AI tussen snelheid en wijsheid: de menselijke maat als harde randvoorwaarde

René de Baaij
partner bij Uātēs, The art of leadership

Inleiding

Kunstmatige intelligentie versnelt processen en vergroot onze mogelijkheden, maar zonder duidelijke grenzen schuift efficiëntie al snel over wijsheid heen. De vraag is niet wat kan, maar wat moet en wenselijk is. Wie de menselijke maat centraal zet, ontwerpt en gebruikt AI op een manier die waardigheid, autonomie en verbondenheid versterkt.

AI is in korte tijd doorgedrongen tot zorg, rechtspraak, onderwijs en bedrijfsvoering. Algoritmen analyseren gedrag, doen voorspellingen en sturen keuzes. De belofte is reëel: sneller, nauwkeuriger, toegankelijker. Tegelijk schuurt er iets. Wie bepaalt hoe beslissingen tot stand komen? Welke vooroordelen liften mee in data en modellen? Wat gebeurt er met vakmanschap als we taken uitbesteden aan systemen die sneller en overtuigender lijken dan wij? De kern is geen technisch detail maar een maatschappelijke keuze: hoe houden we de menselijke maat leidend in systemen die door schaal en rekenkracht ons beoordelingsvermogen kunnen verdringen?

Dit artikel stelt één vraag centraal: wat hebben leiders, ontwerpers en professionals vandaag te doen om AI te richten op menselijkheid en rechtvaardigheid, zodat innovatie geen botsing wordt met de waarden die werk en samenleving dragen?

Spanning die er toe doet

Mijn stelling: “In het AI-tijdperk is ‘meer controle door systemen’ een verleiding. Wat werkt is ‘meer verantwoordelijkheid door mensen’. Efficiëntie zonder ethiek is versnelling zonder richting.”

Algoritmen optimaliseren op wat je ze vraagt en op wat je ze voedt. Wanneer het doel impliciet snelheid of kosten wordt, krijgt alles wat lastig meetbaar is – menselijke waardigheid, context, relationele schade – te weinig gewicht. Bias in trainingsdata versterkt bestaande patronen: ongelijkheid wordt rationeel verpakt. Intussen werkt psychologische besmetting: als systemen overtuigend spreken, nemen we hun uitkomst eerder over, ook wanneer de onderliggende aannames onduidelijk zijn. “Het model zal wel gelijk hebben” is een stille verschuiving van verantwoordelijkheid. Leiderschap dat werkt, draait dit om: maak

doelen expliciet, scheid hulpmiddel en besluit, en laat zien waar de grens loopt tussen advies en oordeel. Zo wordt wijsheid niet ingehaald door snelheid. Dat is niet alleen een morele keuze; toonaangevende kaders vragen aantoonbaar menselijk toezicht, uitlegbaarheid en herstelpaden.

Toepassing

Leg vooraf vast welke beslissingen principieel in menselijke handen blijven en op basis van welke waarden. Spreek per toepassing drie expliciete criteria af, bijvoorbeeld rechtvaardigheid, proportionaliteit, herstelbaarheid, en verwijs er in elk besluit naar. Houd ‘mens in de lus’ niet alleen procedureel maar reëel: professionals krijgen tijd en taal om af te wijken, en afwijkingen worden niet afgestraft maar onderzocht. De EU-AI-verordening dwingt dit de komende jaren bovendien stapsgewijs af, met hogere eisen voor hoog-risico-systemen en verplicht menselijk toezicht (Inwerkingtreding 1 augustus 2024; eerste verboden praktijken sinds 2 februari 2025).

Onderstroom zichtbaar maken

Onder techniek bewegen rollen, loyaliteiten en projecties. Bestuurders kunnen al te veel hoop projecteren op technologie als ‘neutrale scheidsrechter’. Ontwikkelteams voelen loyaliteit naar snelheid en elegantie, gebruikers naar gemak, toezichthouders naar zekerheid. Zonder taal voor die onderstroom ontstaat schijnobjectiviteit (“het systeem zegt het”) of defensie (“wij kunnen hier niets meer aan doen”). Benoem wie welke verantwoordelijkheid draagt, waar belangen botsen, en welke waarden zwaarder wegen dan metrics. Dat maakt het gesprek eerlijk en voorkomt dat het morele oordeel verdampt in een zwarte doos. In de rechtspraak is dit expliciet gemaakt in Europese beginselen: grondrechten, non-discriminatie, kwaliteit en veiligheid, transparantie en gebruikerscontrole.

