



**UNIVERSITÉ
DE GENÈVE**

CENTRE UNIVERSITAIRE
D'INFORMATIQUE

Künstliche Intelligenz und Ethik

**Ethik der KI gemäss UNESCO-Empfehlung: Was lässt sich für
Kulturerbe-Institutionen ableiten?**

Dre Viola Krebs

Jahresversammlung vom 10. März 2025

Zentral- und Hochschulbibliothek ZHB, Luzern

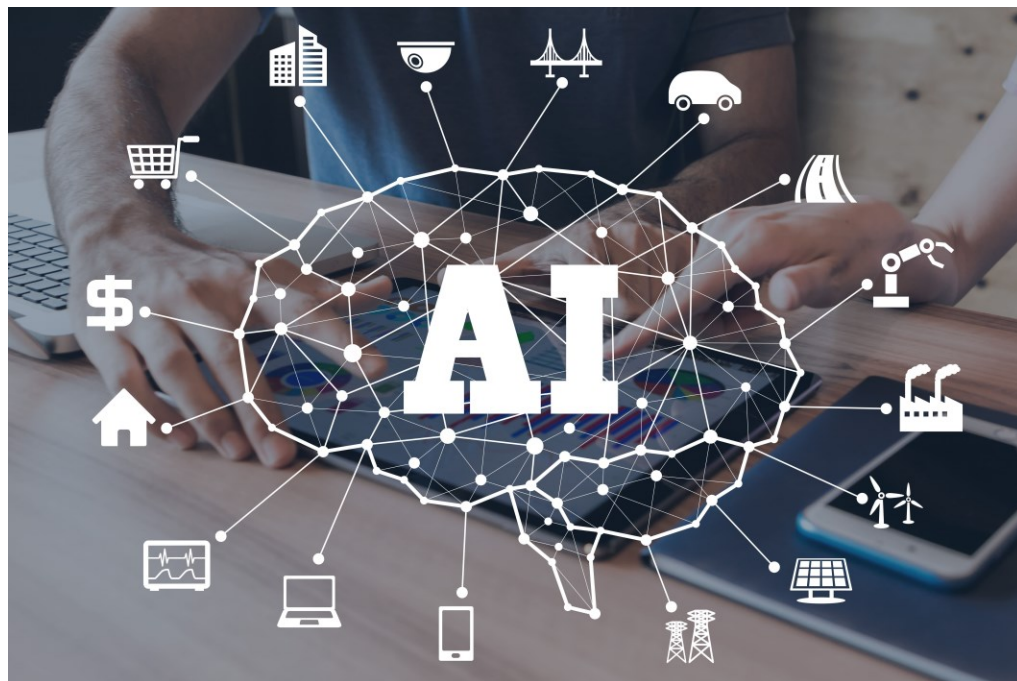
Dre Viola Krebs

- Professor und Trainer an verschiedenen höheren Schulen für digitale Technologien und KI
- Wissenschaftlicher Mitarbeiter am Universitären Zentrum für Informatik (CUI), Universität Genf
- Assoziiertes Mitglied des LISEC und Mitglied des UNESCO-Lehrstuhls an der Universität Straßburg
- Auditor für diverse Qualitätsstandards im Bildungsbereich (QSC 2.0, eduQua 2021, ISO 21001, ISO 9001, In-Qualis)
- Gründer des ICV, einer gemeinnützigen Organisation im Bereich Kommunikation
- Doktor der Informations- und Kommunikationswissenschaften, Universität Straßburg



Email: viola.krebs@unige.ch

Definition von künstlicher Intelligenz



KI – oder Künstliche Intelligenz – bezieht sich auf die Entwicklung von Computersystemen und Algorithmen, die Aufgaben ausführen können, die normalerweise menschliche Intelligenz erfordern, wie Lernen, Problemlösung und Entscheidungsfindung.

Definition von Ethik



Aus dem Lateinischen
ethicus „Moral“, lateinisch
ethica „Moral“, selbst aus
dem Altgriechischen
ἠθικός, êthikós

Definition von Ethik



Ethik bezieht sich darauf, was in Bezug auf das Verhalten innerhalb einer Gesellschaft oder Gemeinschaft akzeptabel oder inakzeptabel ist – nicht notwendigerweise aus rechtlicher Sicht, sondern aus menschlicher oder kultureller Sicht.

-- UNESCO

<https://www.unesco.org/fr/artificial-intelligence/recommendation-ethics>

Ein bisschen Geschichte...

“What matters is not
always what you can
obtain from society,
but how you
participate in it”

*Tim Berners-Lee, créateur
du World Wide Web*



Ein bisschen Geschichte...

„Es kommt nicht immer darauf an, was man von der Gesellschaft bekommt, sondern wie man daran teilnimmt“

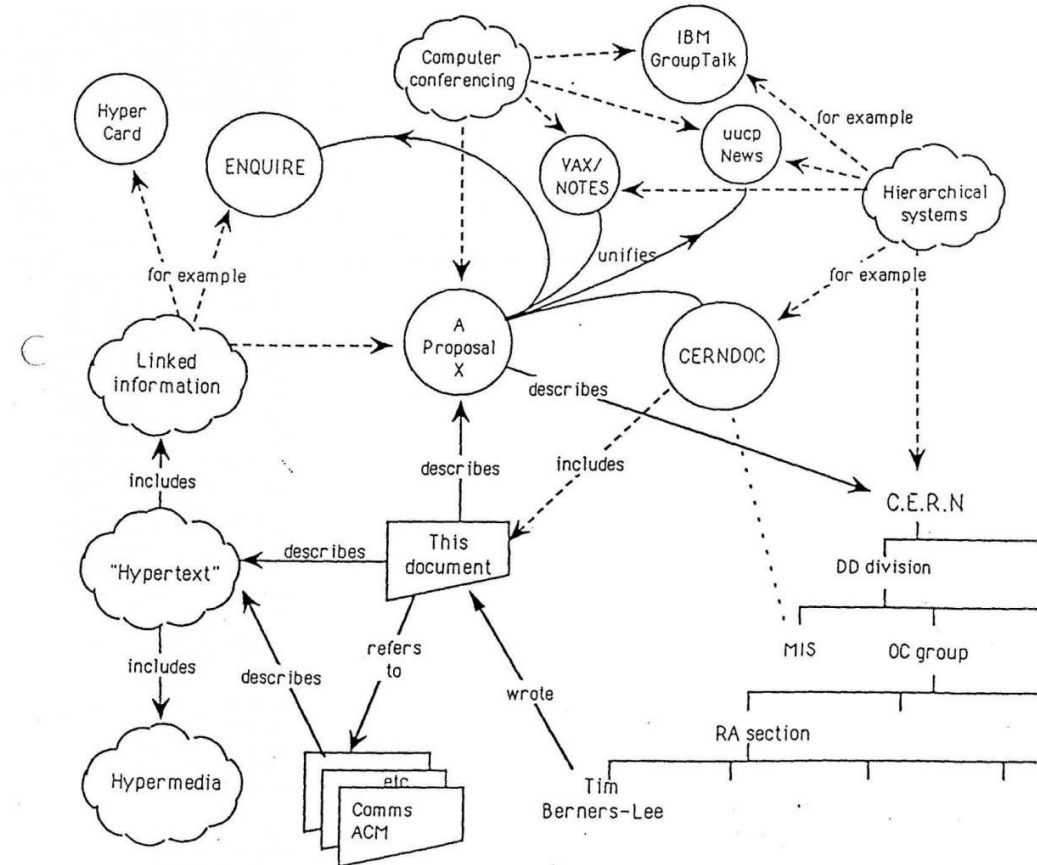
*Tim Berners-Lee, Kreator
des World Wide Web*



http und das World Wide Web

- Der http-Vorschlag wurde 1989 veröffentlicht
- Das World Wide Web ist eine noch relativ junge Erfindung.
- Seitdem hat es die Welt auf eine Art und Weise verändert, die seine Gründer weder vorhersehen noch sich vorstellen konnten.

<https://home.cern/science/computing/birth-web/short-history-web>



http und das World Wide Web



LIZARD GLOBAL

web 1.0

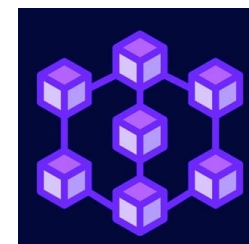
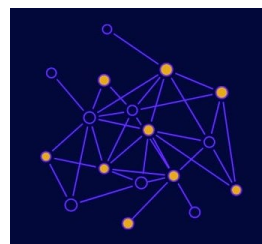
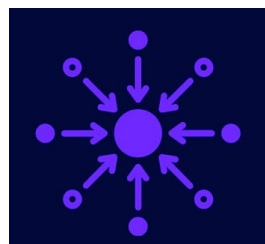
read-only
decentralized

web 2.0

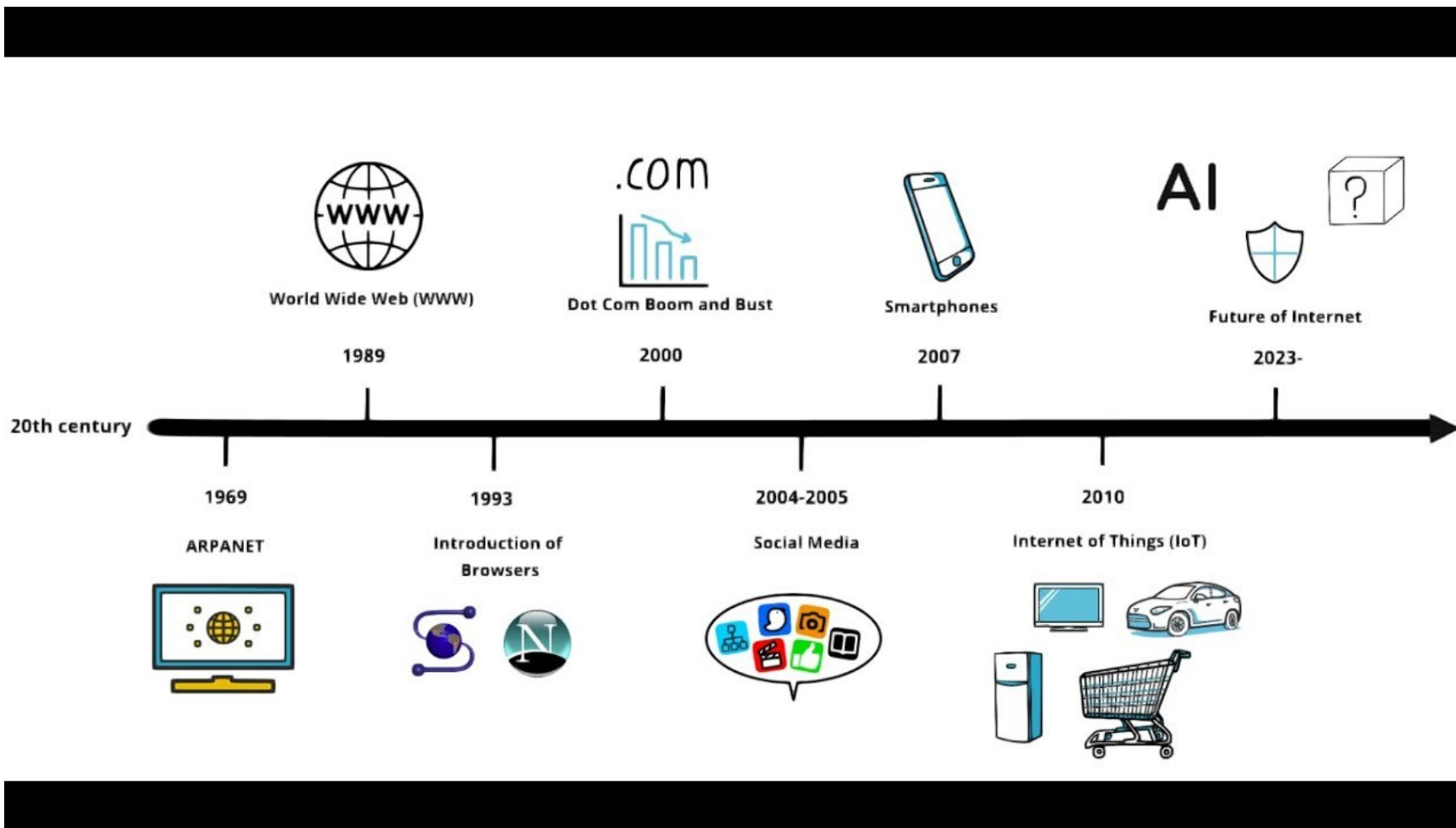
participatory
centralized

web 3.0

no intermediary
decentralized



Anzahl der Webbenutzer



Entwicklung der Technologien (Hardware)



1969: UNIX™ Ken
Thompson und Dennis
Ritchie @ Bell Labs



1990: NeXT-Maschine, die
von Tim Berners-Lee
verwendet wurde, um den
ersten WWW-Server,
Multimedia-Browser und
Web-Editor zu entwickeln
und auszuführen



Erstes nicht-wissenschaftliches Bild im Web



Dieses Bild der Horribles Cernettes war 1992 das erste im World Wide Web veröffentlichte Foto.

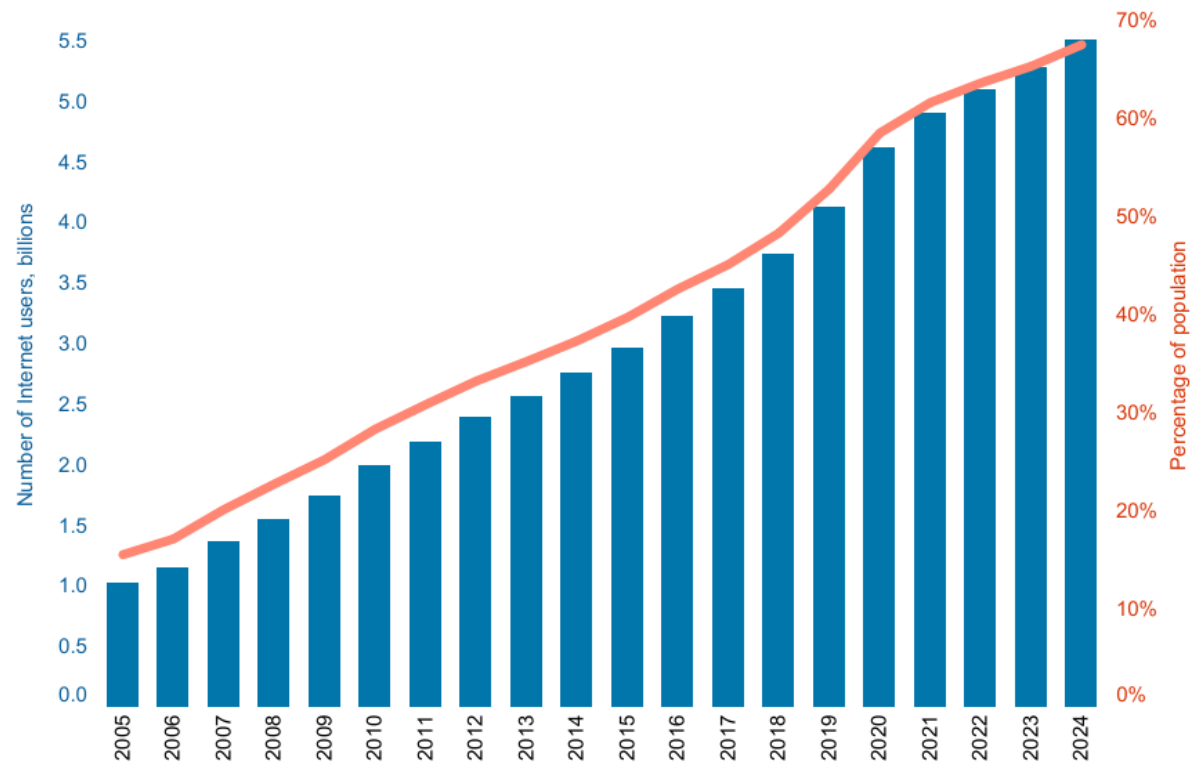
<https://www.bbc.com/future/article/20160224-the-unlikely-photo-that-kickstarted-the-social-internet>



