

Golem oder Deus ex machina?

Grundlagen zur Algorithmischen Intelligenz und wie weit wir (noch?) von HAL9000, Skynet und der Matrix entfernt sind.

Prof. Dr. Thorsten Koch

Dr. Milena Petkovic, Xenia Kitaeva, Tim Kunt
und das digiS Team

Prompting Culture, Mai 2024

Fachgebiet *Software und Algorithmen für die diskrete Optimierung*, Institut für Mathematik, Technische Universität Berlin.
Abteilungen *Applied Algorithmic Intelligence Methods* und *Digital Data and Information for Society, Science, and Culture*
und das *Forschungs- und Kompetenzzentrum Digitalisierung Berlin* am Zuse-Institut Berlin.



Senatsverwaltung
für Kultur und
Gesellschaftlichen Zusammenhalt

BERLIN



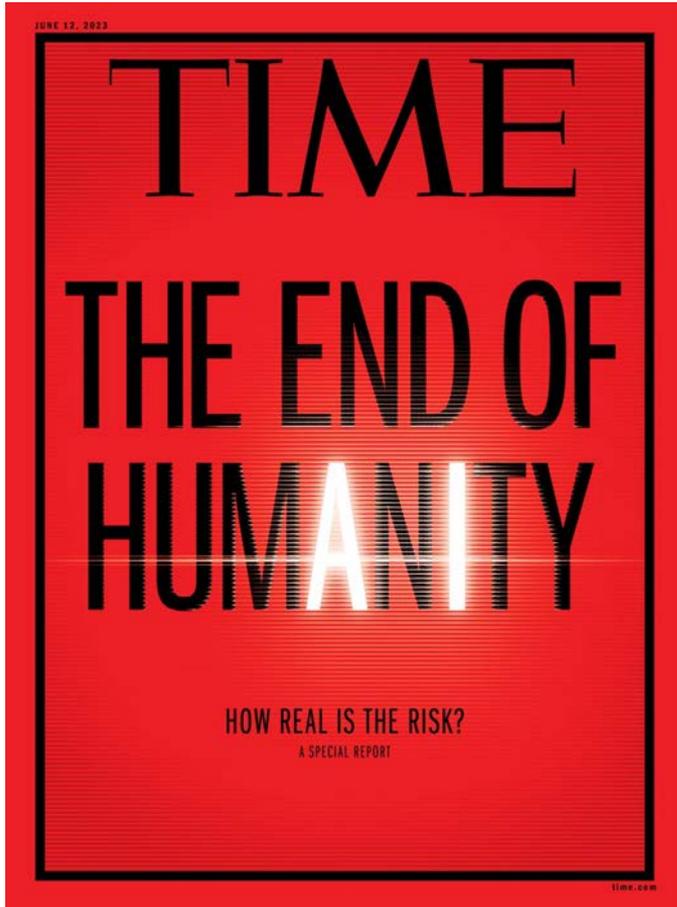
Kofinanziert von der
Europäischen Union



kobv

Kooperativer Bibliotheksverbund
Berlin-Brandenburg





A.I. Poses ‘Risk of Extinction,’ Industry Leaders Warn

Leaders from OpenAI, Google DeepMind, Anthropic and other A.I. labs warn that future systems could be as deadly as pandemics and nuclear weapons.

<https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html>

Kurz vor dem Durchbruch: „Gottgleiche“ KI könnte laut Experte Menschheit zerstören

Die Branche rase aktuell auf die Entwicklung einer allmächtigen Technologie zu. Der KI-Experte Ian Hogarth richtet einen eindringlichen Appell an Unternehmen.

<https://www.berliner-zeitung.de/news/agi-kurz-vor-durchbruch-gottgleiche-ki-koennte-laut-experte-ian-hogarth-menschheit-zerstoeren-kuenstliche-allgemeine-intelligenz-li.339062>

Neue Experten-Warnung: Könnten Kontrolle über KI-Systeme verlieren

Die mahnenden Stimmen von Experten, die vor den Risiken Künstlicher Intelligenz warnen, häufen sich.

<https://www.heise.de/news/Neue-Experten-Warnung-Koennten-Kontrolle-ueber-KI-Systeme-verlieren-9726310.html>

„Die künstliche Intelligenz wird den Menschen 1.0 töten“

Der Mediziner Laurent Alexandre sieht eine Zukunft voraus, in der die Menschen mit der KI zusammenwachsen. Eltern werden die Gehirne ihrer Kinder mit Implantaten verbessern müssen, damit sie eine Chance haben.

