

*In gesprek met Steven Dorrestijn*

# Een **ethische** blik op AI

Auteur: **Sibrenne Wagenaar**

In een tijd waarin kunstmatige intelligentie (AI) en ChatGPT steeds meer verweven raken met ons dagelijks leven en werk, kunnen we makkelijk voorbijgaan aan de ethische implicaties van deze technologieën. Ik duik zelf graag en met veel enthousiasme in de wereld van AI-tools en hun eindeloze mogelijkheden. Ik hou van experimenteren, uitproberen en verkennen welke mogelijkheden dit allemaal biedt. Misschien heb jij dat ook?

Kritische geluiden uit de krant of op LinkedIn maken wel dat ik af en toe bij de ethische kant van AI stilsta. Maar dit is vaak van relatief korte duur. Terwijl die kritische geluiden zeker belangrijk zijn om in ogenschouw te nemen, denk maar aan het gegeven dat AI-systemen vooroordelen kunnen versterken die aanwezig zijn in het trainingsmateriaal. Of het idee dat AI-systemen vaak toegang nodig hebben tot

grote hoeveelheden persoonlijke gegevens om effectief te werken. Laat staan de black box van AI: hoe komt ChatGPT tot een bepaalde beslissing of uitkomst? Dat weten we eigenlijk niet.

Wat zijn belangrijke ethische vragen die we onszelf kunnen stellen? Steven Dorrestijn is lector Ethiek & Technologie aan de Saxion



Hogeschool, met een speciale focus op de impact van technologische ontwikkelingen op de samenleving en hoe deze kunnen worden ontworpen en gebruikt op een ethisch verantwoorde manier. Ik sprak hem over de invloed van AI en ChatGPT op leren en ontwikkelen, en over manieren om hier op een meer ethische manier over in gesprek te gaan.

### Wat willen we dat AI-systemen ons vertellen?

Steven Dorrestijn: "AI, en met name Large Language Models als Bing en ChatGPT, werken met immense datasets die onschatbare inzichten kunnen bieden. Daar maken velen van ons graag en gretig gebruik van. Echter, de vraag rijst of data ons echt alles kunnen vertellen wat we moeten weten. Het is essentieel om de waarde van data kritisch te bekijken en te begrijpen dat algoritmes niet neutraal zijn; ze zijn zo goed als de principes en de gegevens waarop ze zijn getraind."

"De huidige AI-ontwikkelingen gaan allemaal over data. We hebben de neiging om te denken dat data ons alles kunnen vertellen. Maar is dat wel zo? Dat is ontzettend moeilijk te checken. Zit er echt zoveel waarde in data? Wat willen we ons dan laten voorzeggen, en wat niet? Geloven we wat data ons vertellen

*"Het is essentieel om de waarde van data kritisch te bekijken en te begrijpen dat algoritmes niet neutraal zijn."*



### Over Steven Dorrestijn...

Dr. Steven Dorrestijn is lector Ethiek & Technologie aan de Hogeschool Saxion. Eerder werkte hij op de Universiteit Twente, als universitair docent bij Industrieel ontwerpen en bij de afdeling filosofie waar hij in 2012 promoveerde. In 2006 studeerde Dorrestijn een jaar filosofie in Parijs met een beurs van de Franse regering. Eerder studeerde hij werktuigbouwkunde en filosofie in Twente.

Met zijn onderzoeksgroep ontwikkelt hij werkvormen om ethiek en technologie praktisch te maken, zoals de Product Impact Tool en de Ethical Readiness Check. Samenleving en digitalisering vormen een belangrijk focusgebied. [www.stevendorrestijn.nl](http://www.stevendorrestijn.nl)

en volgen we zomaar op wat AI ons voorzeft? Dat zijn belangrijke vragen om bij stil te staan. En ik zie dat velen van ons zich die tijd niet gunnen."

Wat willen we dat deze systemen ons vertellen? En wat ook niet? Dit zijn best fundamentele vragen maar stellen we ons deze vragen wel voldoende? Volgens Steven is dit geen luxe maar een noodzaak. We

hebben niet puur te maken met een nieuwe technologische ontwikkeling, maar het gaat ook om de invloed ervan op ons leven en onze maatschappelijke waarden.

