Data Ethics and AI

Nikolay Voropayev

28. Mai 2024

Zusammenfassung

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

1. KI is nicht wirklich intelligent
2. KI wird uns nicht ausloeschen wie in der Terminator Franchise.(Wikipedia, 2024c)
3. KI soll nicht nur in den haenden von Big-Tech Firmen ueberlassen werden, sondern sollte open-source gehalten werden.(Wikipedia, 2024b)
4. Datenschutz im zusammenhang mit KI is umsomehr 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 Neuronalen Netzwerken. (IBM, 2024c)

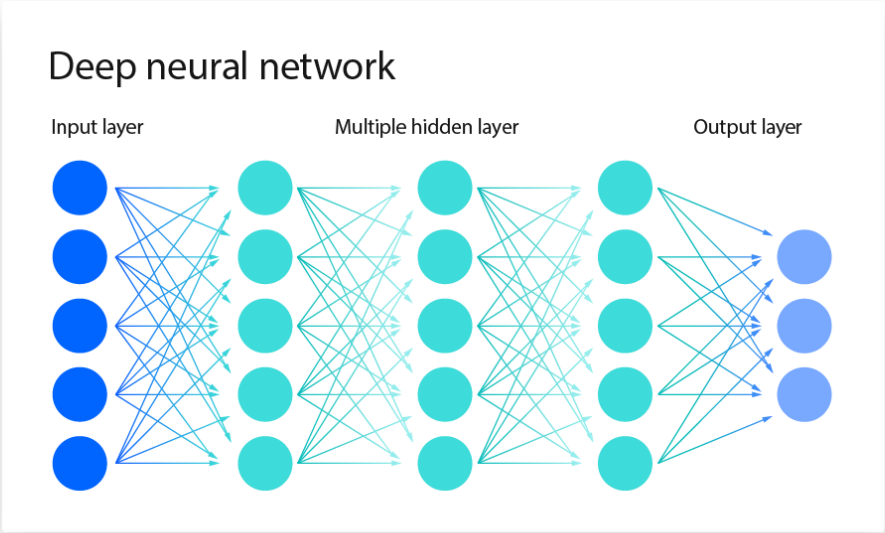

Abbildung 1.1: Neurale-Netzwerk-Grafik, IBM

Ich werde in dieser Arbeit nicht in die Mathematischen details eingehen, auch nicht den Unterschied zwischen KI und "Machine Learning"(IBM, 2024b) erklaeren, da dies fuer diese Arbeit nicht besdonders wichtig ist. Einfach erklaert beruht Machine Learning mehr auf menschlichen Eingriffen, wahernd "tiefes"Lernen auf vielen SSchichten"der KI beruht. Was mit Schichten gemint ist, wird weiter unten erklaert. Auch wie diese Neuronalen Netzwerke funktionieren wird auf der IBM-website (IBM, 2024a) gut erklaert. Ich bin kein KI-Ingineur und kann dewswegen nur

die Erklaerungen von anderen Zitieren oder mein oberflaechliches Verstandnis davon in Worte fassen.

1.1.1 Was sind Neuranale Netzwerke?

Einfach erklaert, 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, haengt 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 fuer spaeter wichtig.
Falls es schwer faellt, dies im Textformat zu verstehen, kann dieses Video (atomic frontier, 2021) dabei helfen. Der betreiber dieses Youtube-Kanals hat auch selber seine eigene KI trainiert.

1.2 Was ist Datenschutz und warum ist es wichtig?

Per mirriam-webster (mirriam 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 ueber 10kg pro kopf pro jahr betraegt (statista.com, 2023), sondern es geht um Informationen, wie Lokationsdaten, die durch Bluetooth, WLAN oder Mobilfunknetze durch triangulation ausgerechnet werden koennen. Oder Kaufgewohnheiten durch die Ausgabensdaten der Kredikarten. Viele solche Daten koennen aus anderen "herausgelesen"werden. Manche schlussfolgerungen zu schliessen ist es jedoch nicht moeglich fuer ein klassisches Computer-Programm.

Datenschutz ist der Prozess der minimierung von der Menge dieser Daten, welche aufgezeichnet werden. Zum beispiel durch das verwenden von verschluesselung kann der Inhalt einer Nachricht unzugaenglich gemacht werden. Dabei sind aber die Meta-Daten (**metadata**) nicht unbedingt unsichtbar. Wenn die IP-Adresse, von der eine Nachricht gesendt worden ist und die Uhrzeit zugaenglich ist, kann genau bestimmt werden, wer diese Nachricht gesendt hat. Dies ist ein sehr grobes beispiel und Meta-Daten werden von Gerichten benutzt um zu entscheiden, ob Jemand Verhaftet werden sollte oder nicht.

Das am weitesten verbreitete benutzen von Mata-Daten ist das sogennante Geofencing.(**geofencing**)