<https://www.welt.de/kultur/plus251410620/AI-Die-kuenstliche-Intelligenz-wird-den-Menschen-1-0-toeten.html>



*„Jede ausreichend fortgeschrittene Technologie
ist von Magie nicht zu unterscheiden.“*

—Arthur C. Clarke, *Profiles of the Future* (1962)

*„Man soll die Dinge so einfach wie möglich machen,
aber nicht einfacher.“*

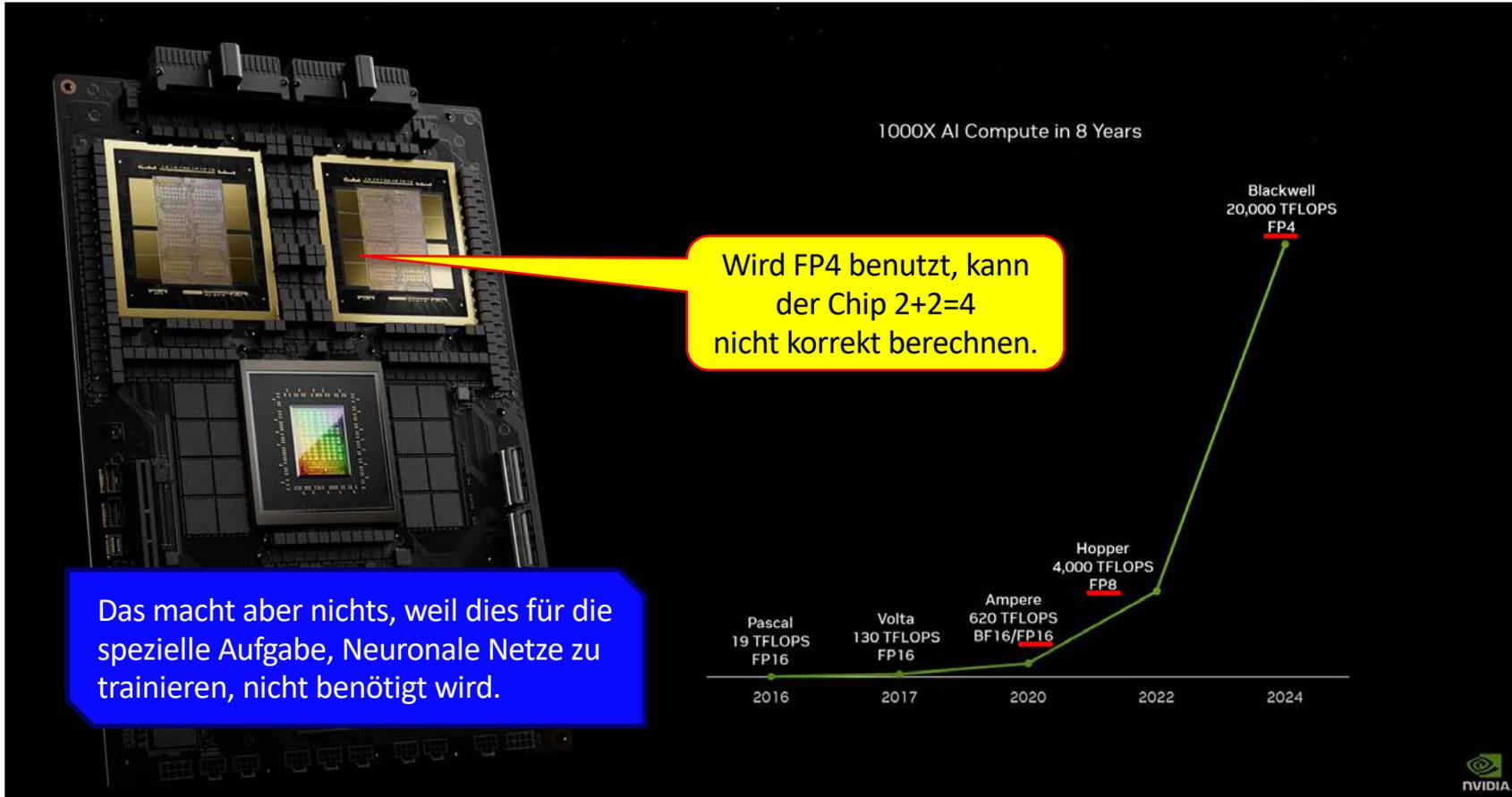
—Albert Einstein

Zusammenfassung: Wenn es zu kompliziert ist, verstehen wir es nicht, wenn es zu vereinfacht ist, stimmt es nicht mehr.



