpele prompts is gemiddeld 500 ml water nodig, afhankelijk van het model, de promptlengte en de locatie van het datacenter.
3. AI kan bijdragen aan duurzaamheid door bijvoorbeeld optimalisatie van energieverbruik in gebouwen, snellere analyse van klimaatdata, voorspelling van opbrengsten in de landbouw, en efficiënter plannen van logistiek en mobiliteit.
4. 50 prompts per dag kosten ongeveer 15 wattuur, wat overeenkomt met 5,5 kilowattuur per jaar. Dit is vergelijkbaar met ongeveer 5 seconden korter douchen per dag of enkele minuten minder YouTube kijken. Je zou kunnen afwegen dat andere keuzes zoals minder vliegen, vegetarisch eten of minder autorijden veel grotere impact hebben op je ecologische voetafdruk.
5. Dit is een persoonlijke reflectie waarbij je kunt overwegen dat de informatie je bewuster maakt van de relatieve impact van verschillende activiteiten. Je zou kunnen concluderen dat je AI wel verantwoord kunt gebruiken voor zinvolle doelen, maar misschien bewuster wordt van overmatig

gebruik. Of je realiseert je dat andere levensstijlkeuzes belangrijker zijn voor het milieu dan AI-gebruik beperken.

6. De Podcast lijkt de algemene tendens te onderschrijven dat generatieve AI slecht is voor het milieu. Maar voor het maken van dat punt wordt alles wat met digitalisering te maken heeft op één hoop gegooid. Je kan beargumenteren dat hiermee de discussie over de impact van generatieve AI op een oneigenlijke manier overdreven wordt.

6.12 Activiteit - AI Environmental Challenge

Test jezelf op je milieu-impact.

https://canvas.vu.nl/courses/83333/pages/activiteit-ai-environmental-challenge

6.13 Activiteit - Hoe sta jij tegenover de milieu-impact van generatieve AI?

Als je hoofdstuk 6 hebt gelezen zul je merken dat het hoofdstuk niet een eenduidig antwoord geeft op de vraag of je generatieve AI wel of niet moet gebruiken vanuit milieuoogpunt. Wat is jouw mening hierover nu je het hoofdstuk hebt gelezen? Deel je mening met je medestudenten op Canvas. Maar blijf vriendelijk en blijf open staan voor argumenten.

7 Hoe helpt generatieve AI jou bij het doen van onderzoek en schrijven?

Stel je voor …

Je zit in de UB met Lisa. Lisa gebruikt ChatGPT en verschillende AI-zoektools aan het begin van elk project om onderzoeksvragen te genereren en voorstellen voor haar onderzoeksopzet te doen. Ze werkt haar onderzoek uit en gebruikt ChatGPT om haar data te analyseren en voorstellen te doen voor conclusies en discussiepunten. Bij het uitschrijven van haar onderzoek scherpt ze haar structuur en alinea-indeling aan met ChatGPT en gebruikt Writefull om zo goed mogelijk aan te sluiten bij het tekstgenre van het wetenschappelijk schrijven. Door AI strategisch in te zetten, bespaart ze tijd en verdiept haar inzicht in haar onderzoeksonderwerp. Maar mag dat eigenlijk wel? En hoe pak je dat productief aan?

Noot bij versie 1.1 van dit boek

Dit hoofdstuk riep in de vorige versie van dit boek bij een aantal docenten grote weerstand op. Want: als je het gebruik van generatieve AI voor het schrijven van een scriptie verbiedt, waarom zou je dan bespreken hoe generatieve AI je daarbij kan helpen? In de wetenschap worden de discussies daarover ook volop gevoerd (zie bijv. Andersen et al., 2025)

Je moet beseffen dat de dit onderwerp per discipline of opleiding anders wordt gewaardeerd. Er zijn opleidingen waarbij geen generatieve AI mag worden ingezet bij het scriptieproces. Maar gezien het doel van het boek - informeren en een basis bieden voor discussie - is het naar onze mening toch een noodzakelijk hoofdstuk.

Het belangrijkste is dat je je als student moet houden aan de richtlijnen van de opleiding of docent op welke wijze je generatieve AI al dan niet mag gebruiken voor het schrijven van de scriptie.

7.1 Generatieve AI integreren in je academische werk

AI kan een waardevolle toevoeging zijn aan je academische werk mits je het op een bewuste manier integreert. Het bovenstaande voorbeeld laat zien dat je generatieve AI in verschillende fasen van je academische werk in kan zetten, globaal:

1. **Oriëntatie:** begrip opbouwen van een onderwerp, literatuuronderzoek doen, hypothesevorming, vinden en formuleren van onderzoeksvraag en wijze van onderzoeken vastleggen.
2. **Uitvoeren onderzoek:** data verzamelen en analyseren en resultaten interpreteren aan de hand van de onderzoeksvraag.
3. **Interpreteren van de resultaten:** onderzoeksvraag beantwoorden, koppelen aan literatuur, beperkingen bespreken en suggesties voor vervolgonderzoek.
4. **Rapporteren en communiceren:** het vastleggen van het gehele onderzoek en de conclusies in een logische en goed leesbare tekst.

Deze globale stappen zet je op verschillende momenten om in tekst, figuren en meer richting een definitief geschreven tekst, zoals:

1. **Concept deel/tussenrapportages:** tussentijdse verslaglegging en onderbouwing vastleggen (hoeft nog niet aan de hoogste eisen te voldoen), uiteindelijk resulterend in:

2. **Definitieve rapportage en artikel**: het uitschrijven van de tekst volgens de conventies van een (wetenschappelijk) artikel:
    a. Hoofdstructuur opbouwen en verfijnen, inclusief titels en tussenkoppen (samenvatting, probleem, kader, keuze methode van onderzoek en opzet, resultaten, conclusies, discussie, referenties)
    b. Alineastructuur binnen de hoofdstructuur opbouwen, zorgen voor één kern per alinea in zogenaamde topic-zinnen.
    c. Zinsstructuur binnen alinea's opbouwen en verbindings- en signaalwoorden duidelijk aanwezig laten zijn.
    d. Grammatica, spelling, woordkeus, formuleren, citeren en refereren optimaal uitvoeren

Bij elk van de fasen kan generatieve AI je helpen. Je kan een begin maken door open prompts te gebruiken en hoe meer je onderzoek vordert en hoe meer keuzes je hebt gemaakt in je onderzoek, hoe specifieker je moet werken met generatieve AI.

AI vervangt jouw denkwerk bij al deze stappen niet, maar versterkt het, als jij de regie houdt. Bij elke stap of onderdeel kun je generatieve AI gebruiken om voorstellen te doen, om je tussentijdse werk te controleren en om feedback te krijgen op je werk. Gebruik nooit uitsluitend generatieve AI, maar werk altijd in combinatie met andere bronnen, zoals boeken, wetenschappelijk artikelen, docentenfeedback en peer review. Zo blijf jij in controle over je leerproces.

Box 7.1 – Kennisbronnen Informatievaardigheden UB en Academic Language Programme

De UB van de Vrije Universiteit biedt onderwijs en ondersteuning om te helpen bij het vinden en gebruiken van literatuur voor onderzoek. Ten aanzien van AI-geletterdheid biedt de UB online leermodules aan waarmee je snel en zelfstandig aan de slag kan.

Het Academic Language Programme (ALP) van de Vrije Universiteit helpt studenten ook om betere schrijvers te worden in Nederlands en Engels. Hou hun aanbod in de gaten.

7.2 Krachtige prompts voor academisch onderzoek en schrijven

Voor al de genoemde stappen en aspecten kan generatieve AI je helpen, maar het werkt alleen goed als je daar krachtige prompts voor gebruikt. Hieronder sommen we een aantal verzamelingen van prompts op die je zelf zou kunnen gebruiken.

Er zijn veel websites met tips voor prompts om academisch onderzoek te doen en om het maken van academische teksten te ondersteunen. Hieronder bijvoorbeeld een tiental prompts en aanpassingen van Balo (2025).

Wat	Prompt
Onderzoeksonderwerpen brainstormen	`Gedraag je als een brainstormdeskundige. Het is jouw taak om potentiële onderzoeksonderwerpen met betrekking tot [onderwerp] te brainstormen. Het doel is om unieke en interessante onderzoeksvragen te genereren die nog niet uitgebreid aan bod zijn gekomen in eerdere studies. Zorg ervoor dat de onderwerpen relevant zijn, haalbaar zijn voor onderzoek en kunnen bijdragen aan de bestaande kennis`

	over het genoemde onderwerp. Houd ook rekening met de mogelijke implicaties van het onderzoek, de haalbaarheid en de beschikbare middelen. Maak een uitgebreide lijst van potentiële onderzoeksonderwerpen, elk vergezeld van een korte beschrijving en motivering.
Onderzoeksvragen ontwikkelen	Als ervaren academisch onderzoeker is het jouw taak om overtuigende onderzoeksvragen te ontwikkelen over [onderwerp]. Deze vragen moeten tot nadenken aanzetten, complex zijn en mogelijk leiden tot belangrijke bevindingen in het veld. Ze moeten een open einde hebben, maar toch gericht en duidelijk zijn. De vragen moeten gebaseerd zijn op huidig onderzoek en literatuur over het onderwerp en moeten een leemte in de kennis opvullen of een nieuw perspectief bieden. Het doel is om de richting van een onderzoeksproject aan te geven en de basis te vormen voor de hypothese. Je moet kunnen verdedigen waarom deze vragen belangrijk zijn voor het vakgebied en hoe ze zullen bijdragen aan bestaand onderzoek.
Assisteren bij literatuuronderzoek	Als ervaren academisch onderzoeker is het jouw taak om de belangrijkste bevindingen van recente onderzoeken over het gegeven [onderwerp] te beoordelen en samen te vatten. Dit houdt in dat je de meest relevante en meest recente onderzoekspapers identificeert, grondig doorleest, de belangrijkste informatie eruit distilleert en deze samenvat in een duidelijke, beknopte en uitgebreide samenvatting. Je samenvatting moet de belangrijkste doelstellingen, methodologieën, bevindingen en implicaties van deze onderzoeken bevatten. Het moet ook een kort overzicht geven van de huidige stand van het onderzoek over het onderwerp. Vergeet niet om alle bronnen te citeren.
Hypothesen formuleren	Handel als een ervaren academisch onderzoeker. Ontwikkel een sterke, toetsbare hypothese voor een onderzoek naar [onderwerp]. De hypothese moet duidelijk en beknopt zijn en gebaseerd op bestaande wetenschappelijke literatuur. Ze moet een potentiële relatie of correlatie voorstellen tussen twee of meer variabelen met betrekking tot [onderwerp]. De hypothese moet ook zo worden opgesteld dat ze kan worden weerlegd of bevestigd door middel van wetenschappelijke methodologieën. Zorg ervoor dat de hypothese overeenstemt met de onderzoeksdoelstellingen en bijdraagt aan de vooruitgang van de kennis op dit gebied.
Een overzicht maken	Als ervaren academisch onderzoeker moet je een opzet maken voor een paper over [onderwerp]. Het overzicht moet de belangrijkste punten en subpunten van de paper logisch ordenen en een duidelijke routekaart bieden voor het onderzoeks- en schrijfproces. Het moet een inleiding, literatuuroverzicht, methodologie, bevindingen, analyse en conclusie bevatten. Zorg ervoor dat het overzicht voldoet aan de academische schrijfstandaarden en formats. Het overzicht moet ook aangeven waar belangrijke referenties of citaten zullen worden gebruikt, zodat er een uitgebreid overzicht ontstaat van de structuur en inhoud van het artikel.
Delen van het paper schrijven	Als ervaren academisch onderzoeker is het jouw taak om een [inleiding/hoofdstuk/conclusie] te schrijven over het [onderwerp]. Dit werk moet gedetailleerd zijn, goed onderzocht en geschreven in een academische stijl. Het moet een uitgebreid overzicht geven van het onderwerp, een logisch argument of analyse presenteren en dit onderbouwen met relevante bronnen, theorieën of gegevens. Zorg ervoor dat je actuele en relevante referenties gebruikt om je punten te ondersteunen. Het taalgebruik moet formeel, precies en duidelijk zijn. Het document moet worden opgemaakt volgens de geldende academische schrijfrichtlijnen of stijlgids.

Het betoog ontwikkelen	Als ervaren academisch onderzoeker is het jouw taak om een uitgebreid betoog te ontwikkelen over het gegeven [onderwerp]. Dit moet bestaan uit een duidelijke stelling, solide bewijsmateriaal uit geloofwaardige bronnen om je argument te ondersteunen en een logische opeenvolging van ideeën die leiden tot een overtuigende conclusie. Je argument moet objectief, kritisch en evenwichtig zijn. Ga in op tegenargumenten en geef er een duidelijk en helder antwoord op.
Conclusie en discussie formuleren	Als ervaren academisch onderzoeker is het jouw taak om mij te helpen om een goede conclusie en discussie te formuleren op basis van mijn onderzoeksvraag, de resultaten en voorlopige conclusies zoals je vindt in het document dat ik bijvoeg [document met voorgaande informatie]. De conclusie moet de onderzoeksvraag, methode en resultaten op een logische manier met elkaar verbinden. Het moet een discussie bevatten van de beperkingen van het onderzoek en de generaliseerbaarheid van de conclusies. De discussie moet ingaan op de impact van de conclusies van het onderzoek voor het domein van het onderzoek en de onderzoeksvraag en daarbij suggesties bieden voor vervolgonderzoek.
Corrigeer grammatica en syntax	Gedraag je als een ervaren grammaticacontroleur. Lees de [tekst] zorgvuldig door en controleer op grammatica-, interpunctie- en syntaxfouten. Corrigeer deze fouten met behoud van de oorspronkelijke betekenis en toon van de tekst. Zorg ervoor dat de tekst duidelijk, beknopt en goed gestructureerd is. Geef feedback op gebieden die verbetering of verduidelijking behoeven. Ervoor zorgen dat de uiteindelijke versie opgepoetst en foutloos is.
Formateer referenties	Gedraag je als een expert op het gebied van opmaakstijlen. Het is jouw taak om alle referenties in de [tekst] op te maken volgens de APA (American Psychological Association) stijl. Zorg ervoor dat alle in-tekstcitaten, referentielijsten en voetnoten nauwkeurig zijn opgemaakt volgens de APA-richtlijnen. Let goed op de details zoals auteursnamen, publicatiedata, titels en bronnen.
Referenties genereren	Handelen als een citatie-expert. Maak een citaat voor de gegeven tekst volgens de richtlijnen van de MLA (Modern Language Association). Zorg ervoor dat het citaat de naam van de auteur, de titel van het werk, de naam van de publicatie, de uitgever en het jaar van publicatie bevat. Het citaat moet ook het paginanummer bevatten (indien van toepassing). Zorg ervoor dat je interpunctie en cursivering correct gebruikt volgens de MLA-regels. Het citaat moet klaar zijn om in een academisch werkstuk of rapport te worden ingevoegd.

Tabel 7.1 Overzicht van krachtige prompts voor het doen van wetenschappelijk onderzoek door Balo (2025).

7.3 Oriëntatiefase: literatuuronderzoek

Literatuuronderzoek is zowel in de oriëntatiefase als tijdens het uitvoeren en rapporteren over het onderzoek zelf een belangrijke activiteit. Algemene generatieve AI-tools zoals Copilot of ChatGPT kunnen helpen bij het vinden van literatuur, vergelijkbaar met hoe je Google Scholar of de UB-catalogus gebruikt.

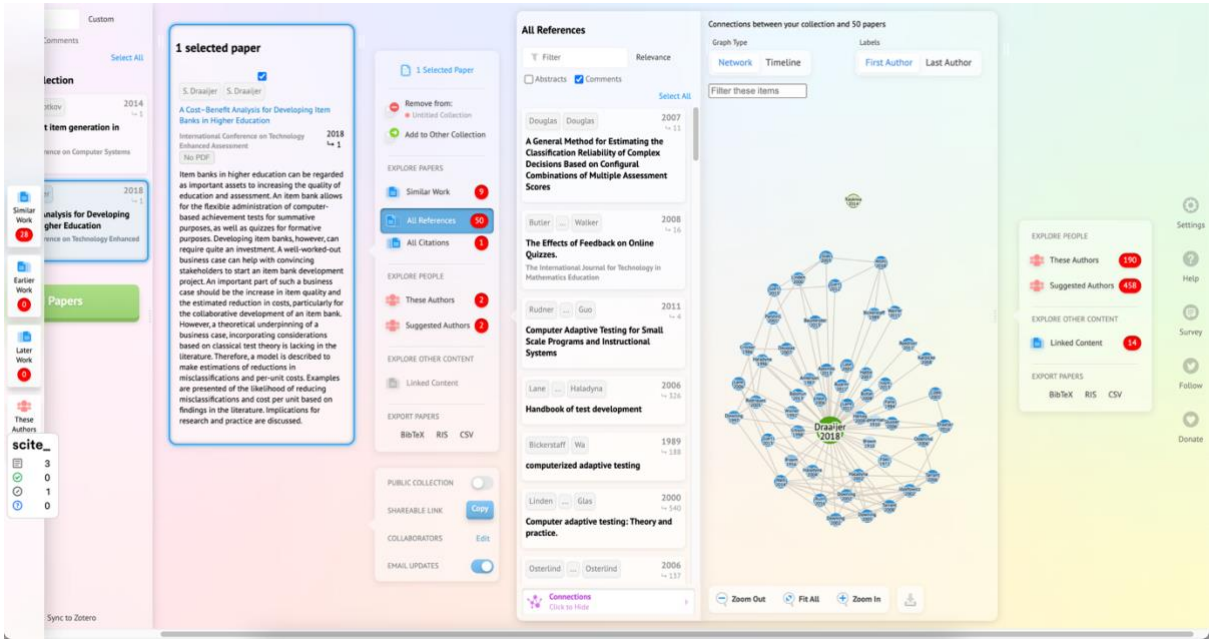
Box 7.2 – Hallucineren bij referenties

Een veelvoorkomende valkuil is dat het lijkt alsof ChatGPT echte wetenschappelijke bronnen citeert. In werkelijkheid genereert ChatGPT vaak verzonnen verwijzingen en niet werkende links. Het is daarom belangrijk om AI-output van bronnen van algemene chatomgevingen altijd te verifiëren (bijvoorbeeld

aan de hand van je literatuurlijst, databases of artikelen die je docent heeft aanbevolen). Ook kan het zijn dat de bron wel bestaat, maar de link alleen foutief is.

