

KI, Daten und Ethik

Nikolay Voropayev

31. Mai 2024

Zusammenfassung

In diesem Dokument wird grob erklart 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, 2024d)
3. KI soll nicht nur in den haenden von Big-Tech Firmen ueberlassen werden, sondern sollte open-source gehalten werden.(Wikipedia, 2024c)
4. Datenschutz im zusammenhang mit KI is unsomehr wichtig als normalerweise.
5. Was fuer Schaeden kann KI verursachen?
6. Was gibt es fuer Alternativen der Big-Tech KI modelle?

Ich moechte auch akzentieren, dass beim Schreiben dieser Arbeit keine KI benutzte wurde. Werder zum nachschlagen, noch fuer das Schreiben selber. Alle Saetze dieser Arbeit und jede Zeile Code, welcher von mir geschrieben wurde, ist komplett von mir geschrieben, dies heisst auch, dass jegliche Tools ausser der Eingebauten Vorschlaege des VSodium-Flatpaks nicht verwendet wurden.

Inhaltsverzeichnis

1	Einleitung	3
1.1	Was is KI?	3
1.1.1	Was sind Neuranale Netzwerke?	4
1.2	Was ist Datenschutz und warum ist es wichtig?	4
1.2.1	Warum ist Datenschutz wichtig?	4
2	KI, Daten und Ethik	6
2.1	KI und Daten	6
2.2	Skynet im echten leben	7
2.3	Arbeitsplaetze und KI	8
3	Alternativen	9

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. KIs bestehen aus sogenannten Neuronalen Netzwerken. (IBM, 2024c)

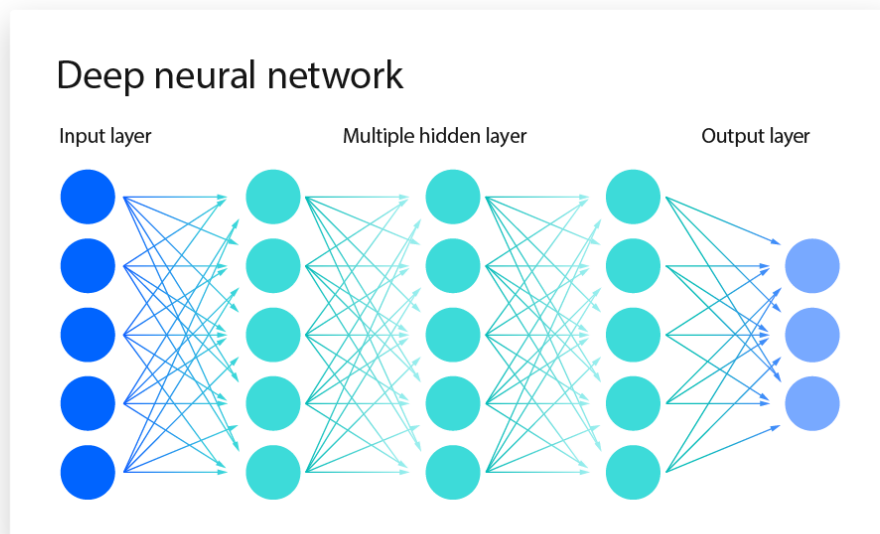


Abbildung 1.1: Neurale-Netzwerk-Grafik, IBM

Ich werde in dieser Arbeit nicht in die mathematischen Details eingehen, auch nicht den Unterschied zwischen KI und „Machine Learning“ (IBM, 2024b) erklaren, da dies fuer diese Arbeit nicht besonders wichtig ist. Einfach erklart beruht Machine Learning mehr auf menschlichen Eingriffen, wahernd „tiefes“ Lernen auf vielen „Schichten“ der KI beruht. Was mit Schichten gemeint ist, wird weiter unten erklart. Auch wie diese Neuronalen Netzwerke funktionieren wird auf der IBM-Website (IBM, 2024a) gut erklart. Ich bin kein KI-Ingenieur und kann dewswegen nur

die Erklärungen von anderen Zitieren oder mein oberflächliches Verständnis davon in Worte fassen.

