

Data Ethics and AI

Nikolay Voropayev

27. Mai 2024

Zusammenfassung

In diesem Dokument wird grob erklärt wie KI funktioniert, es werden die Gefahren von KI analysiert, logisch behandelt und schlussfolgerungen gezogen, welchen beweisen sollen, dass:

1. KI ist nicht wirklich intelligent
2. KI wird uns nicht auslöschen wie in der Terminator Franchise.(Wikipedia, 2024c)
3. KI soll nicht nur in den Händen von Big-Tech Firmen überlassen werden, sondern sollte open-source gehalten werden.(Wikipedia, 2024b)
4. Datenschutz im Zusammenhang mit KI ist umso mehr wichtig als normalerweise.

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Was is KI?

KI steht fuer "Kuenstilche Intelligenz", jedoch sieht KI gar nicht so aus wie ein menschliches Gehirn, welches aus Milliarden von Neuronen besteht. KI's bestehen aus sogennanten Neuralen Netzwerken. (IBM, 2024c)

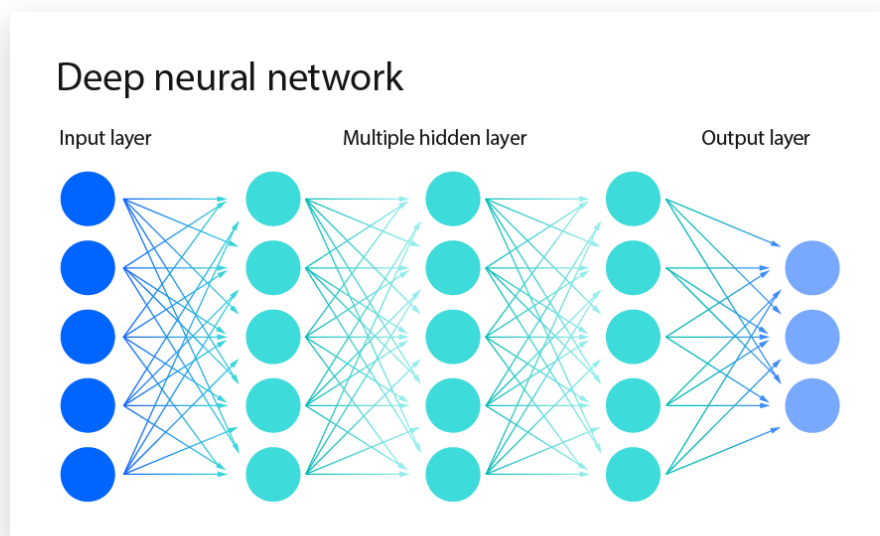


Abbildung 1.1: Neurale-Netzwerk-Grafik, IBM

Ich werde in dieser Arbeit nicht in die Mathematischen details eingehen, auch nicht den Unterschied zwischen KI und "Machine Learning"(IBM, 2024b) erklaren, da dies fuer diese Arbeit nicht besonders wichtig ist. Auch wie diese Neuralen Netzwerke funktionieren wird auf der IBM-website (IBM, 2024a) gut erklart.

1.1.1 Was sind Neurale Netzwerke?

Einfach erklärt, haben Neurale Netzwerke wie in der Abbildung Schichten. In jeder dieser Schichten gibt es Schnittpunkte. Wenn ein bestimmter Input einen Schnittpunkt aktiviert, sendet dieser einen bestimmten Output weiter. Wie stark dieser Output gewichtet ist, und wie er verarbeitet wird, hängt von dem Netzwerk ab. Das Wichtigste ist aber, dass man nicht wissen kann, was in diesen Netzwerken passiert, und warum ein bestimmter Input so wahrgenommen wird, wie er wird. Dies ist für später wichtig.

Falls es schwer fällt, dies im Textformat zu verstehen, kann dieses Video (atomic frontier, 2021) dabei helfen.

1.2 Was ist Datenschutz und warum ist es wichtig?

Per merriam-webster (merriam webster, 2024) sind Daten faktuelle Informationen. Jedoch wenn wir von Daten im Bezug auf Datenschutz sprechen, sind nicht einfach Statistiken zum Schokoladen-Konsum des Durchschnittlichen Schweizers, welches über 10kg pro Kopf pro Jahr beträgt (statista.com, 2023), sondern es geht um Informationen, wie Lokationsdaten, die durch Bluetooth, WLAN oder Mobilfunknetze durch triangulation ausgerechnet werden können. Oder Kaufgewohnheiten durch die Ausgabensdaten der Kreditkarten. Viele solche Daten können aus anderen "herausgelesen" werden. Manche Schlussfolgerungen zu schließen ist es jedoch nicht möglich für ein klassisches Computer-Programm.