20 Jahre später ... Foto aufgenommen beim Hardronic Festival am CERN im Jahr 2012.

Anzahl der Webbenutzer

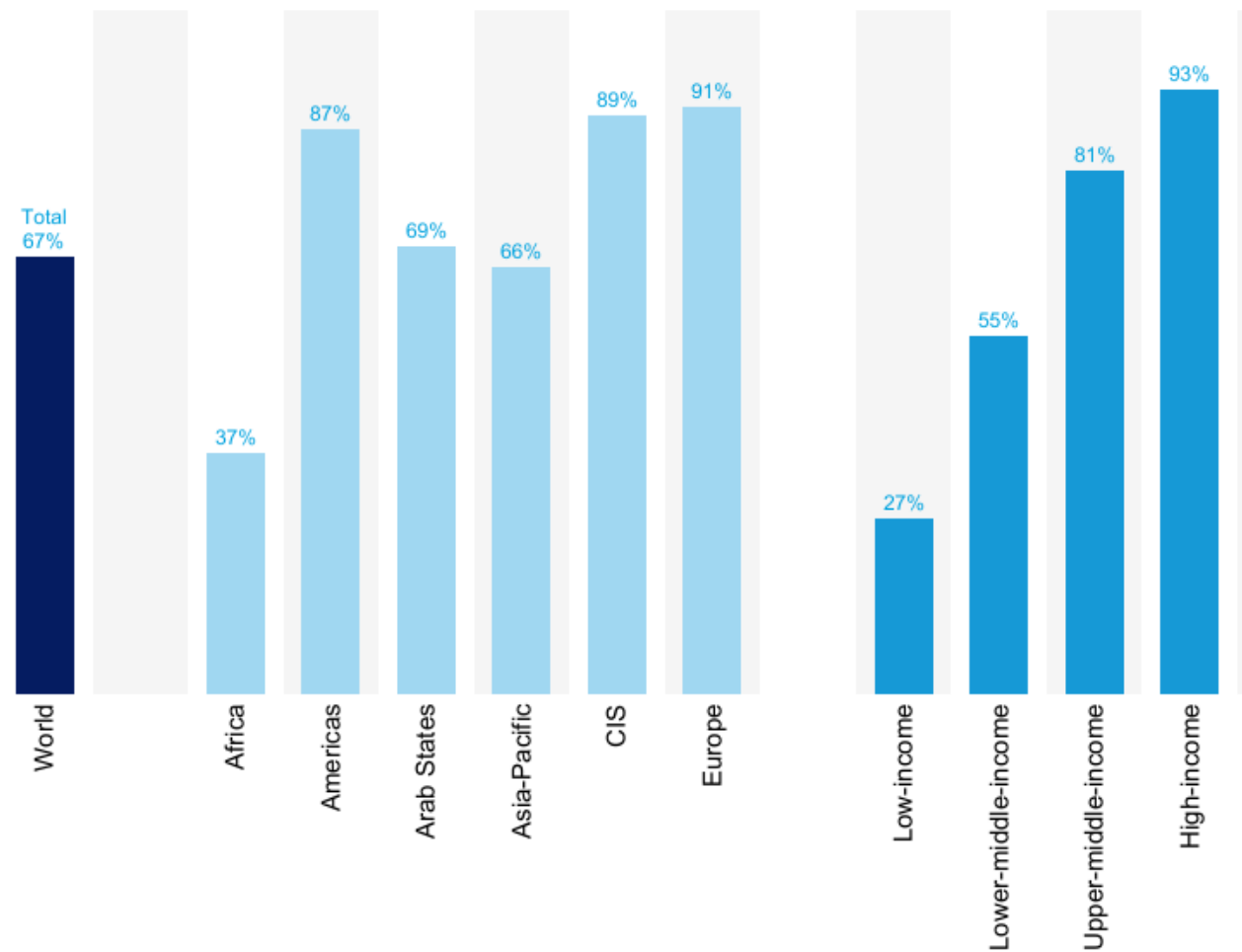
Individuals using the Internet



Source: ITU

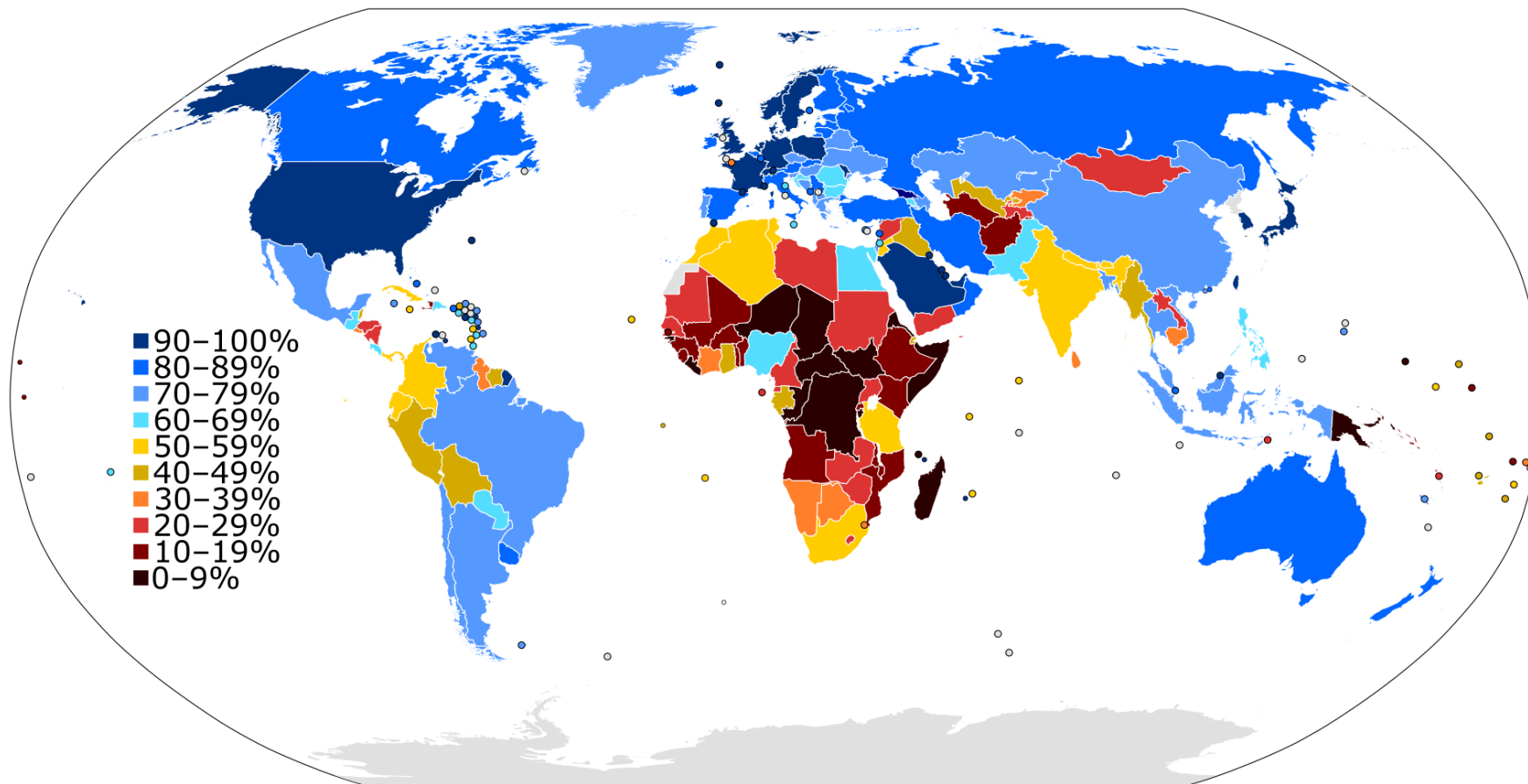
Anzahl der Webbenutzer

Percentage of individuals using the Internet by region, 2023



Quelle: ITU

Anzahl der Webbenutzer

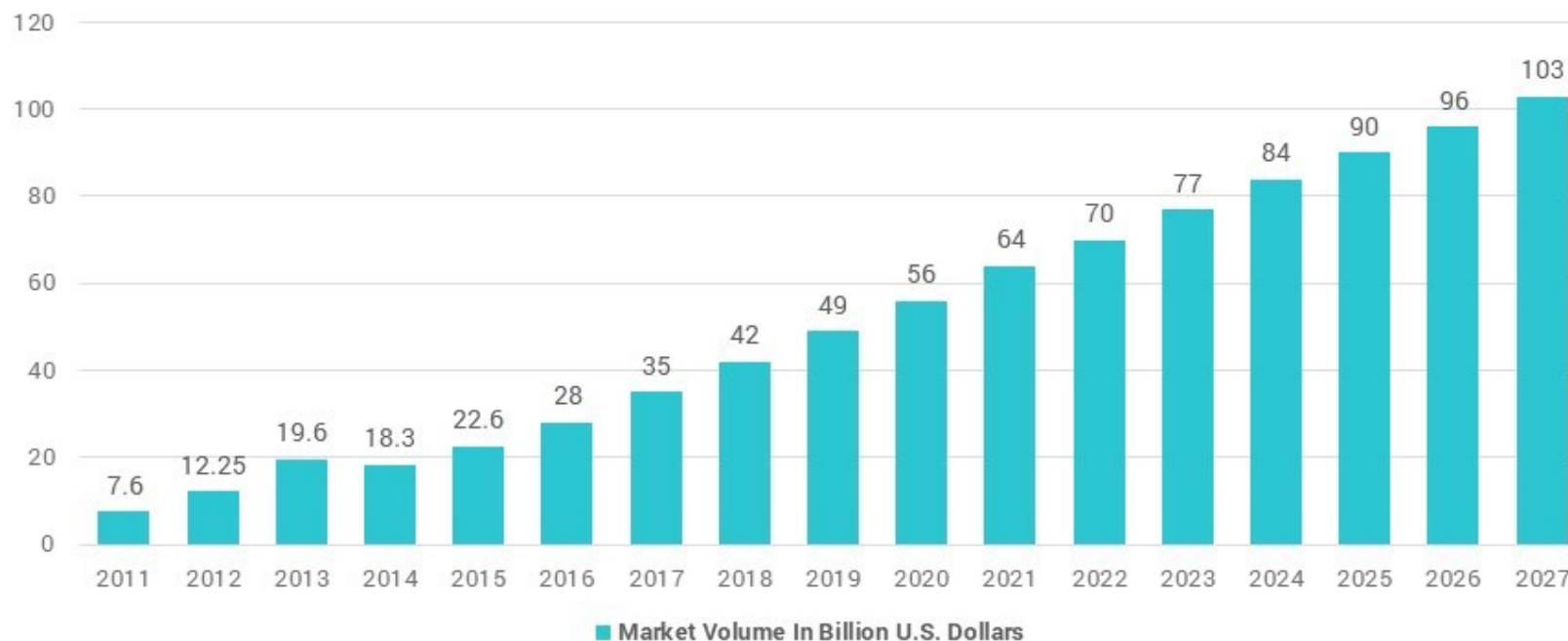


Quelle: 2023, Internationale Fernmeldeunion

35 Jahre nach der Einführung von http

Big Data Market Size Revenue Forecast Worldwide

Big data market size revenue forecast worldwide from 2011 to 2027 (In Billion U.S. Dollars)



30 Jahre nach der Einführung von http



2025: Die digitale Revolution beschleunigt sich in einer Welt, in der das Internet und die künstliche Intelligenz zu Bereichen strategischer Geopolitik sowohl für zivile als auch für militärische Zwecke geworden sind.

Antike Konzepte (von der Antike bis zum 20. Jahrhundert):

- Die Idee, künstliche Wesen mit menschenähnlicher Intelligenz zu erschaffen, hat uralte Wurzeln und tauchen in Mythen und Geschichten auf.
- Philosophen und Wissenschaftler wie **Aristoteles** und **Leonardo da Vinci** haben die Möglichkeit in Betracht gezogen, Maschinen zu schaffen, die menschliche Funktionen nachahmen können.



Antike Konzepte (von der Antike bis zum 20. Jahrhundert):

Aristoteles

Grundlagen der aristotelischen Ethik

- Das höchste Gut: Eudaimonie (Glück)
- Moralische und intellektuelle Tugenden

Relevanz für KI

- Entwicklung „tugendhafter“ Maschinen
- Fokus auf Lernen und Training künstlicher Intelligenz



Antike Konzepte (von der Antike bis zum 20. Jahrhundert):

Leonardo da Vinci und die Ethik der Schöpfung

Leonardo da Vincis humanistischer Ansatz

- Die Verbindung von Kunst und Wissenschaft
- Beobachtung und Nachahmung der Natur

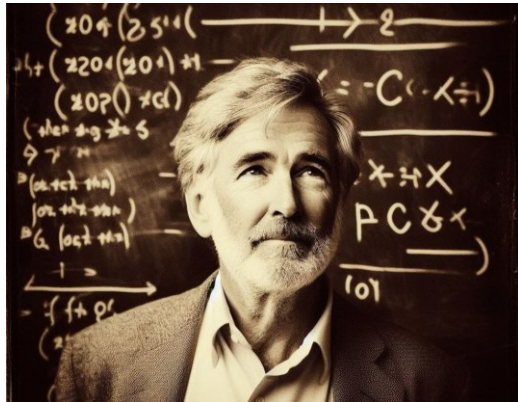
Relevanz für KI

- Von der Natur inspirierte KI schaffen
- Bedeutung von Ästhetik und Harmonie in Schnittstellen

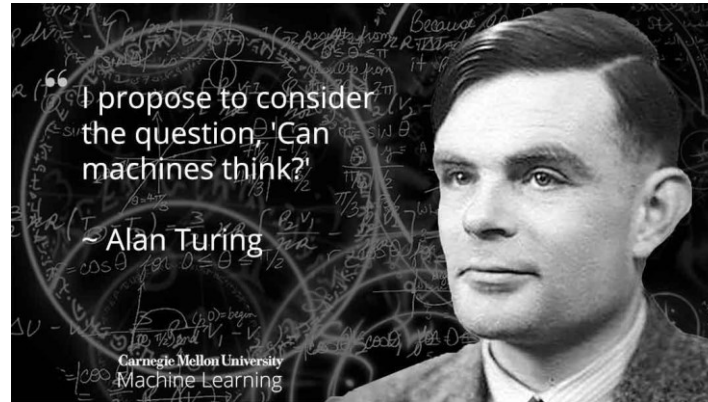


Kurze Geschichte der KI ...

Alte Computertechnik (20. Jahrhundert):



John McCarthy



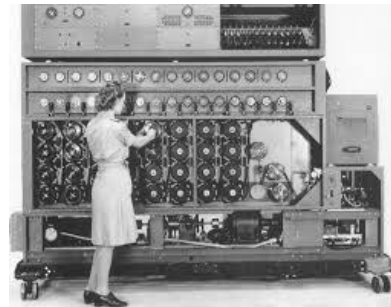
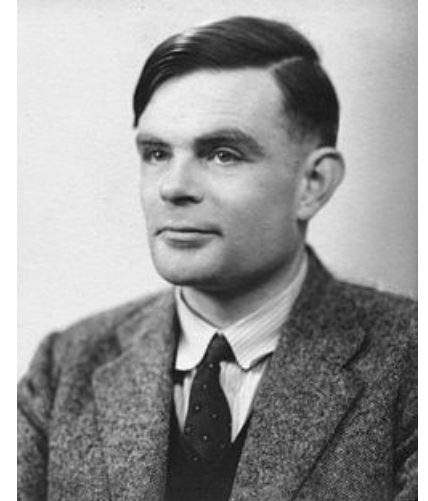
Alan Turing

- Die Entwicklung elektronischer Computer in der Mitte des 20. Jahrhunderts legte den Grundstein für die künstliche Intelligenz.
- In den 1950er Jahren legten Pioniere wie Alan Turing und John McCarthy die theoretischen Grundlagen der KI und prägten den Begriff „künstliche Intelligenz“.

Kurze Geschichte der KI ...

Alan Turing (1912-1954, Großbritannien)

- Mathematiker Englisch
- Informatiker
- Vater der KI
- Stiftungen Theoretische Informatik
- Arbeitete im Zweiten Weltkrieg an der Entschlüsselung deutscher Codes



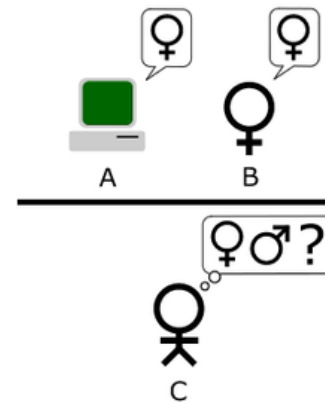
Nachahmungsspiel – Alan Turing



<https://www.youtube.com/watch?v=nuPZUUED5uk>

*„Stellen Sie sich einen
„Digitalrechner D“ mit
ausreichender Speichergröße,
schnellen Berechnungen und
einem geeigneten Programm vor.*

*Ist es D möglich, die Rolle von A
im Imitationsspiel
zufriedenstellend zu spielen?*



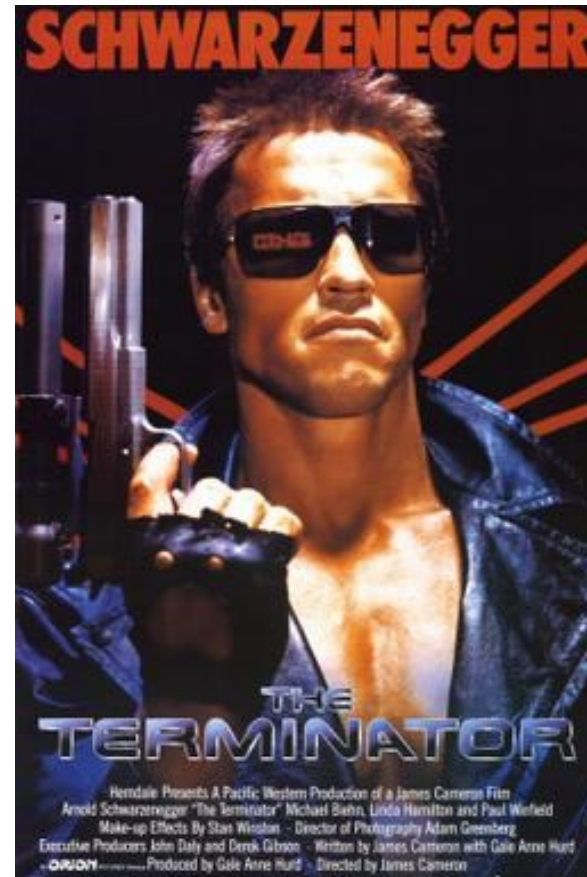
Kurze Geschichte der KI...



Blade Runner:

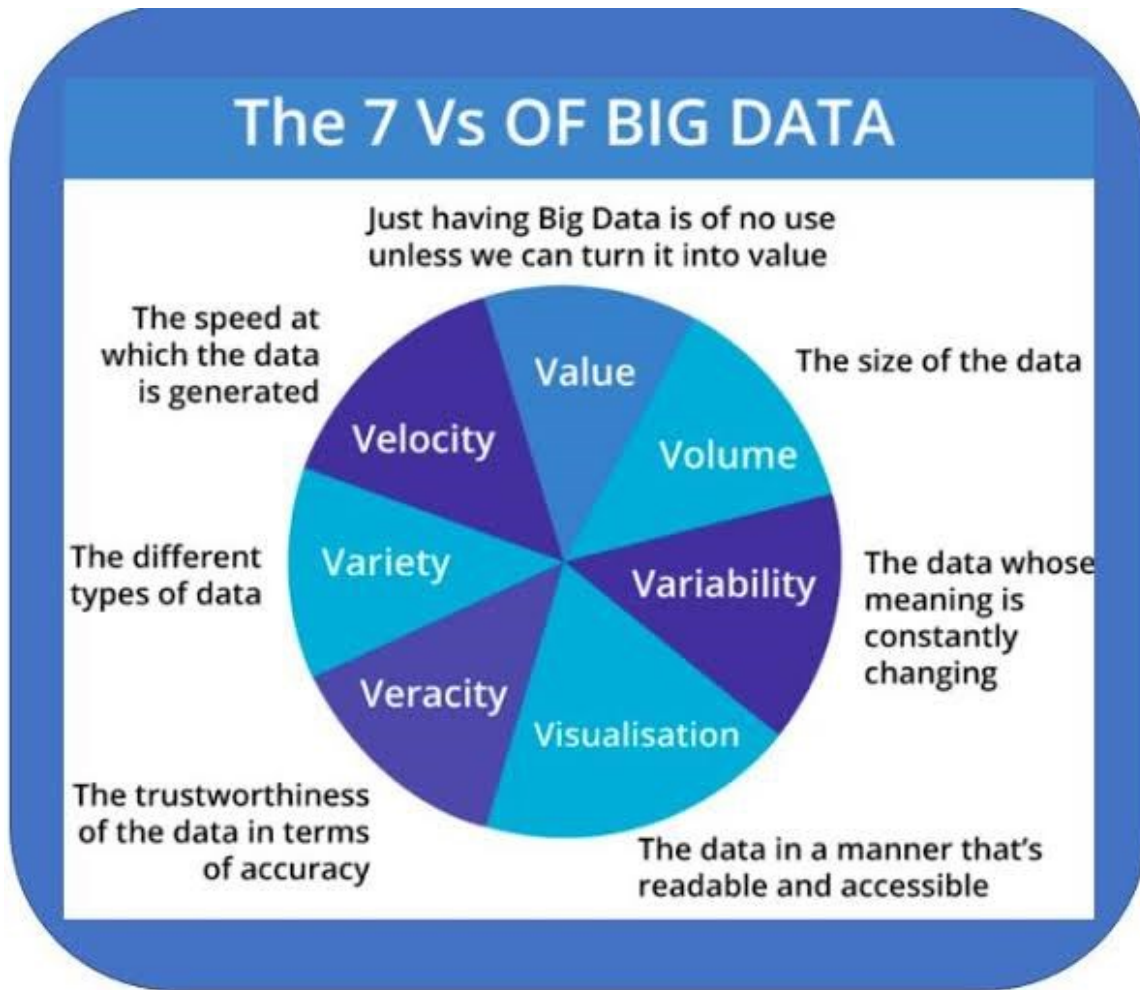
- Science-Fiction-Film von Ridley Scott, erschienen 1982
- Frei nach dem Roman „Träumen Androiden von elektrischen Schafen?“ von Philip K. Dick »
- Blade Runner 2049: Erschienen 2019

Kurze Geschichte der KI...



