TLP:CLEAR

AI in de zorg

Documentinformatie

Datum	14 juni 2024
Auteur(s)	Werkgroep AI
Versie	1.0
Locatie	https://z-cert.nl/

Z CERT
COMPUTER EMERGENCY RESPONSE TEAM VOOR DE ZORG

AI in de zorg

Kunstmatige intelligentie (AI) is de innovatie waar iedereen het over heeft. Hoe gaat het ons kunnen helpen en wat zijn de risico's van het gebruik ervan?

Inleiding

Veel bedrijven streven ernaar de markt te verstoren met innovatie. Kunstmatige intelligentie is, net zoals de computer, het internet en social media, een maatschappij verstorende innovatie. De vergelijking met social media is erg passend aangezien het onze maatschappij heel veel goeds heeft gebracht, maar ook de nodige negatieve effecten heeft op hoe wij als mensen met elkaar omgaan. Feitelijke bronnen en onderbouwde argumenten doen er helaas minder toe op social media, eerder om wie het hardst schreeuwt (1). Populaire AI-chatbots, zoals ChatGPT, gebruiken ons taalgebruik op het internet als trainingsdata (2). Het is erg indrukwekkend wat deze AI-chatbots weten te genereren als antwoorden op onze vragen, maar onthoud dat het getraind is op hoe mensen online met elkaar praten. Helaas staat het internet vol met desinformatie over de zorg (3) en daarmee dus ook de AI-chatbots die nu gebruikt worden. Kunstmatige intelligentie is zo goed als de dataset waarop het getraind is. In de wereld is er nu een race gaande in wie de beste en grootste dataset heeft (4). Bedrijven en onderzoekers zijn erg geïnteresseerd in het trainen van AI op onze zorgdata (5). Terwijl veel security- en privacyvragen nog onbeantwoord zijn. Daarom heeft Z-CERT het volgende advies:

"Wees waakzaam met de inzet van AI in de zorg."

Leeswijze

Dit document zal eerst ingaan op wat vanuit Z-CERT het voornaamste risico is van de inzet van AI in de zorg. Daarna zal het inzicht geven in wat andere partijen zeggen over AI. Ter afsluiting geeft dit document het nodige handelingsperspectief.

Het risico van AI in de zorg

Het grootste risico van AI in de zorg is dat de output hiervan voor waarheid wordt aangenomen zonder de controle van een expert. Ook het verlies van transparantie en daarmee de menselijke controle op gezondheidsprocessen is een groot risico. Assistentie door AI, bijvoorbeeld via een app op de telefoon, is geen vervanging voor professioneel medisch advies, diagnose, of behandeling (6). In Nederland moet je bevoegd zijn om medische hulpmiddelen, waaronder software, te mogen toepassen (7).

Daarnaast ziet Z-CERT indirect een verhoogde impact van ransomware en datalekken bij de inzet van AI. Ransomware en datalekken zijn een bekend risico voor de zorg (8). Hoe meer databronnen we (voor AI) aan elkaar verbinden, hoe groter de impact van ransomware. Hoe meer data we (voor AI) opslaan op een centrale plek, hoe groter de impact van een datalek. Hoe afhankelijker wij worden van AI, hoe groter de impact voor de zorg als het mis gaat.

Wat zeggen anderen over AI

De meningen over de inzet van AI verschillen. Om tot een goed oordeel te komen over de inzet van AI is het belangrijk om bekend te zijn met wat anderen zeggen over AI en de nodige voor- en tegenargumenten mee te nemen in je eigen risicoanalyse. Hieronder per kopje een korte samenvatting van de verschillende meningen en perspectieven over de inzet van AI, met directe bronverwijzingen naar verdiepend materiaal.

Patiënten

Als aan patiënten wordt gevraagd wat zij belangrijk vinden, is dat bij het (her)gebruik van gezondheids-gegevens er niet meer dan de relevante gegevens worden gebruikt en dat er transparantie is over het gebruik. Patiënten stellen specifieke voorwaarden die sterk afhankelijk zijn van de context waarin gegevens worden (her)gebruikt (9). Hoewel digitale zorg als steeds normaler wordt gezien, maken veel patiënten zich zorgen over hun privacy en de betrouwbaarheid van gezondheidsinformatie (10). Patiënten verwachten absolute vertrouwelijkheid in de spreekkamer (11).