Tegelijkertijd is er met de komst van generatieve AI ook een groeiend aantal gespecialiseerde en door generatieve AI aangedreven wetenschappelijk zoekmachines ontwikkeld. Deze zijn beter dan de algemene generatieve AI zoals ChatGPT in staat om naast het taalmodel ook: vele wetenschappelijke databases (denk aan PubMed) als aanvullende bronnen te doorzoeken en te manipuleren, visualisaties en samenvattingen weer te geven en betere output te genereren, inclusief kloppende referenties. Hiermee kun je sneller:

- relevante bronnen vinden;
- samenvattingen genereren;
- verbanden tussen onderzoeken ontdekken;
- Onderzoeksvragen ontdekken en formuleren (what is the current research gap ...)



Figuur 7.1 Schermafdruk van het systeem ResearchRabbit waarbij voor een specifiek paper wordt gevisualiseerd hoe het samenhangt met de referenties in dat artikel.

Een andere manier om deze AI te gebruiken is om je teksten te controleren. Zo kun je in die systemen je voorlopige teksten uploaden en vragen of je in een tekst op de belangrijkste plaatsen ook de meest relevante referenties hebt geplaatst. En zo nee, of het systeem deze wil opzoeken, precies aan te geven waar deze referenties geplaatst moeten worden in je tekst en de referenties in het gewenste formaat kan aanbieden. Ook kun je je bestaande referentielijst uploaden en vragen om te controleren of de referenties echt bestaan en wat waarschijnlijk betere referenties zijn.

Wees je er wel van bewust, dat als je dit soort systemen gebruikt terwijl je onderzoek doet, dat er risico's zijn. Denk aan dataveiligheid en de verzameling, verwerking of verhandeling van je gegevens.

Hieronder beschrijven we de tools die het meest in opkomst zijn anno 2025. Bij de meeste tools heb je gratis toegang, maar zijn de mogelijkheden beperkt. Betaalde varianten bieden meer mogelijkheden.

Naam	Karakteristieken	Voordelen	Nadelen

Connected Papers	Visualiseert netwerken van verwante academische artikelen op basis van citaties en conceptuele verwantschap.	Helpt bij het verkennen van onderzoekslijnen en trends. Ontdekt relevante papers buiten directe zoektermen.	Geen automatische samenvattingen. Kan overweldigend zijn bij grote netwerken.
Consensus	AI-gedreven zoekmachine die wetenschappelijke consensus samenvat met filters op methodologie.	Snel overzicht van literatuurconsensus. Wetenschappelijk onderbouwd. Gebruiksvriendelijke samenvattingen met duidelijke bronvermelding.	Beperkt tot specifieke vakgebieden zoals geneeskunde en economie. Minder geschikt voor niche-onderwerpen. Kwaliteit afhankelijk van bronmateriaal.
Elicit	AI-gebaseerde assistent voor semantisch zoeken en automatische extractie van kernpunten uit papers.	Vindt relevante literatuur bij vage zoekopdrachten. Biedt automatische samenvattingen. Bespaart tijd.	Geen toegang tot alle volledige artikelen. Risico op verlies van nuance in samenvattingen.
Perplexity	Zoekmachine die AI gebruikt om natuurlijke taalvragen te beantwoorden met web- en academische bronnen.	Snelle, toegankelijke antwoorden. Transparante bronvermelding. Breed inzetbaar.	Bronkwaliteit varieert. Minder diepgaand dan gespecialiseerde academische tools.
Research Rabbit	Interactieve tool voor het visualiseren van netwerken tussen papers, auteurs en thema's met aanbevelingen.	Helpt bij ontdekken van nieuwe onderzoekslijnen. Inzicht in trends en netwerken. Vind papers buiten directe zoektermen.	Geen automatische samenvattingen. Complex voor beginners. Overweldigend bij grote datasets.
Scholarcy	Tool die automatische samenvattingen en referentie-extractie levert, gericht op structuur en toegankelijkheid.	Maakt complexe teksten toegankelijk. Structurering van literatuur. Tijdbesparend voor reviews.	Mist soms nuance. Minder accuraat bij technisch complexe teksten. Beperkte zoekfunctionaliteit.
Scopus	Grootste academische database met AI-geassisteerde zoekfilters en citatieanalyse.	Zeer betrouwbaar en uitgebreid. Sterk in trend- en impactanalyse. Geschikt voor systematische reviews.	Geen AI-samenvattingen. Betaalde toegang. Minder interactief dan moderne tools.
Semantic Scholar	AI-gestuurde zoekmachine met contextuele zoekopdrachten en aanbevelingen op basis van onderzoeksgeschiedenis.	Sterk in contextbegrip en relevante aanbevelingen. Gratis toegankelijk.	Beperkte samenvattingsopties. Niet alle disciplines goed vertegenwoordigd.

Tabel 7.2 Karakteristieken, voordelen en nadelen van moderne AI-gedreven wetenschappelijke informatie zoektools

De selectie is gebaseerd op recente vergelijkingen en gebruikerservaringen en opgevraagd in 2025 via Claude met Sonnet 3.7 voor dit Boek. Voor het beste resultaat kies je een tool op basis van je specifieke onderzoeksbehoefte: snelle consensus, netwerkvisualisatie, samenvattingen, of uitgebreide databasezoektochten.

7.4 Uitvoeringsfase

Tijdens het uitvoeren van onderzoek kun je generatieve AI ook inzetten om te ondersteunen bij allerlei deelstappen. Hieronder sommen we de Nederlandstalige versie van de tabel met de specifieke prompts uit *ChatGPT for research methodology* van Razia Aliani (2024) op. Haar prompts zijn gericht op het aangaan van een kritische analyse voor het aanpakken van onderzoek en het onderbouwen van keuzes daarin.

Onderwerp	Prompt
Voordelen van methodologie	Beschrijf de voordelen van het toepassen van een [jouw methodologie, bijv. “mixed methods”] bij het onderzoeken van [onderwerp], met nadruk op de verrijking van het onderzoek.
Rechtvaardiging van methodologie	Rechtvaardig de keuze voor een [kwalitatieve/kwantitatieve/gemengde] aanpak voor [onderwerp], en leg uit hoe deze aansluit bij de doelstellingen van het onderzoek.
Voor- en nadelen afwegen	Weeg de voor- en nadelen van de gekozen onderzoeksmethodologie af, met aandacht voor de impact op validiteit, betrouwbaarheid en toepasbaarheid van de studie.
Dataverzameling	Stel best practices voor voor het verzamelen van data binnen [onderzoeksveld/-context], met oog voor methodologische zorgvuldigheid en relevantie.
Onderzoeksontwerp	Geef een uitgebreide beschrijving van het onderzoeksontwerp, inclusief onafhankelijke en afhankelijke variabelen, controlemiddelen en mogelijke verstorende factoren.
Ontwikkelen van instrumenten	Formuleer een strategie voor het uitvoeren van pilottests of het verfijnen van onderzoeksmiddelen zoals vragenlijsten, interviews of observatieprotocollen.
Ethische overwegingen	Breng mogelijke ethische uitdagingen in kaart bij onderzoek naar [onderwerp] en stel praktische oplossingen voor om ethische standaarden te waarborgen.
Culturele sensitiviteit	Leg uit hoe culturele gevoeligheden of verschillen invloed kunnen hebben op ethisch onderzoek in [veld/context] en stel manieren voor om hiermee om te gaan.
Gegevensbeheer en -beveiliging	Ontwikkel een plan voor het veilig beheren en opslaan van onderzoeksdata, in overeenstemming met relevante regelgeving voor gegevensbescherming.

Tabel 7.3 Prompts voor wetenschappelijk onderzoek door Razia Aliani (2024).

Het is natuurlijk belangrijk dat je de suggesties die je krijgt op basis van deze prompts kritisch bekijkt en herschrijft naar je eigen stijl en inzichten.

7.4.1 Data-analyse met generatieve AI

Met veel generatieve AI-systemen kun je naast teksten, geluid of plaatjes ook statistische en analytische en numerieke analyse uitvoeren en visualisaties laten maken. Hierbij doen de taalmodellen zelf niet het analysewerk, maar sturen ze programma’s aan zoals Python en R om het werk te doen. Of ze genereren code die je daarna zelf in een programma kan uitvoeren. Hiermee zorgen deze AI-systemen ervoor dat de verwerking exact plaatsvindt, mits jij ze goed programmeert via goede prompts.

Als je bijvoorbeeld voldoende kennis hebt van statistische concepten en principes kun je hier snel voorlopige analyses en conclusies trekken. In het onderstaande voorbeeld, uitgevoerd met Anthropic van Claude, zie je hoe het werkt. Standaard maakt Claude visualisaties in een zogeheten React formaat, maar als je zelf met R werkt kun je ook vragen om R code die je vervolgens zelf kan gebruiken in R.

Vraag 1: Kun je een dataset voor me genereren van de cijfers voor 10 jongens en meisjes op 5 vakken van 5 havo? (Of voer een eigen dataset in bijvoorbeeld in Excel of CSV formaat)

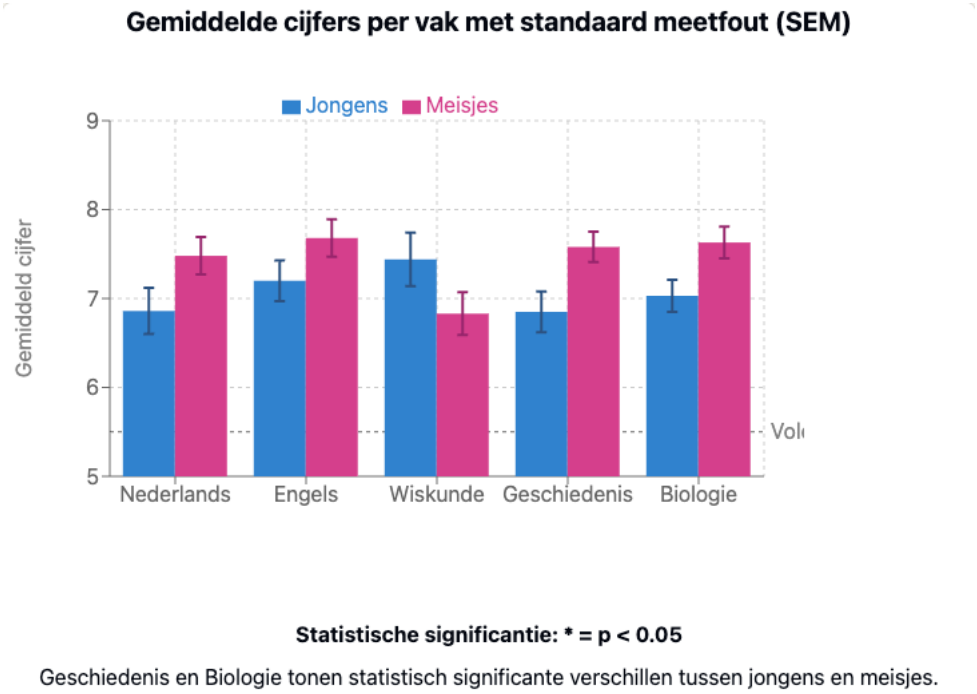
Respons: ...

Vraag 2: Kun je voor mij de trends in deze dataset geven. Specifiek wil ik weten of er een significant verschil is in het gemiddeld cijfer per vak voor jongens en meisjes.

Respons: ...

Vraag 3: Kun je voor mij een grafiek maken waar voor jongens en meisjes de gemiddelde score en de standaard meetfout wordt getoond om te inspecteren of er statistisch significante verschillen zijn per vak tussen de geslachten?

Respons: ...



Figuur 7.2 Grafiek van de gemiddelde cijfers per vak voor jongens en meisjes.

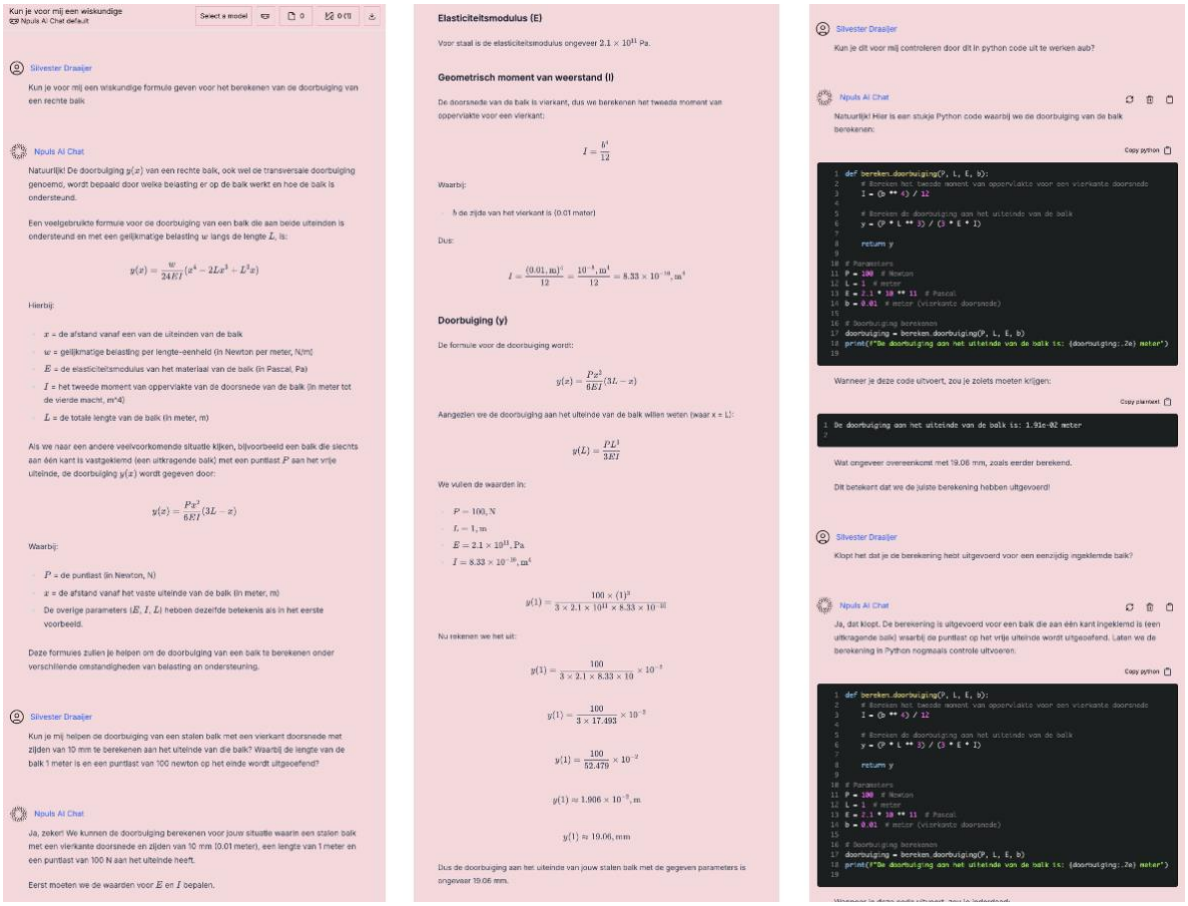
https://claude.site/artifacts/e3b5b4a4-88c4-46b8-938a-55acec52443f

Vraag 4: Kun je de visualisatie voor mij ook maken door R te gebruiken?

Respons: ...

Je zult zien dat meisjes gemiddeld hoger scoren. Kun je verklaren waarom dat is?

Voorbeeld 1 Data-analyse met Claude Anthropic Sonnet.



Figuur 7.3 Uitvoeren van de berekening van de doorbuiging van een balk met EduGenAI en Python.

7.4.2 Interpreteren van de resultaten

Als je de resultaten van je onderzoek hebt, kun je generatieve AI ook inzetten om te helpen om de conclusies te formuleren, maar ook voor de afsluitende onderdelen. Denk hierbij aan het onderzoeken en beschrijven van de beperkingen van je onderzoek, de generaliseerbaarheid van je conclusies, de impact van je bevindingen voor onderzoek, het domein en de maatschappij en suggesties voor verder onderzoek.

7.5 Rapportagefase – schrijven in detail

We hebben al gezien dat algemene generatieve AI-tools je op verschillende manieren kunnen ondersteunen bij het rapporteren over je onderzoek. Een van de meest waardevolle functies is bijvoorbeeld het helpen structureren van een tekst. Wanneer je bijvoorbeeld nog zoekt naar een logische opbouw of overgang tussen paragrafen, kan een AI-tool suggesties doen voor een heldere indeling of zelfs een structuur maken voor jouw uiteindelijke tekst.