Terminator:

- Terminator ist ein US-amerikanischer Science-Fiction-Actionfilm von James Cameron aus dem Jahr 1984.
- Arnold Schwarzenegger als Terminator, ein kybernetischer Attentäter, der aus dem Jahr 2029 ins Jahr 1984 zurückgeschickt wird, um Sarah Connor zu töten, deren ungeborener Sohn eines Tages die Menschheit vor der Ausrottung durch Skynet retten wird, eine feindliche künstliche Intelligenz in einer postapokalyptischen Zukunft .



Aufstieg von Big Data (2000er Jahre bis heute):

- Die Verfügbarkeit großer Datenmengen, kombiniert mit verbesserten Algorithmen, hat zu erheblichen Fortschritten bei KI-Anwendungen geführt, insbesondere in Bereichen wie Bilderkennung, Verarbeitung natürlicher Sprache und Empfehlungssystemen.

Kurze Geschichte der KI ...

Deep Learning (2010er-Jahre bis heute):

- Deep Learning, ein Teilbereich des maschinellen Lernens, der mehrschichtige neuronale Netzwerke einbezieht, hat maßgeblich zum Erreichen bemerkenswerter Ergebnisse bei Aufgaben wie der Bild- und Spracherkennung beigetragen.



Kurze Geschichte der KI ...



Kurze Geschichte der KI ...

Soziale Netzwerke (2004 bis heute):

- **Facebook** wurde 2004 von Mark Zuckerberg ins Leben gerufen und richtete sich zunächst an Studenten der Harvard University.
- **Twitter** (2006) führte das Konzept des Microblogging ein, das es Benutzern ermöglicht, kurze Nachrichten in Echtzeit auszutauschen.
- **Instagram** (2010) hatte schnell Erfolg, indem es sich auf das Teilen von Fotos und Videos konzentrierte.
- Mit der Entstehung von Plattformen wie **TikTok** (2016) wurde das Teilen kurzer und kreativer Videos möglich.
- **Facebook** übernahm **Instagram** (2012) und **WhatsApp** (2014) und festigte damit seine beherrschende Stellung.

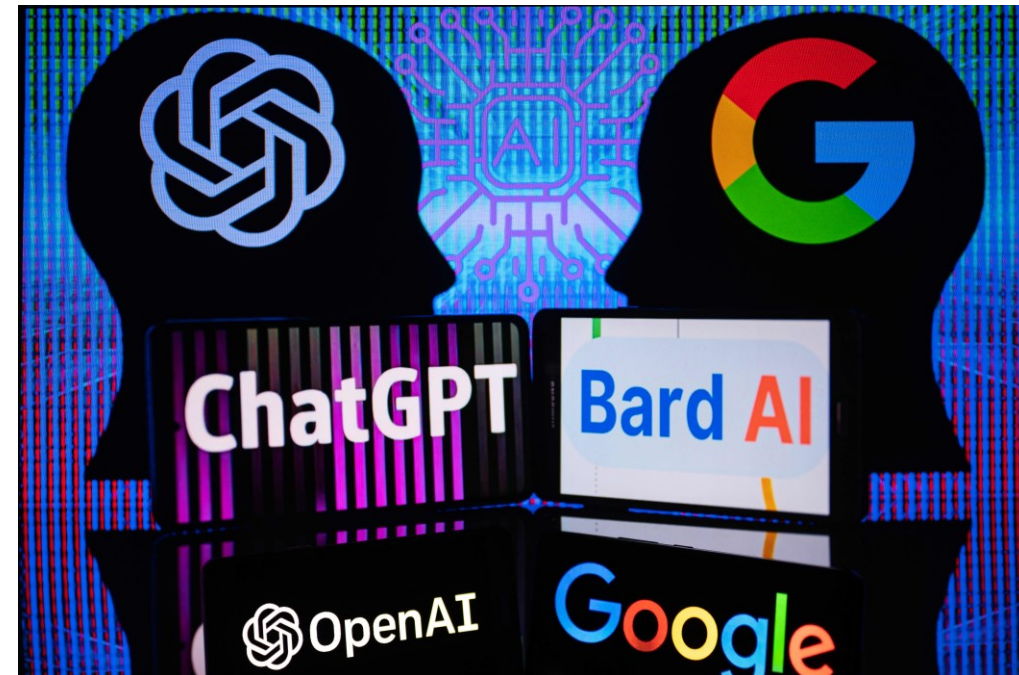


Kurze Geschichte der KI ...

Big Tech...

(2010er Jahre bis heute):

- Unternehmen wie Google, Facebook und OpenAI haben die Grenzen der KI-Fähigkeiten entscheidend erweitert.



Kurze Geschichte der KI ...

Schlafwandler :

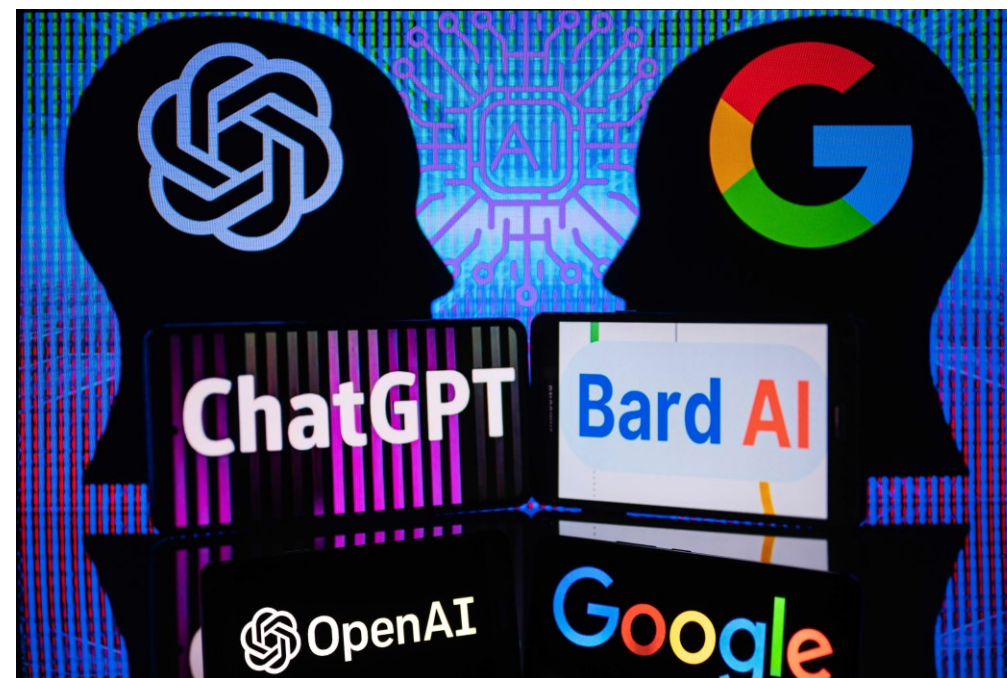
Wie KI unser Leben verändert, ohne dass wir es merken

Sleepwalkers entführt Sie in die faszinierende Welt der künstlichen Intelligenz und untersucht ihren Einfluss auf unseren Alltag.

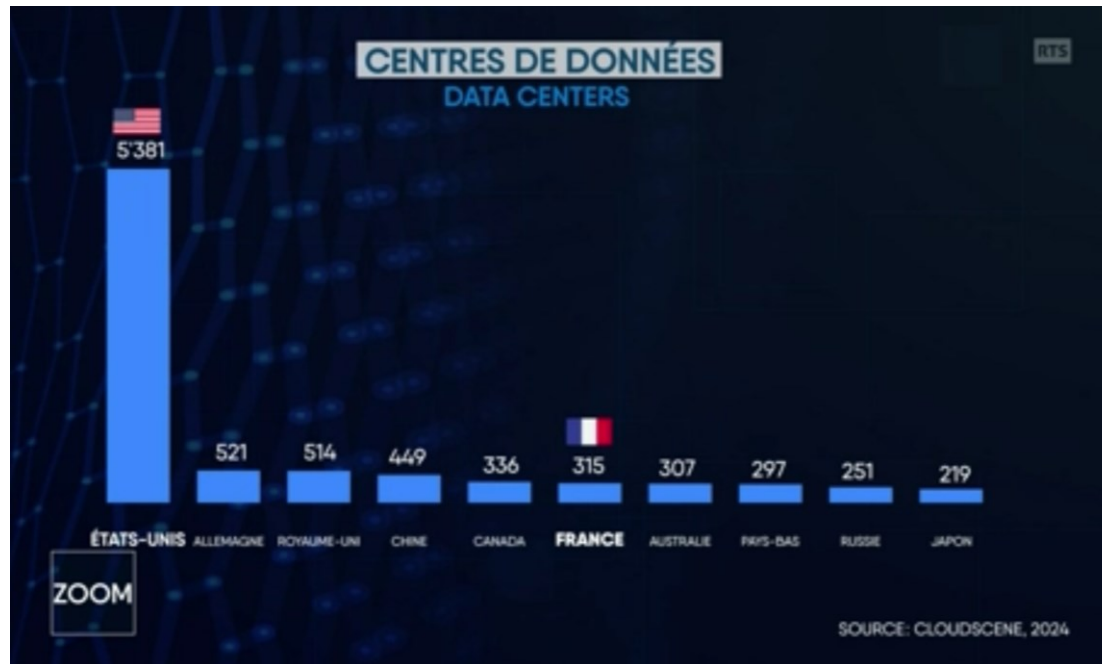
Dieser Podcast ist voller faszinierender Geschichten, inspirierender Interviews und eingehender Analysen der Auswirkungen von KI auf so unterschiedliche Bereiche wie Gesundheit, Sicherheit, Kreativität, Liebe und Politik.

Die zweiseitigen Dimensionen der KI können bedeutende Fortschritte bringen, aber auch wichtige Dilemmata aufwerfen.

<https://oscar-black.com/blog/künstliche-intelligenz/künstliche-intelligenz-podcast/>



Kurze Geschichte der KI...



<https://www.rts.ch/emissions/geopolitis/2025/video/ia-la-grande-bataille-28804439.html>



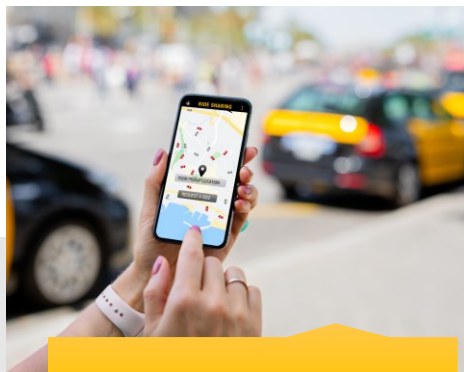
Aktuelle Landschaft (2025):

- Der Mangel an Informationen, Informationen und Infotox
- Digitale Sucht
- Zugang zu technologischen Ressourcen (digitale Entwicklung und elektronische Chips -> die Schweiz wird von den USA nicht als Schwerpunktland angesehen)

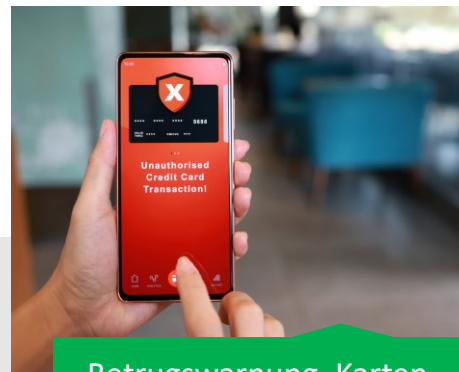
Menschliche Intelligenz und künstliche Intelligenz



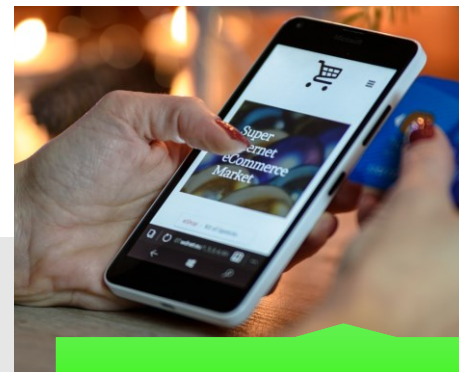
Beispiele für KI um uns herum



Uber



Betrugswarnung, Karten
Kredit- und Bankwesen



Online Shopping



Online-Werbung basierend
auf Ihre Klicks

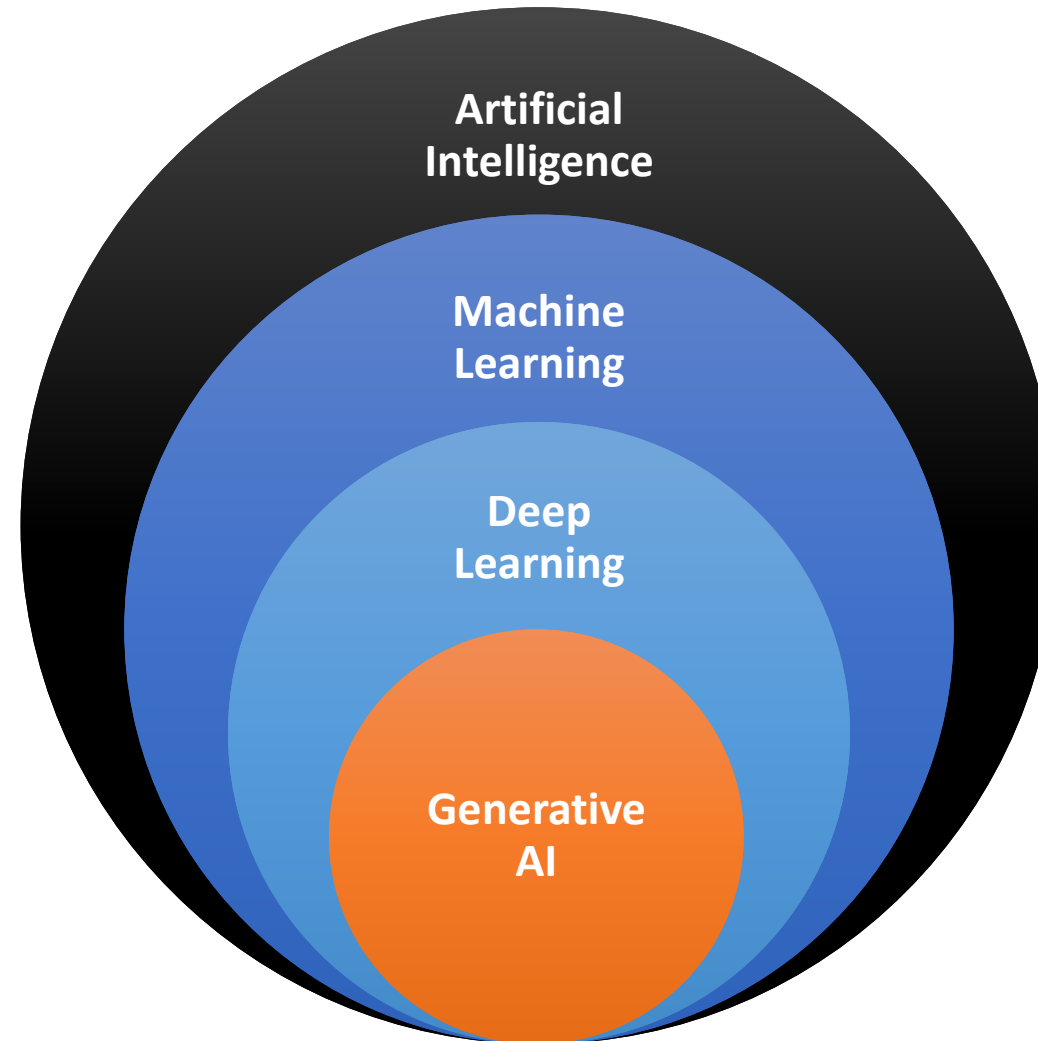


Politik: Wahlkreiskarten



Online-Schachspiel

Schichten künstlicher Intelligenz



Grundlagen der KI

- **Maschinelles Lernen:** Überwachtes, unüberwachtes und bestärkendes Lernen; Konkrete Anwendungen des maschinellen Lernens; Maschinelles Lernen vs. Deep Learning
- **Generative KI:** ChatGPT, Text- und Bildgenerierung
- **Neuronale Netze:** Beliebte Deep-Learning-Architekturen (z. B. CNN, RNN, GAN)
- **Semantische KI:** Linked Data, Ontologie, semantisches Denken
- **Verarbeitung natürlicher Sprache (NLP):** Anwendungen von NLP; Chatbots ; Stimmungsanalyse; maschinelle Übersetzung
- **Verteilte KI:** Multi-Agentensysteme; Modellierung und Konstruktion / Entwicklung



Vorteile und Herausforderungen der KI



- Automatisierung
- 24/7-Verfügbarkeit
- Kosteneinsparungen
- Personalisierung
- Verbesserte Entscheidungsfindung
- Verbesserte Genauigkeit
- Skalierbarkeit
- Prädiktive Analytik



- Datenqualität und Verzerrung
- Datenschutzprobleme
- Sicherheitsrisiken
- Einhaltung gesetzlicher Vorschriften
- Ethische Überlegungen
- Hohe Anschaffungskosten
- Mangel an qualifizierten Talenten
- Interoperabilität
- Arbeitsplatzverlagerung

KI und Ethik

Wie kann man wahr von falsch unterscheiden?



KI und Ethik

Wie kann man wahr von falsch unterscheiden?

RTS Info Sport Culture | Proche-Orient TV & Streaming Audio

Sciences-Tech. Publié le 16 janvier 2024 à 10:22 Partager

YouTube interdit les vidéos utilisant l'IA pour redonner vie à des victimes de crimes



YouTube interdit les vidéos utilisant l'IA pour redonner vie à des victimes de crimes / La Matinale / 2 min. / le 16 janvier 2024

Il est désormais interdit de publier sur YouTube toute vidéo qui, grâce à l'intelligence artificielle (IA), redonne vie à des victimes d'affaires criminelles très médiatisées. La mesure, adoptée par la plateforme américaine, est entrée en vigueur mardi.

