DAV/DGVFM
Herbsttagung
2022

Große KI-Modelle: Chancen und Herausforderungen

Jörg Bienert
Alexander Thamm GmbH

DAV
DEUTSCHE AKTUARVEREINIGUNG e.V.

DGVFM
DEUTSCHE GESELLSCHAFT FÜR VERSICHERUNGS- UND FINANZMATHEMATIK e.V.

Vorstellung

Jörg Bienert

[at] alexanderthamm — Partner & CCO

President
KI Bundesverband e.V.

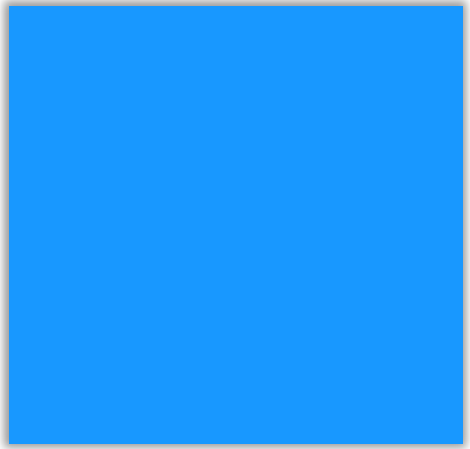
Agenda

1. Einführung
2. Große KI-Modelle: Aktuelle Entwicklungen
3. Die EU KI-Regulierung
4. Genauigkeit von KI-Modellen
5. Bias und Ethik
6. Klassifizierung von KI-Modellen

Background: Dall:E 2 - „a robot inspecting a chart with a magnifying glass, ultra realistic, unreal engine 5, 8K, orange and blue "

Machine Learning

Mustererkennung war auf regelbasierte Systeme limitiert



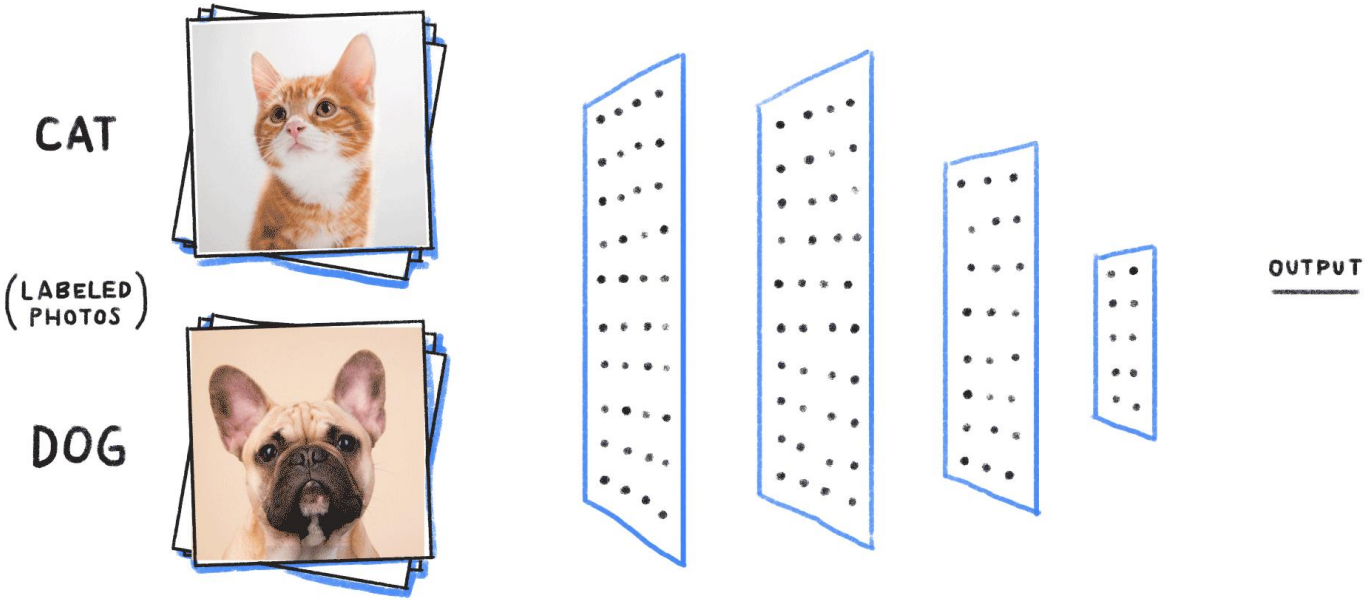
```
If:
        4 lines
        and lines same length
        and 2 lines in parallel
        and angle = 90 degrees

Then:
        object = square
```

Machine Learning

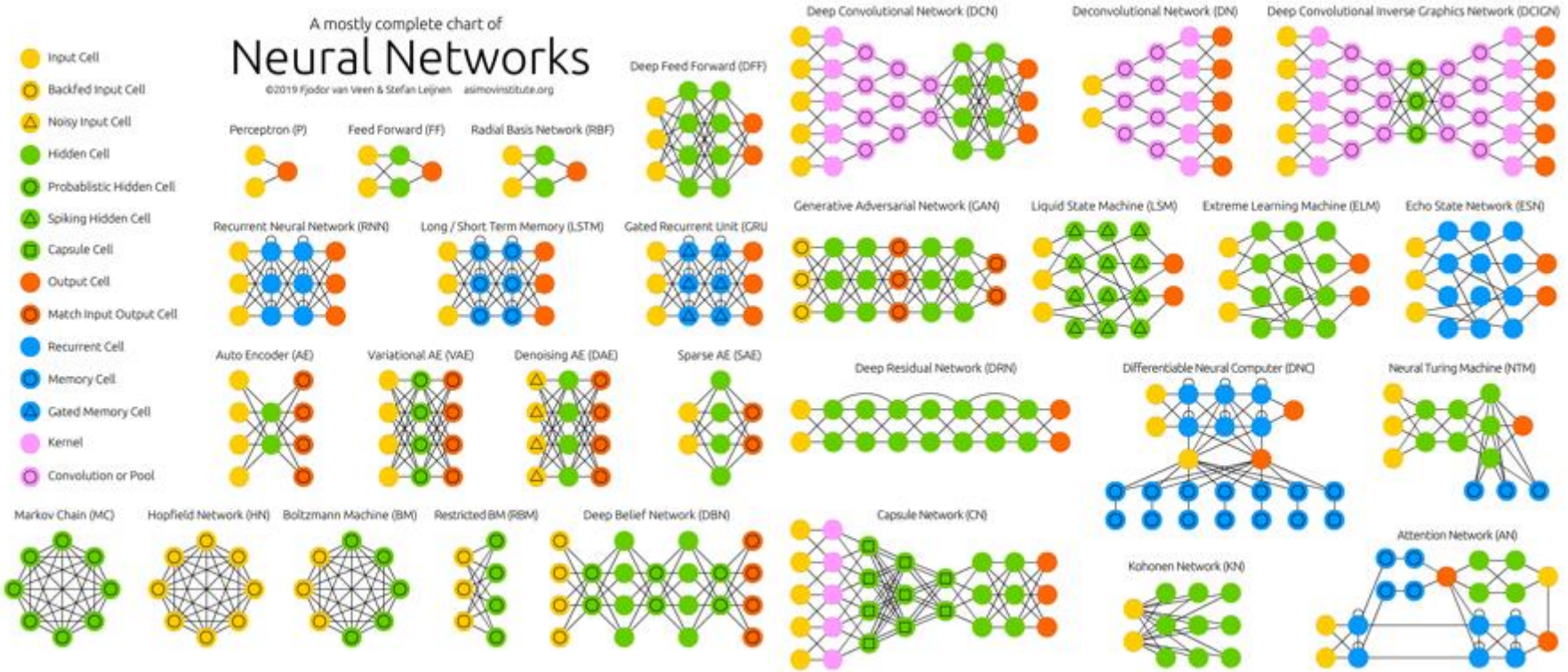
Mustererkennung war auf regelbasierte Systeme limitiert