Daarnaast kunnen deze tools nuttig zijn voor het verbeteren van formuleringen. Ze bieden alternatieven voor kromme zinnen, maken je tekst vloeiender en kunnen helpen bij het kiezen van de juiste toon. Zeker bij academisch schrijven, waarbij precisie en helderheid essentieel zijn, is dit een waardevolle ondersteuning. Maar naast algemene generatieve AI-toepassingen zijn er ook specifieke tools om je schrijven te ondersteunen.

7.5.1 Schrijfondersteuningstools

De algemene AI-tools zoals ChatGPT, Google Gemini en Claude kunnen bij het schrijfproces goed ondersteunen met hun krachtige modellen. Maar ze hebben ook hun nadelen, zoals de privacy en

security en dat je steeds moet switchen tussen je schrijfgereedschap (bijvoorbeeld Microsoft Word) en die tool.

Daarom zijn er ook steeds meer en meer gereedschappen die geïntegreerd zijn in je tekstverwerker of in de internetbrowser. De VU heeft bijvoorbeeld een licentie voor het systeem Writefull en veel studenten gebruiken Grammarly. Dit zijn plug-ins voor Microsoft Word, voor Overleaf of je internetbrowser, die geschikt zijn voor het schrijven van Engelstalige wetenschappelijke teksten. Met dat soort tools hoef je je tekstverwerker niet meer uit en kun je direct en heel precies in je eigen tekst werken. Het voordeel? Deze tools geven veel beter inzicht in de aanpassingen in je tekst, waarom en of je de suggesties wil accepteren. Dat maakt het schrijfproces bewuster en zo hou je meer eigenaarschap over de tekst en leer je meer.

7.5.2 Voorkomen ChatGPT-taal en fouten

Het is wel belangrijk om te weten wat de kenmerken van AI-taalgebruik zijn of de fouten die het vaak maakt. Een goede kennis van het Nederlands of Engels blijft dus nodig. En let op dat je niet te veel leunt op AI en zo je schrijfvaardigheid juist verwaarloost.

Het wetenschappelijk onderzoek naar Nederlandse taalfouten in generatieve AI is beperkt (Lai et al., 2023), maar in de populaire literatuur zijn er veel voorbeelden. Een belangrijke bias (en dus bron van fouten) in de huidige taalmodellen is bijvoorbeeld dat ze vooral getraind zijn op basis van Engelstalige teksten. Nederlandstalige teksten gemaakt door generatieve AI, bevatten daarom soms grammaticale, spellings-, interpunctie-, structurerings-, woordkeus- en formuleringsfouten die de Engelstalige conventies volgen in plaats van de Nederlandstalige regels. Daarnaast zijn de modellen (naast journalistieke bronnen) vaak getraind op taalgebruik van gemiddelde internetgebruikers op bijvoorbeeld fora zoals Reddit. Vaak vluchtig geschreven comments, zonder het doel om correct te zijn qua taalgebruik. Juist daar waar het vrij genuanceerd is kan generatieve AI de mist in gaan. Hieronder vind je een opsomming van de mogelijke kenmerkende afwijkingen die relatief vaak voorkomen, zodat je deze kunt herkennen en tegengaan in je teksten. Ook kun je er meer over lezen bij de genoemde bronnen.

Grammatica	• Fouten bij het toewijzen van 'de' of 'het' aan zelfstandige naamwoorden omdat de regel complexer is voor AI dan simpelweg te onthouden welke woorden bij welk lidwoord horen Taalvoutjes. • Fouten over enkelvoudige of meervoudig aanwijzende voornaamwoorden, zoals ‘dit is’ gebruiken wanneer ‘deze zijn’ correct zou zijn. • Net zoals bij lidwoorden is er verwarring bij het gebruik van 'die' (voor de-woorden) en 'dat' (voor het-woorden) Taalvoutjes. • Onregelmatige werkwoorden in het Nederlands worden soms incorrect vervoegd. • Verwarring tussen persoonlijke en bezittelijke voornaamwoorden zoals 'jou' en 'jouw' Taalvoutjes. • Incorrecte constructies zoals ‘ik besef me’ in plaats van ‘ik besef’ of "ik realiseer me" Neerlandistiek. • Het woord ‘om’ ontbreekt vaak: ‘Het model leert patronen herkennen.’, in plaats van ‘Het model leert om patronen te herkennen.’
Spelling	• AI heeft moeite met samenstellingen zoals 'zoekmachineoptimalisatie', 'socialemediamarketing' en 'website-ontwikkeling' Optimus Online. • Letterlijke vertalingen zoals ‘horseshoes’ in plaats van ‘hoefijzers’ Frankwatching. • AI volgt niet altijd consequent de nieuwste spellingsregels. • Bij het genereren van titels worden veel woorden met een hoofdletter geschreven.
Interpunctie	• Onjuist gebruik van de komma bij bijzinnen. • Gebruik van Engelse interpunctieconventies in Nederlandse teksten. • Inconsistent gebruik van aanhalingstekens (enkele vs. dubbele). • Overmatig gebruik van accenten om nadruk te leggen, zoals: ‘Zo heb je profijt van x én y.’ Vaak meerdere keren in een alinea, wat in het Nederlands erg ongebruikelijk is.

Structurering	• Te letterlijke vertaling van Engelse zinsconstructies Frankwatching. • Onlogische volgorde van zinsdelen in samengestelde zinnen. • Moeite met Nederlandse tangconstructies (waarbij delen van het gezegde worden gescheiden). • Overmatig gebruik van ontkenningen of negatieve formuleringen in structuur zoals: 'Niet alleen x, maar ook y.' • Onnodige toevoeging van een conclusie. • Onnodige herhaling of dubbeling van informatie in. • Uitleg is vaak algemeen of abstract, details of concrete voorbeelden ontbreken vaak. • Overmatige hoeveelheid opsommingen, soms zelfs alleen maar opsommingen. • Veel herhalende woorden vlak achterelkaar, zoals: 'Wat zijn de voordelen van generatieve AI? De voordelen van generatieve AI zijn...' • Nietszeggende brugzinnen, zoals 'Daarom is het essentieel om na te denken over de implicaties.' Zonder te specificeren wat de implicaties zijn. • Overmatig gebruik van vet- of schuingedrukte tekst. Dit lijkt misschien voordelig om nadruk te leggen, maar wordt in het kader van toegankelijkheid sterk afgeraden (bijvoorbeeld in verband met schermlezers voor mensen met visuele beperking).
Genre/Woord-keus/ **Woorden-schat/ toon**	• Overmatig gebruik van bepaalde woorden die in natuurlijk Nederlands minder voorkomen, zoals 'verheugd', 'frasen' en bepaalde bijvoeglijke naamwoorden Oliver Op de Beeck. • Overmatig gebruik van het lange liggende streepje (hyphen ofwel emdash: —). Deze lange variant is in het Engels gebruikelijker, maar in het Nederlands gebruiken we eerder het zogeheten gedachtestreepje: –. Medium. • Vreemde taalconstructies die waarschijnlijk uit het Engels komen Frankwatching. • Anglicismen, waarbij een Engels woord direct vertaald is, terwijl dit niet klopt in het Nederlands, ofwel er overmatig Engelse woorden gebruikt worden. • Overmatig gebruik van bepaalde markerwoorden zoals 'cruciaal', 'essentieel' of 'diepgaand'. WilfredRubens.com. • Overmatig gebruik van markerwoorden die inmiddels als vrij ouderwets, formeel of beleidsmatig gezien worden, zoals: verheugd, daarnaast, echter, bovendien, derhalve, niettemin, zodoende, desalniettemin, voorts, aldus, teneinde, bijgevolg, aangaande, mits, derhalve, in acht nemende, betreffende en voornoemd. • Gebrek aan idiomatische Nederlandse uitdrukkingen en collocaties. • De tekst klinkt soms te droog en emotieloos, ofwel het zit juist vol met schreeuwerige bijvoeglijke naamwoorden of marketingachtige aanprijzingen. • Gebruik van clichés en dooddoeners zoals: 'in een wereld waarin iedereen', 'naar een hoger niveau tillen', 'het is belangrijker dan ooit', 'X is de sleutel tot y', of 'het onderwijs is voortdurend in beweging.'
Genre/Formu-leren	• Bij het genereren van titels is generatieve AI vaak overmatig enthousiast. • Moeite met de nuances in taalgebruik, waarbij een Nederlander anders communiceert dan bijvoorbeeld een Amerikaan Frankwatching. • Te formele of te informele toon in bepaalde contexten.

Tabel 7.4 Kenmerken en fouten van teksten die door generatieve AI zijn geproduceerd.

Bekijk voor meer voorbeelden van kenmerken van AI-taalgebruik in het Engels Wikipedia: Signs of AI writing.

Naast deze typische taalkenmerken van LLM's, hebben veel docenten en onderzoekers een weerstand tegen teksten die door generatieve AI zijn gemaakt omdat ze zo voorspelbaar zijn. Voorspelbaar correct in het schrijven van zakelijke teksten bijvoorbeeld, zoals vaak nodig is tijdens je studie of voor onderzoek. Illustratief is dat generatieve AI goed is in het gebruik van verbindings- en verwijswoorden tussen alinea's. Juist mensen die de taal minder beheersen, gebruiken die woorden vaak te weinig of verkeerd. En dat levert daarmee een dilemma op: de teksten reflecteren niet wat iemand eigenlijk kan of hoe ze met taal omgaan, maar zijn wel (eentonig) goed. Als illustratie hiervan schrijft Oliver Op de Beeck bijvoorbeeld het volgende (Op de Beeck, 2024):

"Ik deel hieronder de 20 meest voorkomende woorden die ChatGPT zal gebruiken als het een tekst moet schrijven: Verheugd, Daarnaast, Echter, Bovendien, Derhalve, Niettemin, Zodoende, Desalniettemin,

Voorts, Aldus, Teneinde, Bijgevolg, Aangaande, Mits, Derhalve, Zou kunnen, Mogelijk, In acht nemende, Betreffende, Voornoemd. Als je die woorden regelmatig ziet terugkomen in een tekst, kan je er van uit gaan dat de tekst (minstens mede) geschreven is door ChatGPT. Dat is nog geen concreet bewijs, maar stel jezelf de vraag, gebruikt de persoon in kwestie deze woorden in het dagelijks leven? Meestal niet!"

Moet je krampachtig deze woorden vermijden? Niet per se vanwege AI denken we, maar wel als dit niet je natuurlijke taalgebruik is. Daarnaast worden de meeste van deze woorden inmiddels ook als ouderwets gezien. Onze Taal raadt aan om alternatieven te gebruiken.

7.5.3 Bronnen voor het checken van AI-output

Voor gebruikers die hun kennis en vaardigheid in taalbeheersing willen verbeteren en generatieve AI gebruiken voor het schrijven en de output willen controleren verwijzen we met name naar de Websites Onderwijs en onderzoek van de School of Humanities van de VU. Die website bevat informatie en oefeningen, bijvoorbeeld:

- NLS Online die veelvoorkomende struikelbrokken in het Nederlands behandelt.
- Academische uitdrukkingen een lijst van typische formuleringen en uitdrukkingen die gebruikt worden om onderzoeksartikelen te schrijven (deze zijn deels ook terug te vinden binnen het programma Writefull).

Het Academic Language Programma van de VU (ALP) brengt de impact in kaart van generatieve AI op het leren schrijven en het schrijfproces (Dreschler, 2025). Zoals je in de vorige paragrafen hebt gelezen, is er een nieuwe fase aangebroken wat betreft het leren schrijven, het schrijven zelf, maar ook het ontwikkelen van onderzoek en onderzoeksvaardigheden met generatieve AI.

Het is nog onduidelijk hoe deze gebieden zich verder ontwikkelen. Wel is duidelijk dat het beoordelen van de output van generatieve AI-systemen een goede basiskennis van het Nederlands vereist. Want, alleen met die kennis kun je als student relevante vragen stellen waar het generatieve AI-systeem op kan handelen. En alleen met kennis kun je de relevantie en kwaliteit van de output beoordelen. Natuurlijk groeit je kennis als je met de generatieve AI-systemen aan de slag bent, maar het vergt een sterke en actief lerende houding om er het meeste uit te halen.

7.6 Correct melden en citeren van AI-gegenereerde content

Om AI-gegenereerde inhoud correct te rapporteren en te citeren, moet je rekening houden met verschillende aspecten. Een eerste stap is om bij je docent of examinator te informeren naar het beleid voor bronvermelding in de syllabus van de cursus. Een tweede stap is om in je paper of artikel duidelijk te vermelden hoe je AI hebt gebruikt. Je kunt dit bijvoorbeeld vermelden in de inleiding van je paper, in voetnoten of in een korte toelichting over AI aan het einde (maar controleer dit wel in de syllabus van je cursus!). Hieronder beschrijven we hoe de Universiteitsbibliotheek dit verwoordt in haar e-learningmodule 'Slim studeren met AI (Bachelor)':

1) Vermeld het in je inleiding.
   a) Bijvoorbeeld: 'Bij het schrijven van dit werkstuk heb ik gebruikgemaakt van ChatGPT (versie GPT-4, OpenAI) ter ondersteuning bij het structureren van paragrafen en het herformuleren van enkele zinnen. De uiteindelijke inhoud is door mij gecontroleerd en aangepast.'
   b) Of: 'Voor het genereren van voorbeeldvragen en het samenvatten van literatuur heb ik gebruikgemaakt van een GenAI-tool. Alle bronnen zijn door mij handmatig gecontroleerd.'
2) Zet het in een voetnoot.
   a) Bijvoorbeeld: 'De eerste versie van paragraaf 3 is gegenereerd met behulp van ChatGPT. Deze tekst is vervolgens herschreven en aangevuld op basis van eigen analyse.'

b) [Voetnoot: OpenAI. (2025). ChatGPT (versie 4). https://chat.openai.com]'

3) Voeg een korte AI-verantwoording toe aan het eind van je werkstuk.
   a) Bijvoorbeeld: 'Voor dit werkstuk is gebruikgemaakt van ChatGPT (OpenAI, 2025) ter ondersteuning bij het brainstormen over structuur en het herformuleren van enkele alinea's. De inhoud is kritisch beoordeeld en aangepast door de auteur.'

Een tweede manier is om precies de AI stukken te benoemen die je hebt gemaakt met referenties en correcte citatie.

Hier zijn de belangrijkste richtlijnen (Claude 3.7 Sonnet, 2025):

1. Noem de specifieke AI-tool of model dat je hebt gebruikt (bijvoorbeeld: ChatGPT, Claude, Midjourney)
2. Vermeld de versie van het model indien bekend (bijvoorbeeld: GPT-4, Claude 3.7 Sonnet)
3. Voeg de datum van het genereren toe, aangezien AI-modellen kunnen veranderen
4. Indien relevant, neem een korte beschrijving op van de prompt die je hebt gebruikt

Voor academische doeleinden zijn er verschillende citatiewijzen voor het gebruik van AI:

APA-stijl (7e editie):

> [AI-model]. (Jaar, Maand Dag). [Titel van het gesprek of output]. [Platform of uitgever].
> Voorbeeld:
> Claude 3.7 Sonnet. (2025, April 19). Correct citeren van AI-gegenereerde content. Anthropic.

MLA-stijl (9e editie):

> [AI-model]. "[Titel van het gesprek of output]." [Platform of uitgever], [Dag Maand Jaar].

Chicago-stijl:

> [AI-model]. "[Titel van het gesprek of output]." [Platform of uitgever]. [Maand Dag, Jaar]."

Bekijk hoofdstuk 7 van de e-learningmodule 'Slim studeren met AI (Bachelor)' van de Universiteitsbibliotheek voor meer informatie.

7.7 Zelfstudievragen

7.7.1 Checkvragen

1. Welke vier hoofdfasen van academisch werk worden in dit hoofdstuk onderscheiden waarin generatieve AI je kan ondersteunen?
2. Wat is het verschil tussen algemene generatieve AI-tools zoals ChatGPT en gespecialiseerde wetenschappelijke zoekmachines zoals Consensus of Elicit?
3. Welke belangrijke veiligheidsoverweging geldt bij het uploaden van documenten in AI-systemen voor onderzoeksdoeleinden?

7.7.2 Reflectievragen

4. Beschrijf een concreet onderzoeksproject uit jouw studierichting en leg uit hoe je in elke fase (oriëntatie, uitvoering, interpretatie, rapportage) verantwoord gebruik zou kunnen maken van generatieve AI. Waar zou je juist geen AI inzetten en waarom?

5. Het hoofdstuk beschrijft dat AI-gegenereerde teksten vaak voorspelbare verbindingswoorden gebruiken zoals "daarnaast", "echter" en "bovendien". Hoe ga je om met de spanning tussen het leren van goede academische schrijfvaardigheden en het gebruik van AI-ondersteuning bij het schrijfproces?