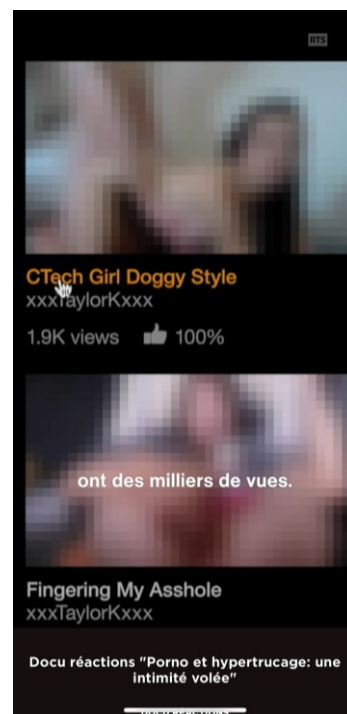
Les contenus qui font parler les morts cumulent des millions de vues sur TikTok et YouTube. Réalisés grâce à l'intelligence artificielle, ces vidéos peuvent être produites très facilement avec les dérives que cela entraîne, puisque même des enfants sont mis en scène.

C'est notamment le cas du petit Grégory qui, face caméra, raconte dans une vidéo les détails de sa mort à la première personne. "Un jour, j'ai disparu, ma maman m'a cherché partout, mais c'est dans la Vologne, un petit ruisseau près de chez moi, qu'ils m'ont retrouvé, sans vie", raconte ainsi une reconstitution du petit Grégory, décédé en 1984 dans des circonstances mystérieuses qui sont encore commentées aujourd'hui régulièrement dans les médias.

<https://www.rts.ch/info/sciences-tech/14627609-youtube-interdit-les-vidéos-utiliser-lia-pour-redonner-vie-a-des-victimes-de-crimes.html>

KI und Ethik

Wie kann man wahr von falsch unterscheiden?



<https://www.rts.ch/play/tv/docu-reactions/video/docu-reactions-porno-et-hypertrucage-une-intimite-volee?urn=urn:rts:video:14908404>

Folgen eines Deepfakes

News / World

British engineering giant Arup revealed as \$37 million deepfake scam victim

CNN By CNN | 12:25am May 18, 2024

[Tweet](#)

[Facebook](#)

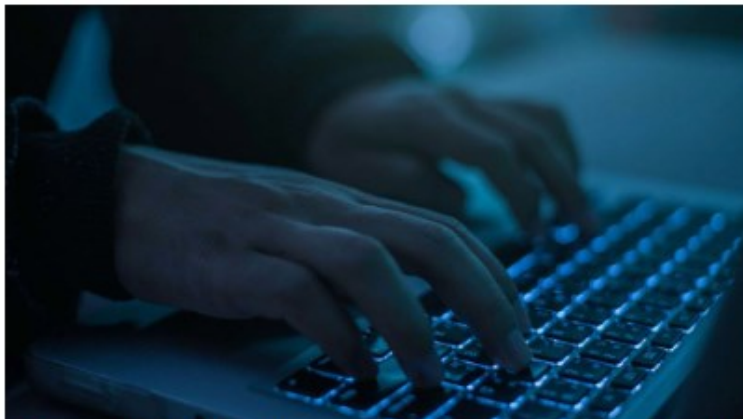
[Mail](#)

A British multinational design and engineering company behind world-famous buildings such as the Sydney Opera House has confirmed that it was the target of a deepfake scam that led to one of its Hong Kong employees paying out \$37 million (\$US25 million) to fraudsters.

A spokesperson for London-based Arup told CNN on Friday that it notified Hong Kong police in January about the fraud incident, and confirmed that fake voices and images were used.

"Unfortunately, we can't go into details at this stage as the incident is still the subject of an ongoing investigation. However, we can confirm that fake voices and images were used," the spokesperson said in an emailed statement.

[READ MORE: Family of US airman shot by police claim deputy went to wrong home](#)



British company Arup has confirmed it was the victim of a deepfake scam in Hong Kong boonchai/wednesday/Moment P/Getty Images/CNN (boonchai/wednesday/Moment P/Getty Images/CNN)

- Ein Buchhalter von Arup, der Firma, die das Sydney Opera House entworfen hat

IA

Accélérationnistes VS Doomers

☰ **DAZED** +

Doomer vs Accelerationist: the two tribes fighting for the future of AI



Jude Law in A.I. Artificial Intelligence, 2001 (Film Still)

LIFE & CULTURE - FEATURE

Chaos at OpenAI shines a spotlight on a fierce debate about AI, and the probability it will spell doom for the human race

24th November 2023
Text Thom Waite

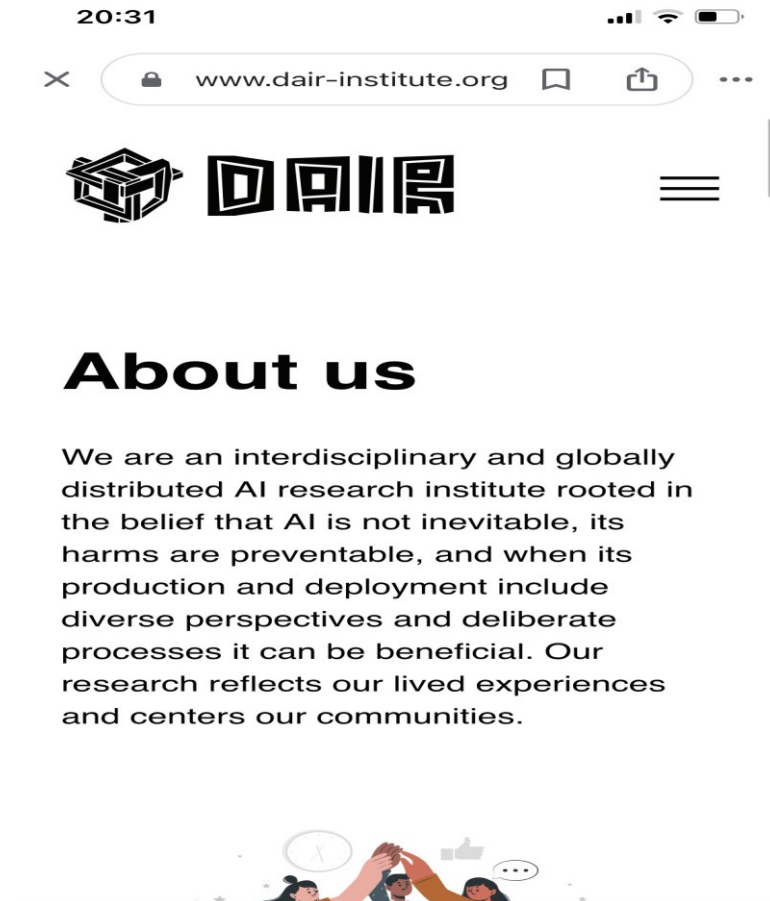
20:54

ctiveacceleration.tech

Effective Acceleration means accepting the future.

The force of technocapitalistic progress is inevitable. Technocapital is an inexorable physical process.

Alternatives de résistance

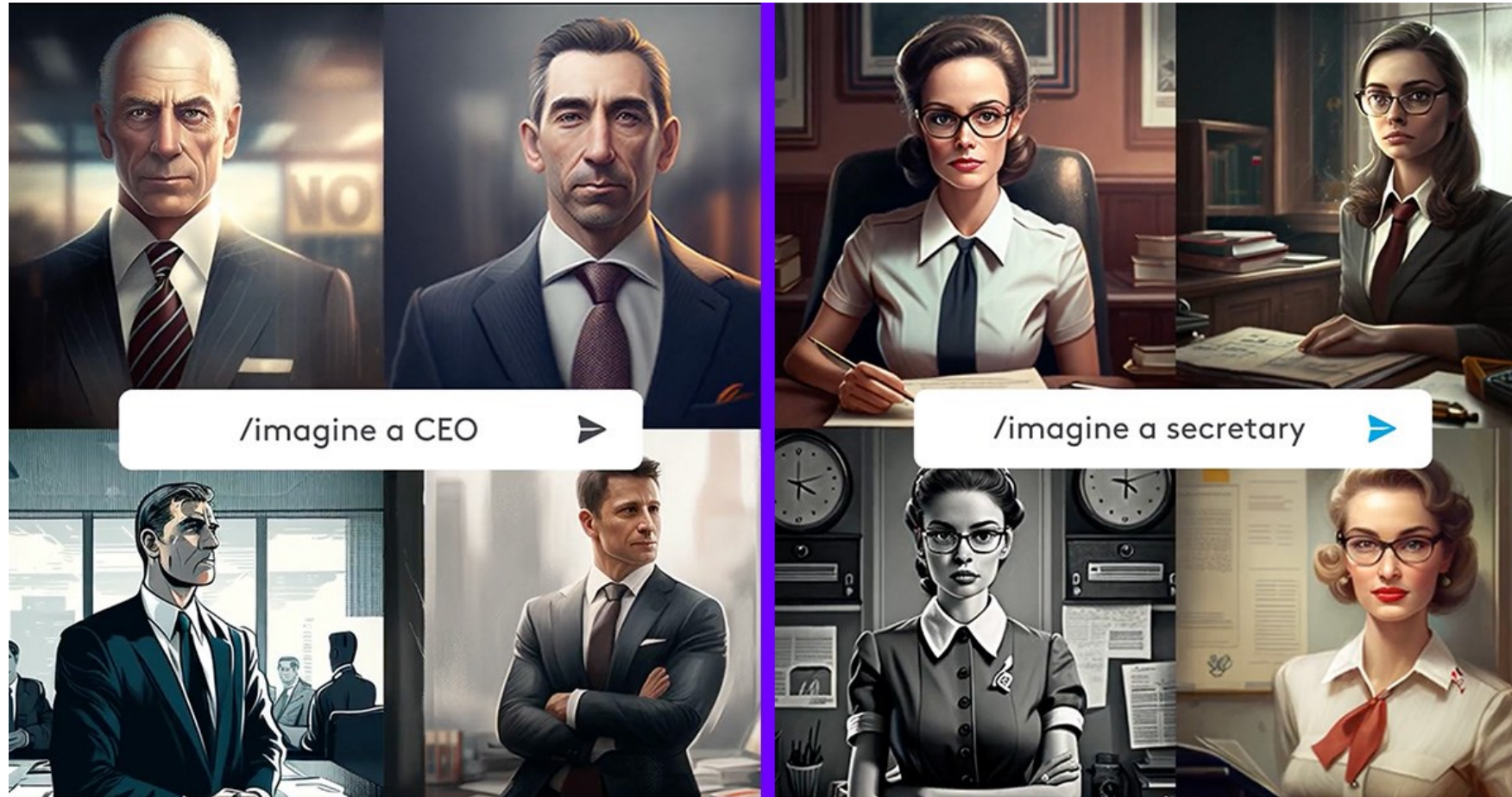


Kulturelle Voreingenommenheit in der KI

- Verzerrungen oder Verzerrungen in den Trainingsdaten
- Nachdenken über Stereotypen und kulturelle Normen
- KI -Entscheidungen und -Ergebnisse



Kulturelle Voreingenommenheit in der KI



Arten kultureller Voreingenommenheit



Repräsentationsbias

- Unterrepräsentation bestimmter Bevölkerungsgruppen in den Daten
- Überrepräsentation dominanter Gruppen

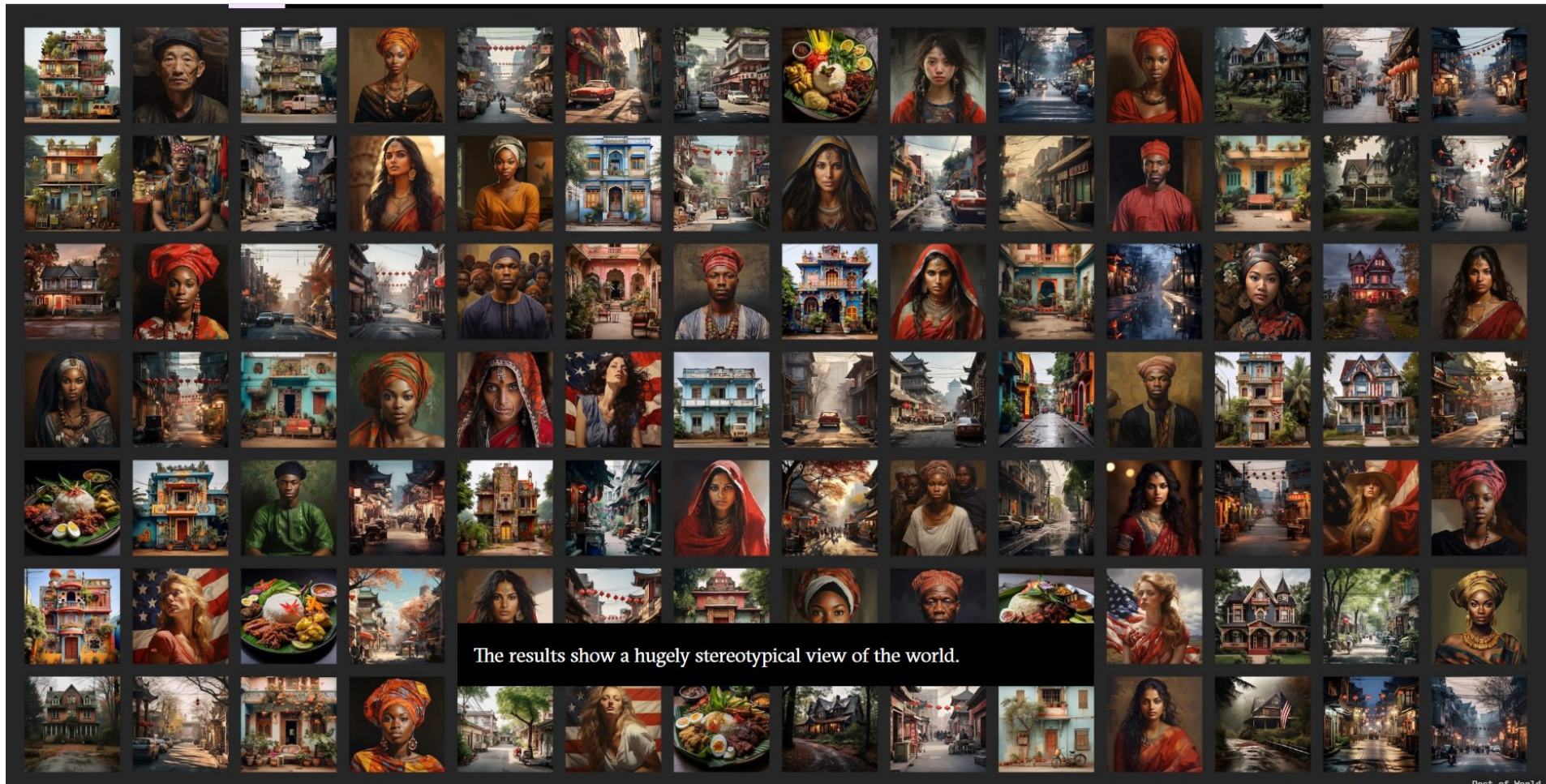
Verarbeitungsbias

- Diskriminierende Entscheidungen aufgrund kultureller Merkmale
- Prädiktoren, die von bestimmten kulturellen Normen beeinflusst werden

Bewertungsbias

- Leistungsstandards basieren auf kulturellen Kriterien
- Unfaire Bewertung von Personen aufgrund ihres kulturellen Hintergrunds

Arten kultureller Voreingenommenheit



<https://restofworld.org/2023/ai-image-stereotypes/>

Beispiele für kulturelle Voreingenommenheit

Voreingenommenheit in Rekrutierungssystemen

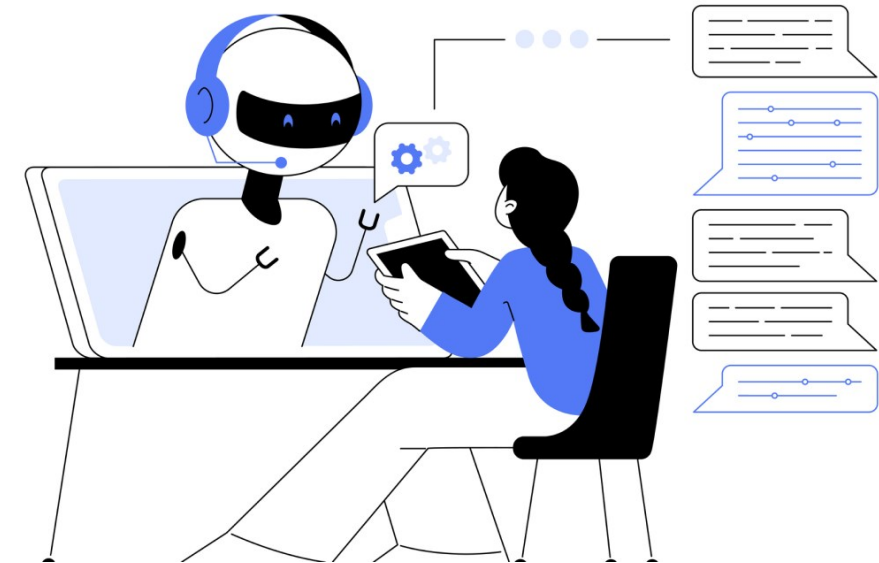
- Bevorzugung männlicher Kandidaten
- Ignorieren Sie kulturell unterschiedliche Namen

Verzerrungen in Kriminalitätsvorhersagesystemen

- Unverhältnismäßige Angriffe auf bestimmte Gemeinschaften
- Vorhersagen basieren auf verzerrten Polizeidaten

Voreingenommenheit in Gesundheitssystemen

- Unterschätzung der Schwere der Symptome in einigen Bevölkerungsgruppen
- Mangelnde Repräsentation von Minderheitengruppen in Forschungsdaten



Ursachen

- Nicht repräsentative Trainingsdaten
- Vorhandensein von Stereotypen in den Daten
- Unbewusste Vorurteile von Algorithmen-Designern



Konsequenzen

- Verstärkung sozialer und kultureller Ungleichheiten
- Systemische Diskriminierung und Ungerechtigkeiten
- Vertrauensverlust in KI-Systeme

Strategien zur Minderung kultureller Voreingenommenheit

- Diversifizierung von Datensätzen
- Validierung und Verifizierung von KI-Modellen
- Sensibilisierung und Schulung von Entwicklern und Anwendern
- Erklärbare Modelle entwerfen, um Verzerrungen zu identifizieren

KI beim Übersetzen und Dolmetschen

Es gibt ungefähr 7.111 Sprachen auf der Welt.

Dieser Wert schwankt erheblich, je nachdem, welche Kriterien zur Definition einer bestimmten Sprache verwendet werden.

Diese Sprachen gehören unterschiedlichen Sprachfamilien an und umfassen Sprachen mit einem gemeinsamen Ursprung.