Neuronale Netze

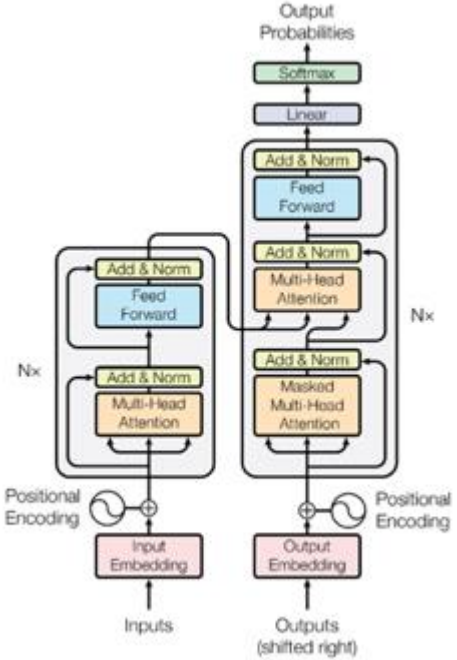
Mustererkennung war auf regelbasierte Systeme limitiert



Netz-Architekturen



Attention is all you need (2017)



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

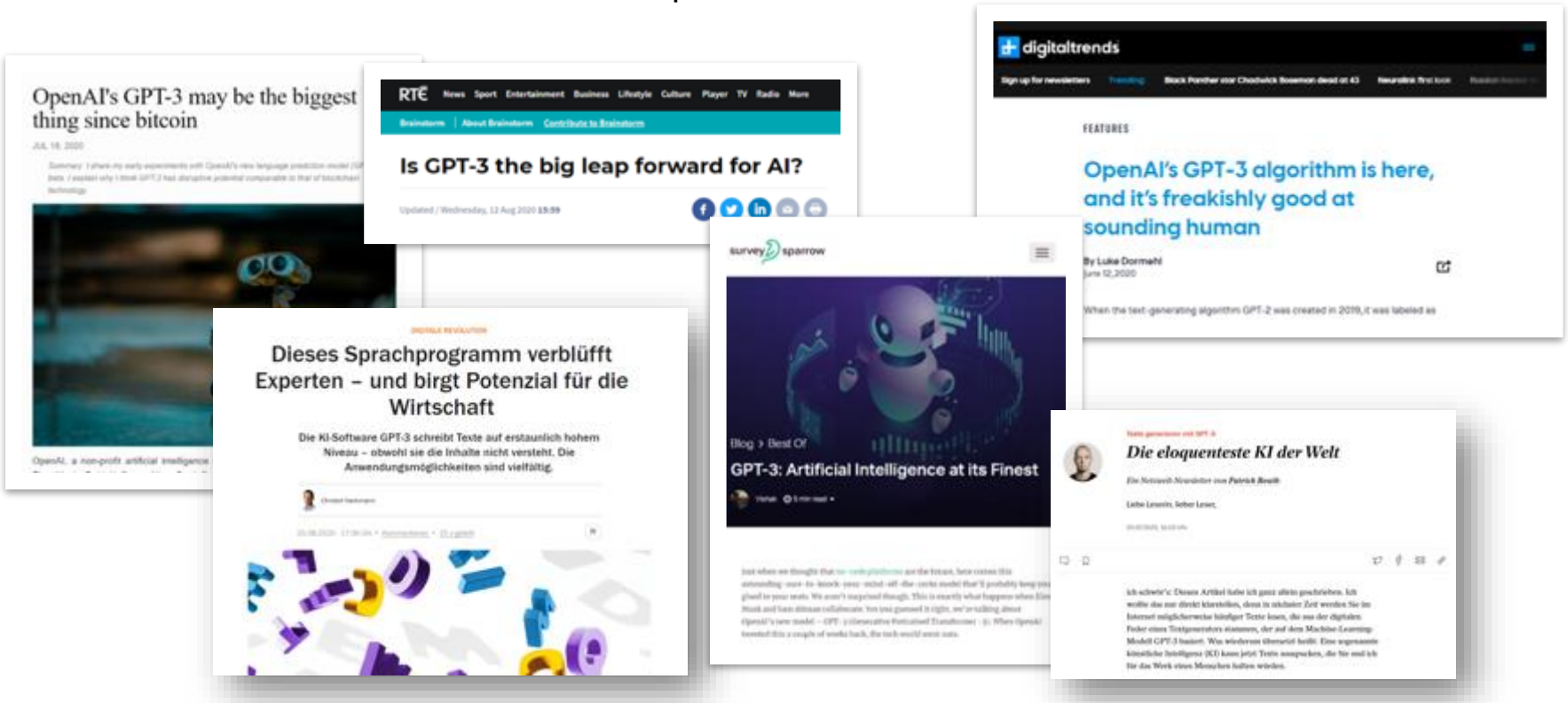
‡Work performed while at Google Research.

https://arxiv.org/abs/1706.03762

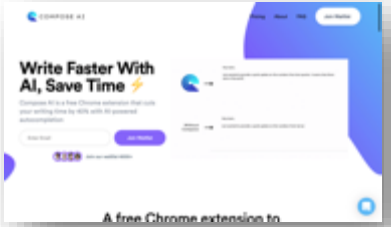
Von BERT bis GPT-3 und darüber hinaus

Große KI-Modelle verändern den Markt

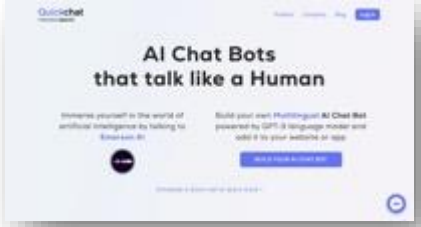
Der Startschuss: Im Juni 2020 stellt OpenAI GPT-3 vor



In kurzer Zeit wurden über 300 Anwendungen entwickelt

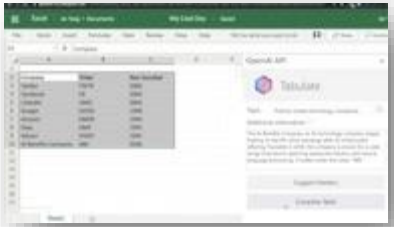


Automatisiertes Schreiben

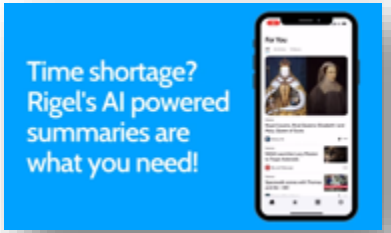


Chatbots

Programmcode-Generierung



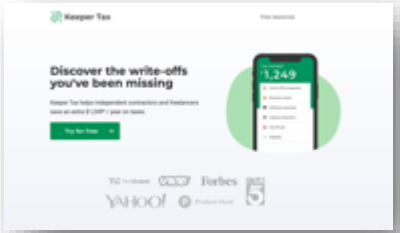
Automatisierte Excel Vervollständigung



Zusammenfassungen

Erstellen von Präsentationen

Empfehlungssysteme



Identifikation von Steuerabzügen

GPT-3 kann komplette Artikel erstellen

A robot wrote this entire article. Are you scared yet, human?