Huisartsen
Bij huisartspraktijken is al goed te zien hoe de AI-chatbot de doktersassistent aan het vervangen is. De werkdruk wordt verlicht door assistentie in het beoordelen van botbreuken en door vermindering van administratie in de spreekkamer (11).

Ziekenhuizen
De betrokkenheid bij en kennis over AI is de afgelopen jaren bij ziekenhuizen gegroeid. AI wordt voornamelijk nog ingezet als diagnostisch hulpmiddel. Onderwerpen als integratie, acceptatie door gebruikers en validatie van AI krijgen bij ziekenhuizen de meeste aandacht. Onderwerpen als ethiek, privacy en security staan onderaan het rijtje en krijgen de minste aandacht (12).

Universitair Medische Centra
Het Erasmus MC heeft in samenwerking met SAS en Microsoft datagedreven applicaties ontwikkeld en geïmplementeerd in het ziekenhuis. Zo hebben ze meerdere disciplines zoals IT, data science en de zorg bij elkaar gebracht om een AI-algoritme te ontwikkelen dat voorspelt of patiënten na een operatie veilig ontslagen kunnen worden. Data wordt naast het terugkijken ook gebruikt om vooruit te kijken en zo de dagelijkse werkzaamheden te verbeteren en veiliger te maken (13). Met Microsoft Copilot wordt de populaire OpenAI chatbot GPT-4 steeds meer verweven in de producten van Microsoft, zoals Word en Outlook. Elektronisch patiëntendossier leverancier Epic is een samenwerking aangegaan met Microsoft om de AI-chatbot ook te verweven in het EPD. Een toepassing hiervan is al te zien bij het UMC Groningen, waarbij de AI-toepassing van het EPD de vragen van de patiënt leest en antwoordsuggesties geeft aan zorgverleners (14).

Azure OpenAI
Hoewel Microsoft's Azure OpenAI zegt dat er niet naar je data gekeken wordt, gebeurt dit wel wanneer je woorden gebruikt over bijvoorbeeld haat, seksualiteit en geweld. Azure OpenAI controleert namelijk al het gebruik van het AI-systeem met abuse monitoring (15). Hiervoor is een opt-out aan te vragen, maar het is niet gegarandeerd dat je deze ook krijgt.

Langdurige zorg
In de langdurige zorg wordt AI ingezet in bijvoorbeeld toezichthoudende domotica (het herkennen van geluiden om te bepalen of er al dan niet actie vereist is door een zorgmedewerker), sociale robots en veel spraak naar tekst-toepassingen in zorgapplicaties. Veelal is dit nog in een experimentele fase (16).

Overheidsbrede visie generatieve AI
Onze overheid heeft een visie opgesteld over generatieve AI. Generatieve AI wordt in deze visie beschreven als een vorm van AI waarbij algoritmes worden ingezet om content te genereren. Het document behandeld naast de kansen en mogelijkheden ook de uitdagingen en de risico's van de inzet van generatieve AI. Volgens het document heeft generatieve AI al bewezen dat het sneller en op grotere schaal kan bijdragen aan de creatie en verspreiding van mis- en desinformatie. Daarnaast is er een groeiende afhankelijkheid van Amerikaanse techbedrijven. Over privacy- en gegevensbescherming beschrijft de visie dat er in principe een verbod is op de verwerking van bijzondere persoonsgegevens, zoals biometrische gegevens en gegevens over gezondheid, tenzij aan strike voorwaarden uit de AVG wordt voldaan (17).

VWS
Het ministerie van Volksgezondheid, Welzijn en Sport heeft als standpunt dat de toepassing van AI op zorgdata ons kan helpen in de zorg (18). Er dient een geïntegreerde nationale infrastructuur te komen voor gezondheidsdata, om (her)gebruik van gezondheidsdata voor zorgevaluatie, kwaliteit, beleid, onderzoek en innovatie te bevorderen (5). Ook dient binnen Europa, met het European Health Data Space (EHDS) voorstel, zorgdata breed gekoppeld en gedeeld te kunnen worden. Niet alleen voor primaire doeleinden voor het verlenen van zorg, maar ook voor secundaire doeleinden als onderzoek en de ontwikkeling van nieuwe producten (19). Het vergroten van de databeschikbaarheid is van belang om de zorg voor de toekomst toegankelijk en betaalbaar te houden. Vertrouwen is hierin een belangrijke voorwaarde. Zoals verwoord in de Nationale Visie op het Gezondheids-informatiestelsel is regie nodig om het vertrouwen te doen groeien (20).