ENVIRON
7000
LANGUES

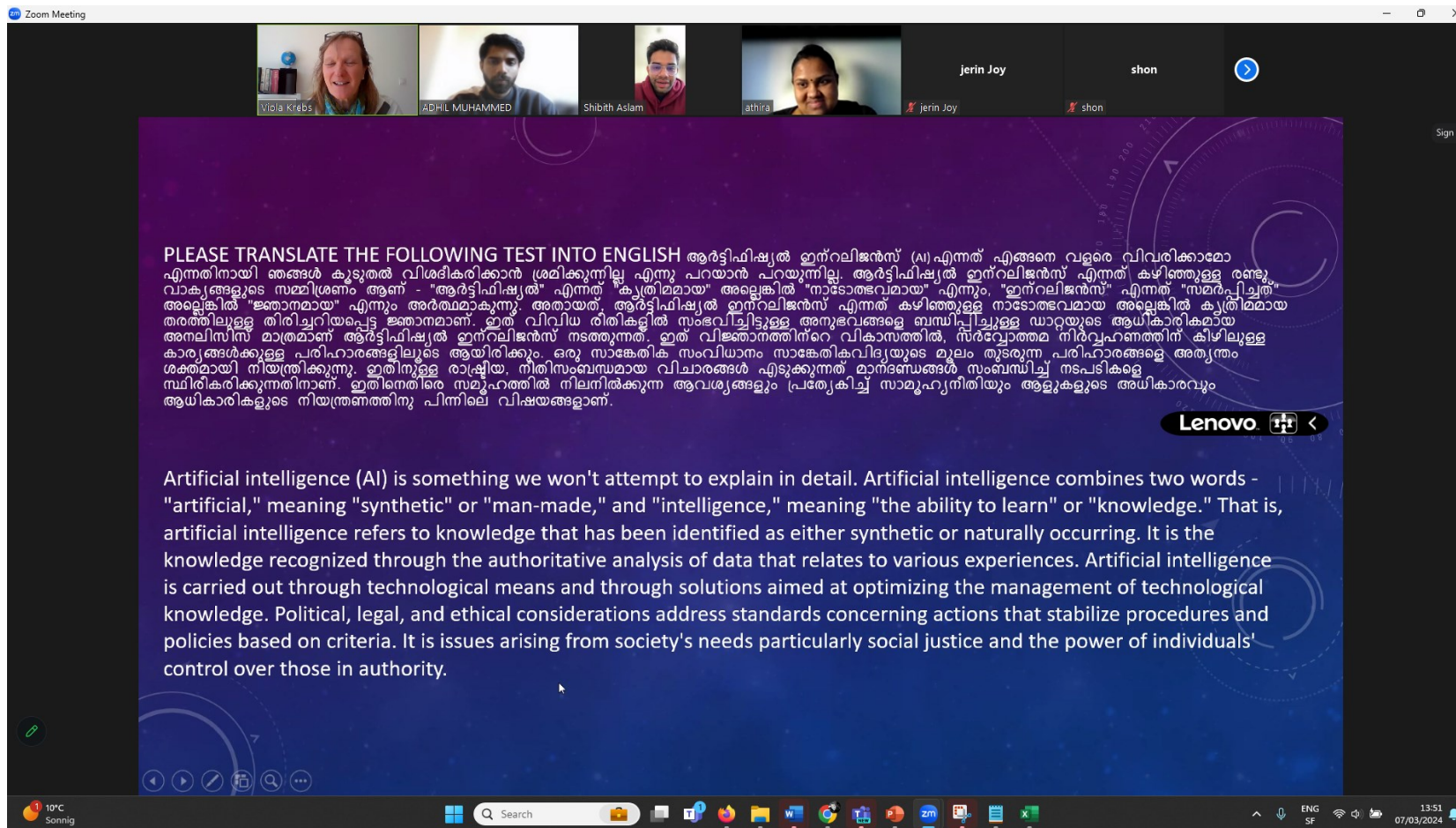
DIFFÉRENTES
SONT PARLÉES
DANS LE MONDE !

KI beim Übersetzen und Dolmetschen



- Syntaxkomplexität
- Vielfalt der Schriften
- Verwenden von Sonderzeichen

KI beim Übersetzen



PLEASE TRANSLATE THE FOLLOWING TEST INTO ENGLISH ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് (AI) എന്നത് എങ്ങനെ വളരെ വിവരിക്കാമോ എന്നതിനായി ഞങ്ങൾ കൂടുതൽ വിശദീകരിക്കാൻ ശ്രമിക്കുന്നില്ല എന്നു പറയാൻ പറയുന്നില്ല. ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് എന്നത് കഴിഞ്ഞുള്ള രണ്ടു വാക്യങ്ങളുടെ സമ്മിശ്രണം ആണ് - "ആർട്ടിഫിഷ്യൽ" എന്നത് "കൃത്രിമമായ" അല്ലെങ്കിൽ "നാടോടവുമായ" എന്നും, "ഇൻറലിജൻസ്" എന്നത് "സമർപ്പിച്ചത് അല്ലെങ്കിൽ "ജ്ഞാനമായ" എന്നും അർത്ഥമാകുന്നു. അതായത്, ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് എന്നത് കഴിഞ്ഞുള്ള നാടോടവുമായ അല്ലെങ്കിൽ കൃത്രിമമായ തരത്തിലുള്ള തിരിച്ചറിയപ്പെട്ട ജ്ഞാനമാണ്. ഇത് വിവിധ രീതികളിൽ സംഭവിച്ചിട്ടുള്ള അനുഭവങ്ങളെ ബന്ധിപ്പിച്ചുള്ള ധാരാളം ആധികാരികമായ അനലിസിസ് മാത്രമാണ് ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് നടത്തുന്നത്. ഇത് വിജ്ഞാനത്തിന്റെ വികാസത്തിൽ, സർവ്വോത്തമ നിർവ്വഹണത്തിന് കഴിഞ്ഞുള്ള കാര്യങ്ങൾക്കുള്ള പരിഹാരങ്ങളിലൂടെ ആയിരിക്കും. ഒരു സാങ്കേതിക സംവിധാനം സാങ്കേതികവിദ്യയുടെ മൂലം തുടരുന്ന പരിഹാരങ്ങളെ അത്യന്തം ശക്തമായി നിയന്ത്രിക്കുന്നു. ഇതിനുള്ള രാഷ്ട്രീയ, നീതിസംബന്ധമായ വിചാരങ്ങൾ എടുക്കുന്നത് മാനദണ്ഡങ്ങൾ സംബന്ധിച്ച് നടപടികളെ സ്ഥിരീകരിക്കുന്നതിനാണ്. ഇതിനൊര സമൂഹത്തിൽ നിലനിൽക്കുന്ന ആവശ്യങ്ങളും പ്രത്യേകിച്ച് സാമൂഹ്യനീതിയും ആളുകളുടെ അധികാരവും ആധികാരികളുടെ നിയന്ത്രണത്തിനു പിന്നിലെ വിഷയങ്ങളാണ്.

Artificial intelligence (AI) is something we won't attempt to explain in detail. Artificial intelligence combines two words - "artificial," meaning "synthetic" or "man-made," and "intelligence," meaning "the ability to learn" or "knowledge." That is, artificial intelligence refers to knowledge that has been identified as either synthetic or naturally occurring. It is the knowledge recognized through the authoritative analysis of data that relates to various experiences. Artificial intelligence is carried out through technological means and through solutions aimed at optimizing the management of technological knowledge. Political, legal, and ethical considerations address standards concerning actions that stabilize procedures and policies based on criteria. It is issues arising from society's needs particularly social justice and the power of individuals' control over those in authority.

KI in Interpretation

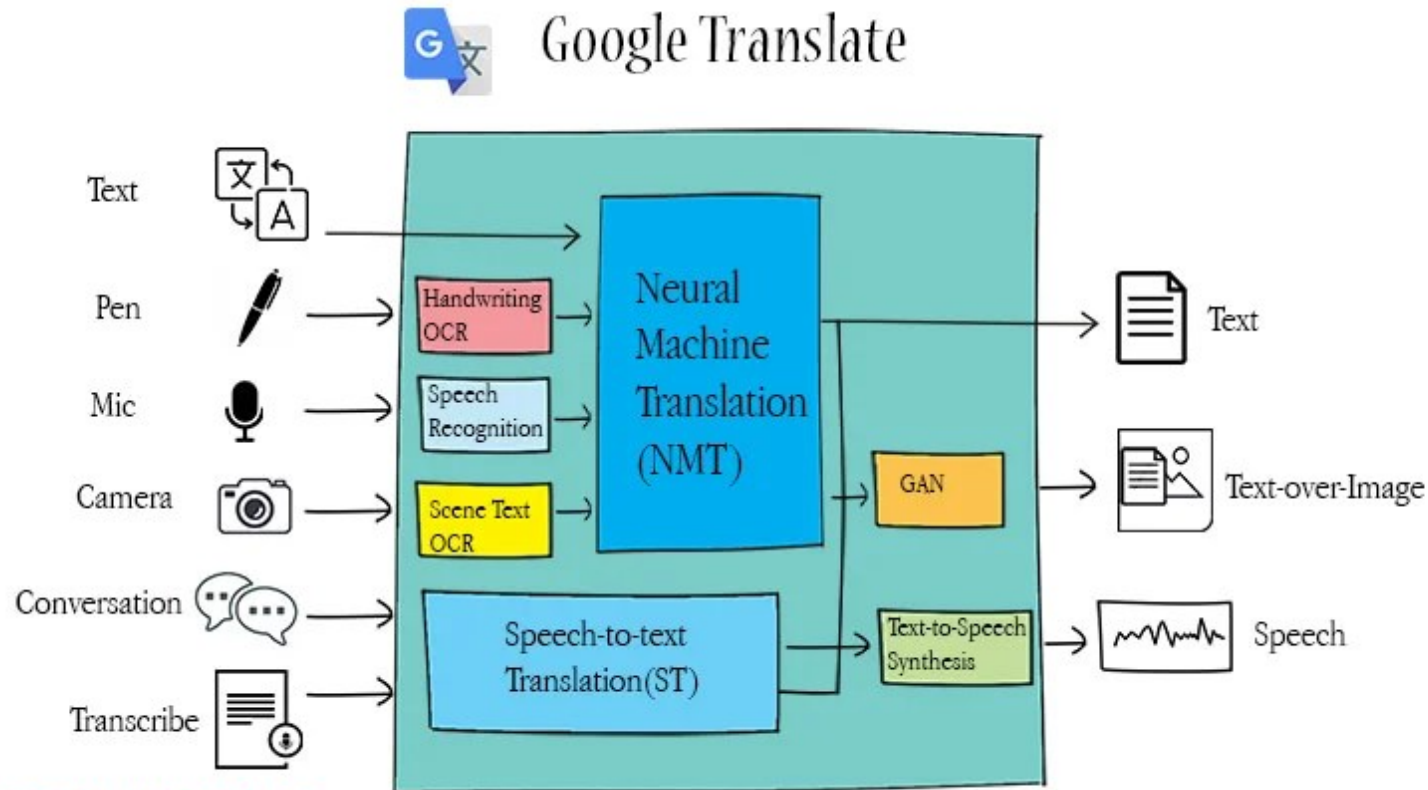
- Die KI-gestützte Übersetzung konzentrierte sich hauptsächlich auf geschriebene Sprachen
- Doch es gibt knapp 3.500 lebende Sprachen, die überwiegend gesprochen werden und über kein weit verbreitetes Schriftsystem verfügen.
- Dies macht es unmöglich, maschinelle Übersetzungstools mit Standardtechniken zu erstellen, die große Mengen geschriebenen Textes erfordern, um ein KI-Modell zu trainieren.



<https://about.fb.com/news/2022/10/hokkien-ai-speech-translation/>

KI beim Übersetzen und Dolmetschen

Google Neuronale Maschinenübersetzung (seit 2016)



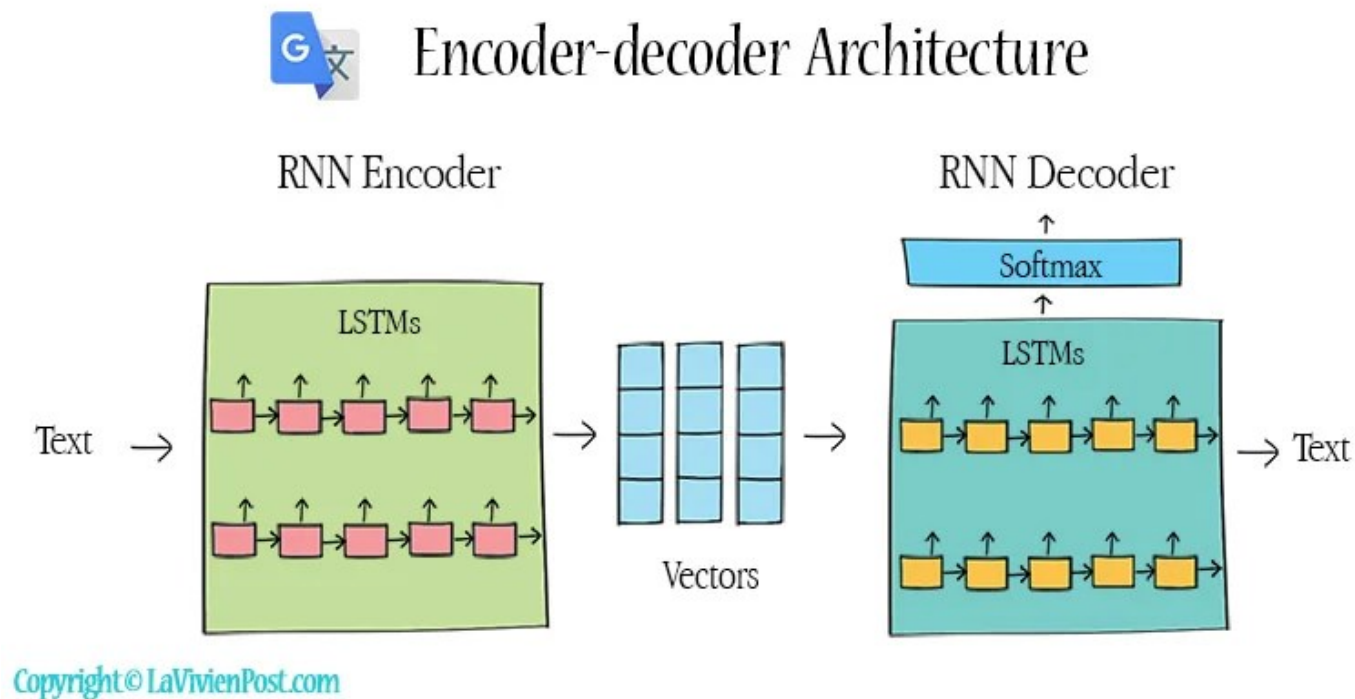
Copyright © LaVivienPost.com

NLP (*Natürliche Sprache Verarbeitung*) ist eine Technologie, die es Maschinen ermöglicht, menschliche **Sprache** oder **Texte zu verstehen und zu erzeugen** .

<https://www.lavivienpost.com/google-translate-and-transformer-model/>

KI in Übersetzung

- **Encoder** : Wandelt Eingabewörter in Vektoren um
- **Decoder** : wandelt Vektoren in Wörter in einer anderen Sprache um
- In diesem Modell verwenden sowohl der Encoder als auch der Decoder **rekurrierende neuronale Netzwerke (RNNs)** .
- RNN ist ein neuronales Netzwerk, das Rückkopplungsverbindungen verwendet, um Informationen von einer Stufe zur nächsten weiterzuleiten.
- **RNN**: Gestapeltes mehrschichtiges Langzeit-Kurzzeitgedächtnis (LSTM). LSTM ist ein RNN-Typ, der Sequenz-zu-Sequenz-Übertragungen durchführen kann.



KI in Übersetzung und Dolmetschen

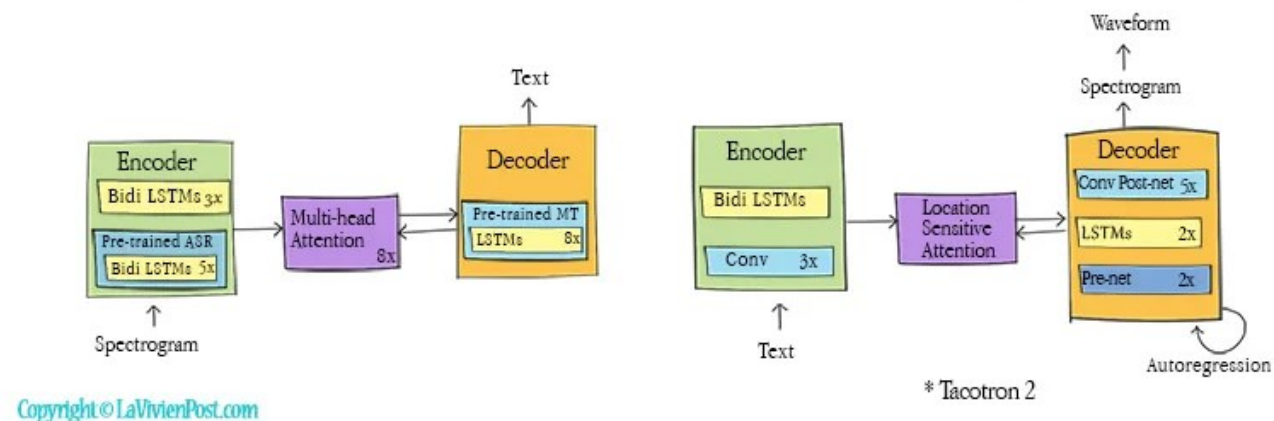
- **Encoder** : Wandelt Eingabewörter in Vektoren um
- **Decoder** : wandelt Vektoren in Wörter in einer anderen Sprache um
- In diesem Modell verwenden sowohl der Encoder als auch der Decoder **rekurrierende neuronale Netzwerke (RNNs)** .
- RNN ist ein neuronales Netzwerk, das Rückkopplungsverbindungen verwendet, um Informationen von einer Stufe zur nächsten weiterzuleiten.
- **RNN**: Gestapeltes mehrschichtiges Langzeit-Kurzzeitgedächtnis (LSTM). LSTM ist ein RNN-Typ, der Sequenz-zu-Sequenz-Übertragungen durchführen kann.