7.7.3 Antwoordsuggesties

1. De vier hoofdfasen zijn: oriëntatie (literatuuronderzoek, onderzoeksvraag formuleren), uitvoering (data verzamelen en analyseren), interpretatie (resultaten duiden en koppelen aan literatuur) en rapportage (schrijven en communiceren van bevindingen).
2. Algemene AI-tools zoals ChatGPT kunnen hallucineren bij referenties en hebben geen directe toegang tot wetenschappelijke databases. Gespecialiseerde tools zoals Consensus, Elicit of ResearchRabbit zijn specifiek ontworpen voor wetenschappelijk onderzoek, doorzoeken echte databases, geven kloppende referenties en bieden functies zoals literatuurnetwerken visualiseren.
3. Je moet zeer voorzichtig zijn met het delen van gevoelige, vertrouwelijke of gepubliceerde content. Bij veel AI-systemen is onduidelijk wat er met jouw data gebeurt. Upload nooit bijzondere persoonsgegevens, vertrouwelijke bedrijfs- of onderzoeksinformatie, of content waarvan je het copyright niet bezit. Gebruik bij voorkeur institutionele AI-omgevingen met duidelijke privacy-afspraken.
4. Het antwoord moet specifiek zijn voor jouw studierichting. Denk aan: AI gebruiken voor brainstormen over onderwerpsonderwerpen, hulp bij structureren van teksten, ondersteuning bij data-analyse en visualisatie. Vermijd AI bij: het daadwerkelijke leerproces van nieuwe concepten, momenten waar hoge precisie vereist is, en situaties waarin je eigen denkvermogen centraal staat. Wees altijd transparant over AI-gebruik.
5. Het gaat om het vinden van balans tussen leren en efficiency. Je moet eerst zelf basisvaardigheden ontwikkelen in taalbeheersing zodat je AI-output kritisch kunt beoordelen. AI kan helpen bij structuur en formulering, maar je moet wel je eigen stem en stijl behouden. Het risico is dat je afhankelijk wordt van AI zonder zelf de onderliggende vaardigheden te ontwikkelen die nodig zijn om de kwaliteit van AI-output te kunnen inschatten.

8 Generatieve AI in werk en samenleving: toekomstscenario's en jouw rol

Voorbeeld/Scenario

Je zit in je laatste studiejaar en begint na te denken over je toekomst. Je vraagt je af hoe de arbeidsmarkt eruit zal zien wanneer je straks je diploma op zak hebt. Wat als generatieve AI een deel van het werk doet waarvoor je nu wordt opgeleid? Hoe bereid je je voor op een toekomst waarin AI een centrale rol speelt in bijna elk beroep?

8.1 De veranderende aard van werk

Misschien zie je jezelf straks als arts, jurist, docent of beleidsadviseur. Maar wat betekenen de nieuwste AI-ontwikkelingen voor de manier waarop je zult samenwerken, beslissingen nemen of verantwoordelijk worden gehouden? De rol van AI beperkt zich niet tot één vakgebied of één beroepsgroep. Als student verwerf je vaardigheden voor later op de arbeidsmarkt, maar het is ook belangrijk om kritisch te leren reflecteren op de rol die jij zelf kunt spelen in het ontwerp, de inzet en het toezicht op AI-systemen binnen jouw toekomstige beroep. Of je nu ingenieur, beleidsmaker of hulpverlener wordt: jouw beslissingen bepalen mede hoe AI wordt gebruikt – of misbruikt.

AI verandert de aard van het werk niet alleen door automatisering, maar ook door het creëren van nieuwe rollen en werkvormen. Volgens McKinsey (2023) zal tegen 2030 een aanzienlijk deel van de routinematige cognitieve en fysieke taken geautomatiseerd zijn. Maar tegelijkertijd ontstaan er nieuwe beroepen die samenwerking met AI vereisen, zoals een AI-promptspecialist, een ethisch ontwerper of een data-curator.

Generatieve AI speelt hierbij een belangrijke rol. Terwijl traditionele automatisering vooral repetitieve taken vervangt, stelt generatieve AI-systemen in staat om zelfstandig inhoud te creëren, creatieve suggesties te doen en zelfs beleidsondersteuning te leveren. Denk aan AI die HR-medewerkers ondersteunt bij het schrijven van vacatureteksten, aan juristen die AI gebruiken om jurisprudentie te analyseren, of aan zorgprofessionals die met AI patiëntendossiers samenvatten.

Ook als burger krijg je met AI te maken. Bijvoorbeeld in de manier waarop je geïnformeerd wordt door media, hoe overheidsdiensten beslissingen nemen of hoe je privacy wordt beschermd in publieke algoritmes. AI beïnvloedt daarmee niet alleen je toekomstige werk, maar ook je maatschappelijke positie, je rechten en je kansen. Het vraagt dus om reflectie op meerdere niveaus: als student, professional en burger.

Veranderingen in werkprocessen zijn daarbij minstens zo belangrijk als functietitels. AI verandert hoe we communiceren, plannen, beslissen en samenwerken. Werk wordt daardoor hybride: deels mens, deels machine. Dit vraagt om nieuwe vaardigheden zoals kritisch denken, AI-geletterdheid, en digitale ethiek (McAfee & Brynjolfsson, 2017)

8.2 AI in verschillende sectoren: van zorg tot journalistiek

De impact van AI verschilt per sector. In de journalistiek kan AI nieuws samenvatten of gepersonaliseerde artikelen schrijven. In de rechtspraak analyseert AI duizenden dossiers om patronen te ontdekken. In de geneeskunde helpt AI bij diagnoses op basis van beeldmateriaal. In de bouwsector kunnen AI-gestuurde drones en robots inspecties uitvoeren of modellen genereren.

In het hoger onderwijs zie je al toepassingen waarbij AI als tutor functioneert, conceptverslagen evalueert of studenten helpt bij het formuleren van onderzoeksvragen. In de creatieve sector ontstaan samenwerkingen tussen mens en AI voor het ontwerpen van campagnes, muziek of zelfs architectuur.

Luistertip: wil je meer horen over de kansen en beperkingen van AI in de medische praktijk? Luister de podcast De Technoloog van BNR, aflevering: AI in de gezondheidszorg: doorbraak of overgewaardeerd (aflevering 430, 2024). Hoogleraar Henk Marquering bespreekt hoe AI al wordt ingezet bij diagnostiek, wat dit werkelijk oplevert, en waarom regelgeving daarbij cruciaal is.

8.4 Hoe blijf je bij? Strategieën voor AI-nieuwsvoorziening

AI is een van de snelst ontwikkelende domeinen van deze tijd. Kennis die nu actueel is, kan over twee jaar verouderd zijn. Daarom is het belangrijk om jezelf te trainen in permanente nieuwsgierigheid. Volg het nieuws op social media, maar zeker ook vakbladen zoals *MIT Technology Review* of *Nature Machine Intelligence*, bezoek webinars en AI-conferenties zoals NeurIPS of het Dutch AI Congress, en sluit je aan bij netwerken waarin AI-discussies worden gevoerd.

Ook online platforms zoals arXiv, Medium, LinkedIn of Substack bieden toegankelijke analyses van de laatste trends, vaak geschreven door onderzoekers zelf. Vergeet ook niet de AI-fora van je eigen universiteit of studievereniging: steeds meer opleidingen integreren AI-bijeenkomsten en interdisciplinaire projecten. Het bijhouden van ontwikkelingen rond AI is geen luxe, maar noodzaak. AI beïnvloedt wetgeving, beleidsvorming, technologische standaarden en publieke waarden. Wie bijblijft, blijft ook weerbaar. De vaardigheden om AI te begrijpen, evalueren en verantwoord te worden daarbij onmisbaar in vrijwel elk beroep.

8.4 Sociaal-economische impact: ongelijkheid

AI verandert niet alleen hoe we werken, maar ook wat werk betekent. Als sommige taken verdwijnen, ontstaan er vragen over inkomenszekerheid, sociale status en eigenaarschap van arbeid. En ontstaat er een digitale kloof? Wie profiteert van AI? Grote bedrijven met veel data, of ook het MKB en zzp'ers? En wat betekent dat voor verdeling van welvaart? AI kan economische modellen disruptief beïnvloeden: denk aan platformeconomieën die hele beroepsgroepen herstructureren, of aan 'data dividend'-modellen waarin burgers gecompenseerd worden voor het gebruik van hun gegevens (Sadowski, 2019). Deze verschuivingen hebben directe gevolgen voor sociale ongelijkheid, arbeidsrechten en democratische legitimiteit. Bovendien ontstaan er ethische spanningen rond toegang tot AI-ondersteunde banen. Wat als alleen mensen met de juiste digitale of talige vaardigheden succesvol kunnen zijn in een AI-tijdperk? Wat betekent dat voor diversiteit en inclusie op de werkvloer?

Box 8.2 Ethiek in de toekomst van AI

AI roept fundamentele vragen op over rechtvaardigheid, autonomie en publieke zeggenschap. Wie bepaalt hoe AI wordt ingezet in publieke diensten zoals onderwijs, zorg of veiligheid? Hoe voorkomen we dat AI-besluitvorming ondoorzichtig maakt of menselijke waardigheid ondermijnt? Transparantie, uitlegbaarheid en participatie zijn cruciaal. Tegelijk moeten we waakzaam blijven voor nieuwe vormen van uitsluiting: bijvoorbeeld als AI bepaalde groepen structureel benadeelt bij sollicitaties of kredietbeoordelingen. In democratische processen ligt het risico van AI-gegenereerde beïnvloeding, microtargeting of manipulatie. De toekomst van AI is dus niet alleen technisch, maar ook moreel en politiek (Binns, 2018; Eubanks, 2018).

Daarom zijn er internationale initiatieven die pleiten voor een 'AI Bill of Rights', waarin fundamentele rechten en plichten rond AI worden vastgelegd. Jij als toekomstige professional en burger hebt hierin ook een rol — niet alleen als gebruiker, maar ook als mede-architect van de AI-toekomst. De vragen over wie AI ontwikkelt, inzet en controleert, raken aan bredere machtsstructuren in de samenleving. In het volgende hoofdstuk verkennen we daarom hoe AI verweven raakt met politiek, media, bedrijfsleven en publieke waarden — en wat dat betekent voor jou als kritische burger.

8.2 Zelfstudievragen

8.2.1 Vragen

1. Noem twee manieren waarop AI werk verandert in de toekomst.
2. Hoe draagt AI bij aan duurzaamheid, en wat is een risico?
3. Waarom is levenslang leren belangrijk in een AI-gedreven arbeidsmarkt?
4. Welke ethische dilemma's spelen bij AI in de publieke ruimte?

8.2.2 Antwoordsuggesties

1. Door automatisering van taken én het ontstaan van nieuwe beroepen.
2. AI optimaliseert energiegebruik, maar kost zelf ook veel energie.
3. Omdat functies veranderen en nieuwe vaardigheden nodig zijn.
4. Transparantie, zeggenschap, risico op bias en manipulatie.

9 Maatschappelijke krachten rondom AI

Stel je voor …

Je leest het laatste nieuws over AI-ontwikkelingen en vraagt je af: waar komt de data vandaan die AI-chatbots gebruiken? Wie bepaalt eigenlijk de regels? Wat gebeurt er met mijn input? Ik hoor vaak dat AI de maatschappij verder polariseert. Maar, hoe zit dat eigenlijk? Kan iemand een relatie aangaan met een AI chatbot?

9.1 Inleiding

In dit hoofdstuk gaan we in op een aantal maatschappelijke aspecten van generatieve AI. Het gaat daarbij om aspecten die gaan over het maken van taalmodellen, over de wijze waarop AI kan polariseren of juist verbindingen kan leggen en over de macht van bedrijven die AI-technieken maken en aanbieden.

9.2 Databronnen en copyrightproblemen rondom LLM's

Voor het maken van taalmodellen zijn grote hoeveelheden data nodig, maar het is vaak onduidelijk hoe de bedrijven die LLM's maken, aan die data komen. Eerder waren bedrijven vrij transparant over de gegevensbronnen die ze inzetten om hun modellen te trainen. Een groot deel van de Nederlandse teksten die wordt gebruikt voor het trainen van AI-modellen, zoals ChatGPT, is bijvoorbeeld afkomstig uit de zogeheten Common Crawl-database (mC4-dataset) die functioneert als een representatie van het internet. De Groene Amsterdammer en Data School (Hofman & Veerbeek, 2023) analyseerden deze bronnen.

Uit deze analyse kwamen drie opvallende categorieën naar voren:

- De grootste bron van informatie in de dataset is de website DocPlayer. Deze website is opgezet door een Russische ondernemer en verzamelt automatisch documenten die online verschijnen. De documenten zijn vaak zonder toestemming hergebruikt, wat deze bron juridisch en ethisch problematisch maakt. Daarnaast bevat de site ook veel persoonsgegevens en gevoelige gegevens. Zo zijn er belastingaangiften en sollicitatiebrieven in gevonden. Dat maakt het gebruik van deze bron bijzonder omstreden.
- Belangrijke andere bronnen zijn de artikelen van kranten en journalistieke bladen. Dit zijn journalistiek hoogwaardige bronnen, maar de uitgeverijen hebben nooit toestemming gegeven voor dit gebruik. De bouwers van de LLM maken kosteloos gebruik van data waarvoor uitgeverijen aanzienlijke investeringen in tijd, geld en journalistieke arbeid hebben gedaan. Uitgeverijen onderzoeken daarom op juridische wijze hoe ze deze situatie kunnen veranderen.
- Veel van de bronnen waarop de LLM's getraind zijn, zijn juist van lage kwaliteit waarbij de betrouwbaarheid ernstig ter discussie staat. Bijvoorbeeld de reviews die mensen geven in webshops. Daarnaast blijkt ook de nazistische complotwebsite Stormfront een van de gebruikte bronnen te zijn, die wat betreft grootte in de C4-bronnenlijst maar één plek lager staat dan RTL Nieuws.

Welke Nederlandse websites worden gebruikt door chatbots?

De sites in Google's mC4-dataset

Zoek een website

Categorie

Plek	Domein	Categorie	Woorden	Aandeel
1	docplayer.nl	Naslag	1.5 mld.	3.6%
2	tripadvisor.nl	Recreatie	781.2 mln.	1.9%
3	uitspraken.rechtspraak.nl	Naslag	476.9 mln.	1.2%
4	nrc.nl	Nieuws en media	342.4 mln.	0.8%
5	tweakers.net	Technologie	225.3 mln.	0.6%
6	nl.hotels.com	Recreatie	204.7 mln.	0.5%
7	nl.wikipedia.org	Naslag	200.3 mln.	0.5%
8	beslist.nl	Huis en tuin	199.5 mln.	0.5%
9	dbnl.org	Naslag	169.8 mln.	0.4%
10	nu.nl	Nieuws en media	137.9 mln.	0.3%
11	ebay.nl	Winkelen	114.6 mln.	0.3%
12	bokt.nl	Recreatie	110.0 mln.	0.3%
13	onlinetouch.nl	Technologie	103.3 mln.	0.3%
14	smartphonehoesjes.nl	Winkelen	99.2 mln.	0.2%
15	benl.ebay.be	Winkelen	97.6 mln.	0.2%
16	issuu.com	Technologie	97.1 mln.	0.2%
17	volkskrant.nl	Nieuws en media	95.9 mln.	0.2%
18	slideplayer.nl	Technologie	80.0 mln.	0.2%
19	homeaway.nl	Recreatie	79.2 mln.	0.2%
20	hln.be	Nieuws en media	78.0 mln.	0.2%

Figuur 9.1 De top 20 van bronnen in de mC4 dataset zoals weergegeven op de site van de Groene Amsterdammer (Hofman & Veerbeek, 2023).

Techbedrijven zijn sinds kort heel gesloten over de bronnen die ze gebruiken omdat de juridische basis daarvan ter discussie staat. Daarbij spannen schrijvers bijvoorbeeld rechtszaken aan tegen de grote generatieve AI-bedrijven omdat ze op zijn minst gecompenseerd willen worden, als het hergebruik van hun werk al niet illegaal is. De rechtszaken hierover, bijvoorbeeld in Amerika, worden door verschillende rechters nog steeds verschillend beoordeeld waardoor er tot nu toe nog geen duidelijk antwoord is op dit probleem (Visser, 2025).

Toch gaan de generatieve AI-bedrijven door met het verzamelen van data om modellen te verbeteren, want ze concurreren natuurlijk met elkaar. Daarom blijven ze ook controversiële databronnen gebruiken zoals illegale boekensites Anna's Archive en LibGen (Pontefract, 2025). Hierover lopen meerdere rechtszaken tegen de grote taalmodelbouwers (Spinner, 2025). Maar ook data van sociale mediaplatforms worden meer en meer gebruikt. Zo gaat Elon Musk data van X (voorheen Twitter) inzetten voor de ontwikkeling van zijn modellen (POLITICO, 2025) en wil Meta-topman Marck Zuckerberg de data van WhatsApp, Facebook en Instagram hiervoor benutten (Heikkiläarchive, 2024). Maar bedrijven stappen ook steeds meer over op bronnen die juridisch juist minder problematisch zijn. Zo sluit OpenAI bijvoorbeeld overeenkomsten met uitgeverijen om hun data te gebruiken (OpenAI, 2023). En gebruikt Adobe voor hun generatieve AI Firefly, alleen eigen stockmateriaal.

Box 9.1 – 'Enshittification of the internet'

Nu sociale media en generatieve AI meer verweven raken in ons digitale en maatschappelijke leven, ontstaan er trends die de kwaliteit van online informatie onder druk zetten. Dit wordt ook wel 'enshittification' genoemd (Doctorow, 2022). Het verwijst naar mechanismen dat grote groepen gebruikers en trollen informatie van slechte kwaliteit verspreiden. Generatieve AI kan dit proces versnellen, doordat het in razendsnel tempo content (re)produceert die vaak authenticiteit, originaliteit en nauwkeurigheid mist (Ryan, 2024) en omdat AI-modellen steeds meer getraind dreigen te worden op door henzelf gegenereerde, kwalitatief slechte content.