“ChatGPT is absoluut fascinerend: een computer die tegen en met ons praat. Een paar maanden geleden heb ik mijn lectorale rede gehouden. Toen kwam de vraag even in me op: zal ik ChatGPT gebruiken? Dan heb ik vast de eerste met AI geschreven lectorale rede. Bovendien kan het me helpen: ik kan er van alles mee opzoeken, stukjes tekst uitwerken. Misschien is het interessant? Ik heb ervoor gekozen om het niet te doen. Heel bewust. Met een lectorale rede vind ik het belangrijk om te laten zien hoe ik mezelf heb ontwikkeld.”

### Goed gebruik

Waar zet je AI wel bij in en waarbij niet? Wat vinden we ‘goed gebruik’? Hoe gebruiken we AI op een manier die ten goede komt aan iedereen, zonder in te boeten op onze waarden? Vanuit een ethisch perspectief belangrijke vragen. Vragen die je zowel op persoonlijk, organisatie- als maatschappelijk niveau kunt stellen: Welke impact heeft het gebruik van AI op privacy en gegevensbescherming? Hoe waarborg je de betrouwbaarheid van de informatie die je van ChatGPT krijgt? Hoe draagt je gebruik van ChatGPT bij aan de verspreiding van vooroordelen? In hoeverre maak je je afhankelijk van ChatGPT bij werk dat eigenlijk kritisch denken vereist? “Het is waardevol om het gesprek over ‘goed gebruik’ met elkaar te voeren. Dat vraagt wel om een pas op de plaats, terwijl velen van ons het gevoel hebben op een snel rijdende trein te zitten.”

Naast goed gebruik, duikt verkeerd gebruik van AI inmiddels ook al regelmatig op, denk aan deepfakevideo’s van bekende

Nederlanders, audioberichten, of met AI gegenereerde muziek met stemmen van beroemde artiesten die hier geen toestemming voor hebben gegeven.

### Doel en middel

“Voor ethische reflectie op techniek is het goed om eens kritisch te kijken naar doel en middel. Generatieve AI is als middel over ons uitgestort zonder dat er een duidelijk doel of een vraagstelling aan ten grondslag lag. Een basale ethische vraag is: hebben we dit nodig? Zo ja, waarvoor? Deze vraag stellen we ons niet. Eigenlijk zouden we een stapje terug moeten doen: is dit een goed middel, voor een goed doel?”

*“Een basale ethische vraag is: hebben we dit nodig?”*

### Impact

Vervolgens is de vraag of een technologie goed uitpakt. Wat is de impact van generatieve AI zoals ChatGPT? Zijn er onbedoelde en negatieve neveneffecten? Zeker zeven belangrijke thema’s komen naar voren, te weten: impact op toetsing, leren en werk, beschikbaarheid en eerlijke verdeling, macht, copyright en milieu.

### Impact op toetsing

“Het gebruik van AI in het onderwijs heeft veel impact en vraagt om aanzienlijke veranderingen, met name op het gebied van toetsing. De traditionele manier van toetsen komt onder druk te staan door de mogelijkheden die AI biedt. Studenten hebben met AI toegang tot een breed scala aan

informatie en kunnen complexe vragen vlot beantwoorden. Dit maakt dat traditionele toetsvormen, gericht op het verwerken van bronnen en schrijven van verslagen, niet meer betrouwbaar functioneren, omdat studenten dit onbedoeld en ongezien door AI kunnen laten doen.”

### **Impact op leren en ontwikkelen**

“Wanneer de deadline voor een essay nadert, een project ingediend moet worden, of een onderzoeksvoorstel gereed moet zijn, is de verleiding groot om een taalmodel zoals ChatGPT in te zetten als hulpmiddel. Tegelijkertijd hechten we veel waarde aan ons persoonlijk leerproces en streven we ernaar onze kennis en inzichten te verdiepen. Hoe kunnen we de voordelen van dergelijke technologieën benutten zonder ons eigen leer- en ontwikkelproces te ondermijnen?”

