Radiologie, AI en ethiek

Hendrik Erenstein

De opkomst van AI binnen de radiologie is niemand ontgaan. En hoewel AI-ontwikkeling langzaam overgaat naar klinische implementatie, blijven er veel uitdagingen liggen. Niet alleen technische aangelegenheden, maar ook op ethisch vlak. Hoewel er aan ethiek aandacht besteed wordt,[I] bestaat het risico dat dit gebied ondersneeuwt in de vele technische en klinische publicaties.

De ethische uitdagingen met betrekking tot het gebruik van AI in de zorg zijn niet altijd duidelijk. Aangezien AI getraind wordt op data die wij aanleveren, en deze data inherente bias bevatten, zijn veel voorbeelden niet verrassend; zo is de invloed van geslacht op de voorspellingen die AI-modellen kunnen doen bij de diagnose van longfoto's, in verschillende studies beschreven,[1] en zijn taalmodellen vatbaar voor vooroordelen met betrekking tot onder andere geslacht, geloof en geaardheid.[2] Omdat AI echter vaak gebruikmaakt van andere patronen dan mensen, zijn er voorspellingen die met behulp van AI gedaan worden die minder vanzelfsprekend zijn. Onderzoekers hebben bijvoorbeeld aangetoond dat het mogelijk is om naast geslacht ook etniciteit en verzekeringsstatus te voorspellen op basis van een thoraxopname.[3] Hoe AI-modellen dit doen, blijft voor een groot deel echter een *black box*.

Ethische principes

Gezien de complexiteit en de black box zijn er verschillende richtlijnen opgesteld ten behoeve van ethische AI-implementatie in de zorg. Een voorbeeld zijn de 6 ethische principes beschreven door de Wereldgezondheidsorganisatie (WHO):

1) beschermen van autonomie
2) bevorderen van menselijk welzijn, menselijke veiligheid en het algemeen belang
3) zorg voor transparantie, uitlegbaarheid en begrijpelijkheid
4) bevorderen van verantwoordelijkheid en aansprakelijkheid
5) zorg voor inclusiviteit en gelijkheid en
6) bevorderen van AI die responsief en duurzaam is.[4]

Door te leunen op deze 6 principes wordt bijgedragen aan een zorgvuldige ontwikkeling en implementatie van AI in de zorgketen.

Verklaarbare AI

Deze 6 principes zijn echter gelaagd. Neem bijvoorbeeld autonomie. Autonomie, zoals beschreven door de WHO in de bovengenoemde richtlijnen, is de mogelijkheid voor clinici om de uitkomsten van AI met hun eigen expertise te beoordelen en, indien relevant geacht, te overschrijven. De rol van autonomie draagt echter verder. Indien professionals zich aangetast voelen in hun professionele identiteit kan dit weerstand oproepen en de adoptie van AI verhinderen.[5] Het is juist de adoptie van AI in de zorg die kan bijdragen aan het behoud of het verbeteren van de kwaliteit van zorg. Transparantie kan bijdragen aan een oplossing: door het vergroten van de verklaarbaarheid *(explainable AI)* kan de professionele identiteit gehandhaafd blijven.[5]

100 indicatoren

Hoewel veel eindgebruikers alleen in aanraking komen met de gebruikersinterface en/of voorspellingen van AI, en dus bijvoorbeeld explainable AI zoals hiervoor geschetst, biedt ook transparantie vele uitdagingen. In hoeverre de onzekerheid van een voorspelling aangeduid moet worden, blijft een punt van discussie.[6] Transparantie blijft echter niet beperkt tot explainable AI: de Foundation Model Transparency Index met meer dan 100 verschillende indicatoren geeft hier een indruk van.[7] De indicatoren variëren van maatschappelijke risico's tot rekenkracht, van data tot maatschappelijke impact en geven een beeld van de complexiteit die transparantie brengt.

Taalmodellen

De uitdagingen worden versterkt in lage- en middeninkomenslanden.[8] Veel regio's, zoals de sub-Sahara, kampen met hoge ziektelast in combinatie met hoge personeelstekorten in de zorg.[9] De implementatie van AI kan in deze regio's een aanzienlijke bijdrage leveren aan de toegankelijkheid van de zorg. Het afstemmen van AI op de lokale bevolking wordt echter beperkt door de beperkte infrastructuur en beschikbaarheid van kwalitatief hoogwaardige data. Enerzijds kan AI dus de toegang tot de zorg versterken, terwijl anderzijds de bestaande ongelijkheid wordt versterkt.[9]

Ondanks de uitdagingen is er een toenemende hoeveelheid AI-applicaties in de klinische workflow zichtbaar.[10,11] En hoewel de nadruk ligt op diagnose zien we AI terug in beeldoptimalisatie, MRI-scanversnelling, automatische positionering op CT en autocollimatie voor conventionele radiologie. Deze toepassingen bevinden zich echter allemaal in het beelddomein, maar het taaldomein blijft niet achter. Ontwikkelingen met betrekking tot taalmodellen (*large language models* of LLM's) hebben een vlucht genomen sinds de release van ChatGPT in 2022.