1.2.1 Warum ist Datenschutz wichtig?

Dies ist ein sehr komplexes Thema und es gibt endlos Informationen dazu. Da es aber immer gut ist eine eigene Meinung zu bilden, werde ich einfach nur Beispiele bringen, warum Datenschutz wichtig sein kann.

Es gibt zahlreiche Vorfälle bei denen zum Beispiel Versicherungs-Premien von Autofahrern in den Vereinigten Staaten höher wurde, weil sie stark gebremst haben, und wahrscheinlich ein KI entschieden hatte, dass dieser Autofahrer schlecht Auto fahren kann. Auch wenn er in einer Situation bremste, in der er einen Unfall verhinderte. CNN hat dieses Problem publiziert. (CNN, 2024b)

Aber dies ist nur eine Art, auf die unregulierter oder inkompetenter Umgang mit Daten verheerende Folgen hat, es gibt, wie ich schon erwähnt habe, endlos solche Beispiele.

Es gibt sehr viel Videos auf Youtube darüber, welche viel besser als ich es erklären könnte, zu diesem Thema Erklärungen bereitstellen, und dieser Youtube-Kanal (One, 2024a) lässt für alle Informationen Quellen, dadurch ist es einfach die Informationen zu verarbeiten und sie können überprüft werden. Ich würde empfehlen die unten gelisteten Videos anzuschauen.

Apple and Google contact tracing is a dystopian nightmare (One, 2020)

Google vs DuckDuckGo | Search engine manipulation, censorship and why you should switch (One, 2018)

Google will spy on you in physical stores – Can businesses really do anything? (One, 2017)

Your Car Is a Better Spy than Facebook (One, 2021c)

Don't use WhatsApp! (One, 2021a)

Your Keystrokes Are In The Cloud (One, 2021d)

Falls man die Zeit hat, würde ich auch empfehlen, 1984 von George Orwell und 451 Fahrenheit von Ray Bradbury zu lesen, es sind sehr spannende Bücher und sie haben mit der Thematik vieles gemeinsam. Netterlich ist dies alles nicht nötig, einfach um zu verstehen, warum Datenschutz wichtig ist, Die wichtigsten Quellen werden später angegeben und werden auch markiert

sein, das sie fuer ein verstaendnis noetig sind. Aber das wissen vom monumentalem Aussmass dieser Probleme ist nuetzlich.

Kapitel 2

KI, Daten und Ethik

2.1 KI und Daten

Wie schon frueher erwaeht, stellt KI neue moeglichkeiten for, Informationen zu verarbeiten. Ungluecklicherweise bedeutet dies, dass eine KI Daten viel besser verarbeiten kann. Die Social-Media-Plattform Reddit zum Beispiel verkauft alle Benutzer-Generierte Daten an OpenAI, eine KI-firma, welche hauptsaechlich Microsoft gehoert. (OpenAI, 2024a) Hier von der offiziellen Website von OpenAI. (OpenAI, 2024b)

Technologien entwickeln sich sehr schnell, viel schneller als entsprechende Gesetze eingefuehrt werden koennen. So wurden zum Beispiel Deep-Fakes (Wikipedia, 2024a), welche massive schaden verursacht haben (CNN, 2024a), erst vor kurzen von Politikern als Problem anerkannt.

Das schockierendste ist, das es schon jetzt Socia-Credit-Scores gibt, die nicht von einem Diktaturstaat, sondern von Geschaeften, Banken und Versicherungsfirmer benutzt werden, um Kunden auf ihre "Wertigkeit" zu gradieren. All dies moeglich durch KI. Das Problem damit, KI zu erlauben dies zu tun ist das KI mit denen Daten arbeitet, die es bekommt. Dies ist spaeter wichtig. (One, 2021b) Dieses Video ist fuer das Verstaendnis notwendig

Es gibt auch eine grosse Luege, welche alle Big-Tech KI-Firmer immer wieder erzaehlen. Naemlich sollte KI gefaehrlich sein, und es sollten nur bestimmte Firmer die erlaubnis haben, KI zu erstellen. Dies sollte durch Gesetze geregelt werden, die es schwer machen, ein eigenes, open-source KI-Modell selber herzustellen, und ja, dies ist moeglich, und sogar relativ einfach, denn es gibt videos, in denen Youtuber eine einfache KI trainieren, um zum beispiel das bestmoegliche keyboard-layout fuer bestimmte kriterien. Deswegen wollen Big-Tech Firmer den Zugang ueber Lizenzen sperren. (atomic frontier, 2023)

Es gibt sogar ein Dokument, welches von einem Google-Mitarbeiter geschrieben wurde, in welchem beschreiben wird, wie open source modelle besser sind, als die, welche von Big-Tech erstellt werden und wie es unmoeglich ist, dagegen fair zu gewinnen. (Guardian, 2023)

(One, 2024b) Dieses Video ist fuer das Verstaendnis notwendig. Es beschreibt die Luege von Big-Tech sehr genau und erklaert auch warum es eine Luege ist und nicht die Wahrheit.