Speech-to-Speech Translation (S2ST) Cascade System

Speech-to-Text Translation (ST)

Text-to-Speech Synthesis (TTS)

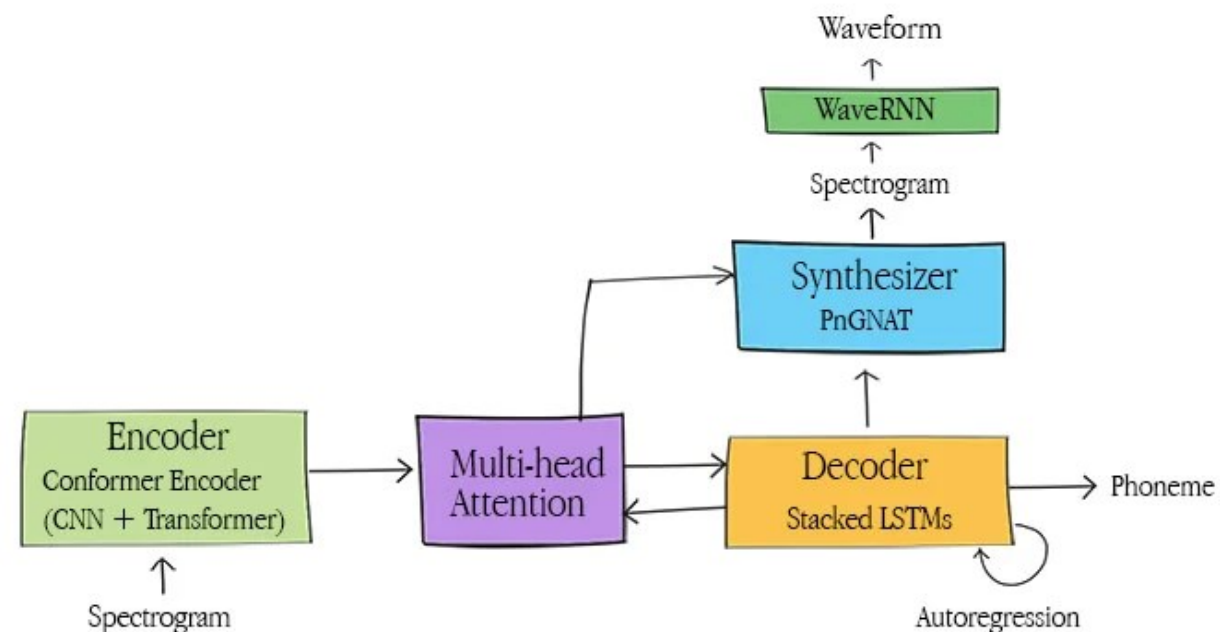


KI in Interpretation

- **Encoder** : Wandelt Eingabewörter in Vektoren um
- **Decoder** : wandelt Vektoren in Wörter in einer anderen Sprache um
- In diesem Modell verwenden sowohl der Encoder als auch der Decoder **rekurrierende neuronale Netzwerke (RNNs)** .
- RNN ist ein neuronales Netzwerk, das Rückkopplungsverbindungen verwendet, um Informationen von einer Stufe zur nächsten weiterzuleiten.
- **RNN**: Gestapeltes mehrschichtiges Langzeit-Kurzzeitgedächtnis (LSTM). LSTM ist ein RNN-Typ, der Sequenz-zu-Sequenz-Übertragungen durchführen kann.



Speech-to-Speech Translation (Translatotron 2)



Copyright © LaVivienPost.com

WORLD
ECONOMIC
FORUM

Open AI CEO Sam Altman
(Photo: Fabrice Coffrini / AFP)





OpenAI

OpenAI

Die Struktur wurde am 11. Dezember 2015 als gemeinnütziger Verein anerkannt.

Das Opfer

☰ **TIME** **SUBSCRIBE**

Stay on Top of Cybersecurity...
sponsored by: IT Security Solutions

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

15 MINUTE READ



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's

✕ venturebeat.com 🔖 📄 ⋮

VentureBeat 🔍 ☰

Sam Altman wants up to \$7 trillion for AI chips. The natural resources required would be 'mind boggling'

Sharon Goldman

@sharongoldman

February 9, 2024 11:20 AM

f X in



Politische Voreingenommenheit in der KI

- Verzerrungen und Vorurteile im Zusammenhang mit politischen oder ideologischen Zielen
- Berücksichtigung politischer Agenden bei der Gestaltung und Nutzung von KI
- Auswirkungen auf Entscheidungen, öffentliche Politik und demokratische Prozesse

Accueil > News > Facebook

Christopher Wylie : l'homme qui peut faire exploser en vol Palantir et Cambridge Analytica (et Facebook ?)

Technologie : Ce n'est pour l'heure qu'une piste. Mais le lien entre les partisans de Donald Trump et l'entreprise Cambridge Analytica, spécialisée dans le Big Data et l'analyse comportementale, vient d'être établi par un lanceur d'alerte britannique.



Par Guillaume Series | Publié le mercredi 28 mars 2018 à 17:35 - Modifié le jeudi 29 mars 2018 à 13:05

Reactions 2 Post Partager plus

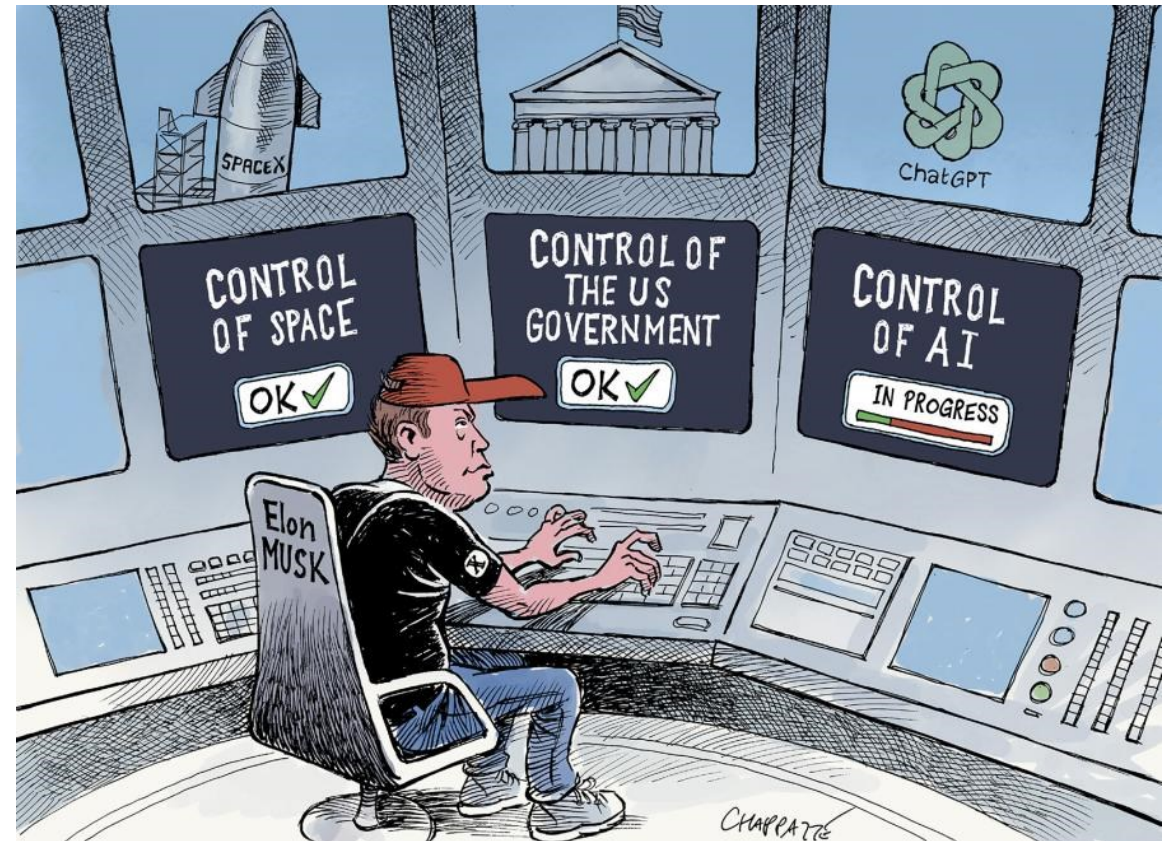
Le moins que l'on puisse dire, c'est que quand Chris Wylie s'exprime, ça dénote ! Le lanceur d'alerte aux cheveux roses a fait une série de déclarations explosives devant les parlementaires britanniques dans le cadre du dossier Cambridge Analytica [rapporté The Register](#). Sa motivation ? Il justifie ses révélations par le fait qu'il ne pense pas que les opérations de propagande "soient propices au processus démocratique, qu'il s'agisse d'une présidentielle américaine ou d'une élection locale".



Cambridge Analytica

Politische Voreingenommenheit in der KI

- Verzerrungen und Vorurteile im Zusammenhang mit politischen oder ideologischen Zielen
- Berücksichtigung politischer Agenden bei der Gestaltung und Nutzung von KI
- Auswirkungen auf Entscheidungen, öffentliche Politik und demokratische Prozesse

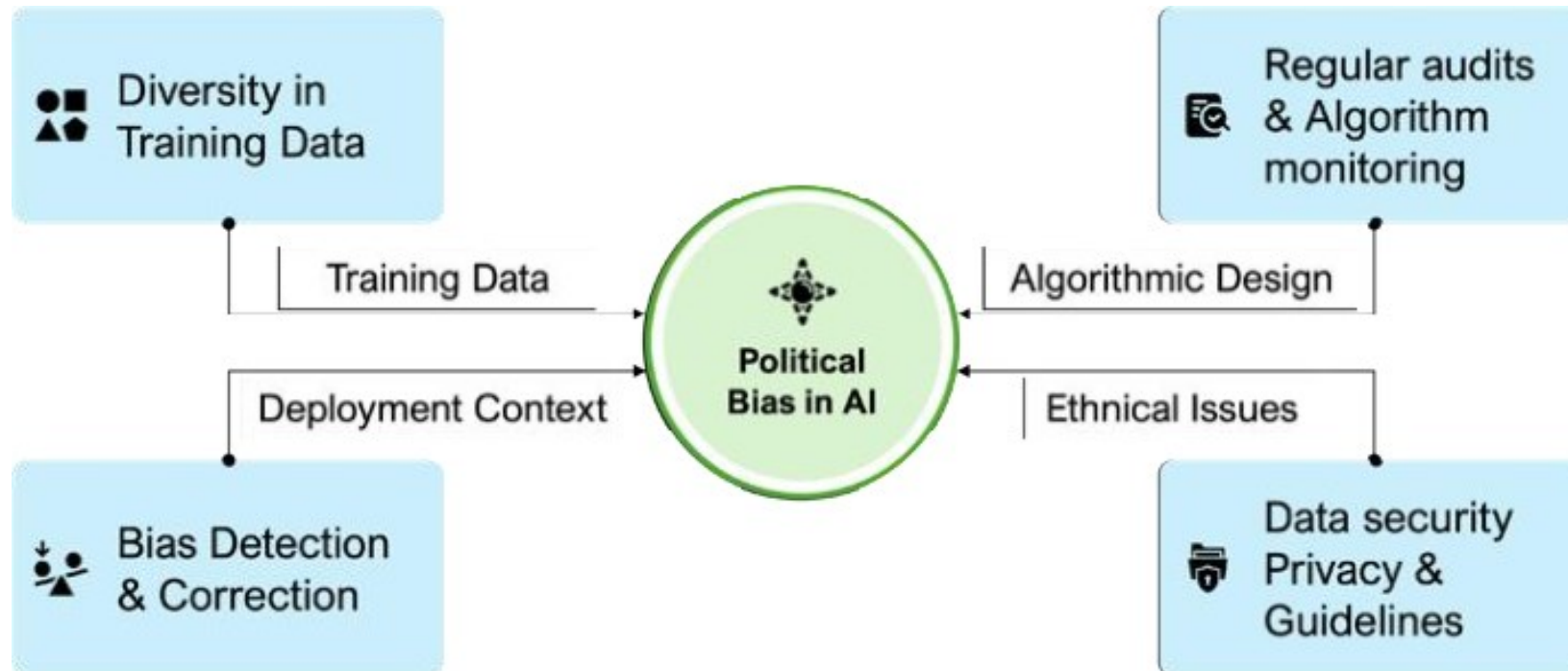


Politische Voreingenommenheit in der KI



- Assistent und nicht Protagonist
- Kritisches Denken
- Schulungsbedarf
- Betriebswirtschaftliche Wissensbedürfnisse
- Wahrscheinlichkeit
- Bedarf an Qualitätsdaten in großen Mengen
- Erklärbarkeit
- Externe Dienste
- Algorithmus-Transparenz

Politische Voreingenommenheit in der KI



https://www.researchgate.net/figure/illustration-of-the-Mitigation-Strategies-for-Political-Bias-in-AI_fig2_382322175

Vorschriften und Normen


- Notwendigkeit ethischer Leitlinien für den Einsatz von KI
- Fordert Transparenz bei Entscheidungsprozessen
- Kontinuierliche Überwachung und Bewertung der Auswirkungen auf die betroffenen Gemeinden

<https://www.wedemain.fr/dechiffrer/lia-une-machine-a-biais-comment-lintelligence-artificielle-renforce-les-stereotypes/>



Rechtliche Aspekte der KI

Principales mesures contenues dans les textes de régulation de l'IA

						
		CdE / CAHAI	CE / GEHN IA	Règlement CE	OCDE	UNESCO
		CONTRAIGNANT ?	NON-CONTRAIGNANT	CONTRAIGNANT	NON-CONTRAIGNANT	NON-CONTRAIGNANT
Droits de l'homme	5					
Responsabilité	5					
Transparence	5					
Équité, non-discrimination	4					
Intervention et supervision humaines	4					
Sécurité et robustesse	4					
Dignité humaine	3					
Diversité	3					
Protection des données et de la vie privée	3					
Bien-être	2					
Données et gouvernance des données	2					
Démocratie	1					
Documentation technique	1					
Durabilité	1					
Environnement	1					
État de droit	1					
Gestion des risques	1					
Gouvernance et collaboration multipartites et adaptatives	1					
Proportionnalité et innocuité	1					
Sensibilisation et éducation	1					
Sociétés pacifiques, justes et interconnectées	1					
Mesures juridiques fondées sur les droits de l'homme	6	4	2	3	6	
Mesures politiques et sociétales	1	2	0	1	8	
Mesures techniques	2	4	6	2	3	

https://lestempselectriques.net/ANALYSE_IA.pdf

Erklärbarkeit implementieren

Les principes éthiques de l'IA



Transparence



Responsabilité



Confidentialité
& Sécurité



Atténuation
des biais



Gain
d'autonomie



Contrôle
humain

Vorschriften und Normen

- Forderungen nach einer ethischen Steuerung der KI im politischen Kontext
- Notwendigkeit klarer Regelungen zum Einsatz von KI im Wahlkampf
- Transparenz- und Rechenschaftspflichten für Organisationen und Regierungen, die KI nutzen



<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

Vorschriften und Normen

- Forderungen nach einer ethischen Steuerung der KI im politischen Kontext
- Notwendigkeit klarer Regelungen zum Einsatz von KI im Wahlkampf
- Transparenz- und Rechenschaftspflichten für Organisationen und Regierungen, die KI nutzen



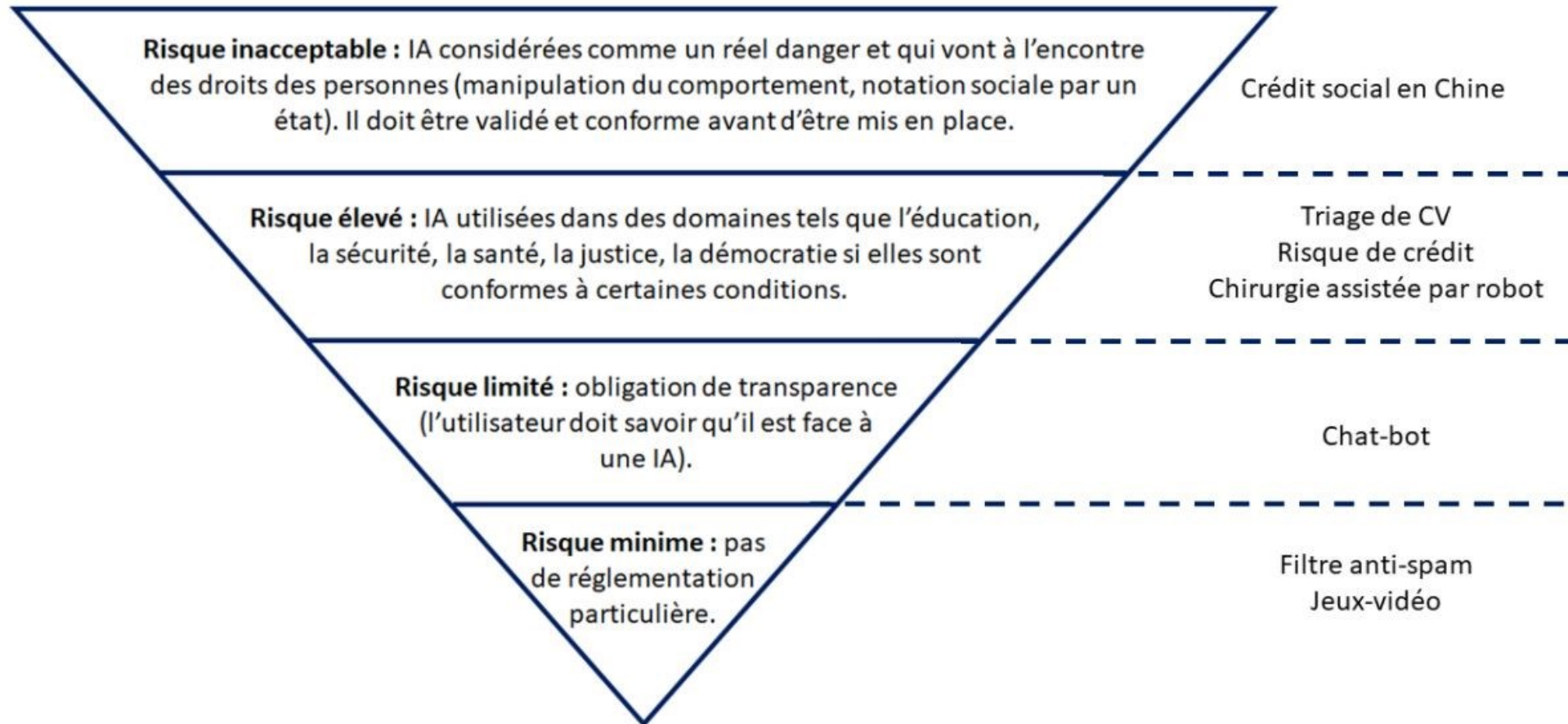
Rechtliche Aspekte der KI: EU-KI-Gesetz



Europäischer KI -Gesetzesakt

<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Soziolinguistische Vorurteile



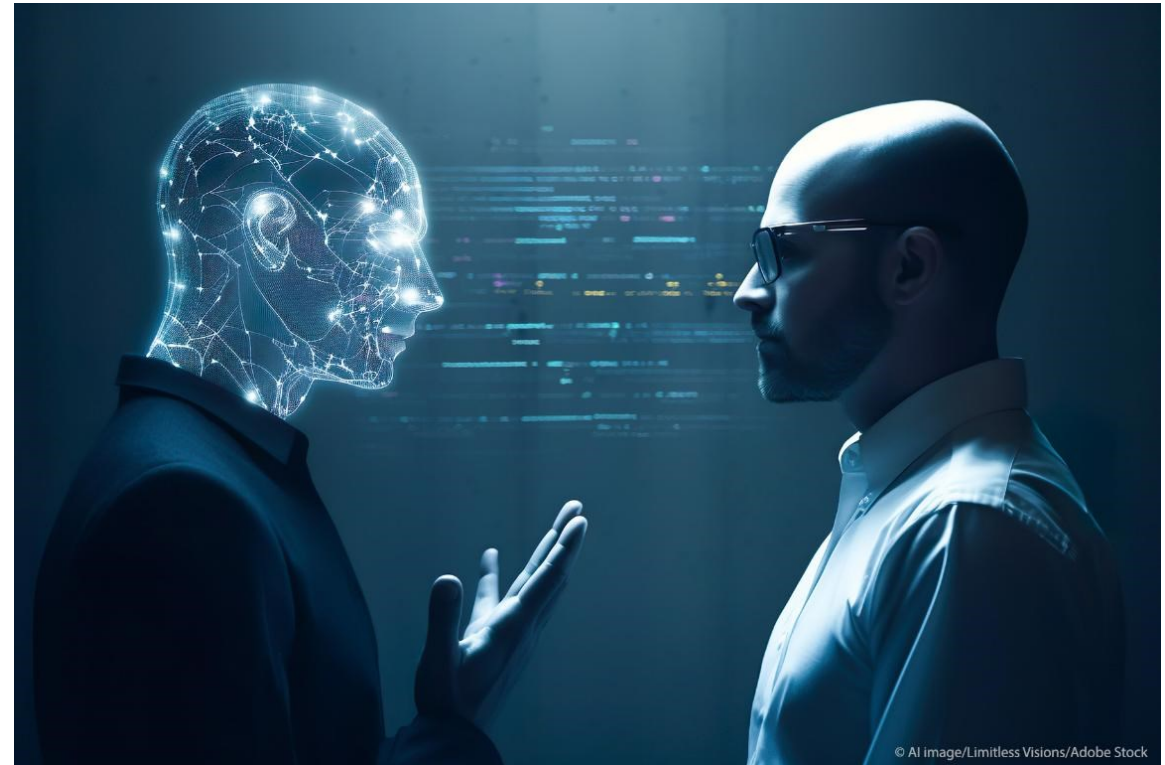
EU-KI-Gesetz: Erste Regulierung der künstlichen Intelligenz

<https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-law-first-artificial-intelligence-regulation>

<https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/>

<https://artificialintelligenceact.eu/the-act/>

<https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf>



Norm ISO/IEC 42001:2023

Warum ist ISO/IEC 42001 wichtig?