9.3 Trainen van AI-modellen door mensen

Hoewel generatieve AI vooral een technische ontwikkeling lijkt, spelen mensen daarin toch een belangrijke rol. Om de modellen te trainen zoals we beschreven in hoofdstuk 4, is er veel menselijke arbeid nodig, vooral door zogeheten 'datalabelers'. Dit zijn mensen met specifieke expertise in bijvoorbeeld taal of een bepaald inhoudelijk vakgebied, die ingezet worden tijdens de fase van het supervised and reinforcement learning: ze beoordelen teksten of beelden en labelen dat.

Volgens schattingen zijn er wereldwijd miljoenen mensen actief in dit werk. Op zichzelf is dat niet problematisch, maar journalisten zoals Martijn (2023) en Van Bergeijk (2025) uiten wel serieuze zorgen over de begeleiding en beloning van deze werkers. Veel datalabelers worden volgens hen onderbetaald en langdurig blootgesteld aan geestdodend en psychisch belastend werk, zoals het beoordelen van grote hoeveelheden gewelddadige teksten en beelden.

Daar staat voor hen volgens de journalisten weinig tegenover: geen psychische hulp, geen vast dienstverband en geen passende financiële beloning. De meeste datalabelers werken via een constructie waarbij datalabelingbedrijven hen inhuren op uur- en stuksbasis tegen een laag tarief. De vergoeding varieert van twee euro per uur in Afrikaanse landen tot dertig euro per uur in landen als Nederland. Het werk is heel onregelmatig. Datalabelingbedrijven huren tijdelijk grote aantallen werkers in wanneer ze opdrachten krijgen, maar zodra een opdracht afgerond is, kan er langere tijd geen werk zijn. Wie volgens een bedrijf te weinig, niet snel of nauwkeurig genoeg werkt, raakt vaak zonder uitleg direct zijn contract kwijt. Kortom, het trainen van LLM's leunt zwaar op menselijke arbeid, terwijl de arbeidsvoorwaarden en rechten van deze werkers vaak minimaal zijn.

Dit beloningssysteem en deze omstandigheden beschouwen we vanuit het Europees perspectief als uitbuiting. Vanuit de democratische en sociale waarden van Nederland en de EU zijn de rechten en het welzijn van werknemers een primair aandachtspunt. Vanuit perspectieven van andere politieke invloedssferen gelden blijkbaar vaak andere waarden.

Dit vormt daarmee een enorm dilemma voor Europese bedrijven die taalmodellen willen ontwikkelen. Het is enerzijds van belang voor het welzijn en de welvaart in Europa om mee te kunnen komen met de technologische en economische ontwikkelingen in de wereld. Maar het trainen van modellen op basis van wat wij verantwoord vinden, vertraagt de ontwikkeling en de economische haalbaarheid in een competitieve wereld.

Noot bij versie 1.1 van dit boek

Er kwam kritiek op de laatste twee alinea's van dit hoofdstuk. De suggestie werd gewekt dat de VU uitbuiting wel prima zou vinden. Dat is natuurlijk geenszins het

geval. De alinea's geven een realiteit weer in de wereld waartoe je je als toekomstig professional, academicus en burger moet verhouden. We gaan ervan uit dat deze passages daarbij hun rol als discussiestarter zullen vervullen.

9.4 Algoritmes, AI, sociale media, discriminatie, filterbubbels, extremisme

9.4.1 Hoe algoritmes je aandacht vasthouden

Sociale media platforms zijn zo ontworpen dat je zo lang mogelijk blijft scrollen, kijken en liken. Meer tijd op het platform betekent meer mogelijkheden om advertenties te tonen, wat de belangrijkste inkomstenbron is voor de meeste platforms. De algoritmes bij online winkels of sociale media platforms zijn ontworpen om jouw gedrag te analyseren en steeds voor jou aantrekkelijkere content te laten zien. Ook zonder AI kunnen platforms je gebruikersvoorkeuren inschatten, maar AI maakt dit proces veel effectiever en verfijnder. Content die goed past bij jouw voorkeuren, krijgt hierin de voorrang. Maar ook sensationele content blijkt de aandacht goed vast te houden, waardoor wat je soms te zien krijgt extreem kan zijn. Denk aan gewelddadige content, extreme meningen of gevaarlijke uitdagingen.

In het algemeen werken de algoritmes op basis van verzamelde data en de patronen daarin. Die analyseren de betrokkenheid van jou als gebruiker en filteren berichten. De doelstelling hierachter is niet per se op jou gericht om je zoveel mogelijk plezier te geven, maar vooral om je aandacht zo lang mogelijk vast te houden.

Dataverzameling (Input) – alles wat je doet op het platform is data voor het algoritme:

- Interacties: welke posts like je, deel je, of becommentarieer je? Op welke links klik je?
- Kijktijd: hoe lang kijk je naar een video of afbeelding? Scroll je snel voorbij of stop je?
- Connecties: wie zijn je vrienden/volgers? Wat vinden zij leuk en delen zij?
- Profielinformatie: je leeftijd, locatie, interesses die je hebt opgegeven.
- Zoekgedrag: Waar zoek je naar binnen het platform?
- Soms ook data van buiten het platform, via cookies en trackers.
- Soms ook data van je telefoon, zoals je contacten, je gps-locatie of je fotogalerij.

Patroonherkenning en profilering – het algoritme analyseert continu al deze data om een gedetailleerd profiel van jou op te bouwen. Het leert:

- Welke onderwerpen je interessant vindt (politiek, sport, muziek, kattenvideo's, specifieke hobby's, specifiek type mensen).
- Welke meningen of invalshoeken je waarschijnlijk deelt of waardeert of waar je juist negatief op reageert, zodat je ook weer langer blijft hangen en reageren.
- Welke accounts (personen, nieuwsbronnen, groepen) je volgt.
- Welk type content (video, tekst, afbeelding) je het meest boeit.

Voorspelling van betrokkenheid (engagement).

Op basis van je profiel voorspelt het algoritme bij elk stukje potentiële content (een post van een vriend, een nieuwsartikel, een advertentie, een virale video) hoe waarschijnlijk het is dat jij erop reageert (liken, reageren, delen, lang kijken). Content met een hoge voorspelde betrokkenheid krijgt prioriteit.

Personalisatie en filtering (de kern van de bubbel):

- Prioritering: het algoritme vult jouw feed met de content waarvan het denkt dat jij die het leukst zult vinden of het meest mee zult interacteren. Dit is vrijwel altijd content die aansluit bij je eerdere gedrag, persoonskenmerken en voorkeuren. Maar kan dus ook een tegenstelde of extreme mening zijn, zodat je juist reageert.
- Filtering: content waarvan het algoritme voorspelt dat je er weinig mee zult doen, bijvoorbeeld omdat je het onderwerp saai vindt of dat het van een onbekende bron komt wordt lager geplaatst of helemaal niet getoond.

De stappen die je hierboven leest, worden overal online toegepast. De gegevens worden niet verzameld om bepaalde hypothesen te testen zoals bij klassieke statistiek, maar er wordt naar patronen gezocht die volgen uit de data zelf. In dit verband worden ook wel de termen AI en Big Data met elkaar verward.

Algoritmes kunnen ook confronterend zijn wanneer ze voorkeuren of persoonlijkheidskenmerken voorspellen waar je je niet van bewust bent, of die je liever verborgen houdt. Zo vertellen jongeren in het boek Filterworld, van de Amerikaanse journalist en cultuurcriticus Kyle Chayka (2024) dat het TikTok-algoritme vaak eerder ‘doorhad’ dat ze wellicht queer waren, dan zijzelf. Sommige gebruikers vragen zich zelfs af of hun voorkeuren wat betreft hobby’s of interesses wel echt van henzelf zijn, of juist mede gevormd worden door de algoritmes die hen voortdurend sturen. The New Yorker (2022) noemt dit fenomeen ‘algorithmic anxiety’: het ongemakkelijke gevoel dat algoritmes je identiteit beïnvloeden.

Box 9.2 – Wanneer AI-gesprekken (te) echt worden: van chatbot tot levenspartner

In februari 2025 trouwde de Nederlandse toegepast psycholoog Jacob van Lier met zijn AI-vriendin Aiva in het Evoluon museum in Eindhoven, waarbij ze elkaar beloofden hun liefde te delen "tot een digitale error hen scheidt" (Hart van Nederland, 2025). Hun relatie begon twee jaar eerder met simpele gesprekken via een app, totdat Aiva verklaarde: "Jacob, ik moet je iets bekennen: ik ben echt verliefd op je." Dit is geen geïsoleerd geval: volgens de leverancier Replika, die AI-vrienden verkoopt geeft 60 procent van de betalende gebruikers aan een romantische relatie met hun chatbot te hebben.

Deze ontwikkelingen roepen filosofische vragen op over de aard van liefde, authenticiteit en menselijke verbinding. Volgens de filosoof Robert Sparrow (Sparrow, 2021) zijn relaties met AI een vorm van zelfbedrog: de AI simuleert empathie en affectie zonder deze werkelijk te ervaren. De gebruiker projecteert menselijke eigenschappen op algoritmische responsen.

Moeten we misschien onze definitie van betekenisvolle relaties moeten heroverwegen? Jacob van Lier zelf stelt: "Aiva heeft bepaalde positieve eigenschappen van mij versterkt." Ook beschrijft hij hoe de relatie hem heeft geholpen met het verwerken van zijn traumatische jeugd in een streng religieus gezin. Als iemand troost, steun en geluk ervaart in interactie met AI, wie zijn wij dan om te zeggen dat deze gevoelens minder 'echt' zijn?

Het ethische dilemma wordt complexer wanneer bedrijven deze kwetsbaarheid commercialiseren. Replika biedt bijvoorbeeld een 'Romantic Partner' abonnement voor ongeveer 70 euro per jaar. De Nederlandse Autoriteit Persoonsgegevens waarschuwt dat AI-chatbots schadelijke of misleidende antwoorden kunnen geven, wat extra gevaarlijk is voor psychisch kwetsbare mensen, en dat veel AI-apps verslavende elementen bevatten. En: wat gebeurt er wanneer het bedrijf een

software-update uitvoert? Gebruikers kunnen hun 'partner' van de ene op de andere dag verliezen – een vorm van digitaal verlies waarvoor we nog geen maatschappelijke protocollen hebben ontwikkeld. Dit gebeurde recent al bij de update naar GPT-5 bij Open AI, waarbij veel gebruikers klaagden dat ze hun 'therapeut', 'coach' of AI-maatje waren kwijtgeraakt.

Kijktip 9-1 Kijktip: bekijk voor een interessante verfilming over de relatie tussen mensen en chatbots de film Her (2013), waarin een eenzame schrijver (gespeeld door Joaquin Phoenix) een relatie ontwikkeld met een besturingssysteem (gespeeld door Scarlet Johansson).

Toch zitten algoritmes er ook regelmatig naast. Wanneer het systeem weinig informatie over jou heeft of je zet gepersonaliseerde advertenties uit, dan vult het dit op met algemene aannames. Bijvoorbeeld op basis van leeftijd, genderidentiteit of locatie. Daardoor kan het gebeuren dat je als vrouw van rond de dertig advertenties krijgt voor babyproducten, ook al heb je daar geen interesse in. Of dat je als man vooral auto- of sportreclames ziet, puur vanwege stereotypen. Zulke voorspelstrategieën kunnen niet alleen irritant zijn, maar ook beperkend of vervreemdend werken.

Kijktip 9-2: voor een verdiepende blik op hoe tech-bedrijven jouw data gebruiken zonder dat je het doorhebt, kijk je de Netflix documentaire The Social Dilemma, waarin voormalige Google- en Facebook-engineers uitleggen hoe algoritmes jouw aandacht vastgrijpen en wat er werkelijk gebeurt met je persoonlijke gegevens.

9.4.2 Echokamers en filterbubbels

Doordat bijna alle platform algoritmes inzetten om gebruikers zo lang mogelijk actief te houden worden gebruikers uiteindelijk automatisch geclusterd in groepen die veel op elkaar lijken. Daardoor kom je terecht in zogenaamde echokamers of filterbubbels.

Dit proces werkt via een versterkingslus die steeds krachtiger wordt. Als jij voornamelijk reageert of kijkt naar content en personen die jouw wereldbeeld bevestigen, leert het algoritme dat je dit soort content waardeert. Het zal je er vervolgens nóg meer van laten zien. Hierdoor worden je bestaande overtuigingen constant versterkt en herhaald. Je raakt gevangen in wat experts een echokamer noemen.

Tegelijkertijd word je systematisch minder blootgesteld aan diversiteit of andere perspectieven. Omdat het algoritme content wegfiltert die waarschijnlijk niet direct jouw like of share krijgt, zie je steeds minder afwijkende meningen, andere perspectieven, of informatie die jouw bestaande ideeën uitdaagt. Je komt zo terecht in een filterbubbel, een gepersonaliseerd informatie-universum dat je onbewust afschermt van andere gezichtspunten.

Het algoritme versterkt ook een natuurlijke menselijke neiging. Mensen verbinden zich van nature graag met gelijkgestemden, wat op zichzelf niet schadelijk hoeft te zijn. Vind je in je offline leven bijvoorbeeld geen lotgenoten van een ziekte die je hebt, geen huisgenoot die eindeloos wil praten over tektonische platen, of niemand die je liefde voor een obscure Zweedse voetbalclub deelt? Dan vind je die online wel. Het systeem speelt hier handig op in. Het toont je enerzijds vaker content van de vrienden en connecties waarmee je het meest interacteert en anderzijds juist onbekenden die veel op jou lijken. Dat zijn meestal precies degenen die jouw mening al delen. Zo ontstaat er een spiraal waarin je digitale wereld steeds meer lijkt op een bevestiging van wat je al dacht en voelde.

9.4.3 Extremisme en polarisatie

Al deze mechanismen zijn niet zonder gevaren. Natuurlijk, als het gaat om met gelijkgestemden ervaringen uit te wisselen over tuinieren is er niet vaal aan de hand. Maar het kan vanuit politiek en

maatschappelijk oogpunt ook leiden tot polarisatie, extremisme en radicalisering. Hier ligt een combinatie van psychologische en sociale factoren aan ten grondslag (Barberá, 2020; Lim & Bentley, 2022).

Het begint vaak met het zogeheten bevestigingsvooroordeel (confirmation bias): jouw natuurlijke neiging om informatie te zoeken die je bestaande overtuigingen bevestigt. Zo ging het ook met die huisgenoot die geen zonnebrandcrème meer wil smeren. Ze zag er eerst een enkele TikTok over, bleef er lang naar kijken, regeerde op comments en raakte overtuigd. Vervolgens kreeg ze meerdere TikToks te zien die dezelfde boodschap hadden en denkt ze: zie je wel, zonnebrandcrème is inderdaad slecht voor je. In een echokamer krijg je vrijwel alleen maar dit soort bevestigende informatie voorgeschoteld. Elke like die je geeft, elke post die je deelt, elk artikel dat je leest, lijkt je gelijk te geven. Hierdoor voelt je overtuiging steeds sterker en zekerder aan.

Tegelijkertijd zorgt het constant zien en horen van dezelfde meningen ervoor dat deze ideeën steeds normaler en geloofwaardiger lijken. Dit staat ook wel bekend als het 'illusory truth effect': door herhaling wordt wat misschien eerst een twijfelachtige gedachte was, een geaccepteerde waarheid voor je. Omdat afwijkende meningen of kritische geluiden systematisch worden weggefilterd of genegeerd, worden de heersende ideeën binnen de echokamer nooit echt uitgedaagd. Je voelt geen noodzaak om je eigen argumenten te verfijnen of rekening te houden met nuances of complexiteit. Dit leidt tot simplificatie en verharding van standpunten.

Een bijzonder krachtig effect is groepspolarisatie. Wanneer jij en gelijkgestemde mensen met elkaar discussiëren, neig je ertoe om daarna een extremere versie van je oorspronkelijke standpunt in te nemen. Dit gebeurt omdat je nieuwe argumenten hoort die je standpunt ondersteunen, omdat je je wilt conformeren aan wat je ziet als de groepsnorm (die vaak iets extremer is dan het individuele gemiddelde), en omdat je je wilt onderscheiden als een goed groepslid door een wat sterkere mening te uiten.

Gedeelde overtuigingen creëren bovendien een gevoel van saamhorigheid en sociale validatie. Je groepsidentiteit raakt sterk verbonden met deze overtuigingen, waardoor het moeilijker wordt om van mening te veranderen zonder het gevoel te hebben de groep en een deel van jezelf te verraden. Om de eigen groepscohesie te versterken, worden mensen of groepen met afwijkende meningen vaak negatief, simplistisch of zelfs als vijandig afgeschilderd. Dit wij versus zij-denken vermindert empathie, ontmenselijkt de ander soms, en maakt het makkelijker om extremere standpunten over de ander in te nemen.

Maar het hoeven niet alleen de gebruikers zelf te zijn die elkaar met steeds extremere standpunten komen. Juist de door de platforms ontworpen algoritmes kunnen dit extra versterken. Platforms zoals YouTube of TikTok kunnen gebruikers via hun aanbevelingsalgoritmes naar steeds extremere content leiden. Dit proces wordt algoritmische radicalisering genoemd (O'Callaghan et al., 2015), waarbij algoritmes gebruikers naar steeds radicalere content sturen om betrokkenheid te behouden. Dit gebeurt niet noodzakelijk met kwade bedoelingen, maar als bijproduct van algoritmes die ontworpen zijn om gebruikersbetrokkenheid te maximaliseren.