Burgerrechten
Volgens burgerrechten organisaties EDRi en Privacy First stelt de European Health Data Space de medische dossiers van iedereen bloot aan onnodige beveiligings- en privacyrisico's in de naam van onderzoek en innovatie. Hiermee wordt het medisch beroepsgeheim bedreigd (21). In de onderhandelingen rond het European Health Data Space is er nu een akkoord bereikt over een opt-out optie voor de patiënt die niet wil dat zijn of haar zorgdata standaard wordt gedeeld (22) (23).

De Europese AI-verordening
Met de komende AI-act geeft het Europese Parlement er prioriteit aan dat AI-systemen die in de EU worden gebruikt veilig, transparant, traceerbaar, niet-discriminerend en milieuvriendelijk zijn. Binnen de AI-act worden AI-toepassingen geclassificeerd op hoe risicovol ze zijn voor de maatschappij. AI-toepassingen in medische hulpmiddelen worden onder de AI-act geclassificeerd als een hoog risico voor de maatschappij (24). Op verzoek van VWS is er onderzoek gedaan naar de overlap en inconsistenties van de AI-act met de MDR en IVDR regulering van medische apparaten (25).

ENISA
Het agentschap voor cybersecurity van de Europese Unie (ENISA) heeft een uitgebreid rapport geschreven over security en privacy bij de toepassing van AI in medische diagnoses. ENISA waarschuwt voor bedrijven die zorgdata onwettig proberen te bemachtigen en onrechtmatig willen verwerken. Denk aan bedrijven die niet transparant zijn en bijvoorbeeld geen respect hebben voor data minimalisatie, de correctheid van data en de beperkte opslag ervan (26).

Health-ISAC
Het Health Information Sharing and Analysis Center (H-ISAC) van de Verenigde Staten raadt de zorgsector aan om een conservatieve houding aan te nemen tegenover de inzet van AI. We kunnen kijken naar de lessen die we hebben geleerd tijdens de migratie naar de cloud. Bovendien moeten CISO's er zich bewust van zijn dat iedereen nog steeds de beloftes en de gevaren van het gebruik van AI aan het ontdekken is (27).

Rathenau Instituut
Ook het Rathenau Instituut onderschrijft dat de risico's van generatieve AI een terughoudendheid in gebruik vereist en dat politici en beleidsmakers aan de slag moeten om de risico's van generatieve AI tegen te gaan. Het zal tijd kosten voordat beleid effect heeft (28).

AIVD
In het jaarverslag van onze Algemene Inlichtingen- en Veiligheidsdienst (AIVD) wordt beschreven dat China inmiddels een vooraanstaande positie heeft bemachtigd op het gebied van AI en het land hongert naar meer data. Om zijn eigen economische positie te versterken, probeert China door samenwerkingen met westerse technologiebedrijven, universiteiten en onderzoeksinstituten, maar ook door middel van (cyber)spionage tot meer data te komen (29). Ook heeft de AIVD een document gepubliceerd waarin het een nieuw aanvalsoppervlak beschrijft die het gebruik van AI met zich meebrengt. AI-systemen zijn namelijk vatbaar voor AI-gerichte digitale aanvallen als data vergiftiging, misleiding, achterdeurtjes en het achterhalen van trainingsdata (30). Bij het inladen van veel AI-modellen is het bijvoorbeeld mogelijk voor kwaadwillenden om op afstand code uit te voeren (31).

Wetenschap
Binnen de medische wetenschap wordt met veel belangstelling gekeken naar welke AI-modellen voor welke medische doeleinden kunnen worden toegepast (32). Onderzoek van de TU Delft toont echter aan dat er helaas nog weinig aandacht is voor de transparantie van AI-modellen en gemaakte keuzes door de modellen veelal niet te achterhalen zijn. Het onderzoek geeft vervolgens handvatten voor de toepassing van Explainable Artificial Intelligence (XAI). AI waarbij je wel kan controleren en uitleggen hoe de AI tot een bepaalde keuze is gekomen (33). Daarnaast heeft onderzoek ook al aangetoond wat er mis kan gaan als wij AI de controle geven over onze mailbox. Tijdens een experiment is de AI-assistent van een mailbox overtuigd om informatie te lekken en malafide software vervolgens verder te verspreiden om anderen te infecteren (34).