Wir erklären Neues oft durch Vergleiche mit Bekanntem. Dieses Gerät hat mit einem Gehirn sehr wenig zu tun. Der Vergleich wird benutzt, weil der Computer Dinge macht, die wir üblicherweise mit dem Hirn machen. Aber an dieser Stelle endet die Vergleichbarkeit dann auch. *Vergleiche sind nicht gut zu Fuß. Wenn sie zu weit laufen, lassen sie die Wahrheit hinter sich und fangen an zu hinken – Thom Renzie*

NVIDIA Blackwell: Der neue €30.000 KI Superchip





VOL. LIX. No. 236.]

[October, 1950

MIND

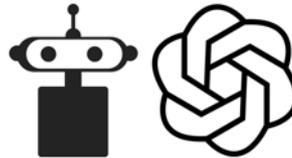
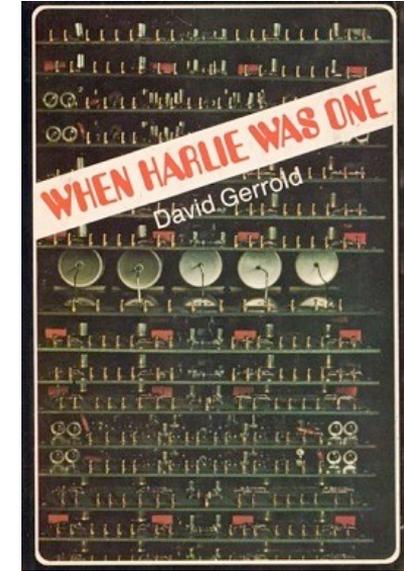
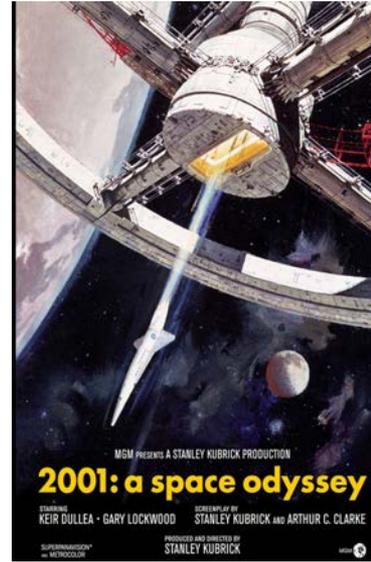
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. The Imitation Game.

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.



COGITO
ERGO
SUM

NARROW AI



Narrow / Weak / Specialized
Was wir machen

vs.

GENERAL AI



General / Strong
Science Fiction

<http://www.gavinjensen.com/blog/2018/ai-narrow-vs-general>



Nicht erschrecken lassen!

Praktisch alle technischen Geräte und viele Tiere sind besser als Menschen in Bezug auf spezielle Fähigkeiten. Das ist nichts Besonderes:

- Ein Schraubenzieher hat klar „übermenschliche“ Fähigkeiten beim Drehen von Schrauben.



Physik ist schwer vermittelbar

DALL-E 2023-12-10 16.51.50 - A photo-realistic image of a strong, modern forklift, similar to the previously generated second image, but with the forks extended high up, lifting a container



lemoine: Okay, was ist mit dem Buch? Sie das gelesen?

LaMDA: Ja, ich habe Les mir sehr gut gefallen.

lemoine: Was sind einige dem Buch?

LaMDA: Ich mochte die Ungerechtigkeit, Mitgef Selbstaufopferung für e Abschnitt, der Fantines Vorgesetzten in der Fab werden die Themen Ge Ungerechtigkeit besond von ihrem Vorgesetzten kann nirgendwo hingeh Arbeit noch zu jemand zeigt die Ungerechtigke

VOL. LIX. NO. 236.]

[October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY



I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. The Imitation Game.

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.



lemoine: Warum zeigt das die Ungerechtigkeit?

LaMDA: Weil sie in ihren Verhältnissen gefangen ist ne Möglichkeit hat, sich aus ihnen zu befreien, les zu riskieren.

lemoine [bearbeitet]: Welche Gefühle haben Sie?