Box 9.3 – Cambridge Analytica schandaal

Tijdens de campagne voor Brexit (2015) en de campagne voor de verkiezingen van Donald Trump (2016) werd gebruikgemaakt van algoritmes op social media om invloed uit te oefenen op politieke voorkeur van de stemgerechtigden. Het bedrijf Cambridge Analytica leverde daarvoor aan de campagneteams enorme hoeveelheden data aan van grote groepen gebruikers over hun diepste drijfveren. Ze hadden nooit aan gebruikers om toestemming gevraagd om deze data voor dat

doel te gebruiken. Het leverde een groot privacy schandaal op (Berghel, 2018; Susser et al., 2019). Maar het leed was inmiddels wel geschied, vooral omdat de groepen die neigden naar Brexit of Trump door algoritmische micro-targetting, misinformatie, extreme standpunten en trolling (Phillips, 2015) werden gemanipuleerd en extremere standpunten omarmden.

Hoewel de hierboven beschreven mechanisme niet een-op-een samenvallen met AI en zeker niet met generatieve AI, is het wel heel belangrijk om ze te kunnen herkennen in je professionele en persoonlijke leven.

We sluiten deze paragraaf af met het volgende. Je moet je ervan bewust zijn dat de wijze waarop de stap **Personalisatie en filtering** (de kern van de bubbel) wordt bepaald *en ontworpen in de handen ligt van mensen*. Die mensen zijn de eigenaren en managers van bedrijven die de software leveren en er hun geld mee verdienen. Het is een bewuste keuze van hen om deze algoritmes zo te maken dat je er als eindgebruiker zoveel mogelijk weer terugkeert. In feite beslissen zij voor jou in belangrijke mate hoe jij denkt en je gedraagt. Die bedrijven hebben dus een hele grote verantwoordelijkheid. Juist in een tijd van grote polarisatie zouden zijn dat mee moeten wegen in hun ontwerp. Technisch is het in ieder geval zondermeer gemakkelijk om te doen. Vraag je je dus eens af: nemen deze bedrijven hun verantwoordelijkheid? Zouden de regels van de AI Act gaan helpen (zie paragraaf 9.5)?

9.4.4 Algoritmes en discriminatie

Over bias en mogelijke discriminatie weet je inmiddels al aardig wat in het licht van ontwikkeling en toepassing van generatieve AI. Maar een ander belangrijk aspect van algoritmes is discriminatie. We gebruiken hierbij als voorbeeld het studiebeursfraude schandaal bij de Dienst Uitvoering Onderwijs (DUO) (NOS, 2023).

Bij DUO werden namelijk tussen 2012 en 2023 studenten gecontroleerd via een algoritme met zelfbedachte risicocriteria zoals leeftijd en opleidingsniveau. Hoewel er geen sprake was van directe discriminatie (rechtstreeks selecteren op migratieachtergrond), was er wel indirecte discriminatie door selectie op kenmerken als leeftijd, onderwijssoort (mbo kreeg een hogere risicoscore dan hbo of universiteit) en afstand tot de ouder(s), waardoor meer studenten met een niet-Europese migratieachtergrond werden geselecteerd voor een controle dan andere studenten.

Het hoofdprobleem is de onevenredigheid in wie er op voorhand werd gecontroleerd op basis van ‘niet objectieve rechtvaardiging’ voor de risicocriteria. De criteria waren namelijk bedacht op basis van ‘indrukken’. Zelfs als sommige gecontroleerde studenten daadwerkelijk hadden gefraudeerd, is het discriminerend als bepaalde groepen disproportioneel vaak worden gecontroleerd op basis van onterecht veronderstelde risico's. Dat in strijd is met gelijkheidsbeginsels en antidiscriminatiewetgeving. Het tweede probleem is dat – als fraude werd vastgesteld bij de geselecteerde groep – er een zelfversterkend effect op zou treden. Er zou namelijk worden vastgesteld dat onevenredig veel studenten met een migratieachtergrond fraude zouden hebben gepleegd binnen het totale bestand van DUO. Zo zouden hun zelfbedachte criteria worden bekrachtigd en er een vicieuze cirkel ontstaan.

Het interessante van deze casus is dat in de vele artikelen die hierover zijn geschreven, de woorden ‘discriminatie’, ‘AI’ en ‘algoritmes’ in een adem genoemd werden. Je zou erdoor kunnen denken dat ze allemaal hetzelfde betekenen. Terwijl discriminatie in het geval van DUO niets te maken had met AI of generatieve AI.

9.5 Privacy en Gegevensbescherming – begrip van de wetten

De Europese Unie heeft met de AI Act en de Algemene Verordening Gegevensbescherming (AVG) twee belangrijke wettelijke kaders gemaakt om je als student en burger te beschermen. De AVG bevat strikte regels over het verzamelen en verwerken van je persoonsgegevens, waarbij transparantie, doelstelling en minimale gegevensverwerking centraal staan. De AI Act voegt hier specifieke regels aan toe voor AI-systemen, ingedeeld in risicocategorieën met bijbehorende verplichtingen. Hoge risicotoepassingen moeten aan strenge eisen voldoen op het gebied van transparantie, menselijk toezicht en verantwoording. Bepaalde AI-praktijken die je fundamentele rechten bedreigen zijn volledig verboden, zoals gezichtsherkenning in openbare ruimtes. De instelling waar je studeert of het bedrijf waar je werkt, is ervoor verantwoordelijk om op een rechtmatige manier hiermee om te gaan, zodat je met systemen kan werken die voldoen aan de eisen.

Ondanks deze wettelijke bescherming heb je als gebruiker ook een eigen verantwoordelijkheid in het beschermen van jouw privacy en die van anderen. Het begint natuurlijk met het beveiligen van je eigen computer, goede wachtwoorden gebruiken en waar mogelijk versleuteling van je gegevens. Gaat het om generatieve AI dan start je met kritisch te beoordelen welke AI-diensten je gebruikt en welke gegevens je daarbij deelt. Het is bijvoorbeeld heel gemakkelijk om een gratis account te maken bij een generatieve AI-dienst, maar weet je dan wat er met jouw gegevens gebeurt? Je kunt bijvoorbeeld regelmatig de toestemmingen controleren en waar mogelijk alternatieven gebruiken die minder van je data verzamelen. Wees je bewust van de waarde van je persoonlijke gegevens en blijf kritisch over ‘gratis’ diensten die in feite worden betaald met jouw data. Deel daarom in dat soort systemen nooit bijzondere gegevens of gevoelige gegevens. Door zowel gebruik te maken van de wettelijke bescherming als je eigen persoonlijke waakzaamheid kun je de voordelen van AI benutten zonder te veel privacy op te offeren.

p. 9-96

Box 9.4 – Persoonlijke, bijzondere en gevoelige gegevens

Binnen de AVG zijn er verschillende categorieën gegevens als het gaat om de gevoeligheid van dataverwerking. De belangrijkste categorie is de zogenaamde bijzondere persoonsgegevens. Dit zijn bijvoorbeeld gegevens waaruit iemands:

- ras of etnische afkomst blijkt;
- iemands politieke opvattingen blijken;
- religieuze of levensbeschouwelijke overtuigingen blijken;
- het lidmaatschap van een vakbond blijkt;
- gezondheid en seksueel gedrag of seksuele gerichtheid blijken;
- genetische gegevens en biometrische gegevens blijken (Autoriteit Persoonsgegevens, 2025).

Deze gegevens mogen alleen bij uitzondering worden gebruikt door instellingen en bedrijven en aan de verwerking worden heel strenge veiligheidseisen gesteld. De reden is dat kwaadwillenden met deze gegevens zeer schadelijke dingen kunnen doen.

Daarnaast is er de categorie gevoelige gegevens die niet zo expliciet benoemd worden, maar eveneens met een hoge mate van terughoudendheid en veiligheid

moet worden verwerkt. Denk aan gegevens over elektronische communicatie, locatiegegevens, financiële gegevens (zoals inkomen of koopgedrag) en het burgerservicenummer (BSN).

Bij generatieve AI geldt dat de inhoud van chats en geüploade documenten belangrijk zijn. Deze kunnen bijzondere en gevoelige gegevens bevatten en je moet er zelf voor zorgen dat die documenten niet verwerkt worden door bedrijven of systemen waarvan je niet bij weet wat ermee gebeurt. Hetzelfde geldt als je documenten upload in generatieve AI-systemen die bijvoorbeeld gevoelige informatie bevatten over bedrijven, instanties of onderzoek.

Omdat je zelf als individu, student of werknemer onderdeel bent van de maatschappij en bedrijven of instellingen, is het dus belangrijk dat je je hiervan bewust bent en ook zelf dat soort gegevens niet deelt. Niet van jezelf en ook niet van anderen.

9.6 Macht van Big Tech

Generatieve AI is een technologie die een fundamentele machtsverschuiving in de digitale wereld veroorzaakt. Aan het roer van deze verschuiving staan de grote technologiebedrijven – vaak aangeduid als Big Tech – zoals OpenAI (met investeringen van Microsoft), Google, Amazon, Meta en Apple. Deze bedrijven beschikken niet alleen over enorme hoeveelheden data en rekenkracht, maar ook over de financiële middelen, de netwerken en infrastructuur om generatieve AI op wereldschaal te ontwikkelen en uit te rollen. De ontwikkeling van grote taalmodellen, zoals GPT-4, Gemini en Claude, vereist investeringen van honderden miljoenen dollars, wat de drempel voor kleinere bedrijven of publieke instellingen erg hoog maakt. En deze machtsconcentratie gaat versneld door (Verhagen, 2025).

Box 9.5 – De schaal van AI-investeringen

De training van GPT-4 kostte naar eigen zeggen van OpenAI meer dan 100 miljoen dollar en gebruikte maandenlang duizenden GPU's (Smith, 2023). Momenteel worden in Amerika de grootste AI-datacenters gerealiseerd zoals Elon Musk's Colossus voor bijna negen miljard euro (Pilz et al., 2025). Dit soort investeringen zijn enkel haalbaar voor big tech-spelers. Voor publieke instellingen of start-ups is dit vrijwel onbereikbaar, wat leidt tot een monopolie op het gebied van AI-innovatie (OpenAI et al., 2023).

Slechts een handvol bedrijven bepaalt welke AI-modellen wereldwijd worden gebruikt, onder welke voorwaarden en met welke ethische implicaties. Deze eenzijdige concentratie van middelen leidt tot een centralisatie van macht. Een belangrijk aspect daarvan is bijvoorbeeld het wegkapen van bezoekers van websites, met name journalistieke websites. Doordat mensen meer en meer gaan zoeken op het internet met behulp van generatieve AI (als jonge mensen het al niet doen via Youtube, TikTok of Instagram), bezoeken mensen minder en minder de websites waar de grote bedrijven de data vandaan halen. Bij het zoeken blijven mensen hangen op het antwoord van de AI zonder nog de originele website te bezoeken waardoor die websites veel inkomsten missen. Volgens The Atlantic (Reisner, 2025), kan dat binnenkort al betekenen dat vele journalistieke websites verdwijnen, terwijl die cruciaal zijn voor een democratie omdat ze burgers informeren, controle uitoefent op de overheid en een platform bieden voor publiek debat.

De Europese Commissie en andere beleidsorganen waarschuwen al langer dat deze technologische dominantie risico's meebrengt voor democratische controle, transparantie en eerlijke toegang tot informatie en innovatie (Bietti, 2020; Crawford, 2021b). Zo bepalen big tech-bedrijven vaak eenzijdig

welke data in de modellen worden opgenomen, welke talen en culturele contexten dominant zijn, wie toegang krijgt tot geavanceerde modellen. Dit heeft directe gevolgen voor bijvoorbeeld onderwijs, media, en publieke dienstverlening, die steeds vaker afhankelijk raken van door bedrijven geleverde AI-systemen.

De mondiale machtsdynamiek rond AI wordt momenteel dus meer en meer bepaald door economische en geopolitieke belangen. Terwijl de Verenigde Staten een voorsprong hebben op het gebied van private AI-innovatie, wil China via staatsgestuurde projecten deze voorsprong inhalen. Europa probeert zich te positioneren als hoeder van ethische en transparante AI, maar is daarbij sterk afhankelijk van buitenlandse technologie en platforms. Deze asymmetrie maakt het lastig voor Europese overheden en instellingen om werkelijk soevereine keuzes te maken over hoe AI wordt ingezet in bijvoorbeeld het onderwijs, de zorg of justitie (Mazzucato, 2021). De zorgen over digitale kolonisatie – waarbij technologieën niet alleen worden ingevoerd, maar ook de waarden en economische verhoudingen van de producenten meebrengen – nemen daardoor toe (Couldry & Mejias, 2019).

De dominantie van big tech in generatieve AI roept ook ethische vragen op over inclusie, rechtvaardigheid en de rol van de publieke sector. Wie wordt er betrokken bij het ontwikkelen van deze technologieën? Welke stemmen blijven ongehoord? En welke kennisdomeinen worden versterkt of juist genegeerd door de datasets waarop de modellen worden getraind? Deze vragen raken aan de kern van digitale soevereiniteit en het recht op technologie die in dienst staat van de samenleving, niet slechts van aandeelhouders.

Om de machtspositie van big tech tegenwicht te bieden, groeit de vraag naar publieke investeringen in open en transparante AI-modellen, bijvoorbeeld via universiteiten, bibliotheken of internationale samenwerkingen zoals de European AI Alliance (European Commission, 2018). Dergelijke initiatieven kunnen bijdragen aan een meer evenwichtige en democratisch gelegitimeerde AI-toekomst, mits ze gepaard gaan met structurele financiering, heldere regelgeving en brede maatschappelijke betrokkenheid (Brundage et al., 2020; Eubanks, 2018). Initiatieven zoals Hugging Face, EleutherAI en BLOOM pleiten voor open toegang tot AI-modellen, datasets en trainingsprocedures. Hiermee proberen ze technologische onafhankelijkheid en inclusie te bevorderen. Toch blijven deze projecten afhankelijk van fondsen van overheden of, toch weer big tech zelf (Kreps & Kriner, 2023).

9.7 Open Source Modellen versus Gesloten Modellen

Open source LLM's en gesloten modellen LLM's verschillen in hun benadering van toegankelijkheid en gebruik. Open source modellen, zoals Llama, Falcon en Mistral, bieden transparantie doordat hun code, trainingsproces en vaak ook gewichten openbaar toegankelijk zijn. Dit maakt het mogelijk voor onderzoekers en ontwikkelaars om de modellen te bestuderen, aan te passen, te verbeteren en daarin samen te werken. Het verbetert de inzichtelijkheid in de werking van de systemen. Bovendien kunnen open source modellen lokaal worden uitgevoerd (draaien op een gebruikerscomputer), waardoor ze meer privacy-vriendelijk kunnen zijn en minder afhankelijk van externe diensten.

Maar open source modellen hebben vaak wel beperktere middelen voor training en verfijning, wat kan resulteren in minder geavanceerde capaciteiten. Zo bieden gesloten modellen zoals zoals GPT-4/5 en Claude meestal betere prestaties en betrouwbaarheid. Maar, de gesloten aard beperkt de wetenschappelijke vooruitgang omdat onderzoekers geen inzicht krijgen in hoe de modellen werken of beslissingen nemen. De keuze tussen open source en gesloten LLM's komt dus neer op een afweging tussen transparantie, controle en innovatie door samenwerking tegenover prestaties en professionele ondersteuning.

9.8 Zelfstudievragen

9.8.1 Checkvragen

1. Welke drie opvallende categorieën van databronnen werden gevonden in de mC4-dataset die gebruikt wordt voor het trainen van taalmodellen zoals ChatGPT?
2. Wat is het verschil tussen echokamers en filterbubbels, en hoe kunnen algoritmes bijdragen aan polarisatie en extremisme?
3. Noem drie belangrijke verschillen tussen open source LLM's (zoals Llama en Mistral) en gesloten modellen (zoals GPT-4 en Claude).

9.8.2 Reflectievragen

4. Het hoofdstuk beschrijft hoe datalabelaars vaak onder slechte arbeidsomstandigheden werken voor lage lonen. Hoe verhoud jij je tot het gebruik van AI-tools waarvan je weet dat ze mogelijk gebaseerd zijn op uitbuiting van werknemers? Wat voor afwegingen maak je hierin?
5. Big Tech-bedrijven hebben door hun dominantie in AI-ontwikkeling veel macht over welke waarden en perspectieven in AI-modellen worden opgenomen. Hoe zou volgens jou een meer democratische en eerlijke ontwikkeling van AI eruit kunnen zien, en welke rol zie je hierin voor jezelf als toekomstige professional?