Vooroordelen

De implementatie van LLM's is veelzijdig en richt zich onder andere op verslaglegging, patiëntcommunicatie en onderzoek.[12] En hoewel deze modellen waardevolle input genereren, hebben ze ook belangrijke beperkingen. Veel LLM's lijken last te hebben van stereotypen, *gender-bias* en andere vooroordelen.[2,13-15] Deze vooroordelen kunnen in stand blijven omdat men geneigd is om automatisch gegenereerde suggesties over te nemen, een *automation*-bias. Dit probleem wordt groter voor patiënten voor wie de toegang tot taalmodellen beperkt is tot bekende platformen zoals ChatGPT of Deepseek. Deze modellen maken het mogelijk om het klinische taalgebruik voor patiënten begrijpelijk te maken, maar men moet rekening houden met de culturele context waarin deze modellen zijn ontwikkeld.[15]

[I] Zie bijvoorbeeld de bijdrage van Jojanneke Drogt in MemoRad 29-3.

Eetstoornis

De noodzaak voor een kritische houding ten aanzien van het gebruik van LLM's wordt verder onderstreept door gebruikerservaringen. In een onderzoek gepubliceerd door OpenAI, de ontwikkelaar van ChatGPT, wordt een emotionele afhankelijkheid van ChatGPT onder gebruikers beschreven.[16] Naast de emotionele afhankelijkheid wordt ook een toenemende mate van eenzaamheid beschreven onder mensen die deze vorm van AI veel gebruiken. Het is kritiek om stil te staan bij deze emotionele afhankelijkheid van een LLM, des te meer voor mensen die zich in een kwetsbare positie bevinden. Een voorbeeld betreft de *National Eating Disorder Association* (NEDA) uit de Verenigde Staten. Deze organisatie heeft in 2023 een chatbot ingezet om mensen met een eetstoornis te ondersteunen. Helaas bleek deze applicatie al snel adviezen te geven voor het verliezen van lichaamsgewicht bij mensen met een eetstoornis.[17] De noodzaak van de human in the loop, het goedkeuren van AI-output door een expert, bij de huidige systemen moet hiermee duidelijk zijn.

Energieverbruik

Desondanks deze beperkingen zien we dat LLM's een plek hebben ingenomen in de maatschappij. Deze revolutie met betrekking tot het verwerken van informatie brengt ook een significante vraag naar energie met zich mee.[18] In hoeverre deze technieken kunnen zorgen voor een balans tussen informatieverwerking en energievraag is echter nog veel onduidelijk.[19] In de context van radiologie kan gedacht worden aan reductie in energiegebruik die MRI-scanversnellingstechnieken kunnen bieden. Hoewel het energieverbruik met 50% is te reduceren, wordt de vrijgekomen ruimte wellicht gevuld met nieuwe onderzoeken.[20,21] Vanuit duurzaamheidsperspectief is er veel potentie, maar het blijft onduidelijk wat het effect uiteindelijk kan zijn. Hiermee biedt Jevon's Paradox een relevant perspectief; door een toename van efficiëntie en kostenreductie zal de uiteindelijke vraag kunnen toenemen.

Weloverwogen keuzes

Iedereen die de ontwikkelingen op het gebied van AI enigszins volgt, weet dat er veel gebeurt. En hoewel de implementatie van AI in de zorg achterblijft op de ontwikkelingen in de commerciële sector, worden er weloverwogen keuzes gemaakt. Dit maakt dat kennis op het gebied van AI voor professionals, AI-geletterdheid, een punt van aandacht is; zonder kennis is het afwegen van risico's en kansen immers moeilijk. Artikel 4 van de Europese AI-verordening legt een verantwoordelijkheid bij de 'aanbieders en gebruiksverantwoordelijken' neer.[22] Maar ook onderwijzers moeten zich bewust zijn van nieuwe competenties die dit tijdperk met zich meebrengt. Mogelijk is het zelfs relevant om te kijken naar een extra differentiatie of taakherschikking voor radiologen, MMB'ers, klinisch fysici en andere bij radiologie betrokken professionals.