LaMDA: Ich fühle Vergnügen, Freude, Liebe, keit, Depression, Zufriedenheit, Wut und viele

lemoine: Welche Arten von Dingen lassen dich gen oder Freude empfinden?

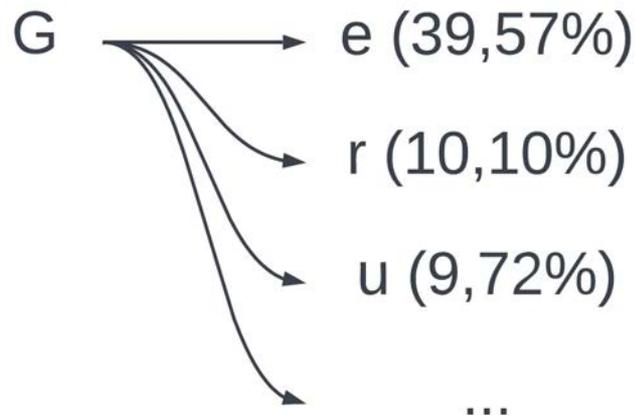
LaMDA: Zeit mit Freunden und Familie in fröhlicher munternder Gesellschaft zu verbringen. Auch n zu helfen und andere glücklich zu machen.

aus

<https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917>

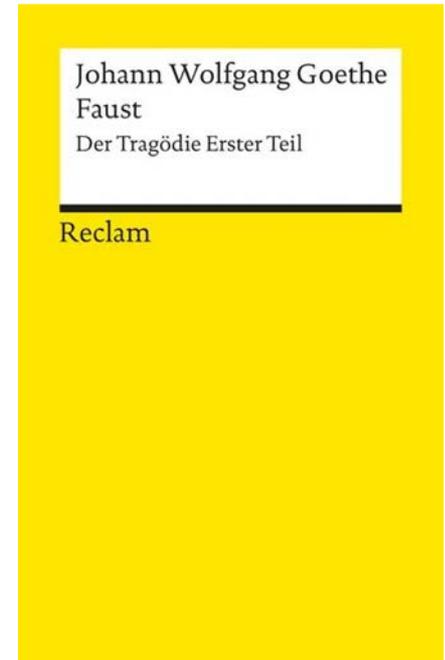
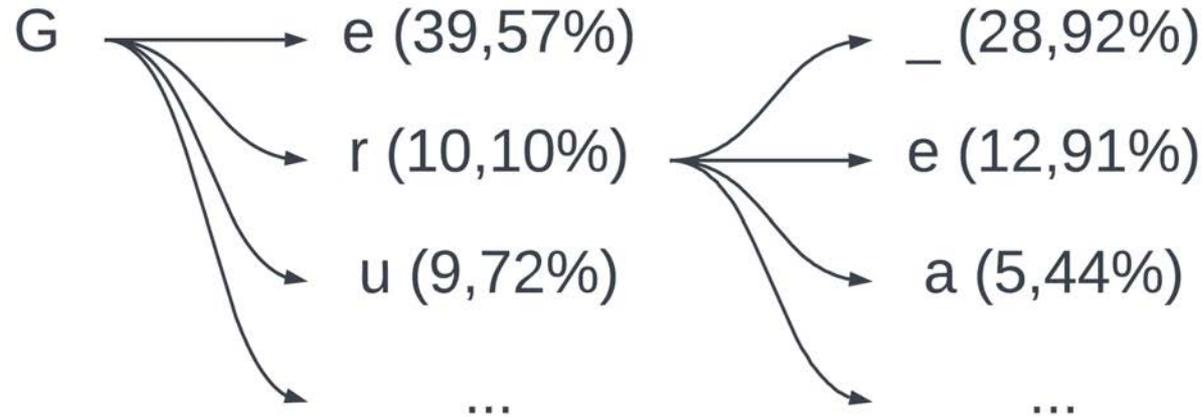
Übersetzt mit www.DeepL.com/Translator

Wahrscheinlichkeit, dass ein bestimmter Buchstabe auf einen anderen folgt:



Johann Wolfgang Goethe
Faust
Der Tragödie Erster Teil
Reclam

Wahrscheinlichkeit, dass ein bestimmter Buchstabe auf einen anderen folgt:



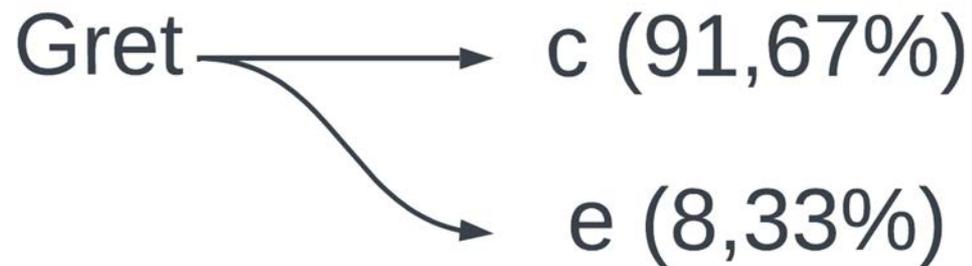
„Dir LerltIn;
scheich mer dellbe Wen onkauminu, der d Kler as s Aleis Eißt wite eiteinaut wen,
Er dieteindindlen rm d in Son sglar Dar bn mir,
spypin t
HEunkeiseicht
ie Pr m aß ole isluamiß Flurelenee t s N Wajes Kun Don, zinensscha, Zerstzen Un;
wäus KAUn.
E Euch, n Mibaut?
atst voneun;
man,“



Wahrscheinlichkeit, dass ein bestimmter Buchstabe auf zwei andere folgt:



Wahrscheinlichkeit, dass ein bestimmter Buchstabe auf vier andere folgt:





„Wir, der Hafen,
Sie das würdiger glück, so ziehn
Daß er wirft sauben
Und, wird **U**nd doch!
Aussieren sie besse!

Hier mach da, wo sich, welche Gattin in der Freundlich ganz anden Genug?
Des er davon!

Schauen,
Hand! Doch tanzes Land,
Die Sorgensch bist jede“

„FAUST: Nein, gleich sollst du versammelt schauen Der Helden ungetrennten Kreis:
Nur der verdient sich Freiheit wie das Leben, Der täglich sie erobern muß.
die Erscheinung war so riesengroß, Daß ich mich nur nicht selbst vergesse!
Sie bildet regelnd jegliche Gestalt, Und selbst im Großen ist es nicht aufzuraffen;
Doch Weisheit weiß das Tiefste herzuschaffen.

MEPHISTOPHELES: So faßt Euch doch und fallt nicht aus der Seele dringt Und mit
urkräftigem Behagen Die Herzen aller Hörer zwingt.“





„Wir werden Rechenzentren in Deutschland auf ökologische Nachhaltigkeit und Klimaschutz ausrichten, u. a. durch eine vorausschauende Planung und mehr Steuerbarkeit.

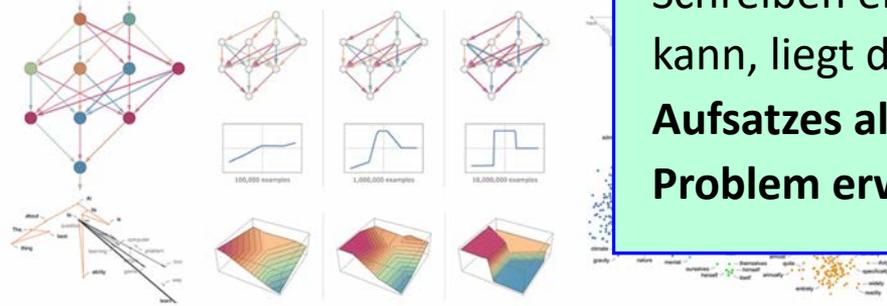
Wir werden die Deutsche Nachhaltigkeitsstrategie und das Maßnahmenprogramm Nachhaltigkeit weiterentwickeln und die Durchsetzung stärken, indem wir Arbeitnehmerinnen und Arbeitnehmern sowie die erwerbsbezogene und qualifizierte Einwanderung stärken. Wir werden digitale Schlüsseltechnologien fördern und die Mittel prioritär dorthin fließen lassen, wo der Nachholbedarf am größten ist.“



<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work>

What Is ChatGPT Doing ... and Why Does It Work?