1.1.1 Was sind Neuronale Netzwerke?

Einfach erklärt, haben Neuronale 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. Der Betreiber dieses Youtube-Kanals hat auch selber seine eigene KI trainiert.

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 (Wikipedia, 2024e) 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.

Datenschutz ist der Prozess der Minimierung von der Menge dieser Daten, welche aufgezeichnet werden. Zum Beispiel durch das Verwenden von Verschlüsselung kann der Inhalt einer Nachricht unzugänglich gemacht werden. Dabei sind aber die Meta-Daten (Wikipedia, 2024) nicht unbedingt unsichtbar. Wenn die IP-Adresse, von der eine Nachricht gesendet worden ist und die Uhrzeit zugänglich ist, kann genau bestimmt werden, wer diese Nachricht gesendet 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 Meta-Daten ist das sogenannte Geofencing. (Wikipedia, 2024b)

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, deswegen werde ich einfach nur Beispiele bringen, warum Datenschutz wichtig sein kann. Es gibt viele Personen, welche sagen „Ich habe nichts zu verstecken.“ Darauf gibt es ein gutes Beispiel, um zu visualisieren, warum das nicht so stimmt. Falls das so ist, wäre diese Person damit Einverstanden, mir Zugang zu allen ihren Konversationen, allen Photos, allen Daten auf ihren Geräten zu geben? Wahrscheinlich nicht, und das ist gut so. Auch wenn eine Firma die Daten nicht weiterverkauft, oder sie von einer KI benutzt werden um passende Werbung für diese Person „auszurechnen“, Hacker bekommen Zugriff die ganze Zeit zu diesen Datenbanken, und diese betreiben dann Identitätsdiebstahl. Das heisst natürlich nicht, dass alle Dateien auf allen Geräten auch so analysiert werden, dieses Beispiel dient nur für das Verständnis.

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, 2024b) lässt für alle Informationen Quellen, dadurch ist es einfach die Informationen zu verarbeiten und sie können auch 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 separat angegeben und werden auch markiert sein, das sie für ein Verständnis nötig sind. Aber das Wissen vom monumentalen Ausmass dieser Probleme ist nützlich.

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, 2024c) 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 KIs prozessieren,

einteilen. In den USA gibt es deswegen ein Wasserproblem in manchen Regionen. (One, 2024a)

Es gab einen ziemlich bekannten Vorfall, bei dem ein US-Resident fotos seines Sohnes wachrend der Covid-19 Pandemie fuer 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 Tuere. (Times, 2023)

2.2 Skynet im echten leben

Wie ich schon vorher erwaeht 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 nichts antun. In ihrer verzweifung 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 erwaeht, 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 ist schon in einer aehnlichen Form 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. (SRF, 2021)

Wie schon vorher erwaeht, es gibt „Social-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 „Subjektive“ 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 Faele finden, in denen die KI komplett grundlos ein Video geloescht wurde, oder jemand der die Rechte zu Copyright-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 „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.

2.3 Arbeitsplaetze und KI

Es ist ein weit verbreiteter Konsensus, das KI Arbeitsplaetze stehlen wuerde und dadurch unethisch ist. Das stimmt nicht, es gibt zwar Arbeitsgeber welche glauben, sie koennten ihre Arbeiter durch KI ersetzen, man darf aber nicht vergessen, dass KI sich nicht unendlich verbessert, und je mehr KI-generierte Materialien im Internet verfuegbar sind, desto langsamer wird die Entwicklung der KI, da diese Daten die KI „vergiften“. Dies passiert, weil ein KI ja menschengemachte Texte imitieren sollte, jedoch wenn diese KI auf vielen KI-generierten Texten trainiert wird, wird es andere KI-generierte Texte imitieren. Also aehnlich wie die kombinierung von verwandter DNA zu fehlbildungen fuehrt, genauso sollte KI nur auf guten Daten trainiert werden.

KI kann sogar Arbeitspleazte schaffen, manche davon sind aber ziemlich unethisch. Amazon versuchte Einkaufslaeden aufzubauen, welche mit Hilfe von KI automatisch die Produkte „einscannnten“ und mann wuerde dann spaeter auf die Credit-Karte oder Debit-Karte eine Rechnung verschickt bekommen.