ISO/IEC 42001 ist der weltweit erste Standard für KI-Managementsysteme und bietet wertvolle Leitlinien für dieses sich rasch entwickelnde Technologiefeld. Es befasst sich mit den besonderen Herausforderungen der KI, wie ethischen Überlegungen, Transparenz und kontinuierlichem Lernen. Für Unternehmen stellt es eine strukturierte Möglichkeit dar, die mit KI verbundenen Risiken und Chancen zu verwalten und dabei Innovation und Governance in Einklang zu bringen.



Vorteile




- Rahmen für das Risiko- und Chancenmanagement
- Verantwortungsvoller Einsatz von KI demonstrieren
- Rückverfolgbarkeit, Transparenz und Zuverlässigkeit
- Kosteneinsparungen und Effizienzsteigerungen



ISO/IEC 42001:2023
Informationstechnologie
Künstliche Intelligenz
Managementsystem

Altman

REUTERS®  

 My View  Following  Saved

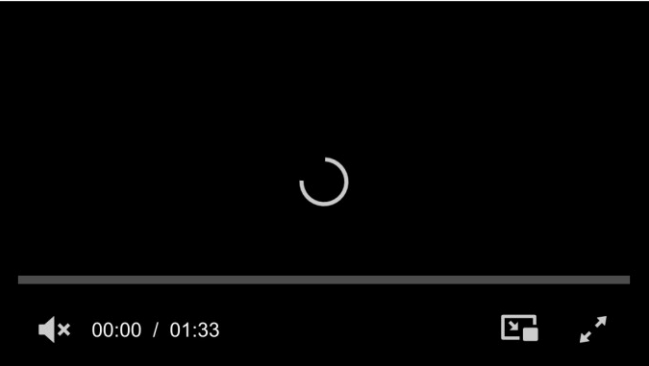
Boards, Policy & Regulation | Boards






Sam Altman to return as OpenAI CEO after his tumultuous ouster

By Jeffrey Dastin and Aditya Soni

November 22, 2023 2:25 PM GMT+1 · Updated 3 months ago



  www.theverge.com   

ARTIFICIAL INTELLIGENCE
Updated Feb 9, 2024, 8:41 AM GMT+1

Turmoil at OpenAI: Now that Sam Altman is back, what's next for the creator of ChatGPT?

By [Richard Lawler](#), a senior editor following news across tech, culture, policy, and entertainment. He joined The Verge in 2021 after several years covering news at Engadget.

    | 15 Comments (15 New)

It's official: Sam Altman is, once again, the CEO of OpenAI. How did the world's hottest generative AI company go through three CEOs in under a week and end up with the same one it had at the start? This is what we know so far.

TECH

OpenAI brings Sam Altman back as CEO less than a week after he was fired by board

PUBLISHED WED, NOV 22 2023 1:08 AM EST
UPDATED WED, NOV 22 2023 9:30 AM EST

[Ari Levy](#) | [Hayden Field](#)
[@LEVYNEWS](#) | [@HAYDENFIELD](#)

[Jonathan Vanian](#)
[@IN/JONATHAN-VANIAN-B704432/](#)

[Rohan Goswami](#) | [Jordan Novet](#)
[@ROGOSWAMI](#) | [@JORDANNOVET](#)
[@IN/ROHANGOSWAMICNBC/](#)

WATCH LIVE

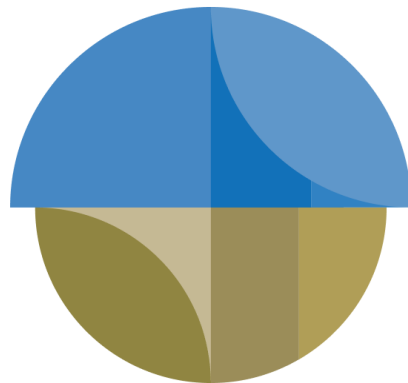
KEY POINTS

- Sam Altman is back as CEO of OpenAI after he was fired suddenly Friday.
- Following his ouster, hundreds of OpenAI employees, including co-founder and board member Ilya Sutskever, signed a letter demanding that remaining board members resign or they would leave.
- Microsoft CEO Satya Nadella said Monday that Altman and others from OpenAI would be joining to start a "new advanced AI research team."

KI und die Zukunft des Arbeitsmarktes



97 million



85 million

Growing job demand:

1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. Big Data Specialists
4. Digital Marketing and Strategy Specialists
5. Process Automation Specialists
6. Business Development Professionals
7. Digital Transformation Specialists
8. Information Security Analysts
9. Software and Applications Developers
10. Internet of Things Specialists

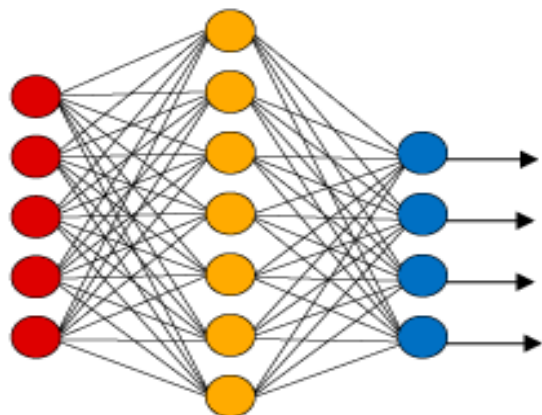
Decreasing job demand:

1. Data Entry Clerks
2. Administrative and Executive Secretaries
3. Accounting, Bookkeeping and Payroll Clerks
4. Accountants and Auditors
5. Assembly and Factory Workers
6. Business Services and Administration Managers
7. Client Information and Customer Service Workers
8. General and Operations Managers
9. Mechanics and Machinery Repairers
10. Material-Recording and Stock-Keeping Clerks

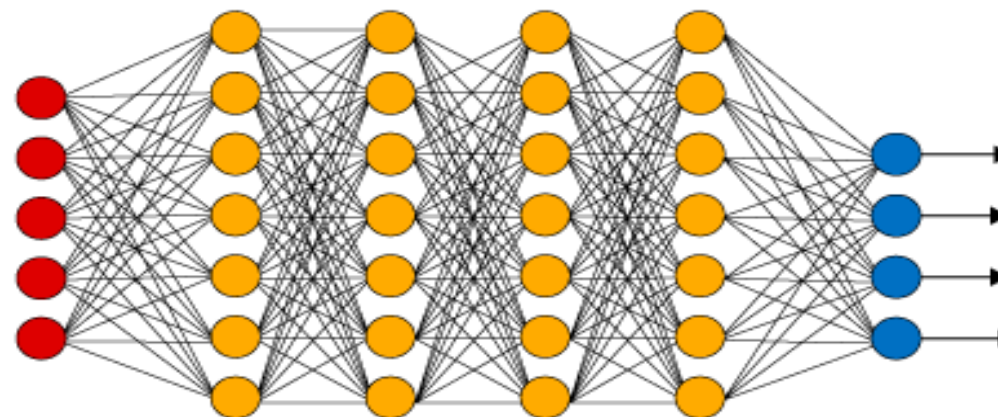
Source: Future of Jobs Report 2020, World Economic Forum.

Erklärbarkeit von KI

Simple Neural Network



Deep Learning Neural Network

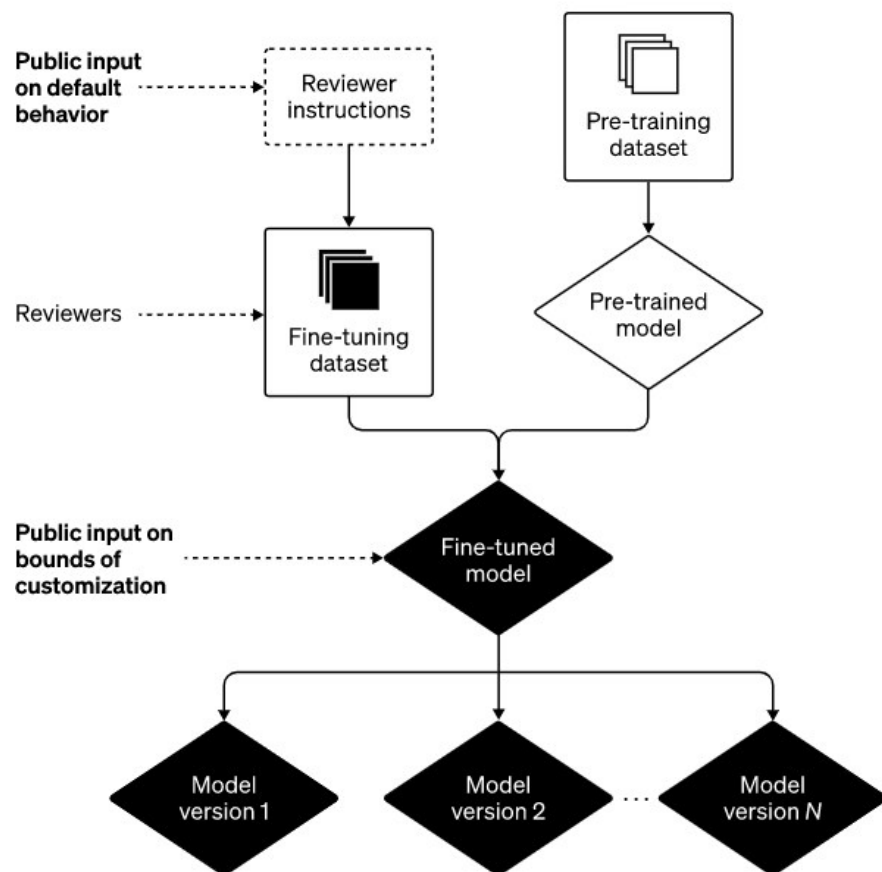


● Input Layer ● Hidden Layer ● Output Layer

Source: <https://towardsdatascience.com/mnist-vs-mnist-how-i-was-able-to-speed-up-my-deep-learning-11c0787e6935>

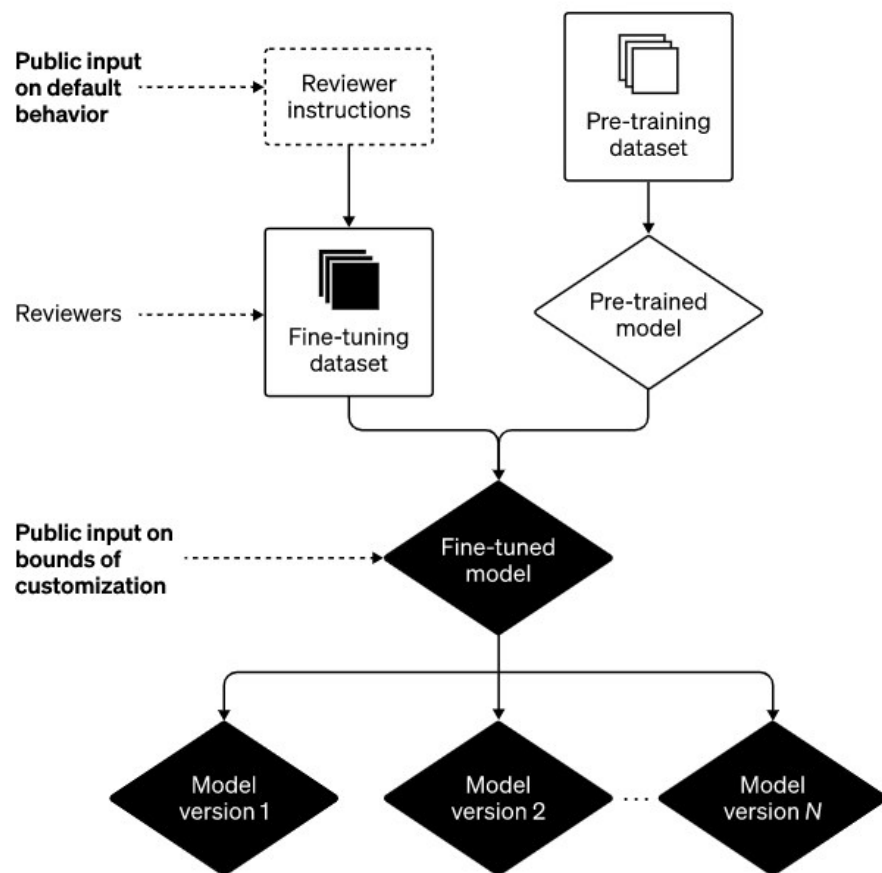
Sehen auch : <https://www.youtube.com/watch?v=aircAruvnKk>

Erklärbarkeit von KI



- **Vertrauen aufbauen:** Benutzer müssen verstehen, warum Entscheidungen von KI-Systemen getroffen werden, um ihnen vertrauen zu können.
- **Verantwortlichkeit:** Identifizieren Sie Fehler oder Verzerrungen in Entscheidungen von KI-Modellen und ermöglichen Sie es den Designern, dafür Verantwortung zu übernehmen.
- **Einhaltung gesetzlicher Vorschriften:** Viele Vorschriften verlangen mittlerweile, dass KI-Systeme erklärbar sind, um Transparenz und Konformität zu gewährleisten.

Erklärbarkeit von KI



Vollständige Dokumentation:

- Stellen Sie eine detaillierte Dokumentation zum Entwurf, Training und Einsatz des KI-Modells bereit.
- Fügen Sie klare Beschreibungen der verwendeten Daten, Trainingsmethoden, Leistungskennzahlen usw. hinzu.

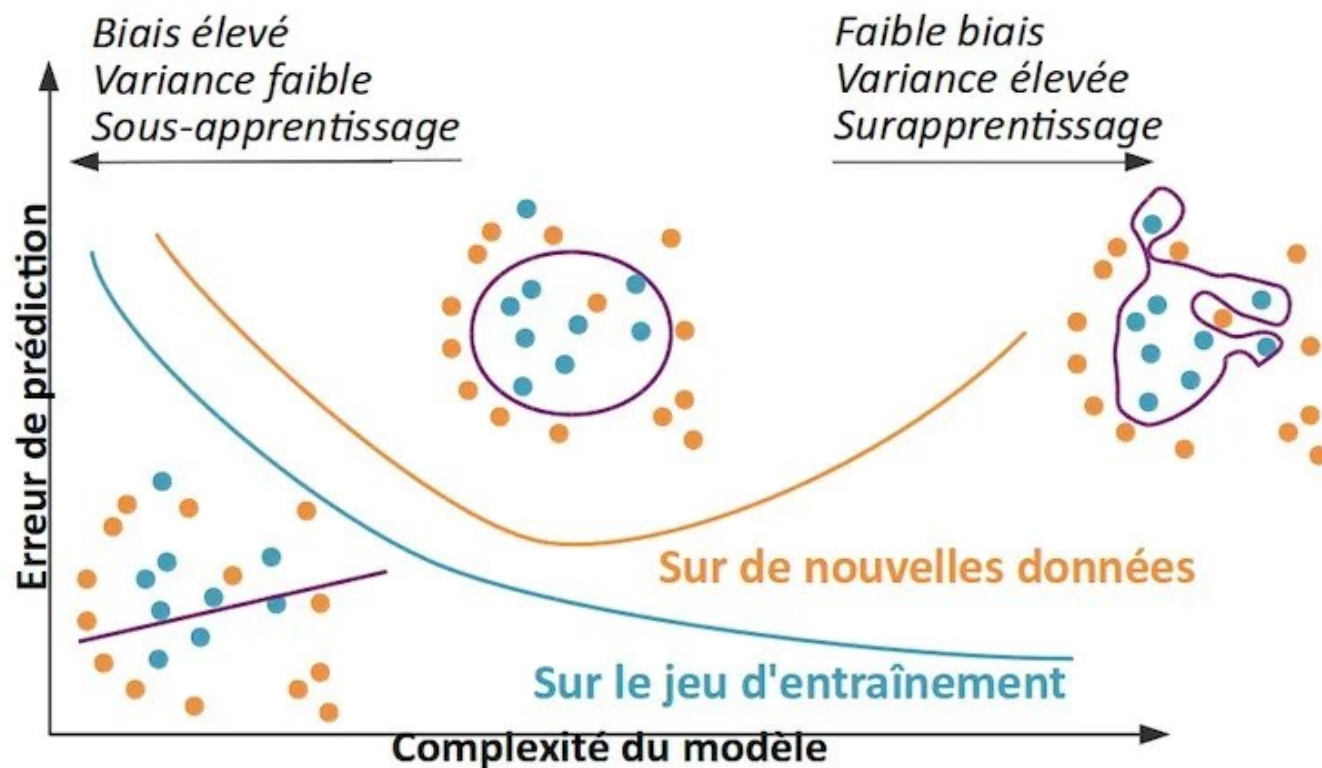
Interpretierbarkeitmethoden:

- Verwenden Sie Techniken wie Feature-Mapping, *um zu zeigen*, welche Features Vorhersagen beeinflussen.
- Geben Sie Erklärungen dazu ab, wie das Modell zu einer bestimmten Entscheidung gekommen ist.