Een praktijkvoorbeeld

Een ziekenhuis onderzoekt AI-ondersteuning bij diagnostiek. Het model herkent patronen op scans en doet aanbevelingen. Artsen waarderen de snelheid, maar vrezen dat twijfel en context minder ruimte krijgen.

Het ziekenhuis definieert een helder doel: snellere en betere triage zonder verlies van medisch oordeel. Er komt een beslisraamwerk met drie vaste vragen bij elk advies: wat is de zekerheid van het model, welke contextinformatie ontbreekt, en wat betekent dit voor risico en herstelbaarheid? Het team maakt een afwijkingslogboek waarin artsen kunnen vastleggen waarom zij afwijken en wat zij anders wegen. Er is een maandelijks interdisciplinair overleg met patiëntenvertegenwoordiging waarin patronen, bias-signalen en effecten worden besproken.

Diagnostische doorlooptijden dalen, maar vooral stijgt het vertrouwen: artsen voelen zich gesteund in plaats van vervangen, patiënten ervaren beter uitgelegde besluiten. Afwijkingen blijken leerzaam: ze leveren concrete verbeteringen op voor model en proces. Dit sluit aan bij

internationale zorgethiek-richtlijnen: technologie ondersteunt, de mens beslist, met aandacht voor rechtvaardigheid, inclusie en herstel.

De drie B's van menselijke maat in AI: bedoeling, begrenzing, bewijslus

Bedoeling maakt expliciet voor welk menselijk goed we de technologie inzetten; niet 'zo snel mogelijk', maar bijvoorbeeld 'rechtvaardige toegang tot zorg' of 'betere onderbouwing van besluiten'.

Begrenzing legt rode lijnen vast: beslissingen die niet worden geautomatiseerd, groepen die extra bescherming krijgen en contexten waarin inzet onwenselijk is.

Bewijslus sluit de cirkel: we toetsen systematisch op bias en effect, bieden bezwaar en herstel en leren zichtbaar van afwijkingen. Deze lens operationaliseert internationale principes: mensgericht, transparant, verantwoord en met continue risicobeheersing.

Wat zegt de wereld van normen en toezicht en wat betekent dat morgen?

Europa heeft een alomvattende AI-verordening. De kern: risicogebaseerde eisen, documentatie en logging, datakwaliteit, menselijke toezichtmechanismen en transparantie. De Europese Commissie heeft een AI Office ingericht voor coördinatie en handhaving. Wie in Europa met AI werkt, borgt nu governance, documentatie en human oversight.

Naast wetgeving bieden internationale kaders houvast voor dagelijks handelen. Het NIST AI Risk Management Framework biedt een praktisch vocabulaire en proces voor risico-identificatie met nadruk op transparantie, validatie, monitoring en mensgerichte doelen. De OECD-AI-principes (2019, geactualiseerd 2024) geven ankerpunten: mensgericht, robuust, transparant, verantwoord en veilig. UNESCO's Aanbeveling over AI-ethiek (2021) verankert mensenrechten en maatschappelijke impact.

In sectoren bestaan aanvullende richtlijnen. De WHO formuleert zes leidende principes voor AI in de zorg. In het onderwijs geeft UNESCO (2023) een mondiale handreiking voor generatieve AI. In de rechtspraak geeft de Raad van Europa (CEPEJ) een ethisch charter met randvoorwaarden voor legitimiteit en rechtsbescherming.

Recht en rechten: waar de grens loopt

De AVG (art. 22) geeft iedere burger het recht om niet uitsluitend aan een geautomatiseerd besluit te worden onderworpen dat rechtsgevolgen heeft of een vergelijkbare significante impact. Dat vraagt om ontwerpkeuzes: betekenisvol menselijk toezicht, uitlegbaarheid voor de betrokkene en reële mogelijkheid tot bezwaar en herstel, niet alleen 'mens in de lus' op papier, maar aantoonbaar in de praktijk.