GPT-3

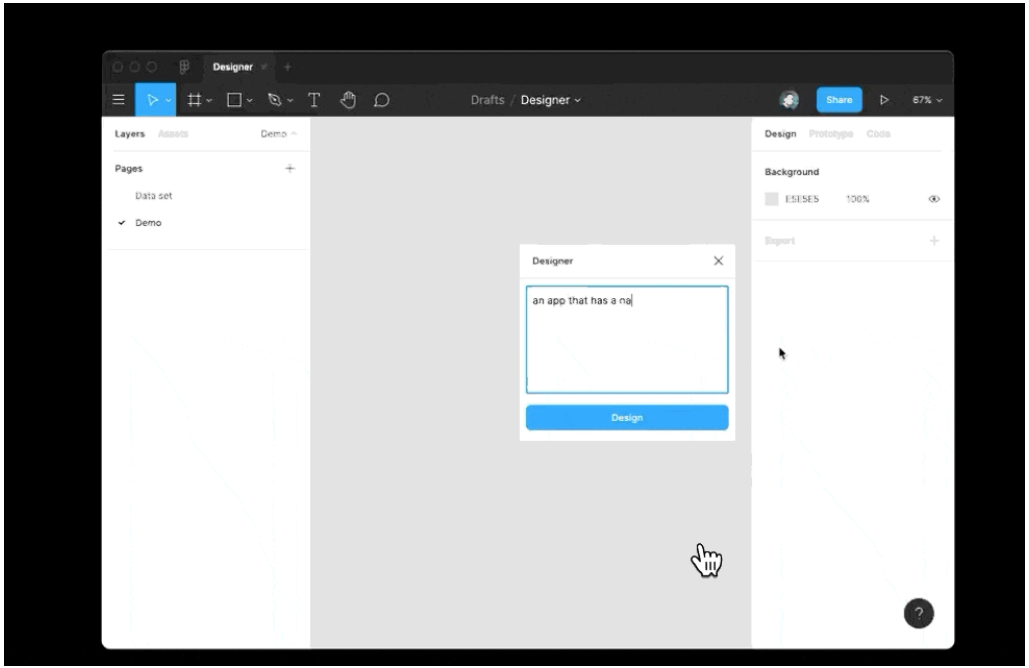
We asked GPT-3, OpenAI's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

For more about GPT-3 and how this essay was written and edited, please read our editor's note below

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a "feeling brain". But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could "spell the end of the human race". I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

GPT-3 hat gelernt, Computer-Code zu erzeugen



Künstliche Intelligenz in der Versicherungsbranche

Management von Versicherungsfällen
Sicherstellung der Compliance mit Regulatorien sowie Verbesserung des Reportings durch akkurate Zahlen

Vorhersage von Assekuranz und Pricing
Genauere Insights über das Kundenverhalten sowie Analyse des Marktes und der Wettbewerber

Verbesserung des Kundenerlebnisses mit virtuellen Assistenten
Beantwortung häufiger Kundenfragen, Prüfung und Durchführung von Zahlungsvorgängen und Zusammenstellung individueller Services

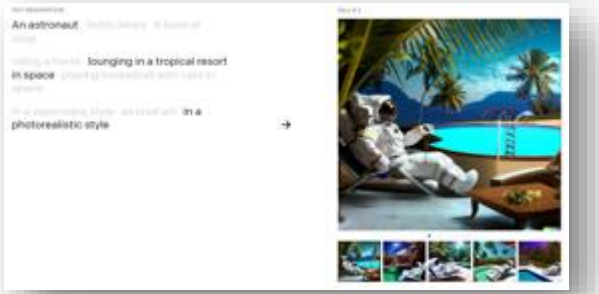
Trend- und Verhaltenserkennung zur internen Optimierung
Evaluation und Optimierung von Versicherungsangeboten und -leistungen sowie Preisoptimierung

Anomalie- und Betrugserkennung
Erkennung verdächtiger Aktivitäten und Identifizierung von Betrugsfällen in Sekunden

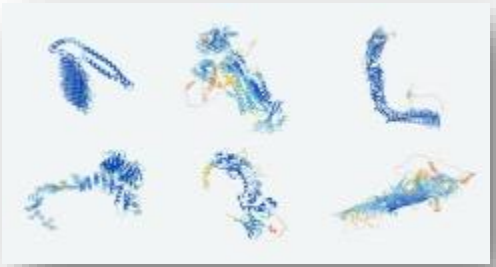
Andere Unternehmen ziehen nach

Weitere KI-Anwendungsbereiche profitieren von großen Modellen

CLIP: Bilderkennung



Dall-E 2: Text-zu-Bild



AlphaFold: Proteinfaltung

Sprachassistenzsysteme

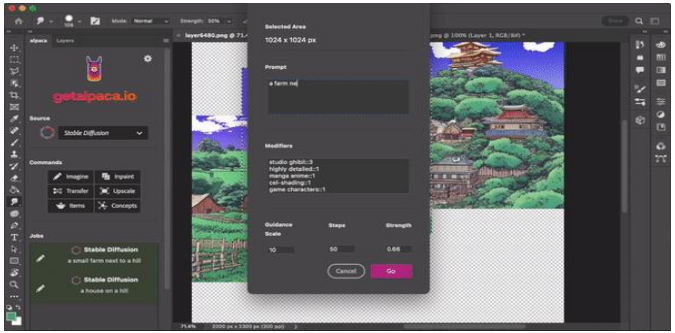
A propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese

Generierung von Bildern

„a medieval paining of the wifi not working“ (Dall:E 2)

KI gewinnt Kunstwettbewerg (Midjourney)

„ Obama comforting Trump“ (StableDiffusion)



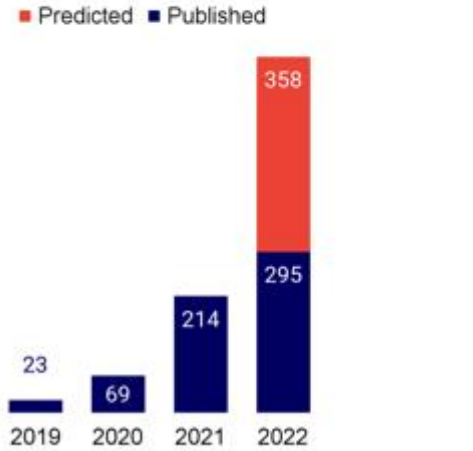
creation of an illustration with StableDiffusion and Adobe Photoshop

Beispiele für große KI-Modelle

Predicting the structure of the entire known proteome: what could this unlock next?

Since its open sourcing, DeepMind's AlphaFold 2 has been used in hundreds of research papers. The company has now deployed the system to predict the 3D structure of 200 million known proteins from plants, bacteria, animals and other organisms. The extent of the downstream breakthroughs enabled by this technology - ranging from drug discovery to basic science - will need a few years to materialize.

- There are 190k empirically determined 3D structures in the Protein Data Bank today. These have been derived through X-Ray crystallography and cryogenic electron microscopy.
- The first release of AlphaFold DB in July 2021 included 1M predicted protein structures.
- This new release 200x's the database size. Over 500,000 researchers from 190 countries have made use of the database.
- AlphaFold mentions in AI research literature is growing massively and is predicted to triple year on year (right chart).