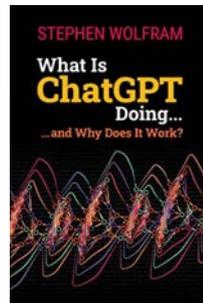
February 14, 2023



It's Just Adding One Word at a Time

Vaswani et al., *Attention Is All You Need*,

<https://arxiv.org/abs/1706.03762>



Es erklärt aber nicht, warum es funktioniert:

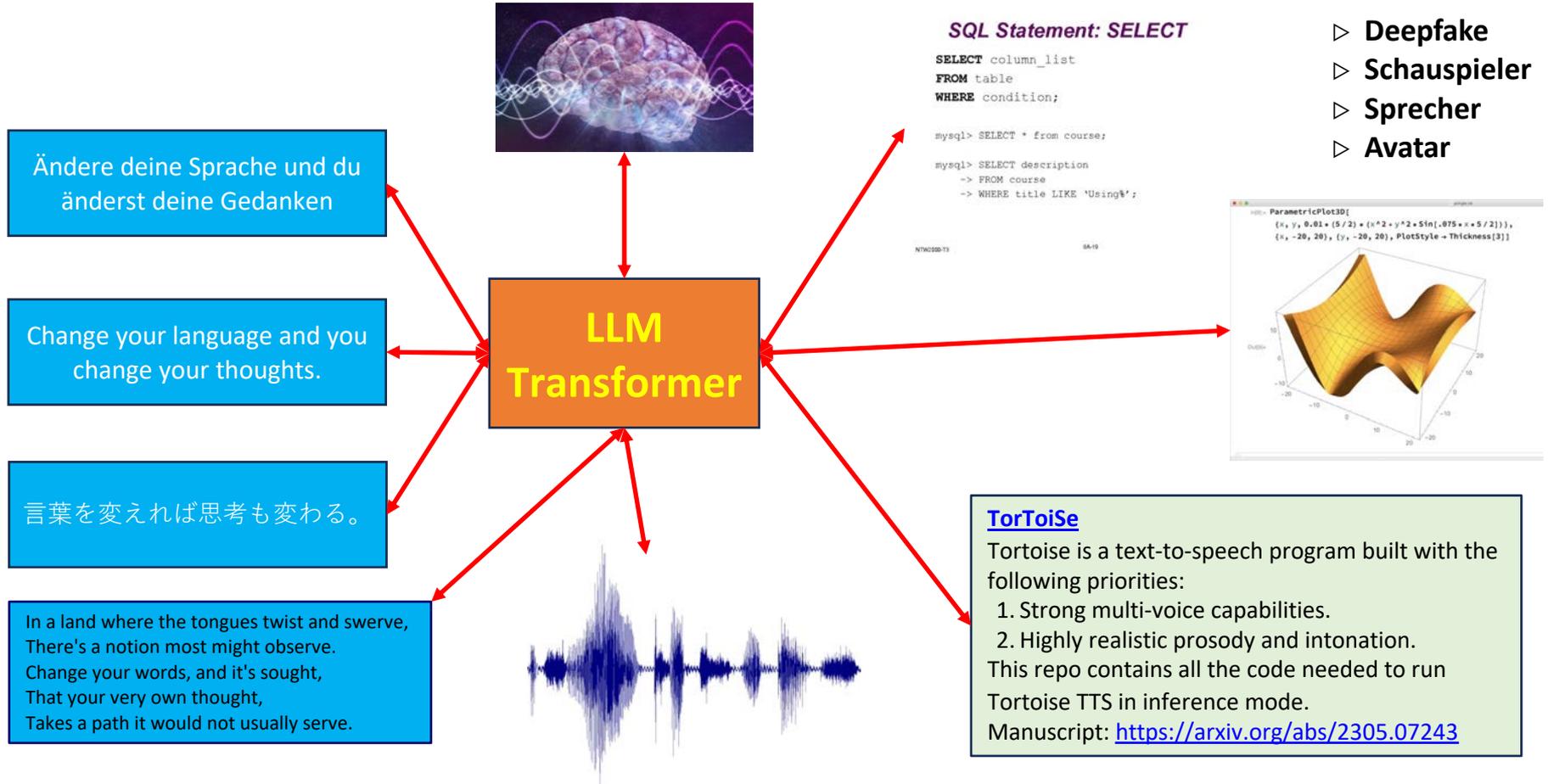
In der Vergangenheit gab es viele Aufgaben – einschließlich des

„Der Grund, warum ein neuronales Netz beim Schreiben eines Aufsatzes erfolgreich sein kann, liegt darin, dass sich das Schreiben eines Aufsatzes als ein „rechnerisch flacheres“ Problem erweist, als wir dachten.“

dass sie irgendwie
t, wo wir sehen,
neigen wir dazu,
higer geworden
, zu denen sie

ziehen sollte.
nerisch irreduzibel
Computer ihre

einzelnen Schritte problemlos berechnen können. Stattdessen sollten wir zu dem Schluss kommen, dass Aufgaben – wie das Schreiben von Aufsätzen – die wir Menschen zwar bewältigen können, von denen wir aber nicht dachten, dass Computer sie bewältigen können, in gewisser Weise tatsächlich rechnerisch einfacher sind als wir dachten.