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 viele Personen, welche sagen Ïch habe nichts zu verstecken."Darauf gibt es ein gutes Beispiel, um zu visualisieren, warum das nicht so stimmt. Falls das so ist, waere diese Person damit Einverstanden, mir zugang zu allen ihren Konversationen, allen Photos, allen Daten auf ihren Geraten zu geben? Wahrschleinlich nicht, und das ist gut so. Auch wenn eine Firma die Daten nicht weiterverkauft, Hacker bekommen zugriff die ganze Zeit zu diesen Datenbanken, und diese betriben dann Identitaetsdiebstahl.
Es gibt zalreiche vorfaelle bei denen zum Beispiel Versicherungs-Premien von Autofahrern in den Vereinigten Staaten hoeher wurde, weil sie stark gebremst haben, und wahrschleinlich ein KI entschieden hatte, dass dieser Autofaherer schlecht auto faheren 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 verherende Folgen hat, es gibt, wie ich schon erwaehnt habe, endlos solche beispiele.
Es gibt sehr viel Videos auf Youtube darueber, welche viel besse als ich es erklaeren koennte, zu diesem Thema erklaerungen bereitstellen, und dieser Youtube-Kanal (One, 2024a) laesst fuer alle Informationen Quellen, dadurch ist es einfach die Informationen zu verarbeiten und sie koennen ueberprueft werden. Ich wuerde 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, wuerde ich auch empfehlen, 1984 von George Orwell und 451 Fahrenheit von Ray Bredberry zu lesen, es sind sehr Spannende Buecher und sie haben mit der Thematik vieles gemeinsam. Netuerlich ist dies alles nicht noetig, einfach um zu verstehen, warum Datenschutz wichtig ist, Die wichtigsten Quellen werden speater 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 erwaehnt, stellt KI neue moeglichkeiten for, Informationen zu verarbeiten. Ungluegcklicherweise bedeutet dies, dass eine KI Daten viel besser verarbeiten kann. Die Social-Media-Platform Reddit zum Beispiel verkauft alle Benutzer-Generierte Daten an OpenAI, eine KI-firma, welche hauptsaechlich Microsoft gehoert. (OpenAI, 2024a) Hier von der offieziellen Website von OpenAI. (OpenAI, 2024b)

Technologien entwicklen sich sehr schnell, viel schneller als entsprechende Gesetze eingefuehrt werden koennen. So wurden zum Beispiel Deep-Fakes (Wikipedia, 2024a), welche massive schaeden 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 Versicherungsfirmen benutzt werden, um Kunden auf ihre "Wertigkeitßu 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-Firmen immer wieder erzaehlen. Naehmlich sollte KI gefaehrlich sein, und es sollten nur bestimmte Firmen 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 realativ 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 Firmen den Zugang ueber Lizensen sperren. (atomic frontier, 2023)
Es gibt sogar ein Dokument, welches von einem Google-Mitarbeiter geschrieben worde, 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 "Kriegeüm 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 Wasserprbolem in manchen Regionen. (**ai-water-wars**)

Es gab einen ziemlich bekannten Vorfall, bei dem ein US-Resident fotos seines Sohnes waehrend der Covid-19 Pandemie fuer den Artzt 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 Tuere. (**google-photos-false-flagging**)

2.2 Skynet im echten leben

Wie ich schon vorher erwaehnt habe, ist KI nicht wirklich Intelligent, aber durch die oben genannten anwendungszwecke kann KI sehr guet sogar Kriege ausloesen. Und zwar echte Kriege.
Sagen wir mal es gibt ein Land, welches Konkurrenz zur derzeitigen Weltmacht leistet. Die Klassischen wege die Konkurrenz zu schwaechen funktionieren nicht, das Land entwickelt sich weiter und Propaganda kann ihm nichgts antun. In ihrer verzweifulng die Spitze zu behalten nimmt die Weltmacht drastische Massnahmen: Sie erlauben KI im Militar zu benutzen, autonome Waffen werden eingesetzt. Es funktioniert perfekt, die Konkurrenten mit ihren fleischigen Soldaten koennen mit der geschwindigkeit, mit der Computer arbeiten nicht mithalten. Einige Jahre spaeter stellt sich jedoch heraus, dass die KI nicht nur Soldaten angegriffen hatte, sonder in den Tausenden zivilisten umbrachte. Dies war zu erwarten, denn wie im ersten Kapitel erwaehnt, wenn KI fehler macht, kann man diese nicht wirklich beheben. Man kann nur das ganze Model weider neu aufbauen. Deswegen sind kleinere Modelle beliebter, vor allem in der Open-source-Welt.

Die geschilderte Situation Toent sehr Dystopisch und man wuerde denken, das koennte nie im echten leben Passieren, nur, es ischt schon Passiert. Die gruende fuer welche die USA ihren Drohnenkrieg startete war zwar nicht so einfach und die diskussion von ethik von krieg und terrorismus ist nicht zewck dieser Arbeit, aber die Tatsache, dass unzaehlige Zivilisten von Machinien umgebracht wurden, weil ein KI so entschied, bleibt stehen. Den Drohnenkrieg zu beschreiben ueberlasse ich professionellen journalisten der SRF. (**drones-srf**)
Wie schon vorher erwaehnt, es gibt SSocial-credit-scores"welche von betrieben personen gegeben werden. Banken sind ein gutes Beispiel dafuer. Das grosse Problem damit ist, dass diese KI's oft wenig Objektiv ist, denn die Daten, aus denen ein theoretisch objektiver Algorythmus wird zum Beispiel rassistisch, oder das Bekannte Amazon-AI, welches benutzt wurde um Arbeits-Kandidaten zu evaluieren. Diese KI "lernte"nach einer Weile, dass Frauen schlechte Arbeiter sind, und nach mehreren versuchen den Algorythmus zu korrigieren, wurde das ganze Programm eingestellt und SSubjektive"Menschen wurden wieder verwendet.
Ein sehr beliebtes Beispiel ist der Youtube-Algorythmus. Nach nur ein paar Minuten nachschlagen kann man buchstaeblich hunderte oder gar tausende Faelle finden, in denen die KI komplett grundlos ein Video geloescht wurde, oder jemand der die Rechte zu Copytight-Geschuetztes 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 ueber Probleme von Menschen entscheiden zu lassen. Das Buch "Watchbirdsïst 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-Ingenineurs 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 Keystroks 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