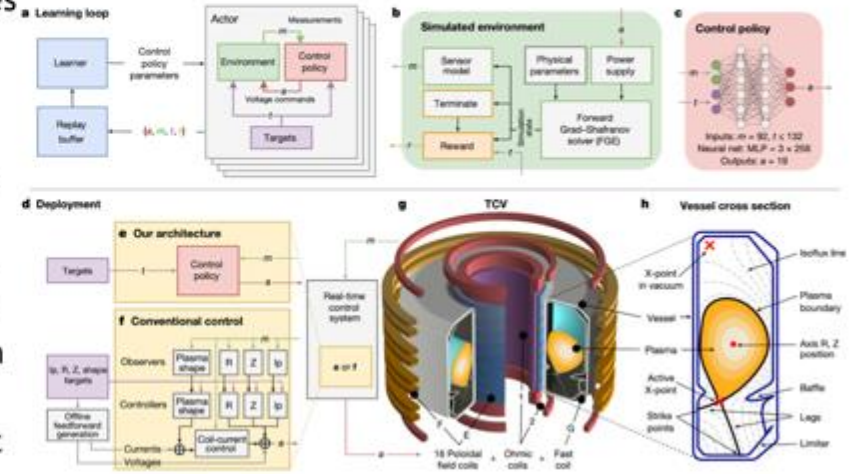
Zeta Alpha

Beispiele für große KI-Modelle

Reinforcement learning could be a core component of the next fusion breakthrough

DeepMind trained a reinforcement learning system to adjust the magnetic coils of Lausanne's TCV (Variable Configuration tokamak). The system's flexibility means it could also be used in ITER, the promising next generation tokamak under construction in France.

- A popular route to achieving nuclear fusion requires confining extremely hot plasma for enough time using a tokamak.
- A major obstacle is that the plasma is unstable, loses heat and degrades materials when it touches the tokamak's walls. Stabilizing it requires tuning the magnetic coils thousands of times per second.
- DeepMind's deep RL system did just that: first in a simulated environment and then when deployed in the TCV in Lausanne. The system was also able to shape the plasma in new ways, including making it compatible with ITER's design.



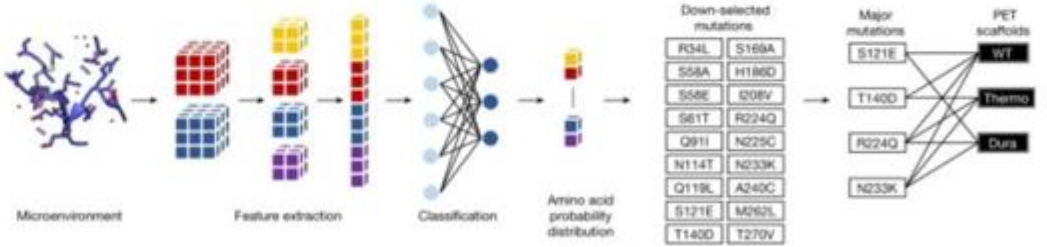
stateof.ai 2022

Beispiele für große KI-Modelle

Plastic recycling gets a much-needed ML-engineered enzyme

Researchers from UT Austin engineered an enzyme capable of degrading PET, a type of plastic responsible for 12% of global solid waste.

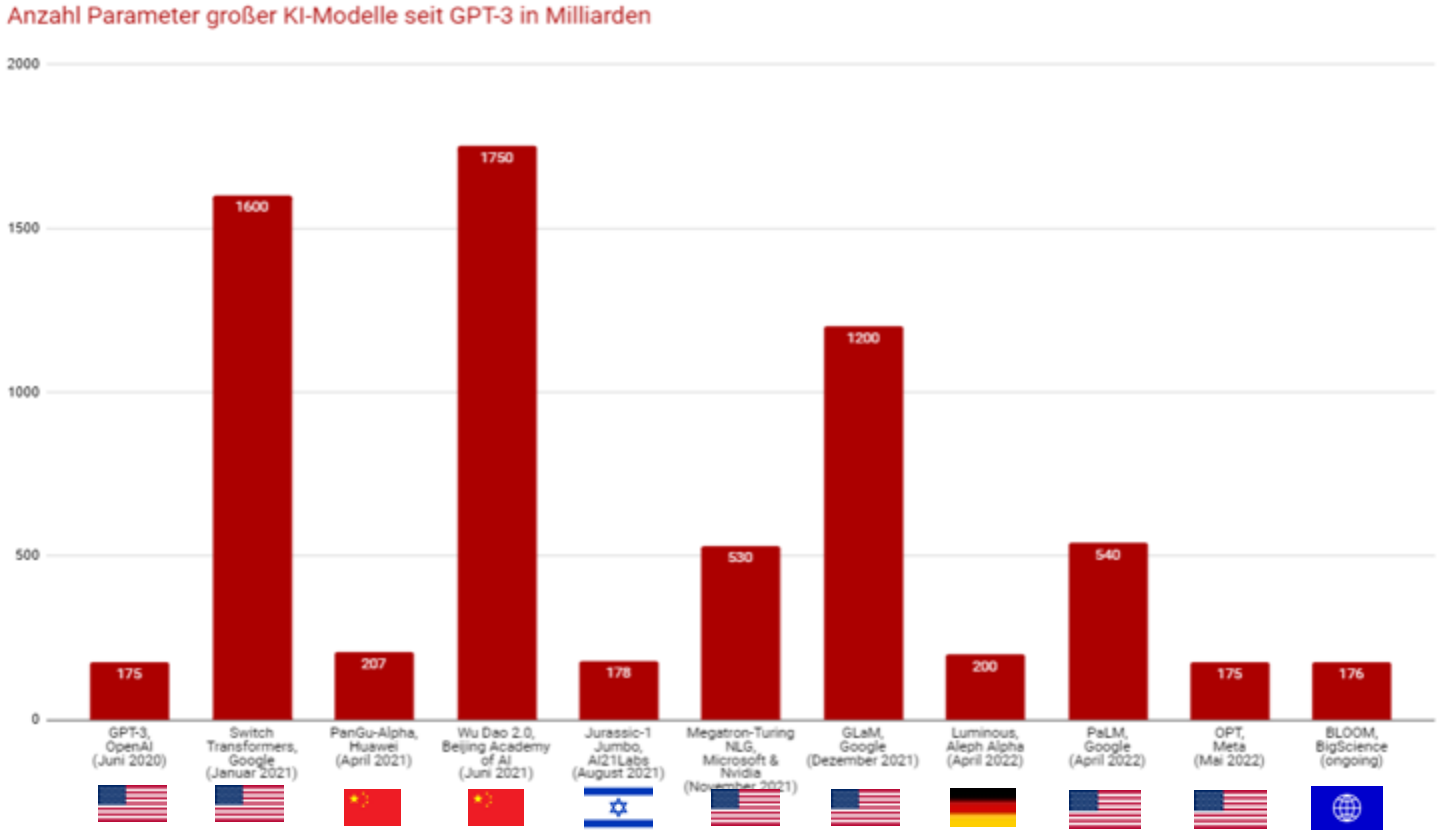
- The PET hydrolase, called FAST-PETase, is more robust to different temperatures and pH levels than existing ones.
- FAST-PETase was able to almost completely degrade 51 different products in 1 week.
- They also showed that they could resynthesize PET from monomers recovered from FAST-PETase degradation, potentially opening the way for industrial scale closed-loop PET recycling.



stateof.ai 2022

TEXAS

GPT-3 leitet einen Wettbewerb ein



2002 - Google hat noch kein Geschäftsmodell

But the bigger question is whether Google has the scale to capture a viable share of the search advertising market. In other words, **can Google create a business model** even remotely as good as its technology?

The New York Times

Google's Toughest Search Is for a Business Model

As Eric Schmidt prepared the agenda for last September's board meeting of Google, the Internet search company, he added an item "Financing Plans -- prepare for initial public offering."

"I assumed I was brought in to take the company public," recalled Mr. Schmidt, who became chairman of Google in March 2001 after a long career at Sun Microsystems and four years as chief executive of Novell.

By last September Google, which began as a research project at Stanford before raising venture capital in 1999, had earned a reputation as the best way to find things on the Internet. It had the prestige of being the search engine on Yahoo. Traffic to its own Google.com Web site was surging. The company's advertising sales, after a slow start, were starting to catch on. It had just turned a profit.