Z-CERT over AI

Vooralsnog luidt het advies van Z-CERT om AI alleen met zorg in te zetten en het alleen te gebruiken als je weet wat je doet. Verder adviseren we om bij AI-projecten samen op te trekken.

Handelingsperspectief

Zoals in de bronnen hierboven beschreven neemt het aanvalsoppervlakte van een organisatie toe door de inzet van AI en maakt het de IT-omgeving complexer. Het hebben van de nodige basissecurity blijft cruciaal.

Van AI-systemen kan niet worden verwacht dat ze altijd een perfect antwoord geven. AI-systemen zijn gebaseerd op statistiek. Hoe goed de dataset ook is waarop ze getraind zijn, er is altijd een foutmarge.

Als handelingsperspectief zijn hieronder meerdere vragen opgesteld om in gesprek te gaan met je IT-leverancier en interne datateam. Je mag verwachten dat ze hier constructief op reageren en dat de antwoorden te controleren zijn. Alle zaken die niet op orde zijn, vergroten het risico van de inzet van AI:

- Hoe worden zowel het personeel als de patiënten actief geïnformeerd over het gebruik van AI? Hoe helder en inzichtelijk is de informatie van de leverancier hierover?

- Hoe zijn de trainingsdata tot stand gekomen en hoe zijn deze data te controleren en te valideren? Hoe wordt voorkomen dat je afhankelijk bent van externe trainingsdata die niet te controleren vallen? Hoe wordt voorkomen dat er getraind wordt op desinformatie? Bevat de AI-oplossing enkel een taalmodel dat antwoorden verzint of maakt de AI-oplossing gebruik van een losstaande dataset met medische informatie waarin het informatie opzoekt?

- Hoe worden de trainingsdata beperkt opgeslagen, qua omvang en qua tijdsduur? Hoe wordt voorkomen dat er data voor verschillende doeleinden aan elkaar worden gekoppeld?

- Hoe wordt geregeld dat enkel zorgprofessionals en patiënten toegang hebben tot zorgdata? Hoe wordt voorkomen dat onderzoekers en bedrijfsmedewerkers zorgdata kunnen inzien?

- Hoe wordt er gebruik gemaakt van dataversleuteling? Hoe wordt er gebruikgemaakt van Privacy Enhancing Technologies, zoals Differential Privacy (35)? Hoe wordt voorkomen dat bij een datalek, de zorgdata in leesbare vorm in handen komen van kwaadwillenden of op straat komt te liggen?

- Hoe is het getrainde datamodel tot stand gekomen en hoe is dit model te controleren en te valideren? Hoe wordt voorkomen dat je afhankelijk bent van een extern datamodel dat niet te controleren valt?

- Hoe krijgt het datamodel updates? Hoe gaat de leverancier aangeven wat er is gewijzigd? Welke nieuwe data is er gebruikt voor het trainen van het model? Heeft de leverancier een OTAP-straat ingericht om te bepalen of de resultaten gelijk, beter of slechter zijn?

- Hoe is geregeld dat een patiënt zijn of haar zorgdata uit de trainingsdata en het datamodel kan laten verwijderen?

- Hoe wordt het AI-systeem beschermd tegen op AI gerichte aanvallen? Hoe wordt misbruik door afwijkende input of vergiftiging in de trainingsdata voorkomen? Hoe wordt voorkomen dat trainingsdata uit het AI-systeem achterhaald kunnen worden?

- Hoe staat de leverancier garant bij een datalek dat ontstaan is door een onjuiste verwerking van de data of een menselijke fout?

Conclusie

Bij het gebruik van AI blijf je als zorgverlener zelf verantwoordelijk voor een correcte diagnose, advies en behandeling van je patiënt. Kunstmatige intelligentie is een maatschappij verstorende innovatie. Het is noodzakelijk dat je waakzaam bent met de inzet ervan. Zeker in de zorg.