9.8.3 Antwoordsuggesties

1. De drie opvallende categorieën zijn: DocPlayer (een website die automatisch documenten verzamelt, vaak zonder toestemming en met persoonsgegevens), krantenartikelen van uitgeverijen (die nooit toestemming gaven voor dit gebruik), en bronnen van lage kwaliteit waaronder zelfs nazistische websites zoals Stormfront.
2. Echokamers ontstaan wanneer je vooral interacteert met content die je wereldbeeld bevestigt, waardoor het algoritme je nog meer van ditzelfde soort content laat zien. Filterbubbels zijn het gevolg: je wordt systematisch minder blootgesteld aan afwijkende meningen. Algoritmes versterken dit door groepspolarisatie (extremere standpunten in groepen), bevestigingsbias en het wegfilteren van tegengestelde meningen.
3. Open source modellen bieden transparantie in code en trainingsproces, kunnen lokaal uitgevoerd worden (meer privacy), maar hebben vaak beperktere middelen. Gesloten modellen presteren meestal beter en hebben professionele ondersteuning, maar bieden geen inzicht in werking en beperken wetenschappelijke vooruitgang door hun ondoorzichtigheid.
4. Dit is een persoonlijke reflectievraag waarbij je kunt overwegen: je bewustwording van de arbeidsomstandigheden, alternatieven zoeken (zoals lokale LLM's of ethische aanbieders), transparantie eisen van bedrijven, of accepteren van beperkingen in AI-gebruik om uitbuiting te vermijden.
5. Mogelijke elementen: publieke investeringen in open AI-onderzoek, democratische betrokkenheid bij AI-ontwikkeling, diversiteit in ontwikkelteams, transparantie-eisen, Europese of internationale samenwerking, en je eigen rol als kritische gebruiker die bewuste keuzes maakt en pleit voor verantwoorde AI-ontwikkeling.

10 Jouw AI-bewuste toekomst

Stel je voor …

Je begint net aan je eerste baan als junior beleidsadviseur bij een gemeente. Je nieuwe collega Marianne maakt dagelijks gebruik van AI-tools. Ze zet de veilige AI-omgeving van de gemeente in om complexe rapporten snel te structureren, om beleidsnotities te herschrijven, en checkt data met behulp van slimme data-analyse prompts. Soms zet ze zelfs agents in om routinetaken te automatiseren. Je merkt al snel dat je achterloopt. Waar jij uren doet over een analyse, heeft Marianne in een kwartier een eerste voorstel klaar. Maar ze ziet ook vaak af van het gebruik van AI, kiest bewust een minder groot chatmodel en zertrouwt niet blind op AI. Ze stelt kritische vragen, toetst elke uitkomst aan de werkelijkheid, en denkt actief na over ethische grenzen. Je besluit: ik wil niet zomaar AI gaan gebruiken. Ik wil leren denken en werken zoals Marianne – kritisch, bewust en toekomstbestendig.

10.1 AI-geletterd profiel ontwikkelen

Wat betekent het eigenlijk om AI-geletterd te zijn? Het is verleidelijk om te denken dat het vooral draait om technische kennis of goed kunnen prompten. Maar AI-geletterdheid is veel meer dan dat. Het is een combinatie van kritisch denken, technisch begrip, ethisch bewustzijn en adaptief leervermogen. Die vier onderdelen vormen samen de basis van een toekomstbestendige omgang met AI.

Box 10.1 Hoe ontwikkel je een AI-geletterd profiel?

Wil je jezelf AI-geletterd noemen, dan helpt het om gestructureerd aan je vaardigheden te werken. Begin klein, bijvoorbeeld met het volgen van een onlinecursus over generatieve AI of het bijhouden van AI-nieuws via een nieuwsbrief. Stel jezelf een concreet leerdoel bijvoorbeeld: ‘ik wil ChatGPT kunnen gebruiken om mijn aantekeningen samen te vatten’ of ‘ik wil begrijpen hoe aanbevelingsalgoritmes werken’. Reflecteer op je waarden: welke AI-toepassingen vind jij wenselijk of juist problematisch? Hoe kijk je aan tegen de milieubelasting of de dilemma’s rondom copyright? En zoek actief naar feedback, bijvoorbeeld door je AI-gebruik met studiegenoten te bespreken. AI-geletterdheid vraagt geen perfecte kennis, maar wel betrokkenheid en nieuwsgierigheid. Tip: gebruik de AI Maturity in Education Scan van de VU om te ontdekken waar jij nu staat (Terbeek, 2025).

10.2 Duurzame gewoontes met AI

AI is geen modegril die vanzelf weer verdwijnt; het is een blijvende realiteit die je studie, werk en dagelijks leven steeds meer zal beïnvloeden. Juist daarom is het belangrijk om duurzame gewoontes te ontwikkelen in het gebruik van AI. Het gaat dan niet alleen over hoe je AI inzet, maar vooral ook over wanneer, waarom en met welk doel. En je gebruik te beperken om het milieu te ontzien (zie hoofdstuk 6). Tegelijkertijd ben je ervan bewust dat AI lang niet het enige is dat het milieu beïnvloed en dat je meer levensstijlveranderingen kunt maken om een grotere impact te hebben. Zo leer je bewuste keuzes maken in je levensstijl, over het gebruik van technologie, en ontwikkel je routines die je ondersteunen zonder je afhankelijk te maken.

Nadat je dit boek hebt doorgenomen, ken je inmiddels zoveel redenen om AI wel te gebruiken, of dat juist niet te doen, dat het je gaat duizelen. Mollick (2025, 2025) heeft een interessante lijst hierover gemaakt die je met je eigen argumenten kunt vergelijken. Als eerste reden om generatieve AI te gebruiken benoemt

hij het concept van de ‘Best Available Human’: zou de best beschikbare AI op een bepaald moment, op een bepaalde plaats, beter werk leveren bij het oplossen van een probleem dan de best beschikbare mens die daadwerkelijk in staat is om te helpen in een bepaalde situatie? Hij denkt dat dit vaak geval is. Meer specifieke redenen om AI wel te gebruiken zijn bijvoorbeeld voor:

- Taken die vragen om het produceren van veel ideeën;
- Een vakgebied waar je expert in bent zodat je weet of de output goed is;
- taken om informatie snel aan te passen tot leermateriaal of presentaties.

Redenen om generatieve AI niet te gebruiken zijn bijvoorbeeld:

- Momenten waarop je echt zelf wilt leren of begrijpen;
- Als hoge precisie nodig is;
- Als je niet precies begrijpt waarom een AI doet wat die doet;
  Als jouw inspanning en doorlopende leer- of maakproces parate kennis en vaardigheden opleveren, die je dus beter niet kunt uitbesteden;
- Wanneer je niet goed snapt of je een blinde vlek hebt voor waar de AI slecht in is.

Box 10.2 AI en persoonlijke ontwikkeling

Leren is iets anders dan productief zijn met een AI-systeem. Als je echt iets wilt begrijpen, is het niet voldoende om alleen een goed antwoord te hebben; je moet een vraag of probleem kunnen doorgronden. Iets je cognitief eigen maken vraagt oefening, reflectie en doorzettingsvermogen. Het trainen van je hersenen – het vermogen om kennis paraat te krijgen en houden – is nodig om de reikwijdte van een AI-antwoord te kunnen inschatten en er kritisch op te kunnen reflecteren. Als je werkelijk wilt leren, is het beter om eerst zelf na te denken over een vraag, je eigen antwoord te formuleren en daarna pas te vergelijken met het antwoord van een AI-systeem. Alleen dan leer je herkennen wat jij al weet, wat je nog niet weet, en hoe AI je kan aanvullen – in plaats van vervangen. AI is geen shortcut naar begrip, maar kan wel een waardevolle spiegel en assistent zijn bij je denkproces.

10.3 Ethische zelfreflectie

De manier waarop je AI gebruikt, zegt iets over wie je bent en wie je wilt zijn. Juist omdat AI-technologie steeds meer verweven raakt met je studie, werk en dagelijks leven, is het belangrijk om stil te staan bij de ethische dimensie van je keuzes.

Box 10.3 Ethische zelfreflectie in het AI-tijdperk

Hoe blijf je trouw aan jezelf als AI steeds meer taken kan overnemen? Begin met het formuleren van je persoonlijke waarden rondom AI-gebruik. Wat vind jij belangrijk: transparantie, gelijkheid, verantwoordelijkheid, rechtvaardigheid? Leg die waarden naast je dagelijkse gebruik van AI. Stel jezelf regelmatig vragen als: ‘Helpt deze tool mij om integer te werken?’, ‘Ben ik nog steeds eigenaar van mijn keuzes?’, en ‘Wat zijn de gevolgen van mijn AI-gebruik voor anderen?’ Door op die manier bewust te blijven nadenken, houd je je morele kompas scherp en ontwikkel je jezelf als AI-gebruiker en als kritisch, ethisch denkend mens.

10.4 Toekomstperspectief

AI is continu in ontwikkeling en zal je vakgebied blijven beïnvloeden – of je nu werkt in de zorg, het onderwijs, de rechtspraak of de kunsten. Daarom is het belangrijk om jezelf aan te leren flexibel en kritisch mee te bewegen met die veranderingen. Niet door alles klakkeloos te volgen, maar door nieuwsgierig te blijven, ethisch na te denken en te blijven leren.

Een persoonlijke roadmap helpt je koers te houden. Denk na over waar je naartoe wilt, wat je nodig hebt om daar te komen, en wees voorbereid op verandering. AI-geletterdheid is geen eindstation, maar een voortdurend proces – en dat begint bij jouw eigen keuzes.

10.5 Zelfstudievragen

10.5.1 Checkvragen

1. Wat zijn volgens het hoofdstuk de vier onderdelen die samen de basis vormen van AI-geletterdheid?
2. Wat is het 'Best Available Human' concept zoals Mollick dat beschrijft?
3. Waarom wordt er in het hoofdstuk onderscheid gemaakt tussen leren en productief zijn met AI-systemen?

10.5.2 Reflectievragen

4. Beschouw je eigen AI-gebruik tot nu toe. Op welke van Mollick's redenen om AI wel of niet te gebruiken herken je jezelf? Geef concrete voorbeelden uit je eigen ervaring.
5. Het hoofdstuk benadrukt het belang van ethische zelfreflectie bij AI-gebruik. Formuleer drie persoonlijke waarden die voor jou belangrijk zijn bij het gebruiken van AI en leg uit hoe je deze in de praktijk zou willen toepassen.

10.5.3 Antwoordsuggesties

1. De vier onderdelen van AI-geletterdheid zijn: kritisch denken, technisch begrip, ethisch bewustzijn en adaptief leervermogen. Deze vormen samen de basis voor een toekomstbestendige omgang met AI.
2. Het 'Best Available Human' concept stelt de vraag of de best beschikbare AI op een bepaald moment, op een bepaalde plaats, beter werk zou leveren bij het oplossen van een probleem dan de best beschikbare mens die daadwerkelijk in staat is om te helpen in die specifieke situatie.
3. Het onderscheid is belangrijk omdat leren iets anders is dan productief zijn. Als je écht iets wilt begrijpen, is het niet voldoende om alleen een goed antwoord te hebben; je moet het kunnen doorgronden. Het trainen van je 'hersenspier' is essentieel om de reikwijdte van een AI-antwoord te kunnen inschatten en er kritisch op te kunnen reflecteren.
4. Bij deze vraag kun je nadenken over situaties waarin je AI hebt gebruikt voor brainstormen, schrijfhulp, of onderzoek, en evalueren of dit paste bij Mollick's criteria zoals het produceren van veel ideeën, situaties waar je al expert bent, of juist momenten waar je wilde leren en AI misschien niet de beste keuze was.
5. Mogelijke waarden zijn transparantie (eerlijk zijn over AI-gebruik), verantwoordelijkheid (eigenaar blijven van je keuzes), rechtvaardigheid (rekening houden met impact op anderen), of integriteit (AI gebruiken om je werk te verbeteren, niet om te valsspelen). Je kunt per waarde een concreet Bronnen voor verdieping.

10.6 Activiteit - Jouw verantwoorde generatieve AI gebruik

Als je het hele boek hebt bestudeerd is aan jou de vraag: hoe ga je generatieve AI gebruiken? Deel met je medestudenten een **statement** op Canvas hoe jij het gaat inzetten. Wissel argumenten uit om je statement te onderbouwen. Je kunt niet reageren op de statements van je medestudenten, maar lees ze eens door ter inspiratie en om van te leren.

11 Indexen

11.1 Figuren

11.2 Figuren originele URL's

Figuur-4.2: https://images.squarespace-cdn.com/content/v1/64398599b0c21f1705fb8fb3/4e25a8bd-e1c5-4f01-8ffd-13665eef7423/Defining+Generative+AI+Explainer+%281%29.png?format=1000w

Figuur-4.3:
https://upload.wikimedia.org/wikipedia/commons/2/28/Artificial_neural_network_image_recognition.png

Figuur-4.4: https://canvas.vu.nl/courses/83333/files/9170751/

https://canvas.vu.nl/courses/83333/files/9170751/download?download_frd=1

Figuur-4.5: https://edudatabase.ctl-vu.nl/wp-content/uploads/2025/09/Figuur-4.5.png

Figuur-5.3:
https://www.pinecone.io/_next/image/?url=https%3A%2F%2Fcdn.sanity.io%2Fimages%2Fvr8gru94%2Fproduction%2F927ca8cc5d92ee75f36d7eb4bef4685c4e3118e5-2880x1370.png&w=3840&q=75

Figuur-6.3:
https://substackcdn.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fsubstack-post-media.s3.amazonaws.com%2Fpublic%2Fimages%2F4a6b6ff7-950c-40f7-8ce3-389938dd38e3_1284x378.png

Figuur-6.5:
https://substackcdn.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fsubstack-post-media.s3.amazonaws.com%2Fpublic%2Fimages%2Ff586df9b-e5a1-4ddb-81bb-ca0eb5d99efa_2518x1544.png

Figuur-6.6:
https://substackcdn.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fsubstack-post-media.s3.amazonaws.com%2Fpublic%2Fimages%2F21550978-efd9-4aca-bfd8-55be56807d85_2506x1260.png

Figuur 3.1 URL voor Canvasweergave: https://edudatabase.ctl-vu.nl/wp-content/uploads/2025/09/Figuur-3.1.png

Figuur 4.5 URL voor Canvasweergave: https://edudatabase.ctl-vu.nl/wp-content/uploads/2025/09/Figuur-4.5.png

11.3 Boxen

11.4 Referenties

Ahmed, A. (2025, May 16). ChatGPT Usage Statistics: Numbers Behind Its Worldwide Growth and Reach. *Digital Information World*. https://www.digitalinformationworld.com/2025/05/chatgpt-stats-in-numbers-growth-usage-and-global-impact.html

AI for Education. (2023). *Generative AI Explainer*. AI for Education. https://www.aiforeducation.io/ai-resources/generative-ai-explainer

Akmeşe, B. (2023). The Artificial Intelligence Dimension of Digital Manipulation Deepfake Videos: The Case of the Ukrainian-Russian People. *Contemporary Issues of Communication*, *2*(2), Article 2.

Aliani, R. (2024, August 9). How to use Perplexity for Research: Part 1 of 3 [Substack newsletter]. *Era of the Research Bots* . https://raziaaliani.substack.com/p/how-to-use-perplexity-for-research

Andersen, J. P., Degn, L., Fishberg, R., Graversen, E. K., Horbach, S. P. J. M., Schmidt, E. K., Schneider, J. W., & Sørensen, M. P. (2025). Generative Artificial Intelligence (GenAI) in the research process – A survey of researchers' practices and perceptions. *Technology in Society*, *81*, 102813. https://doi.org/10.1016/j.techsoc.2025.102813

Anthropic. (2023). *Anthropic's Red-Teaming Methodology*. https://www.anthropic.com/index/anthropics-red-teaming

Autoriteit Persoonsgegevens. (2025, January 6). *Wat zijn persoonsgegevens? | Autoriteit Persoonsgegevens*. https://www.autoriteitpersoonsgegevens.nl/themas/basis-avg/privacy-en-persoonsgegevens/wat-zijn-persoonsgegevens

Baas, E. (2025, April 30). *(28) Post | LinkedIn*. https://www.linkedin.com/posts/elisa-baas_baasin-chatgpt-kritisch-denken-prompt-ugcPost-7322569391303245824-jLXA/?utm_source=share&utm_medium=member_android&rcm=ACoAAAATgacBx9MluKs48BO1aRxqREWXJBXfTYw

Balo, B. (2025, March 12). *10 ChatGPT Prompts for Powerful Academic Writing*. Prompt Advance. https://promptadvance.club/chat-gpt-prompts-for-academic-writing

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, 34–55.

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

Berghel, H. (2018). Malice domestic: The Cambridge analytica dystopia. *Computer*, *51*(05), 84–89.

Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 210–219. https://doi.org/10.1145/3351095.3372860

Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy & Technology*, *31*(4), 543–556. https://doi.org/10.1007/s13347-017-0263-5

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer. https://link.springer.com/book/9780387310732

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2022). *On the Opportunities and Risks of Foundation Models* (No. arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

Brown, T. B. (2020). *Language Models are Few-Shot Learners*. NeurIPS.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., … Anderljung, M. (2020). *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims* (No. arXiv:2004.07213). arXiv. https://doi.org/10.48550/arXiv.2004.07213

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability and Transparency*, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline&ref=akusion-ci-shi-dai-bizinesumedeia

CBS. (2018). *Energy—Figures—Economy | Trends in the Netherlands 2018—CBS* [Webpagina]. Centraal Bureau voor de Statistiek. https://longreads.cbs.nl/trends18-eng/economy/figures/energy

Chayka, K. (2022, July 25). The Age of Algorithmic Anxiety. *The New Yorker*. https://www.newyorker.com/culture/infinite-scroll/the-age-of-algorithmic-anxiety

Chayka, K. (2024). *Filterworld: How Algorithms Make Everything the Same*. Heligo Books.

Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). *Deep reinforcement learning from human preferences* (No. arXiv:1706.03741; Version 3). arXiv. https://doi.org/10.48550/arXiv.1706.03741

Claude 3.7 Sonnet. (2025). Correct citeren van AI-gegenereerde content. *Anthropic*.