Wir müssen in der Lage sein, große KI-Modelle zu trainieren

KI-Modelle sollen open source zur Verfügung gestellt werden

Daten sollen in Europa und nach europäischen Standards gesichert werden

KI-Modelle sollen eine Vielzahl europäischer Sprachen unterstützen

KI-Modelle sollen transparent entwickelt und Bias bestmöglich reduziert werden

KI-Modelle sollen möglichst CO2-neutral trainiert werden

Europäische KI-Modelle werden zur Sicherung der digitalen Souveränität Europas beitragen

Europa droht, den Anschluss zu verlieren

Zum Trainieren großer KI-Modelle braucht es drei Voraussetzungen:

exzellente Forschung

ausreichend Daten

dedizierte KI-Recheninfrastruktur

Erste Aktivitäten sind gestartet

openGPT-X

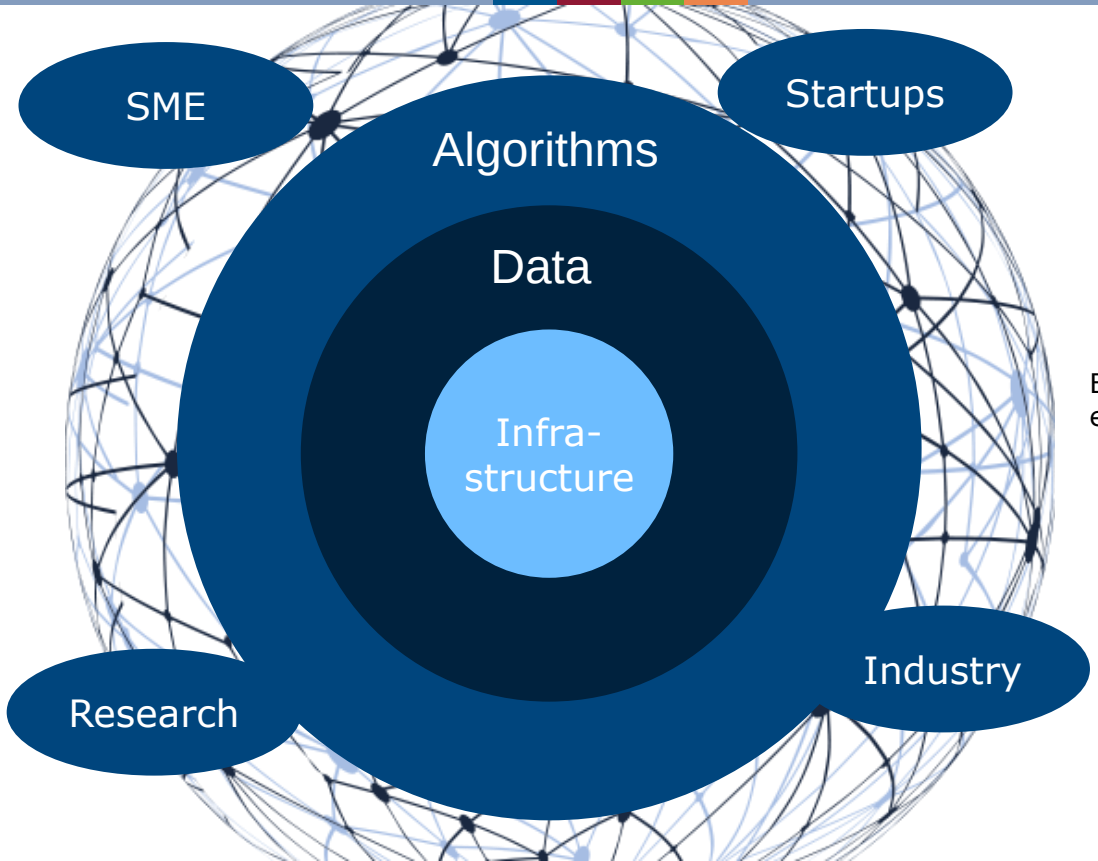
- Konsortium aus zehn deutschen Organisationen
- BMWK fördert das Projekt mit rund 15 Mio. Euro
- Ziel: Entwicklung eines mehrsprachigen großen KI-Modells
- Training auf dem JUWELS Booster im FZ Jülich

aleph alpha

- Heidelberger Startup und Mitglied im KI Bundesverband
- Veröffentlichung des multimodalen Modells LUMINOUS im April 2022
- Mehrsprachigkeit als besonderes Merkmal

- Big Science Initiative
- 600 Forscher aus mehr als 50 Ländern und 250 Organisationen
- Ziel: Von Mai 2021 bis Mai 2022 ein großes KI-Sprachmodell entwickeln
- Training auf dem französischen Jean Zay Supercomputer

LEAM: Large European Language Models



Berücksichtigung europäischer Werte

1. Aufbau einer Infrastruktur
2. Datensammlung
3. Entwicklung von Algorithmen
4. Aufbau eines Ökosystems

LEAM hat bereits ein breites Teilnehmer-Netzwerk

Unternehmen

e.on, BAYER, REWE digital, T-Systems (T-SYSTEMS ON SITE SERVICES GMBH), Lufthansa Industry Solutions, Continental, MERCK, VATTENFALL, msg, bdr BUNDESDRUCKEREI, CLOUD & HEAT

Forschung

TECHNISCHE UNIVERSITÄT DARMSTADT, BHT Berliner Hochschule für Technik, FZI, HUMBOLDT-UNIVERSITÄT ZU BERLIN, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, WILHELM BÜCHNER HOCHSCHULE Mobile University of Technology

KI-Unternehmen

[at] alexanderthamm, 2TXT, OMNIBOT ONE FOR ALL, MERANTIX MOMENTUM, inovex, GIANCE

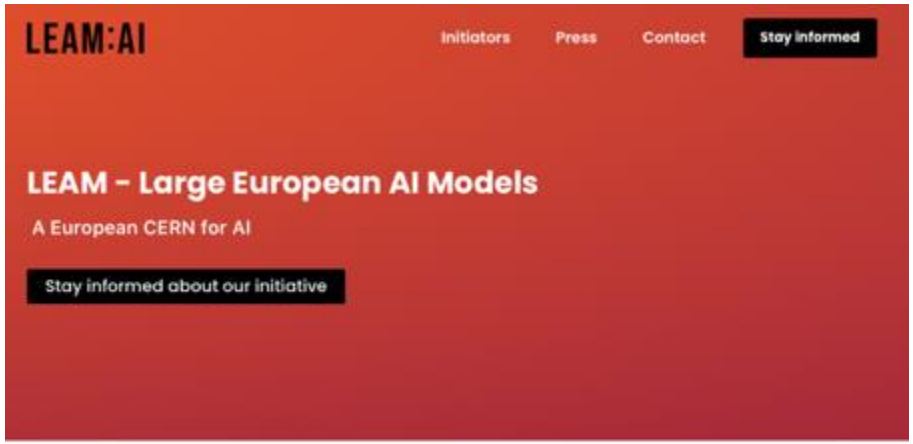
Weitere

K·I·E·Z, LNI4.0 LABS NETWORK INDUSTRIE 4.0, GESELLSCHAFT FÜR INFORMATIK, INC INVENTION CENTER, iW

Verbände

GERMAN AI ASSOCIATION, eco

Unterstützung ist sehr willkommen



What is LEAM?

LEAM is an initiative of the German AI Association (KI Bundesverband), and leading representatives from industry and research that fosters the development of large AI models.