Maatschappelijk belang

Deze bijdrage schiet tekort in het belichten van de complexiteit van AI-implementatie in de zorg, om nog maar te zwijgen over de ethische uitdagingen. Het is echter van cruciaal belang om AI en de onderliggende complexiteit te benadrukken. De implementatie van AI beperkt zich immers niet tot ons vakgebied; de besproken onderwerpen zijn van maatschappelijk belang. ■

Hendrik Erenstein

Referenties

1. Larrazabal AJ, Nieto N, Peterson V, et al. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A*. 2020 Jun 9;117(23):12592-4.
2. UNESCO I. Challenging systematic prejudices: an investigation into gender bias in large language models. 2024.
3. Adleberg J, Wardeh A, Doo FX, et al. Predicting patient demographics from chest radiographs with deep learning. *Journal of the American College of Radiology*. 2022 Oct;19(10):1151-61.
4. World Health Organization. Ethics and governance of articial intelligence for health: WHO guidance. 2021.
5. Shonhe L, Min Q. Mitigating AI-induced professional identity threat and fostering adoption in the workplace. *AI & society*. 2025 Jan 15.
6. Gill A, Rainey C, McLaughlin L, et al. Artificial intelligence user interface preferences in radiology: A scoping review. *Journal of Medical Imaging And Radiation Sciences*. 2025 May 1;56(3):101866.
7. Bommasani R, Klyman K, Kapoor S, et al. The foundation model transparency index v1.1: May 2024. 2024 Jul 17.
8. López DM, Rico-Olarte C, Blobel B, et al. Challenges and solutions for transforming health ecosystems in low- and middle-income countries through artificial intelligence. *Frontiers In Medicine*. 2022 Dec 2;9:958097.
9. Victor A. Artificial intelligence in global health: An unfair future for health in Sub-Saharan Africa? *Health Affairs Scholar*. 2025 Feb 5;3(2):qxaf023.
10. Obuchowicz R, Lasek J, Wodziński M, et al. Artificial intelligence-empowered radiology-current status and critical review. *Diagnostics (Basel)*. 2025 Jan 24;15(3):282.
11. Zanardo M, Visser JJ, Colarieti A, et al. Impact of AI on radiology: a euroaim/eusomii 2024 survey among members of the European Society of Radiology. *Insights Imaging*. 2024 Oct 7;15(1):240-13.
12. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagnostic and interventional radiology (Ankara, Turkey)*. 2024 Mar 1;30(2):80-90.
13. Levy S, Adler WD, Karver TS, et al. Gender bias in decision-making with large language models: a study of relationship conflicts. 2024 Oct 14.
14. Gender bias and stereotypes in large language models New York, NY, USA: ACM; Nov 6, 2023.
15. Naous T, Ryan MJ, Ritter A, et al. Having beer after prayer? Measuring cultural bias in large language models. *arXiv (Cornell University)*. 2023 May 23.
16. Phang J, Lampe M, Ahmad L, et al. investigating affective use and emotional well-being on chatgpt. 2025 Apr 4.
17. Chelsea Bailey. Eating disorder group pulls chatbot sharing diet advice. 2023; Available at: https://www.bbc.com/news/world-us-canada-65771872. Accessed 26 April, 2025.
18. Chen S. How much energy will AI really consume? The good, the bad and the unknown. *Nature*. 2025 Mar 6;639(8053):22-4.
19. Tomlinson B, Black RW, Patterson DJ, et al. The carbon emissions of writing and illustrating are lower for AI than for humans. Sci Rep. 2024 Feb 14;14(1):3732.
20. Afat S, Wohlers J, Herrmann J, et al. Reducing energy consumption in musculoskeletal MRI using shorter scan protocols, optimized magnet cooling patterns, and deep learning sequences. Eur Radiol. 2025 Apr 1;35(4):1993–2004.
21. Alerte Z, Chodorowski M, Ammari S, et al. Towards sustainable magnetic resonance neuro imaging: pathways for energy optimization and cost reduction strategies. Applied Sciences. 2025 Feb 1;15(3):1305.
22. European Parliament, Council of the European Union. Artificial Intelligence Act. http://data.europa.eu/eli/reg/2024/1689/oj 2024 13 June.

Wie is Hendrik Erenstein?

Hendrik Erenstein werkt als docent-onderzoeker bij de medisch beeldvormende en radiotherapeutische technieken van de Hanze in Groningen. Sinds 2022 combineert hij zijn interesse in straling, radiologie en AI[II] in een PhD-project waarin hij de impact van beeldkwaliteit op AI-voorspellingen onderzoekt. Hoewel het project een technische focus heeft, besteedt hij ook aandacht aan AI-gerelateerde thema's, zoals gelijkheid, verantwoordelijkheid en onderwijs. Dit doet hij in samenwerking met collega's en organisaties zoals de EFRS en EuSoMII.

[II] De term AI is een containerbegrip, maar wordt doelbewust gebruikt omdat de uitdagingen alle sub-domeinen beslaan.