*“Hoe kunnen we de voordelen van dergelijke technologieën benutten zonder ons eigen leer- en ontwikkelproces te ondermijnen?”*

### **Impact op werken**

“Een klassiek probleem is dat nieuwe technologie bepaalde werkzaamheden overneemt, wat zorgen over de toekomst van werk en baanverlies en met zich mee kan brengen. Hoewel er banen verdwijnen door automatisering, ontstaan er tegelijkertijd nieuwe banen. De verandering in het werklandschap is aanzienlijk, kijk maar naar

de administratieve beroepen, reclame en marketing. Op deze plekken vraagt het al om een heroverweging van vaardigheden en functies.”

### **Beschikbaarheid en eerlijke verdeling**

“AI-technologieën zijn zeker niet voor iedereen betaalbaar en toegankelijk. Houden we hier bij het ontwerp van AI-oplossingen wel voldoende rekening mee? En welke invloed heeft deze diversiteit en beschikbaarheid op ons onderwijs? Moeten we ons niet nog meer inzetten voor het ontwikkelen van AI-beleid dat discriminatie voorkomt en gelijke kansen bevordert?”

### **Macht**

“Grote bedrijven kunnen aanzienlijke controle uitoefenen over de ontwikkeling en implementatie van AI. Dit brengt als risico met zich mee dat AI wordt ingezet op manieren die vooral de belangen van deze bedrijven dienen. Het is daarom van belang dat er transparantie is en verantwoording wordt afgelegd over hoe AI wordt gebruikt en wie ervan profiteert. We zouden machtsconcentratie willen voorkomen, maar wat is daarvoor nodig? Denk aan regelgevende kaders, opensource AI-projecten en het ondersteunen van kleinere ondernemingen om de concurrentie te bevorderen en innovatie te verspreiden.”

### **Copyright**

“Een specifieke uitdaging ontstaat rondom het correct citeren van bronnen. Hoe gaan we dit goed regelen? Traditioneel verwijzen we naar artikelen of boeken die we in ons academisch werk gebruiken, met duidelijke richtlijnen over hoe dit te doen. Echter, de vraag hoe we de input die we van ChatGPT ontvangen correct kunnen citeren, blijft onbeantwoord. Moeten we de prompts die we gebruiken voor het verkrijgen van informatie vermelden?”

“Bovendien is er veel kritiek op het gebruik van bronnen als input voor generatieve AI. Kunnen we zomaar alle boeken en professionele uitingen als collectief gedachtegoed zien; een soort anonieme tekst, beschikbaar voor training van AI? Velen, vooral ook kunstenaars, denken hier heel anders over en zien dit als een roof van intellectueel eigendom.”

### **Milieu**

“Tot slot vraagt de impact van AI op ons milieu om een ethische verkenning. Bedenk dat een enkele internetzoekopdracht al een verrassend grote ecologische voetafdruk heeft door de energie die datacenters verbruiken. Dit is een aanzienlijke tegenvaller voor de perceptie van AI als een 'schone' technologie. Zijn we ons hier wel voldoende van bewust als we een AI-afbeelding maken, ChatGPT een stuk laten herschrijven of een animatievideo met AI ontwikkelen? Hoe kunnen we bewuster met AI omgaan? Of groenere, meer energie efficiënte AI-modellen ontwikkelen?”

### **Disruptieve technologie**

AI haalt veel overhoop en we kunnen het beschouwen als een disruptieve technologie: het biedt compleet nieuwe mogelijkheden en kan bestaande producten en diensten vervangen. “Hoe fantastisch is het dat je met computers kunt praten? En dat je zonder computervaardigheden gebruik kunt maken van alle technologie?” Zo zijn er nog veel meer nieuwe mogelijkheden te benoemen. En minstens evenveel nadelen.

Veel mensen hebben het gevoel dat de trein niet meer te stoppen is en dat de effecten van AI-technologie nauwelijks nog tegen te houden zijn. “Dat is een zorgwekkende gedachte, want als velen zeggen dat iets

toch niet bij te sturen is, dan gaat dat de werkelijkheid worden. We kunnen het met elkaar ook belangrijk vinden om stil te staan bij de vraag hoe we ervoor kunnen zorgen dat AI-technologieën worden ontwikkeld en gebruikt op een manier die eerlijkheid en gelijkheid bevordert, terwijl tegelijkertijd privacy wordt beschermd en sociale en economische ongelijkheid niet wordt versterkt. Onze gezamenlijke houding is hierin bepalend.”