Van principes naar praktijk: documenteren, toetsen, verbeteren

Twee concrete praktijken maken transparantie en verantwoordelijkheid werkbaar: datasheets voor datasets en model cards voor modellen. Datasheets leggen herkomst, samenstelling,

aannames en beperkingen vast. Model cards documenteren prestaties, intended use en fairness-bevindingen. Daarmee wordt uitleg aan bestuurders, teams en belanghebbenden concreet en wordt auditen mogelijk.

Let op de schaal van taalmodellen

Grote taalmodellen leveren indrukwekkende resultaten, maar brengen reële risico's mee: datalekken, plausibele onzin, versterking van bias en ecologische kosten. Bender & Gebru et al. (2021) waarschuwen voor 'stochastic parrots': overtuigende tekst zonder begrip. Aanbevelingen: documentatie van data, duidelijke grenzen aan gebruik en onderzoeken buiten de route van 'steeds groter'.

Besturing en certificering

Als je structureel wilt borgen dat AI 'onder controle' blijft, helpt een managementsysteem. ISO/IEC 42001:2023 introduceert het AI-managementsysteem (AIMS) met rollen, processen, monitoring en continue verbetering. Geen vinklijstje, wél een manier om richting te geven aan doelen, risico's en verbetercycli die sporen met wet- en regelgeving.

Publieke legitimiteit vraagt publieke dialoog

Draagvlak ontstaat niet in de boardroom alleen. Onderzoek van Ada Lovelace Institute en The Alan Turing Institute laat zien dat burgers AI willen die zichtbaar rechtvaardig, uitlegbaar en toetsbaar is; regulering vergroot comfort, mits die voelbaar is in praktijkervaring. Organiseer daarom participatieve vormen (burgerpanels, beroepsgroepen, cliëntenraden) en maak zichtbaar wat je leert.

Van reflectie naar beweging

Welke beslissingen in onze organisatie zijn te belangrijk om te automatiseren, juist omdat waardigheid, rechtvaardigheid of vertrouwen op het spel staan? Kijk de komende week bewust naar drie momenten waarop jij of je team 'het systeem' als argument gebruikt en onderzoek waarvoor dat een shorthand is: tijdsdruk, onzekerheid of onuitgesproken belangen. Hanteer bij elke nieuwe AI-toepassing één beslischeck: kunnen we in twee zinnen uitleggen aan een kritische buitenstaander hoe dit de menselijke maat versterkt, en welke grens we hebben getrokken waar het systeem níet beslist?

AI wordt richtinggevend waar wij richting geven. Kies vandaag één toepassing, benoem doel en grens, en start met een eenvoudige bewijslus, zodat snelheid en wijsheid elkaar versterken en de menselijke maat geen bijzaak is, maar de randvoorwaarde waaronder innovatie mag versnellen.

Bronnen

- Ada Lovelace Institute, & The Alan Turing Institute. (2023). *How do people feel about AI?*https://www.adalovelaceinstitute.org/wp-content/uploads/2023/06/Ada-Lovelace-Institute-The-Alan-Turing-Institute-How-do-people-feel-about-AI.pdf
- Ada Lovelace Institute, & The Alan Turing Institute. (2025). *How do people feel about AI? 2025 survey findings.*https://attitudestoai.uk/assets/documents/How-do-people-feel-about-AI-2025-Ada-Lovelace-Institute.pdf
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. https://doi.org/10.1145/3442188.3445922
- Council of Europe, CEPEJ. (2018). *European Ethical Charter on the use of artificial intelligence in judicial systems and their environment.*https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment
- European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 22. *Official Journal of the European Union.*https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- European Union. (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). *Official Journal of the European Union.*https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- ISO/IEC. (2023). *ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system.*https://www.iso.org/standard/42001
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., … Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*'19)*, 220–229. https://doi.org/10.1145/3287560.3287596
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1).*https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
- OECD. (2019/2024). *OECD AI Principles.*https://oecd.ai/en/ai-principles
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.*https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
- UNESCO. (2023). *Guidance for generative AI in education and research.*https://unesco.org.uk/site/assets/files/10375/guidance_for_generative_ai_in_education_and_research.pdf
- World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance.*https://www.who.int/publications/i/item/9789240029200
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.*PublicAffairs. https://www.hachettebookgroup.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/

- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press. https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/
- Hildebrandt, M. (2020). *Law for computer scientists and other folk.* Oxford University Press. https://academic.oup.com/book/33735
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://dl.acm.org/doi/fullHtml/10.1145/3458723