Aber es wird noch schlimmer, KI verbraucht unemngen an Wasser, so viel wasser, dass Big-Tech-AI "Kriege" um wasserreserven ausloest. Richter muessen entscheiden, ob sie mehr Wasser and Bauern, Einwohner von Staedten, oder an die Datenzentren, welche die KI's prozessieren, ein-

teilen. In den USA gibt es deswegen ein Wasserpröblem in manchen Regionen. (**ai-water-wars**)

Es gab einen ziemlich bekannten Vorfall, bei dem ein US-Resident fotos seines Sohnes während der Covid-19 Pandemie für den Arzt nahm. Der Mann benutzte Google Photos. Das KI welches illegale Materialien erkennen sollte, entschied die Fotos seien Pornografische Bilder eines Kindes, und nach einer Weile klopfte die Polizei an seiner Tür. (**google-photos-false-flagging**)

2.2 Skynet im echten Leben

Wie ich schon vorher erwähnt habe, ist KI nicht wirklich intelligent, aber durch die oben genannten Anwendungszwecke kann KI sehr gut sogar Kriege auslösen. Und zwar echte Kriege. Sagen wir mal es gibt ein Land, welches Konkurrenz zur derzeitigen Weltmacht leistet. Die klassischen Wege die Konkurrenz zu schwächen funktionieren nicht, das Land entwickelt sich weiter und Propaganda kann ihm nichts antun. In ihrer Verzweiflung die Spitze zu behalten nimmt die Weltmacht drastische Massnahmen: Sie erlauben KI im Militär zu benutzen, autonome Waffen werden eingesetzt. Es funktioniert perfekt, die Konkurrenten mit ihren fleischigen Soldaten können mit der Geschwindigkeit, mit der Computer arbeiten nicht mithalten. Einige Jahre später stellt sich jedoch heraus, dass die KI nicht nur Soldaten angegriffen hatte, sondern in den Tausenden Zivilisten umbrachte. Dies war zu erwarten, denn wie im ersten Kapitel erwähnt, wenn KI Fehler macht, kann man diese nicht wirklich beheben. Man kann nur das ganze Modell wieder neu aufbauen. Deswegen sind kleinere Modelle beliebter, vor allem in der Open-source-Welt.

Die geschilderte Situation tönt sehr dystopisch und man würde denken, das könnte nie im echten Leben passieren, nur, es ist schon passiert. Die Gründe für welche die USA ihren Drohnenkrieg startete war zwar nicht so einfach und die Diskussion von Ethik von Krieg und Terrorismus ist nicht Zweck dieser Arbeit, aber die Tatsache, dass unzählige Zivilisten von Maschinen umgebracht wurden, weil ein KI so entschied, bleibt stehen. Den Drohnenkrieg zu beschreiben überlasse ich professionellen Journalisten der SRF. (**drones-srf**)

Wie schon vorher erwähnt, es gibt "Social-credit-scores" welche von Betrieben Personen gegeben werden. Banken sind ein gutes Beispiel dafür. Das grosse Problem damit ist, dass diese KI's oft wenig objektiv ist, denn die Daten, aus denen ein theoretisch objektiver Algorithmus wird zum Beispiel rassistisch, oder das bekannte Amazon-AI, welches benutzt wurde um Arbeitskandidaten zu evaluieren. Diese KI "lernte" nach einer Weile, dass Frauen schlechte Arbeiter sind, und nach mehreren Versuchen den Algorithmus zu korrigieren, wurde das ganze Programm eingestellt und "subjektive" Menschen wurden wieder verwendet.

Ein sehr beliebtes Beispiel ist der Youtube-Algorithmus. Nach nur ein paar Minuten nachschlagen kann man buchstäblich hunderte oder gar tausende Fälsche finden, in denen die KI komplett grundlos ein Video gelöscht wurde, oder jemand der die Rechte zu Copyright-Geschütztes Material eine Warnung bekam, dass es verboten sei, Copyright-Material zu verbreiten, obwohl diese Person die Rechte hat.