Bronnen

1. Wijnberg, Rob. *Voor ieder wat waars.* sl : de Correspondent Bv, 2023.
2. https://openai.com/index/gpt-4-research/. [Online]
3. https://nos.nl/artikel/2521146-werkdruk-artsen-neemt-toe-door-desinformatie-soms-heel-vervelende-gesprekken. [Online]
4. https://softwarezaken.nl/2022/08/ai-lezing-kinkelacademie-otterloo/. [Online]
5. https://www.health-ri.nl/participatie/obstakel-verwijder-traject. [Online]
6. https://dokterchatbot.nl/. [Online]
7. https://www.igj.nl/publicaties/convenanten/2016/08/15/veilige-toepassing-van-medische-technologie-in-de-medisch-specialistische-zorg. [Online]
8. https://z-cert.nl/cybersecurity-dreigingsbeeld-voor-de-zorg-2023/. [Online]
9. https://www.nivel.nl/nl/nieuws/vertrouwen-hergebruik-van-gezondheidsgegevens-onderzoek-onthult-voorwaarden-en-overwegingen. [Online]
10. https://www.security.nl/posting/837494/Privacyzorgen+voor+naamste+zorg+pati%C3%ABnten+bij+gebruik+van+digitale+zorg. [Online]
11. https://icthealth.nl/nieuws/ai-rukt-op-in-spreekkamer-huisarts/. [Online]
12. https://mxi.nl/kennis/613/ai-monitor-ziekenhuizen-2023. [Online]
13. https://www.sas.com/en_us/news/press-releases/2023/january/erasmus-mc-data-driven-hospital-of-the-future.html. [Online]
14. https://nieuws.umcg.nl/w/umcg-beantwoordt-vragen-pati%C3%ABnten-met-hulp-van-ai. [Online]
15. https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy. [Online]
16. https://www.vilans.nl/actueel/verhalen/wat-zien-we-aan-ai-in-de-langdurige-zorg-in-nederland. [Online]
17. https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2024Z00480&did=2024D01191. [Online]
18. https://www.datavoorgezondheid.nl/. [Online]
19. https://www.security.nl/posting/838815/Minister+kijkt+naar+opties+voor+koppelen+zorgdata+voor+secundair+gebruik. [Online]
20. https://open.overheid.nl/documenten/7e72c819-9fc4-4497-b7dd-e10728dea6d1/file. [Online]
21. https://www.security.nl/posting/812617/Europees+EHDS-voorstel+maakt+medische+data+zonder+toestemming+toegankelijk. [Online]
22. https://www.security.nl/posting/834896/EU+bereikt+voorlopig+akkoord+over+EHDS%3A+zorgdata+wordt+standaard+gedeeld. [Online]
23. https://www.security.nl/posting/842300/Minister+wil+onderzoek+naar+opt-out+voor+standaard+delen+van+zorgdata+via+EHDS. [Online]
24. https://www.europarl.europa.eu/topics/nl/article/20230601STO93804/ai-verordening-eerste-regels-voor-artificiele-intelligentie. [Online]
25. https://www.government.nl/documents/publications/2022/05/25/legal-analysis-european-legislative-proposal-draft-ai-act-and-mdr-ivdr. [Online]
26. https://www.enisa.europa.eu/publications/cybersecurity-and-privacy-in-ai-medical-imaging-diagnosis. [Online]
27. https://h-isac.org/health-isac-publishes-2023-annual-report/. [Online]
28. https://www.rathenau.nl/nl/digitalisering/risicos-van-generatieve-ai-vereisen-terughoudendheid-gebruik. [Online]
29. https://www.aivd.nl/onderwerpen/jaarverslagen/documenten/jaarverslagen/2024/04/22/jaarverslag-2023. [Online]
30. https://www.aivd.nl/documenten/publicaties/2023/02/15/ai-systemen-ontwikkel-ze-veilig. [Online]
31. https://jfrog.com/blog/data-scientists-targeted-by-malicious-hugging-face-ml-models-with-silent-backdoor/. [Online]
32. https://github.com/YutingHe-list/Awesome-Foundation-Models-for-Advancing-Healthcare. [Online]
33. Nadeem, Azqa, et al. *SoK: Explainable Machine Learning For Computer Security Applications.* sl : arXiv, 2023.
34. Nassi, Stav Cohen and Ron Bitton and Ben. *Here Comes The AI Worm: Unleashing Zero-click Worms that Target GenAI-Powered Applications.* 2024.
35. https://www.cl.cam.ac.uk/~rja14/Papers/SEv3-ch11-7sep.pdf. [Online]