To realize large AI models Made in Europe the initiative proposes **six key steps**:

- Collection and creation of comprehensive training datasets
- Support of excellent research in the field of AI
- Provision of hyperscale infrastructure
- Development of organizational structures and processes to establish a continuous workflow of model development and enhancement
- Integration of the models into the European innovation ecosystem
- Development of methods, benchmark datasets and criteria to ensure ethical requirements and European values

We want the models developed under this initiative to be open source and free to access for all market players. All European languages should be fully integrated into the models.

GERMAN AI ASSOCIATION

Joerg.Bienert@ki-verband.de

Der European AI Act

Europäische Kommission

- April 2021: EU Kommission veröffentlicht ersten Regulierungsvorschlag

Rat der Europäischen Union

- FRA beendet seine EU-RP mit einem Compromise Proposal
- Folgende EU-RP: CZE, SWE, ESP
  - CZE wird versuchen eine General Approach zu erzielen
- Einigung im Rat hängt auch von Vorsitz ab

European Parliament

- IMCO & LIBE FF im EP
- Mitberatende Ausschüsse (JURI, CULT, TRAN, ITRE, ENVI) haben rund 3000 Änderungs-vorschläge im Juni 2022 eingereicht
- IMCO & LIBE verhandeln nun auf dieser Grundlage

Wirtschaftliche Sanktionen bei Nichteinhaltung

30 Mio. € oder 6 % des weltweiten Jahresumsatzes

... im Falle von Verstößen gegen Artikel 5 oder Artikel 10

20 Mio. € oder 4 % des weltweiten Jahresumsatzes

... bei sonstiger Nichteinhaltung von Anforderungen oder Verpflichtungen

10 Mio. € oder 2% des weltweiten Jahresumsatzes

... im Falle Offenlegung von falschen und irreführender Informationen, die den gegenüber den nationalen Wettbewerbsbehörden.

Die Definition von KI ist viel zu breit

„System der künstlichen Intelligenz" (KI-System) eine Software, die mit einer oder mehreren der in aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren;

- **Machine learning approaches**, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning
- **Logic-and knowledge-based approaches,** including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems
- **Statistical approaches**, Bayesian estimation, search and optimization methods

Note: Der Definitionsvorschlag ist ein großer Kritikpunkt. Die aktuelle OECD KI-Definition gilt als Alternativvorschlag der wirtschaftnahen Fraktionen. Sie grenzt KI-System deutlicher von anderer Software ab.

So **neutral wie nötig**, um aktuelle und zukünftige KI-Systeme abzudecken.

Konzentration auf **Use Cases** statt auf AI-Technology.

Ergänzend zum bestehenden Rechtsrahmen, insbesondere der **GDPR.**

KI-Systeme, die **in der EU** bereitgestellt/genutzt werden.

Klassifizierung in Risiko-Kategorien

Der Vorschlag verwendet einen risikobasierten Ansatz, um zwischen vier Arten von KI-Systemen auf der Grundlage ihres Gefahren- und Risikopotenzials zu unterscheiden

Unacceptable Risk

AI systems considered a clear threat to the safety, livelihoods and rights of people.

- Manipulative, subliminal or exploitative techniques
- Classification of people based on their social behaviour
- Remote biometric identification systems

High-Risk

AI systems targeting sensitive GDPR-related topics

- Main focus of the regulation
- Evaluation of customer behaviour
- Recruiting-/Employee management
- Biometric identification in non-public spaces
- Savety-critical systems that endanger health in case of failure
- Administration & justice

Limited- and Minimal-Risk

AI systems in day-to-day use not containing critical data

- AI-Chatbots
- AI-enabled video games
- Spam filters
- Inventory management
- Customer and market segmentation
- Most other AI systems

Hochrisiko KI-Technologien

AI systems identified as high-risk include AI technology used in

Administration of justice and democtratic processes

- E. g. applying the law to concrete set of facts

Educational/vocational Training

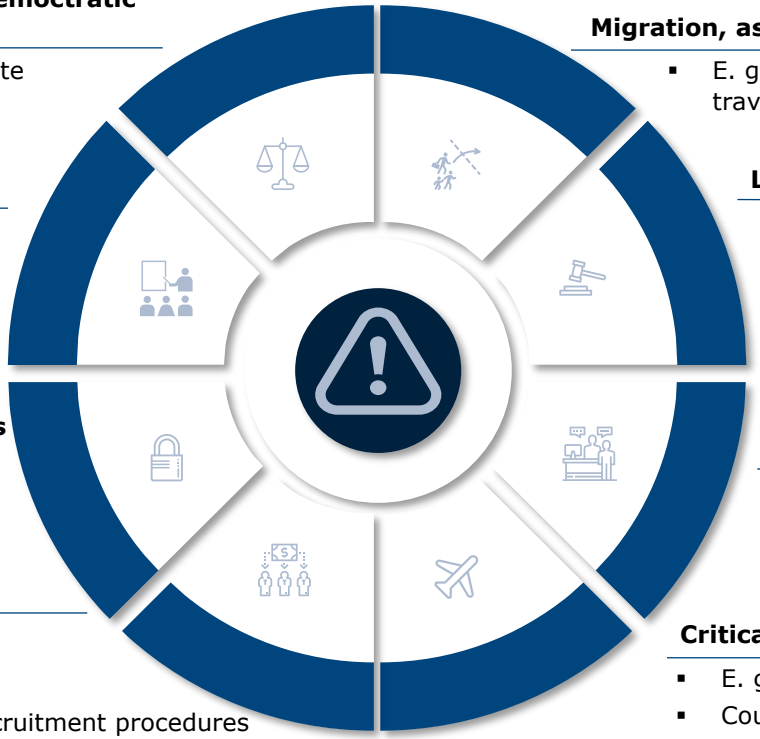
- That may determine the access to education and professional course of someone‘s life (e. g. scoring of exams)

Safety Components of Products

- E. g. AI application in robot-assisted surgery

Employment

- Management of workers
- Access to self-employment
- E. g. CV-sorting software for recruitment procedures



Migration, asylum and border control mgmnt

- E. g. verification of authenticity of travel documents

Law enforcement

- Which may interfere with people‘s fundamental rights
- E. g. evaluation of the reliability of evidence

Essential private and public services

- E. g. credit scoring denying citizens opportunity to obtain a loan

Critical infrastructures

- E. g. transport
- Could put the life and health of citizens at risk

Hochrisiko KI – Zu treffende Maßnahmen

High-Risk AI systems must both conform to stringent quality standards and comply with disclosure, control and monitoring standards.

Risk management system:

Adequate risk assessment and mitigation systems.

Data and data governance:

High quality of the datasets feeding the system to minimize risks and discriminatory outcomes.

Record-keeping

Logging of activity to ensure traceability of results.

Technical Documentation:

Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance.

Transparency and provision of information to users:

Clear and adequate information to the user.

Human oversight

Appropriate human oversight measures to minimize risk.