Kurzgefasst, es ist eine schlechte Idee, KI über Probleme von Menschen entscheiden zu lassen. Das Buch "Watchbirds" ist eine gute Illustration davon. Falls dies immer noch nicht genug Informationen sind, um zu beweisen, dass KI problematisch ist, dann sollte dieses Dokument (Guardian, 2023) eines google-KI-Ingenieurs es beweisen.

Literatur

- atomic frontier. (2021). *This image breaks AI*. Verfügbar 8. April 2021 unter <https://youtube.com/watch?v=p6CfR3Wpz7Y&t=390>
- atomic frontier. (2023). *Why typing sucks now*. Verfügbar 18. November 2023 unter <https://youtube.com/watch?v=188fipF-i5I>
- CNN. (2024a). *British engineering giant Arup revealed as 25milliondeepfakescamvictim*. Verfügbar 17. Mai 2024 unter <https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html>
- CNN. (2024b). *Is your car spying on you?* Verfügbar 23. März 2024 unter <https://youtube.com/watch?v=aHhx8mMUV2o>
- Guardian, T. (2023). *Google engineer warns it could lose out to open-source technology in AI race*. Verfügbar 5. Mai 2023 unter <https://www.theguardian.com/technology/2023/may/05/google-engineer-open-source-technology-ai-openai-chatgpt>
- IBM. (2024a). *What is AI?* Verfügbar 20. April 2023 unter <https://www.ibm.com/topics/artificial-intelligence>
- IBM. (2024b). *What is ML?* Verfügbar 20. April 2023 unter <https://www.ibm.com/topics/machine-learning>
- IBM. (2024c). *what-is-a-neural-network*. Verfügbar 1. Januar 2024 unter <https://www.ibm.com/topics/neural-networks>
- mirriam webster. (2024). *Data definition*. Verfügbar 1. Januar 2024 unter <https://www.merriam-webster.com/dictionary/data>
- One, T. H. (2017). *NO PRIVACY OFFLINE!? Google will spy on you in physical stores – Can businesses really do anything?* Verfügbar 27. Mai 2017 unter https://youtube.com/watch?v=KpplFma_27s
- One, T. H. (2020). *Apple and Google contact tracing is a dystopian nightmare*. Verfügbar 2. Mai 2020 unter <https://youtube.com/watch?v=WRalTWAFBY4>
- One, T. H. (2021a). *Don't use Whatsapp!* Verfügbar 4. Februar 2021 unter <https://youtube.com/watch?v=shpiVm1qpnw>
- One, T. H. (2021b). *Social Scores Are Real And You Have One Too*. Verfügbar 20. September 2021 unter <https://youtube.com/watch?v=VUhKTngpd8c>
- One, T. H. (2021c). *Your Car Is a Better Spy than Facebook*. Verfügbar 9. April 2012 unter https://www.youtube.com/watch?v=WX2SWUMt_fk
- One, T. H. (2021d). *Your Keystrokes Are In The Cloud*. Verfügbar 25. Februar 2021 unter <https://www.youtube.com/watch?v=vCRX0MZm2KI>
- One, T. H. (2024a). *The Hated One*. Verfügbar 27. Mai 2024 unter <https://youtube.com/channel/UCjr2bPAyPV7t35MvcgT3W8Q>
- One, T. H. (2024b). *Why you shouldn't believe the AI extinction lie*. Verfügbar 3. Mai 2024 unter <https://www.youtube.com/watch?v=5NUD7rdbCm8>

- one, T. H. (2018). *Google vs DuckDuckGo / Search engine manipulation, censorship and why you should switch*. Verfügbar 14. Oktober 2018 unter <https://youtube.com/watch?v=SrsCEbi5N7Y>
- OpenAI. (2024a). *Homepage*. Verfügbar 1. Januar 2024 unter <https://openai.com/>
- OpenAI. (2024b). *OpenAI and Reddit Partnership*. Verfügbar 16. Mai 2024 unter <https://openai.com/index/openai-and-reddit-partnership/>
- statista.com. (2023). *Pro-Kopf-Konsum von Schokolade in der Schweiz in den Jahren 2005 bis 2023*. Verfügbar 27. März 2024 unter <https://de.statista.com/statistik/daten/studie/369440/umfrage/pro-kopf-konsum-von-schokolade-in-der-schweiz/>
- Wikipedia. (2024a). *Deepfake*. Verfügbar 26. Mai 2024 unter <https://en.wikipedia.org/wiki/Deepfake>
- Wikipedia. (2024b). *Open-source Software*. Verfügbar 1. Januar 2024 unter https://en.wikipedia.org/wiki/Open-source_software
- Wikipedia. (2024c). *The Terminator*. Verfügbar 20. April 2023 unter https://en.wikipedia.org/wiki/The_Terminator