*“Laten we elkaar goed blijven vragen naar het doel, het waarom van AI.”*

### **Visies op techniek**

Bij de ontwikkeling van nieuwe technieken kunnen we ons afvragen of het echt zo moet zijn, en of het ook anders zou kunnen? “Dit vraagt om een stap terug doen. De een zal dit makkelijker doen of aantrekkelijker vinden dan de ander. Dit heeft te maken met ieders perceptie van technologie.”

“Perceptie – ten aanzien van technologie – kunnen we grofweg in drie categorieën onderbrengen. Zo herkennen we een positief, optimistisch en in extreme vorm utopisch gestemde groep mensen, een groep die ambivalent ten opzichte van nieuwe ontwikkelingen staat, meer pragmatisch van aard, en we kunnen een groep ronduit negatief, pessimistisch gestemde mensen onderscheiden, die in extreme vorm dystopisch zijn.”

**Utopisch**

“Mensen met een positief, optimistische blik zijn enthousiast over de mogelijkheden van technologie en zien vooral positieve veranderingen. Zij zien technologische vooruitgang als een middel om bepaalde problemen op te lossen. Bijvoorbeeld het analyseren van grote datasets voor medisch onderzoek. Voorstanders zijn vaak early adopters van nieuwe technologische tools en zullen nu volop experimenteren met LLM's en generatieve AI. Zo zien ze allerlei mogelijkheden voor AI rondom gepersonaliseerd onderwijs en het stimuleren van creativiteit en innovatie. In een ethisch gesprek gaan zij ervan uit dat AI uiteindelijk enkel goede dingen brengt. Het is logisch dat het er zou komen. Dit kunnen we niet tegenhouden, het is geen keuze. Het is een pad naar de toekomst dat al vastligt.”

**Ambivalent**

“Pragmatische technologiegebruikers benaderen technologie met een praktische insteek. Ze zijn niet per se tegen of voor technologische vooruitgang, maar wegen de voor- en nadelen af. Deze groep maakt gebruik van technologie wanneer het hun leven vergemakkelijkt of verbetert, maar blijft kritisch over de impact en de mogelijke afhankelijkheid ervan. Ze zijn selectief in welke technologieën ze adopteren, gebaseerd op hun specifieke behoeften en de toegevoegde waarde.”

**Dystopisch**

“De derde groep betreft sceptici die kritisch tot pessimistisch staan tegenover de impact van technologie en deze groep is bezorgd over de negatieve gevolgen voor de samenleving, zoals inbreuk op privacy, verlies van banen en het verminderen van menselijk contact.

Hoewel sceptici technologie niet volledig afwijzen, benadrukken ze de noodzaak van voorzichtigheid en bewuste implementatie. Ze pleiten vaak voor strenge regulering en ethische overwegingen. Binnen deze groep vallen ook de uitgesproken tegenstanders van technologische vooruitgang. Zij geloven dat technologie leidt tot meer kwaad dan goed, en ze wijzen vaak op de negatieve aspecten van technologie, zoals machtsverhoudingen en het verlies van traditionele vaardigheden.”

**Ethische reflectie**

“In elke onderwijsomgeving, organisatie of andere setting kom je de verschillende houdingen wel tegen: voorstanders, pragmatische gebruikers, sceptici en critici. Mensen in de techniek zijn vaak optimistisch over techniek, maar er zijn juist ook heel kritische technici die zich verbazen over hoe het grote publiek op alle AI-mogelijkheden springt en over managers die zeggen: ‘We moeten dit wel omarmen.’”

“De voorlopige oplossing kan zijn: ook als je niet zomaar kan stoppen of bijsturen, is het toch belangrijk om ruimte te maken voor een goed gesprek en kritische bijgedachten. Ethiek gaat over de waarden die we belangrijk vinden en in het geval van AI over het verantwoord ontwikkelen en inzetten van AI op een manier die eerlijkheid, privacy en het welzijn van alle betrokkenen respecteert. Eigenlijk is ethische reflectie heel belangrijk voordat een ontwikkeling is gestart. In het geval van AI is aandacht voor kritisch blijven denken cruciaal. Zelfs als de technologie niet te stoppen valt, moeten we insteken op goed gebruik ervan. Laten we elkaar goed blijven vragen naar het doel, het waarom van AI.” ♦

**Sibrenne Wagenaar** is redactielid van TvOO.