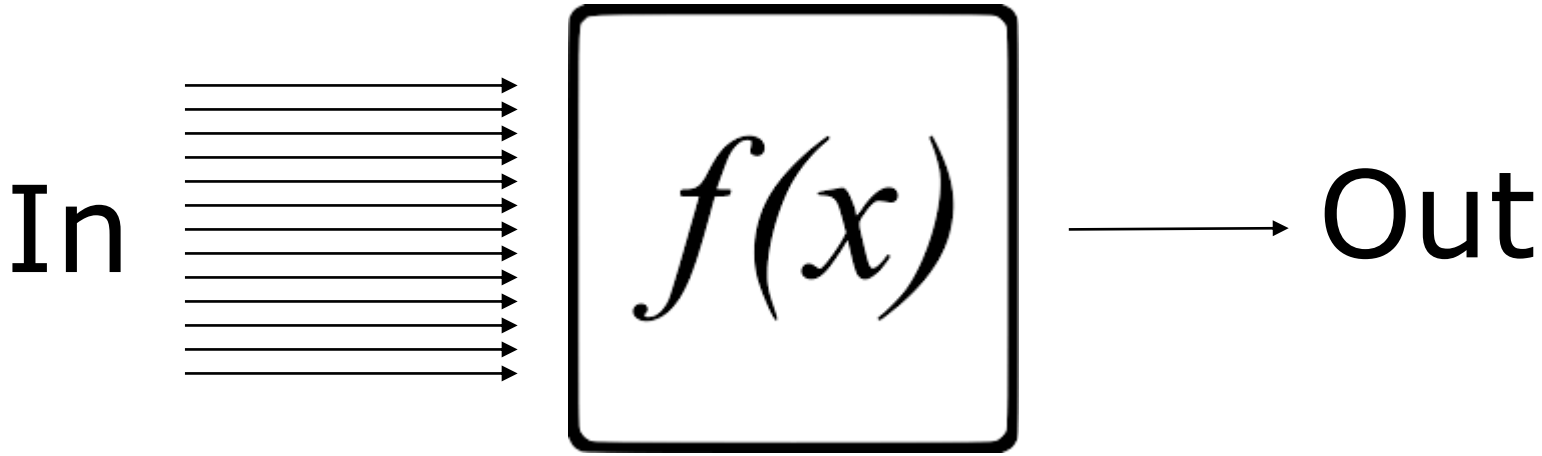
Accuracy, robustness and cybersecurity

High level of robustness, security and accuracy.

Source: Title 3, Chapter 2, Article 9

Blackbox?

Der Output von KI-Modellen basiert auf deren Eingangsvariablen



Hohe Komplexität

Die Anzahl der Eingangsparameter und internen Gewichtungen kann extrem groß sein

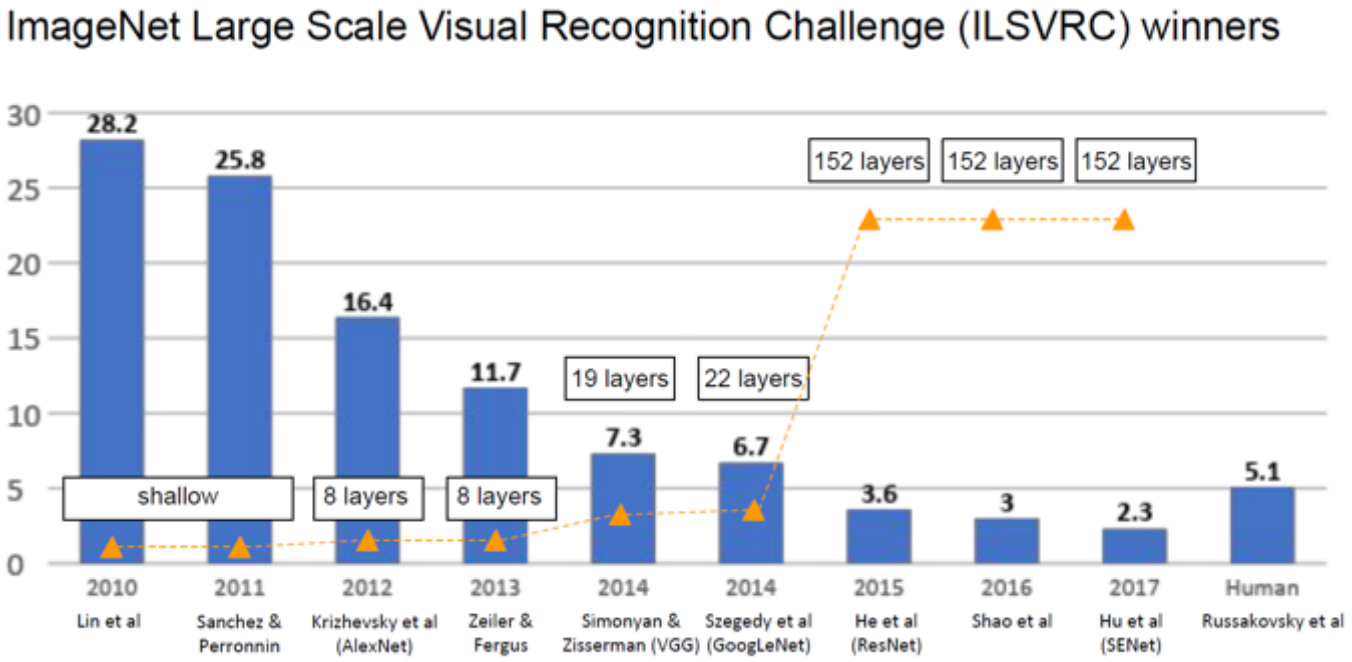
Millions

Trillions

$f(x)$

KI-Modelle werden immer besser

… und übertreffen den Menschen in Sachen Performance



https://kjhov195.github.io/2020-02-10-CNN_architecture_2/

Menschen machen Fehler

Die Welt ist grau

Menschliches Versagen ist verantwortlich für

95%

der **Cyber-Vorfälle**

1 von 11

Krankheitsdiagnosen durch Menschen ist durchschnittlich **fehlerhaft**

Mindestens

90%

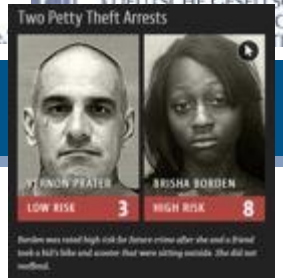
der **Autounfälle** entstehen durch **menschliches Versagen**