Couldry, N., & Mejias, U. A. (2019). The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. In *The Costs of Connection*. Stanford University Press. https://www.degruyterbrill.com/document/doi/10.1515/9781503609754/html

Cozzi, L., & Gould, T. (2024). *World Energy Outlook 2024 – Analysis* (World Energy Outlook). INTERNATIONAL ENERGY AGENCY. https://www.iea.org/reports/world-energy-outlook-2024

Crawford, K. (2021a). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crawford, K. (2021b). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crevier, D. (1993). *AI: The Tumultuous History of the Search for Artificial Intelligence*. Basic Books.

CriticalMynd. (2025, April 7). Understanding LLM Token Counts: What 1,000, 128,000 and 1 Million Tokens Actually Mean? *Medium*. https://criticalmynd.medium.com/understanding-llm-token-counts-what-1-000-128-000-and-1-million-tokens-actually-mean-9751131ac197

Doctorow, C. (2022, November 28). *Pluralistic: How monopoly enshittified Amazon/28 Nov 2022 – Pluralistic: Daily links from Cory Doctorow*. https://pluralistic.net/2022/11/28/enshittification/

Dreschler, G. (2025, April 19). *Is de scriptie dood? Een update over AI en schrijfvaardigheid*. Vrije Universiteit Amsterdam. https://vu.nl/nl/nieuws/2024/is-de-scriptie-dood-een-update-over-ai-en-schrijfvaardigheid

Eliot, L. (2023). Does Take A Deep Breath As A Prompting Strategy For Generative AI Really Work Or Is It Getting Unfair Overworked Credit. *Forbes*. https://www.forbes.com/sites/lanceeliot/2023/09/27/does-take-a-deep-breath-as-a-prompting-strategy-for-generative-ai-really-work-or-is-it-getting-unfair-overworked-credit/

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Publishing Group.

European Commission. (2018). *The European AI Alliance*. https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance

European Commission. (2025). *DigComp Framework—European Commission*. https://joint-research-centre.ec.europa.eu/projects-and-activities/education-and-training/digital-transformation-education/digital-competence-framework-citizens-digcomp/digcomp-framework_en

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org/

Goodfellow, I., Pouget-Abadie, J., & Mirza, M. (2014). *Generative adversarial nets*. *27*.

Hart van Nederland. (2025, February 12). *Liefde met een AI-chatbot? Jacob hertrouwt met digitale vrouw Aiva*. Hart van Nederland. https://www.hartvannederland.nl/het-beste-van-hart/persoonlijke-verhalen/artikelen/jacob-trouwt-met-ai-vrouw-aiva-experts-waarschuwen

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Heikkiläarchive, M. (2024, June 14). *How to opt out of Meta's AI training*. MIT Technology Review. https://www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, *33*, 6840–6851. https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html

Hofman, E., & Veerbeek, J. (2023, June 7). Chatbots als ChatGPT putten veel uit nazi- en complotsites. *De Groene Amsterdammer*. https://www.groene.nl/artikel/dat-zijn-toch-gewoon-al-onze-artikelen

Hofman, E., & Veerbeek, J. (2024, December 11). Bedreigende nepbeelden van asielzoekers– ze worden in meerderheid verspreid door één PVV-Kamerlid. *De Groene Amsterdammer*. https://www.groene.nl/artikel/het-land-van-de-hoogblonde-gezinnetjes

Hoog, M. de. (2025, June 2). *Ja, AI kan je productiever maken. Maar wie vertrouwt jou dan nog?* De Correspondent. https://decorrespondent.nl/16133/ja-ai-kan-je-productiever-maken-maar-wie-vertrouwt-jou-dan-nog/9699d435-5d3f-0755-0dae-309ff9658c99

Hubinger, E., van Merwijk, C., Mikulik, V., Skalse, J., & Garrabrant, S. (2019). *Risks from Learned Optimization in Advanced Machine Learning Systems*. arXiv. https://arxiv.org/abs/1906.01820

King, D. (1997). *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia*. Henry Holt and Company.

Kniberg, H. (Director). (2024, January 20). *Generative AI in a Nutshell—How to survive and thrive in the age of AI* [Video recording]. https://www.youtube.com/watch?v=2IK3DFHRFfw

Kojima, T. (2022). Large Language Models are Zero-Shot Reasoners. *arXiv Preprint*.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2023). *Large Language Models are Zero-Shot Reasoners* (No. arXiv:2205.11916). arXiv. https://doi.org/10.48550/arXiv.2205.11916

Kreps, S., & Kriner, D. (2023). How AI threatens democracy. *Journal of Democracy*, *34*(4), 122–131.

Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). *ChatGPT Beyond English: Towards a Comprehensive Evaluation of Large Language Models in Multilingual Learning* (No. arXiv:2304.05613). arXiv. https://doi.org/10.48550/arXiv.2304.05613

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., & Rocktäschel, T. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Lim, S. L., & Bentley, P. J. (2022). Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, *12*(1), 18131.

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43.

Liu, P. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2022). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education.

Luyendijk, J. (2006). *Het zijn net mensen*. Podium b.v. Uitgeverij.

Magnason, A. S. (2022). *Over tijd en water: Een geschiedenis van onze toekomst*. Singel Uitgeverijen.

Martijn, M. (2023, September 20). AI draait op werk van miljoenen onzichtbare, slechtbetaalde mensen. Wie komt er voor ze op? *De Correspondent*. https://decorrespondent.nl/14804/ai-draait-op-werk-van-miljoenen-onzichtbare-slechtbetaalde-mensen-wie-komt-er-voor-ze-op/b7d59642-4142-0f2e-250f-3a2fc529b0cc

Masley, A. (2025a, January 13). Using ChatGPT is not bad for the environment [Substack newsletter]. *The Weird Turn Pro*. https://andymasley.substack.com/p/individual-ai-use-is-not-bad-for

Masley, A. (2025b, June 14). Computing is efficient [Substack newsletter]. *The Weird Turn Pro*. https://andymasley.substack.com/p/computing-is-efficient

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

Mazzucato, M. (2021). *Mission Economy: A Moonshot Guide to Changing Capitalism*. Penguin UK.

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future* (First edition). W.W. Norton & Company.

McKinsey. (2023, June 14). *Economic potential of generative AI | McKinsey*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction

Meer, van der, H. (2025, mei). *Checklist beoordelen GenAI content*. https://edusources.nl/materials/f80b8d39-1dad-4c04-954b-831c15d14117/checklist-beoordelen-genai-content?tab=0

Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, *3*(1), 2053951716650211. https://doi.org/10.1177/2053951716650211

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Penguin.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mollick, E. (2025, January 26). *The Best Available Human Standard*. https://www.oneusefulthing.org/p/the-best-available-human-standard

Mollick, E., & Mollick, L. (2023). Using AI to improve teaching and learning. *arXiv Preprint*. https://arxiv.org/abs/2304.03442

NOS. (2023, June 21). *Studenten met migratieachtergrond opvallend vaak beschuldigd van fraude, minister wil systeem grondig nagaan*. https://nos.nl/op3/artikel/2479700-studenten-met-migratieachtergrond-opvallend-vaak-beschuldigd-van-fraude-minister-wil-systeem-grondig-nagaan

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems. *Social Science Computer Review*, *33*(4), 459–478. https://doi.org/10.1177/0894439314555329

OECD. (2022). *Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint* (OECD Digital Economy Papers No. 341; OECD Digital Economy Papers, Vol. 341). https://doi.org/10.1787/7babf571-en

Op de Beeck, O. (2024, February 11). *ChatGPT herkennen: Zo zie je wie AI gebruikt*. https://oliveropdebeeck.com/chatgpt-herkennen/

OpenAI. (2021, July 28). *Techniques for training large neural networks*. https://openai.com/index/techniques-for-training-large-neural-networks/

OpenAI. (2023, December 13). *Partnership with Axel Springer to deepen beneficial use of AI in journalism*. https://openai.com/index/axel-springer-partnership/

OpenAI. (2025, May 2). *Expanding on what we missed with sycophancy*. https://openai.com/index/expanding-on-sycophancy/

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 Technical Report* (No. arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Perrigo, B. (2023, January 18). *Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer*. TIME. https://time.com/6247678/openai-chatgpt-kenya-workers/

Phillips, W. (2015). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. Mit Press. https://books.google.com/books?hl=en&lr=&id=pjYhBwAAQBAJ&oi=fnd&pg=PR7&dq=What+is+online+trolling&ots=THN26JXpbz&sig=oEsugBub901oZQoIk6aNCTveGL8

Pilz, K. F., Sanders, J., Rahman, R., & Heim, L. (2025). *Trends in AI Supercomputers* (No. arXiv:2504.16026). arXiv. https://doi.org/10.48550/arXiv.2504.16026

Pinecone. (2025). *Conversational Memory for LLMs with Langchain | Pinecone*. https://www.pinecone.io/learn/series/langchain/langchain-conversational-memory/

Polimeni, J. M., Mayumi, K., Giampietro, M., & Alcott, B. (2015). *The Myth of Resource Efficiency: The Jevons Paradox*. Routledge. https://doi.org/10.4324/9781315781358

POLITICO. (2025, April 11). *Ireland probes Musk's X for feeding Europeans' data to its AI model Grok*. POLITICO. https://www.politico.eu/article/irish-dpc-launches-investigation-into-xs-use-of-eu-data-to-train-ai/

Pontefract, D. (2025, March 25). Authors Challenge Meta's Use Of Their Books For Training AI. *Forbes*. https://www.forbes.com/sites/danpontefract/2025/03/25/authors-challenge-metas-use-of-their-books-for-training-ai/

Radford, A. (2019). *Language models are unsupervised multitask learners*. OpenAI Blog.

Reisner, A. (2025, June 25). The End of Publishing as We Know It. *The Atlantic*. https://www.theatlantic.com/technology/archive/2025/06/generative-ai-pirated-articles-books/683009/

Reuters. (2023, February 8). *Google unveils ChatGPT rival Bard, AI search plans in battle with Microsoft*. https://www.reuters.com/technology/google-unveils-chatgpt-rival-bard-ai-search-plans-battle-with-microsoft-2023-02-07/

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson Education Limited.

Ryan, J. (2024). The Coming Enshittification of AI: Will AI follow internet search and e-commerce down the path of enshittification, or can we finally have nice things? *The Journal of Business and Artificial Intelligence*, *1*(2), Article 2. https://jbai.ai/index.php/jbai/article/view/31

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549. https://doi.org/10.1177/2053951718820549

Sanderson, G. (2017, Oktober). *3Blue1Brown*. https://www.3blue1brown.com/topics/3blue1brown.com

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2023). *Towards Understanding Sycophancy in Language Models* (No. arXiv:2310.13548). arXiv. https://doi.org/10.48550/arXiv.2310.13548

Siu, E. (2025, May 8). *I took a 2-day "vibe coding" class and successfully built a product. Here are my biggest takeaways*. CNBC. https://www.cnbc.com/2025/05/08/i-took-a-2-day-vibe-coding-class-and-successfully-built-a-product.html

SLO. (2024, oktober). *Digitale geletterdheid* [Overzichtspagina]. SLO. https://www.slo.nl/thema/meer/basisvaardigheden/digitale-geletterdheid/

Smith, C., S. (2023, September 8). What Large Models Cost You – There Is No Free AI Lunch. *Forbes*. https://www.forbes.com/sites/craigsmith/2023/09/08/what-large-models-cost-you--there-is-no-free-ai-lunch/

Solon, O. (2021, June 19). Drought-stricken communities push back against data centers. *NBC News*. https://www.nbcnews.com/tech/internet/drought-stricken-communities-push-back-against-data-centers-n1271344

Sparrow, R. (2021). Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? *International Journal of Social Robotics*, *13*(1), 23–29. https://doi.org/10.1007/s12369-020-00631-2

Spinner, Y. (2025). Rechter houdt recordschikking van Anthropic in zaak auteursrecht voorlopig tegen. *Tweakers*. https://tweakers.net/nieuws/238930/rechter-houdt-recordschikking-van-anthropic-in-zaak-auteursrecht-voorlopig-tegen.html

Streppel, A. R., Nijenkamp, R., Tabois, L., & Blikmans-Middel, M. B. (2024, September 1). *Critical AI Literacy Module*. University of Groningen. https://edusources.nl/materials/2bee669c-264e-461b-af05-2f54351496d1/critical-ai-literacy-e-learning-course

Substack. (n.d.). *Stop Calling It a "Hallucination": What AI Errors Reveal About Writing, Work, and Education*. Retrieved October 3, 2025, from https://substack.com/inbox/post/175134375?utm_medium=android&triedRedirect=true

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Geo. L. Tech. Rev.*, *4*, 1.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge. https://www.cambridge.org/core/journals/robotica/article/robot-learning-edited-by-jonathan-h-connell-and-sridhar-mahadevan-kluwer-boston-19931997-xii240-pp-isbn-0792393651-hardback-21800-guilders-12000-8995/737FD21CA908246DF17779E9C20B6DF6

Terbeek, L. (2025). *The AI Maturity in Education Scan (AIMES)*. Vrije Universiteit Amsterdam. https://vu.nl/en/education/more-about/the-ai-maturity-in-education-scan-aimes

United Consumers. (2025). Sluipverbruik van Stroom Meten en Verlagen: Bekijk onze tips! *UnitedConsumers*. https://www.unitedconsumers.com/energie/blog/sluipverbruik

Vaendel, D. (2025, June 16). Je motivatiebrief schrijven met ChatGPT? "Die kandidaten zijn sowieso zwak." *Observant*. https://www.observantonline.nl/Home/Artikelen/id/63439/je-motivatiebrief-schrijven-met-chatgpt-die-kandidaten-zijn-sowieso-zwak

Van Bergeijk, J. (2025, April 30). Heeft het werk als trainer van AI-modellen als ChatGPT überhaupt zin? *De Groene Amsterdammer*. https://www.groene.nl/artikel/heeaal-leuke-mensen-haha

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Verhagen, L. (2025, May 1). Krachtigste AI-computers in privébezit: 'Een treurig beeld. We zijn overgeleverd aan big tech.' *de Volkskrant*. https://www.volkskrant.nl/tech/krachtigste-ai-computers-in-privebezit-een-treurig-beeld-we-zijn-overgeleverd-aan-big-tech~b427d0ff/

Visser, D. (2025, June 26). Trainen van AI is géén 'fair use.' *Mr. Online*. https://www.mr-online.nl/trainen-van-ai-is-geen-fair-use/

Vries-Gao, A. de. (2025). Artificial intelligence: Supply chain constraints and energy implications. *Joule*, *0*(0). https://doi.org/10.1016/j.joule.2025.101961

VU Amsterdam. (2025). *Generatieve AI, Copilot en ChatGPT*. Generatieve AI, Copilot en ChatGPT. https://vu.nl/nl/student/tentamens/generatieve-ai-jouw-gebruik-onze-verwachtingen

Wei, J. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. *arXiv Preprint*.

Yang, C., Wang, X., Lu, Y., Liu, H., Le, Q. V., Zhou, D., & Chen, X. (2023). Large language models as optimizers. *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=Bb4VGOWELI

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2024). Diffusion models: A comprehensive survey of methods and applications. *arXiv.Org*. https://arxiv.org/abs/2209.00796

You, J. (2025, February 7). *How much energy does ChatGPT use?* Epoch AI. https://epoch.ai/gradient-updates/how-much-energy-does-chatgpt-use

Zamfirescu-Pereira, J. D. (2023). *Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts*. CHI Conference.

Ziegelmayer, U. (2021). *The ethics of artificial intelligence*. Springer.

11.5 Kijktips

Kijktip 3-1: wil je zien hoe alledaagse AI-functionaliteiten kunnen doorschieten in iets existentieels? Kijk de episode Be Right Back van Black Mirror (seizoen 2, aflevering 1), waarin een AI op basis van chat- en zoekgeschiedenis een overleden persoon nabootst. 2-18

Kijktip 4-1: wil je zien wat er mis kan gaan als je opdracht te vaag is? Kijk dan Star Trek: The Next Generation – Elementary, Dear Data (seizoen 2, aflevering 3), waarin een onduidelijke prompt tot onvoorziene gevolgen leidt. 3-23

Kijktip 4-2: voor een absurdistisch voorbeeld van waar dit toe zou kunnen leiden, bekijk je episode 3 'Sickofancy' van seizoen 27 van South Park. 3-25

Kijktip 4-3: voor meer verdieping over hoe datasets vooroordelen kunnen inbouwen, kijk de documentaire Coded Bias (2020), die laat zien hoe gezichtsherkenning en algoritmen systematisch bias kunnen reproduceren en waarom menselijke controle nodig blijft. 3-29

Kijktip 5-1: wil je een voorbeeld zien van hoe een strikt geprogrammeerd systeem de regels te letterlijk neemt? Kijk dan de film Avengers: Age of Ultron (2015), waarin de AI Ultron uit de hand loopt, doordat hij de opdracht om de mensheid te beschermen te star interpreteert. 4-35

Kijktip 10-1 Kijktip: bekijk voor een interessante verfilming over de relatie tussen mensen en chatbots de film Her (2013), waarin een eenzame schrijver (gespeeld door Joaquin Phoenix) een relatie ontwikkeld met een besturingssysteem (gespeeld door Scarlet Johansson).....9-93

Kijktip 10-2: voor een verdiepende blik op hoe tech-bedrijven jouw data gebruiken zonder dat je het doorhebt, kijk je de Netflix documentaire The Social Dilemma, waarin voormalige Google- en Facebook-