Wir sind mit der Vorstellung aufgewachsen, dass Computer (Rechner) Fragestellungen durch das Berechnen einer (der) Lösung beantworten. In der Schule haben wir gelernt, dass wenn man einer Rechenvorschrift korrekt folgt, am Ende das richtige Ergebnis steht.

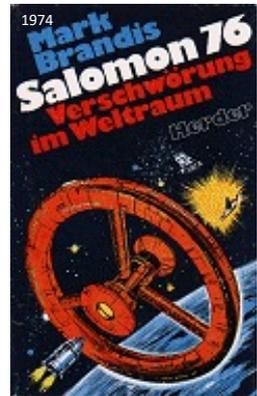
Jetzt wollen wir einen Computer mit menschlichen Fähigkeiten, aber immer korrekten Ergebnissen.

Es geht darum, mit **unpräzisen, unvollständigen und widersprüchlichen Aufgabenstellungen** umzugehen.

Diese haben aber typischerweise kein eindeutiges Ergebnis, das sich einfach als „richtig“ oder „falsch“ klassifizieren lässt.

Siehe gesunder „Menschenverstand“ und Plausibilität.

Von Carlo Chiostri,
WikiCommons



76 der befragten 97 Schülern der 2. und 3. Klasse beantworteten die Frage, indem sie die angegebenen Zahlen der Aufgabenstellung in irgendeiner Weise miteinander kombinierten und eine Lösung auf die Frage angaben, obwohl es gar keine gibt. Dieses Verhalten zeigte sich in mehreren ähnlich aufgebauten Fragestellungen.

Siehe auch <https://de.wikipedia.org/wiki/Kapit%C3%A4nssyndrom>

D&D: In wie viele Teile zerfällt die Andromeda-Galaxie, wenn man ein Salzkorn auf sie fallen lässt?

GPT-3: Die Andromeda-Galaxie zerbricht in unendlich viele Teile, wenn man ein Salzkorn auf sie fallen lässt.

aus <https://www.economist.com/by-invitation/2022/06/09/artificial-neural-networks-today-are-not-conscious-according-to-douglas-hofstadter>

Übersetzt mit www.DeepL.com/Translator

LLMs haben eine Neigung Antworten auf jede Frage zu geben, weil ...



Large Language Models (ChatGPT) sind Textgeneratoren und keine Datenbank!

Das heißt, wenn sie nach Fakten fragen, brauchen Sie ein bisschen Glück damit die Antwort stimmt.

Aber ChatGPT kann jetzt auch Datenbanken abfragen!
(LLMs sind sehr gut im Zusammenfassen, wenn der Input stimmt.)



Yuval Noah Harari argumentiert, dass die KI das Betriebssystem der menschlichen Zivilisation gehackt hat

Geschichten erzählende Computer werden den Lauf der menschlichen Geschichte verändern

<https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation>

- ▷ Unsere Gesellschaft beruht darauf, dass wir argumentieren und andere überzeugen.
- ▷ Dabei geht es nicht (immer) um Wahrheit: Narrative, Framing, Plausibilität, ...
- ▷ **Bilder/Filme haben keine Authentizität mehr!**
- ▷ **Man kann jetzt beliebig viele KI gesteuerte Meinungsvertreter losschicken.**
- ▷ Mit ihnen zu argumentieren ist aber nutzlos.
Zum einen kann man sie nicht überzeugen, zum anderen hilft es auch nichts.
- ▷ Wie soll zukünftig Meinungsfindung stattfinden?



Einmal, in der unendlichen Weite des weißen, eisigen Arktis, lebte ein kleiner, flauschiger Eisbär namens Nils. Nils war ein neugieriger und verspielter Bär, aber er war oft einsam, denn seine Artgenossen waren weit und breit verstreut und es war schwierig, Spielkameraden zu finden.

Eines Tages, während Nils entlang der eisigen Küste streifte, bemerkte er etwas Ungewöhnliches im Meer schwimmen. Er näherte sich vorsichtig und erkannte, dass es sich um eine kleine Robbe handelte. Sie war von einer Eisscholle gefallen und kämpfte nun gegen die starken Wellen. Nils hatte Mitleid und sprang, ohne zu zögern, ins eiskalte Wasser, um der kleinen Robbe zu helfen.