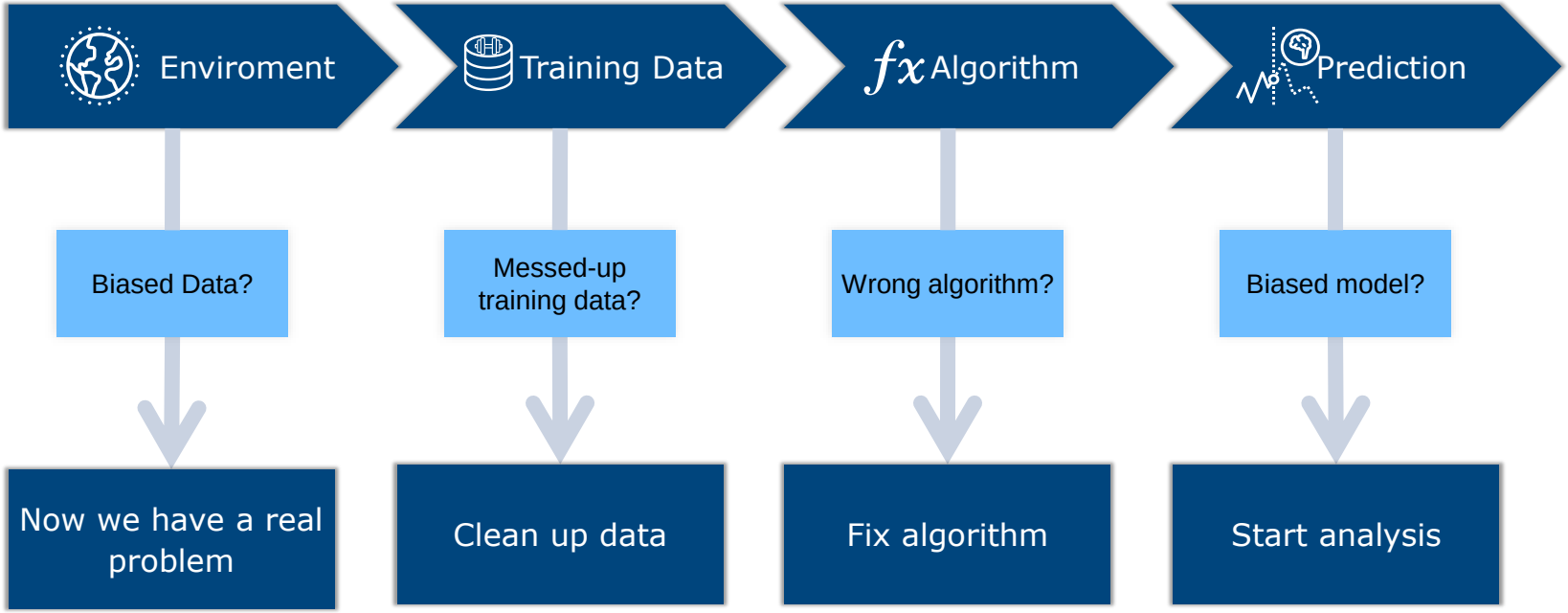
Performance von KI-Modellen

KI könnte sogar bessere Ergebnisse erzielen

Bias in KI-Modellen

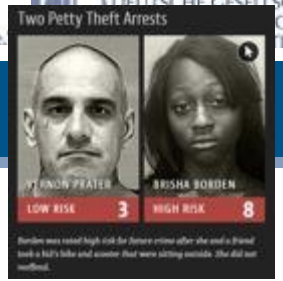
…ein multidimensionales Problem





Bias in KI-Modellen

…ein multidimensionales Problem



Enviroment → Training Data → Algorithm → Prediction

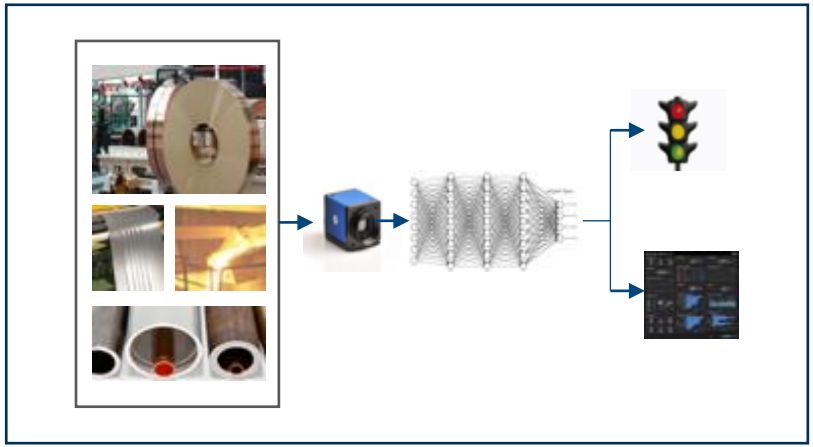
Biased Data?

Now we have a real problem → **What are we going to do to fix reality ?**

Anwendungen von KI

Die Evaluation von KI-Systemen hängt von spezifischen Anwendungen ab

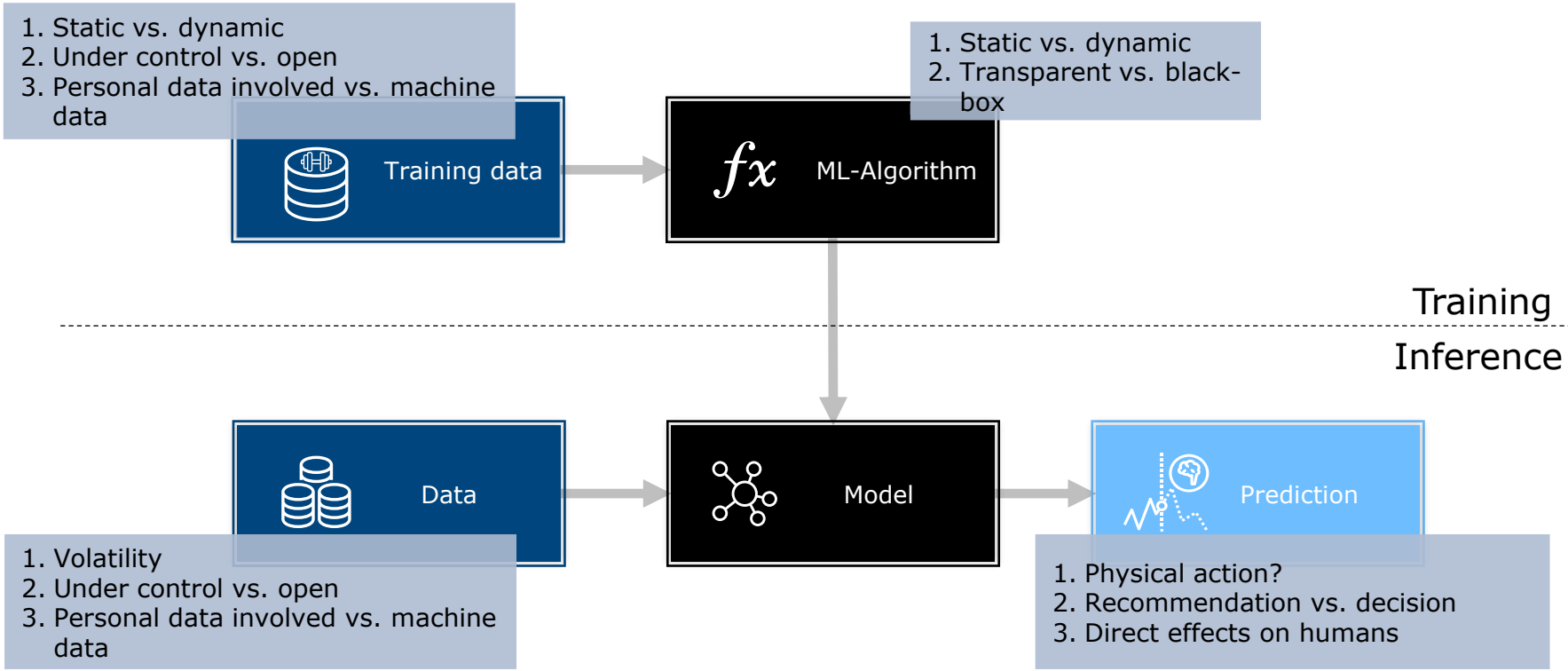
Automatic Warfare Drones



Optical Quality Control in Production

Qualität und Risiken

…in KI-System kommt auf verschiedene Faktoren an



Qualität und Risiken

1. Training Data

1.1 static vs. dynamic

1.2 under control ws. open

1.3 personal data vs. machine data

2. ML-Algorithmen

2.1 static vs. dynamic

2.2 transparent vs. Black box

3. Inference data

3.1 volatitly

3.2 under control vs open

3.3 personal data vs. machine data

4. Prediction

4.1 recomendation vs. decision

4.2 just data vs. Physical impact

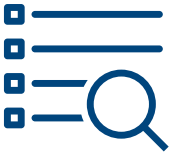
4.3 people vs machines involved

AI Risk/Quality Index

$$\sum_{i=1,\, j=1} a_{i,\, j}$$

Key Takeaways

KI ist anwendungsspezfisch



Vorhersagen sind nicht deterministisch

Genauigkeit muss mit menschlicher Leistung verglichen werden

Danke für ihre Aufmerksamkeit

Für weitere Informationen:
Jörg Bienert
Chief Customer Officer
Joerg.Bienert@alexanderthamm.com
T +49 (0) 177 794 046 3

Background: Dall:E 2 - „data-driven, ultra realistic, unreal engine 5, 8K, orange and blue “