Dieses Modell funktionierte jedoch genau so gut wie die KI, welche Arbeitsplatz-Kandidaten evaluieren sollte. Und zwar nicht wirklich, das ganze Projekt wurde von unterbezahlten Arbeitern aus Indien und anderen Laendern ohne gute Menschenrechte gefuert und wurde nach einer Weile eingestellt. Auch wird dadurch klar, dass das Trainieren von KI auch auf direkte art und weise unethisch sein kann.

Kapitel 3

Alternativen

Also ist alles KI schlecht und sollte verbannt werden? Nein, anstelle das zu tun, ist es am besten kleine KI-Modelle zu promotieren. Ich habe vorher in dieser Arbeit erwahnt, dass KI fuer alle zuganglich gemacht werden sollte. Aber Ich muss dazu etwas hinzufuegen. KI sollte vom Staat reguliert werden, aber nicht auf die Art und weise, auf welche dies jeztgerade von Big-Tech KI-Frimen vorgeschlagen wird. Anstelle diese entscheiden zu lassen, was passt und was nicht, sollte es genau so durchgefuehrt werden, wie andere Gesetze eingefuehrt werden, oder zu mindest in de Schweiz ist dies so. Unzar naehmlich sollte es demokratisch durch abstimmungen entschieden werden. Natuerlich koennen spezialfaelle wie Pornografische Deepfakes illegal sein.

Was kann man also tun um KI gerecht und ethisch korrekt zu halten? Ganz einfach, man solle open-source KI benutzen. Jeder kann ein open-source Modell veraendern oder lokal auf dem eigenem PC laufen lassen. Es dutzende Videos dazu auf Youtube und detaillierete Informationen dazu auf verschiedenen Websites. Zurzeit muss man mehrere Programme installieren, falls man eigene Filter oder abenderungen auf das Modell hinzufuegen will, aber wenn diese Projekte durch das benutzen unterschuetzt werden, werden diese Modelle besser werden. Sogar Meta, frueher als Facebook bekannt, publizierten den code ihrer KI-Modelle, fuer alle fuer gratis zuganglich.

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 \$25 million deepfake scam victim*. 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). *Ai is running out of water*. Verfügbar 4. April 2024 unter https://youtube.com/watch?v=F-6la_I-xkQ
- One, T. H. (2024b). *The Hated One*. Verfügbar 27. Mai 2024 unter <https://youtube.com/channel/UCjr2bPAyPV7t35MvvgT3W8Q>

- One, T. H. (2024c). *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/>
- SRF. (2021). *Der schmutzige Drohnenkrieg der USA*. Verfügbar 20. Dezember 2021 unter <https://www.srf.ch/news/international/tausende-zivile-opfer-der-schmutzige-drohnenkrieg-der-usa>
- 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/>
- Times, T. N. Y. (2023). *A Dad Took Photos of His Naked Toddler for the Doctor Google Flagged Him as a Criminal*. Verfügbar 21. Juni 2023 unter <https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html>
- Wikipedia. (2024a). *Deepfake*. Verfügbar 26. Mai 2024 unter <https://en.wikipedia.org/wiki/Deepfake>
- Wikipedia. (2024b). *Geo-Fence*. Verfügbar 23. April 2024 unter <https://en.wikipedia.org/wiki/Geo-fence>
- Wikipedia. (2024c). *Open-source Software*. Verfügbar 1. Januar 2024 unter https://en.wikipedia.org/wiki/Open-source_software
- Wikipedia. (2024d). *The Terminator*. Verfügbar 20. April 2023 unter https://en.wikipedia.org/wiki/The_Terminator
- Wikipedia. (2024e). *Triangulation*. Verfügbar 5. April 2024 unter <https://en.wikipedia.org/wiki/Triangulation>
- wikipedia. (2024). *Metadata*. Verfügbar 9. Mai 2024 unter <https://en.wikipedia.org/wiki/Metadata>