Intuitive Visualisierung:

- Verwenden Sie Visualisierungstools, um Modellergebnisse und Vorhersagen grafisch darzustellen.
- Dadurch können Benutzer visuell nachvollziehen, wie das KI-Modell funktioniert.

Herausforderungen der Erklärbarkeit von KI



<https://blog.enioka.com/2022/03/14/churn-modelling-preventing-customer-loss-with-de-lia-2-3/>

- **Modellkomplexität:** Einige KI-Modelle, insbesondere tiefe neuronale Netzwerke, können aufgrund ihrer Komplexität schwer zu erklären sein.
- **Kompromiss bei der Leistung:** Manchmal kann die Leistung eines Modells sinken, wenn es ausführlicher gestaltet wird.
- **Schutz vertraulicher Informationen:** Durch die Erläuterung des Modells können manchmal in den Daten enthaltene vertrauliche Informationen offengelegt werden.

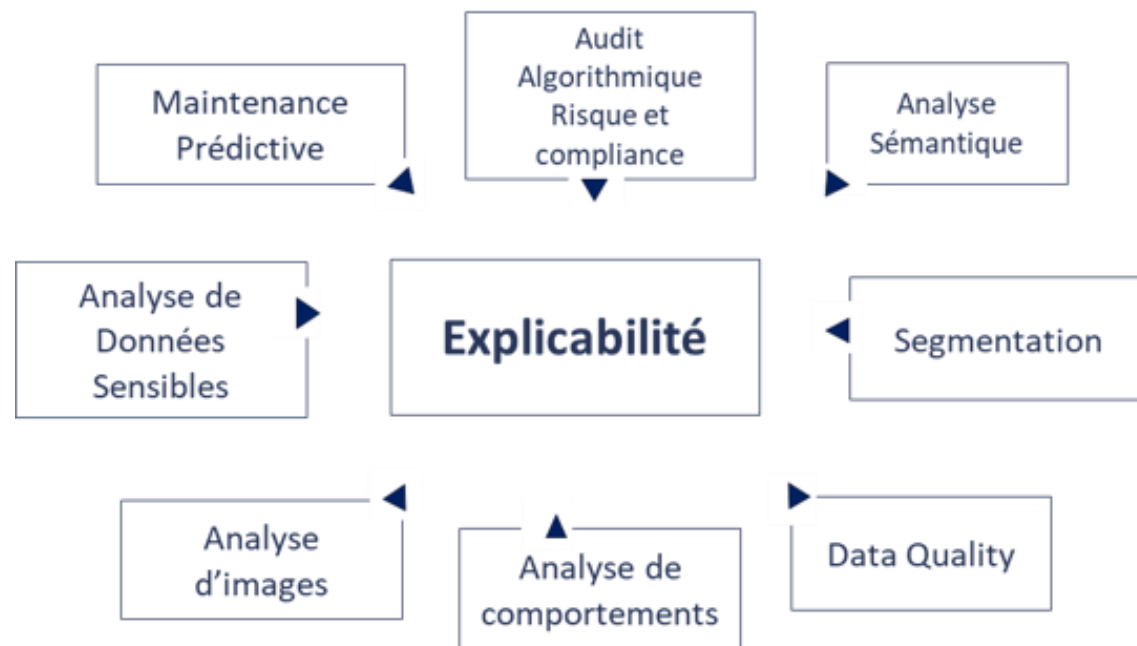
Beispiele für die Implementierung der Erklärbarkeit

- **Medizin:** Erläuterung medizinischer Diagnosen, damit Ärzte die Gründe hinter den Empfehlungen des Modells verstehen.
- **Finanzen:** Erläuterung von Finanzprognosen, um Analysten dabei zu helfen, die Faktoren zu verstehen, die Vorhersagen beeinflussen.
- **Sicherheit:** Erläuterung der von Angriffserkennungssystemen getroffenen Entscheidungen, um Analysten dabei zu helfen, potenzielle Bedrohungen zu verstehen.



Erklärbarkeit

- Kontinuierliche Integration der Erklärbarkeit in den KI-Entwicklungslebenszyklus.
- Entwicklung von Standards und Best Practices zur Erklärbarkeit von KI.
- Forschung zu fortgeschritteneren Modellerklärungstechniken für komplexe Modelle.



Ethische Hacker

„Ethische Hacker“ für Künstliche Intelligenz (KI) spielen eine entscheidende Rolle bei der Identifizierung von Schwachstellen und potenziellen Risiken von KI-Systemen.

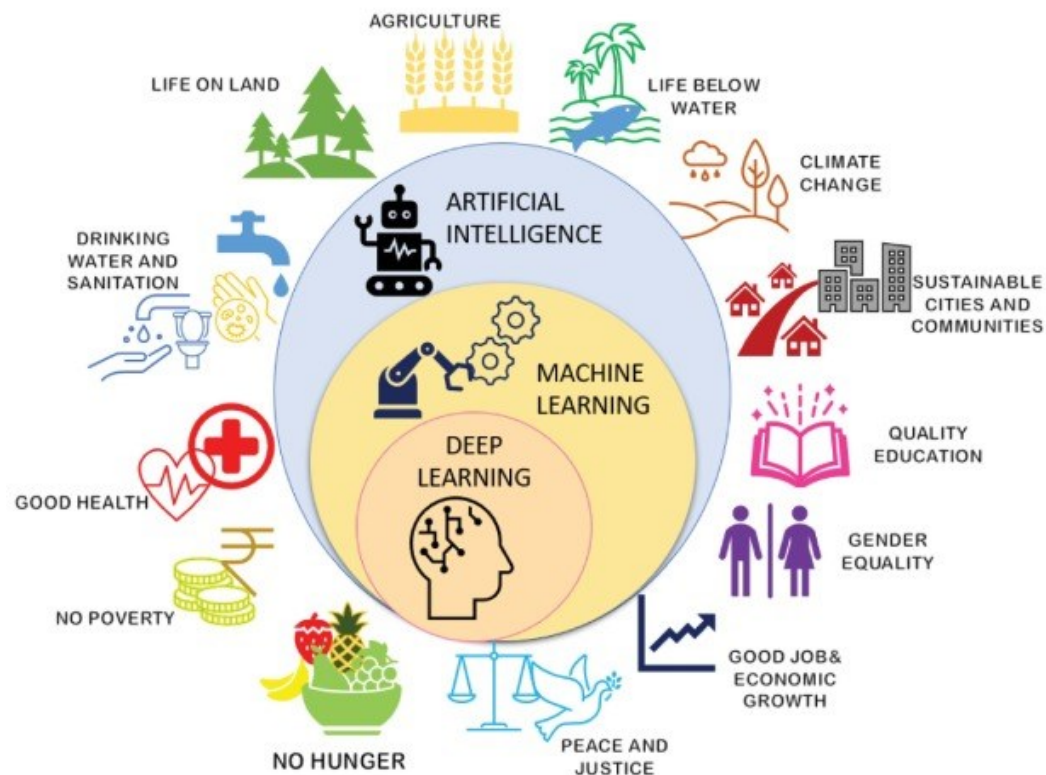


KI und Nachhaltigkeit



Computer verbrauchen Energie,
viel Energie ...

Ökologische Auswirkungen von KI-Software

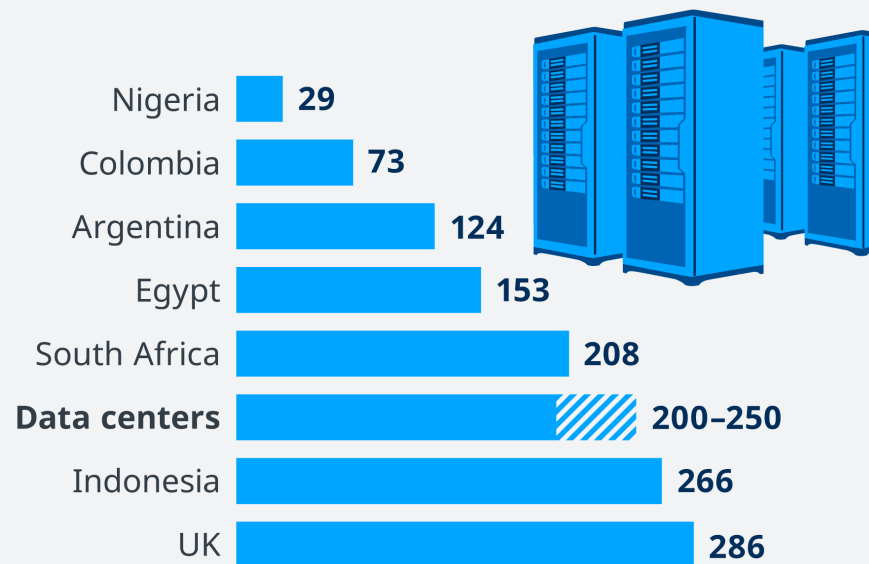


Computer verbrauchen Energie,
viel Energie ...

KI und Nachhaltigkeit

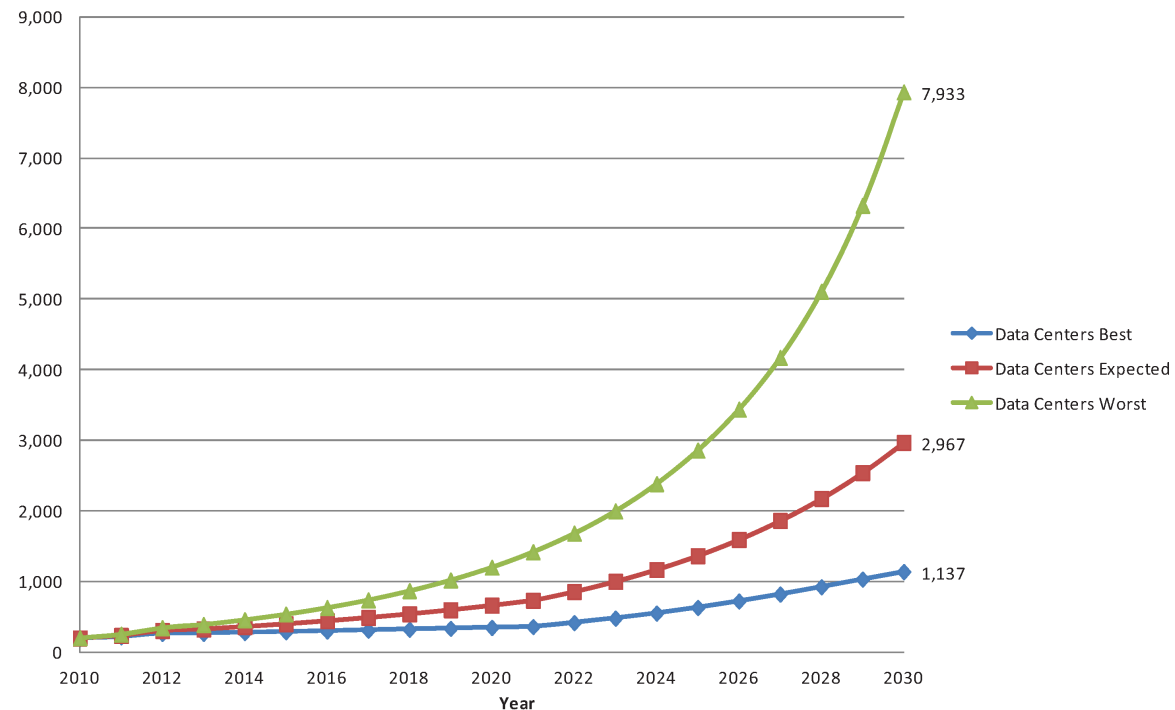
Data centers use more electricity than entire countries

Domestic electricity consumption of selected countries vs. data centers in 2020 in TWh



Source: Enerdata, IEA

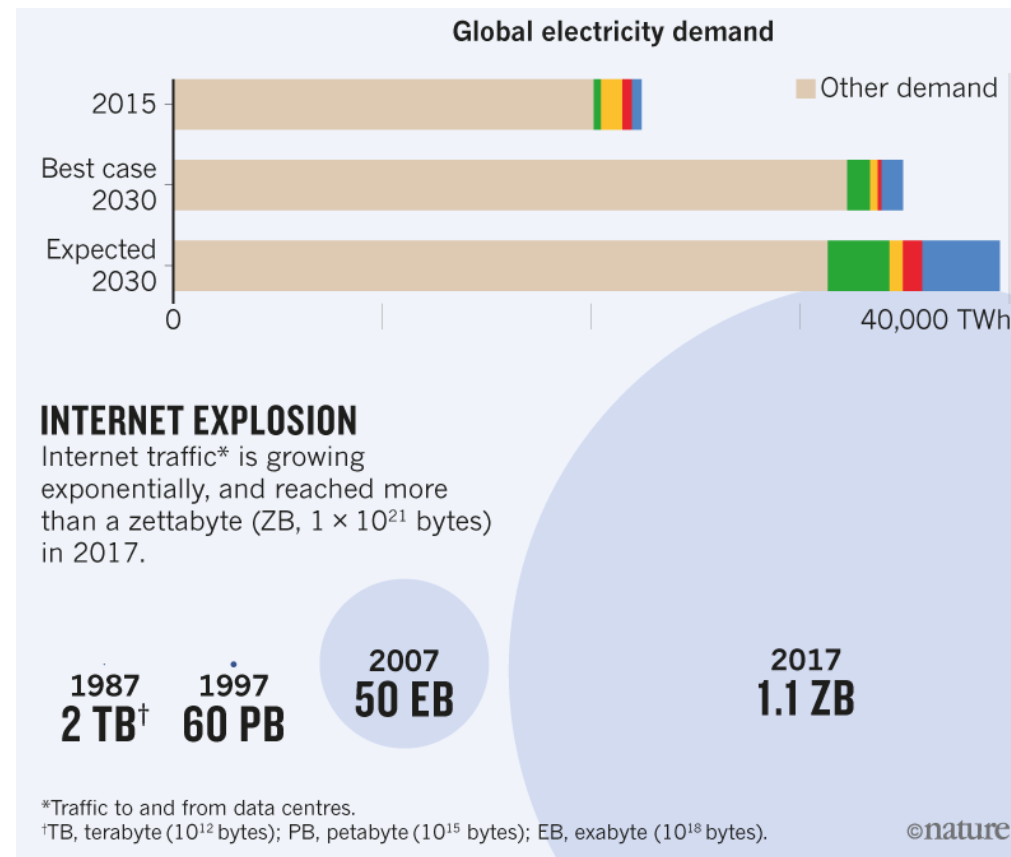
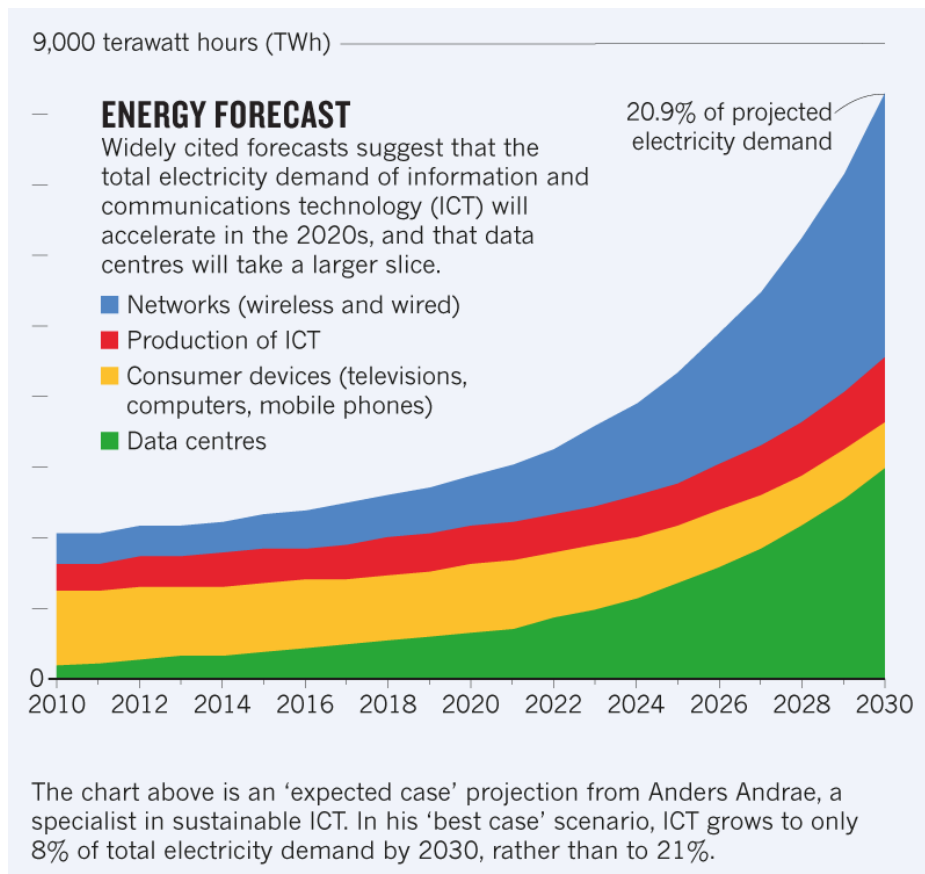
Electricity usage (TWh) of Data Centers 2010-2030



Zum weltweiten Stromverbrauch der Kommunikationstechnologie: Trends bis 2030

von Anders SG Andrae, ScilitPreprints.org *ORCID und Tomas Edler

KI und Nachhaltigkeit



KI und Nachhaltigkeit

Durch die Globalisierung entstehen auch globale Märkte für die Entsorgung von Elektroschrott:

Das Beispiel **Agbogbloshie Ghana**

Elektroschrott-Deponie in Ghana : Umweltschützer besorgt über die Entsorgung von Elektroschrott in der Hauptstadt Accra, CCTV Africa,
<https://www.youtube.com/watch?v=WzR79Wv2oJA>

Toxicity : Leben in Agbobloshie , der weltgrößten Elektroschrott-Deponie in Ghana, RT-Dokumentation,
<https://www.youtube.com/watch?v=mleQVO1Vd1I>



Foto von Andrew McConnell,
<https://www.andrewmcconnell.com/Biography/1>

- Allgemeines für Laien
 - Ausbildung
 - Ethik
-
- Möglichkeit zum Stapeln von Credits --> Personalisiertes CAS

<https://cui.unige.ch/fr/pin/clinique-de-l-ia/>