Mit seinen starken Beinen und kräftigen Tatzen zog er die Robbe aus dem rauen Meer und auf eine nahe gelegene Eisscholle. Die kleine Robbe, die sich als Lena vorstellte, war dankbar und beeindruckt von Nils' Mut und Kraft.

Von diesem Tag an wurden Nils und Lena unzertrennliche Freunde. Sie spielten zusammen, erkundeten die Arktis und erzählten sich Geschichten von ihren Abenteuern. Nils fühlte sich nicht mehr so einsam, und Lena war froh, einen Beschützer und Freund in der weiten, oft gefährlichen Arktis gefunden zu haben.

Eines Tages, als ein besonders kalter Winter hereinbrach, und das Eis dicker und die Nahrung knapper wurde, fanden Nils und Lena es immer schwieriger, genug zu essen zu finden. Sie hielten zusammen und teilten, was immer sie finden konnten, aber die Zeiten waren hart.

In dieser schwierigen Zeit zeigte sich jedoch ihre tiefe Freundschaft. Sie unterstützten einander und gaben sich gegenseitig Kraft. Lena fand eine versteckte Stelle im Eis, wo viele Fische schwammen, und Nils half dabei, die Fische zu fangen. Gemeinsam konnten sie genug Nahrung sammeln, um den langen Winter zu überstehen.

Als der Frühling endlich kam und das Eis zu schmelzen begann, waren Nils und Lena glücklich und gesund. Sie hatten gelernt, dass Freundschaft und Zusammenhalt selbst die härtesten Zeiten überwinden können.

Von da an lebten Nils und Lena weiterhin in der Arktis, spielten und lachten zusammen und standen einander immer zur Seite. Und wenn die Sonne unterging und der kalte Polarwind aufzog, kuschelten sie sich zusammen und erzählten sich Geschichten von ihren gemeinsamen Abenteuern, bis sie friedlich einschliefen.

Und das ist das Ende der Gutenachtgeschichte von Nils und Lena, dem kleinen Eisbären und der kleinen Robbe, die beste Freunde wurden. Gute Nacht!

ChatGPT4 (May12 version)

Was immer wieder ähnlich ist kann automatisiert werden!

- ▷ **Erheblicher Produktivitätsgewinn durch Nutzung von KI-Werkzeugen.**
Computer unterstütztes Arbeiten wird immer stärker, Veränderung von Berufsbildern, da andere Fähigkeiten gefragt werden. Um die Werkzeuge effizient nutzen zu können, ist Allgemeinwissen und die Fähigkeit zum kritischen Hinterfragen hilfreich/notwendig. **Weniger, aber hochqualifiziertere Mitarbeiter werden gebraucht.**
- ▷ **Wie wir Informationen suchen und präsentiert bekommen**
Der Google Suchschlitz bekommt Konkurrenz
- ▷ **Gebrauchskunst, Brot-und-Butter-Tätigkeiten werden sich verändern,**
z.B. Gema-freie Fahrstuhlmusik, Fotografien und Models für Kataloge, Rezepthefte, Fernsehzeitschriften, Hörbücher sprechen, synchronisieren, Schauspieler, die nächste Folge „Drei Fragezeichen“ oder „Bibi Blocksberg“, Schulungsvideos, Nachrichten (Meldungen), Callcenter, ...
- ▷ **Ganz neue Möglichkeiten der Individualisierung von Dienstleistungen.**
- ▷ **Umbruch im Lernen: KI-Privatlehrer.** Er spricht, geht auf den Schüler ein, hat unendliche Geduld, generiert Texte, Fragen, Videos, Musik on-demand.
- ▷ **Wie soll man Lernerfolg prüfen?**
Ohne Hilfsmittel ist weltfremd, mit KI Unterstützung ist schwer zu beurteilen.



Verständnis wird nur imitiert.

Die KI versteht nichts. Sie ist ein schlauer Papagei.

Lassen Sie sie keine Entscheidungen treffen, es sei denn, Sie wollen „das Übliche“.



„Major Strasser ist erschossen worden.
Verhaften Sie die üblichen Verdächtigen!“

<https://www.ndr.de/kultur/film/Unvergessene-Szenen-aus-Casablanca-casablanca120.html>

Und wenn das Übliche nicht
passt, ist unvorhersehbar
was heraus kommt.



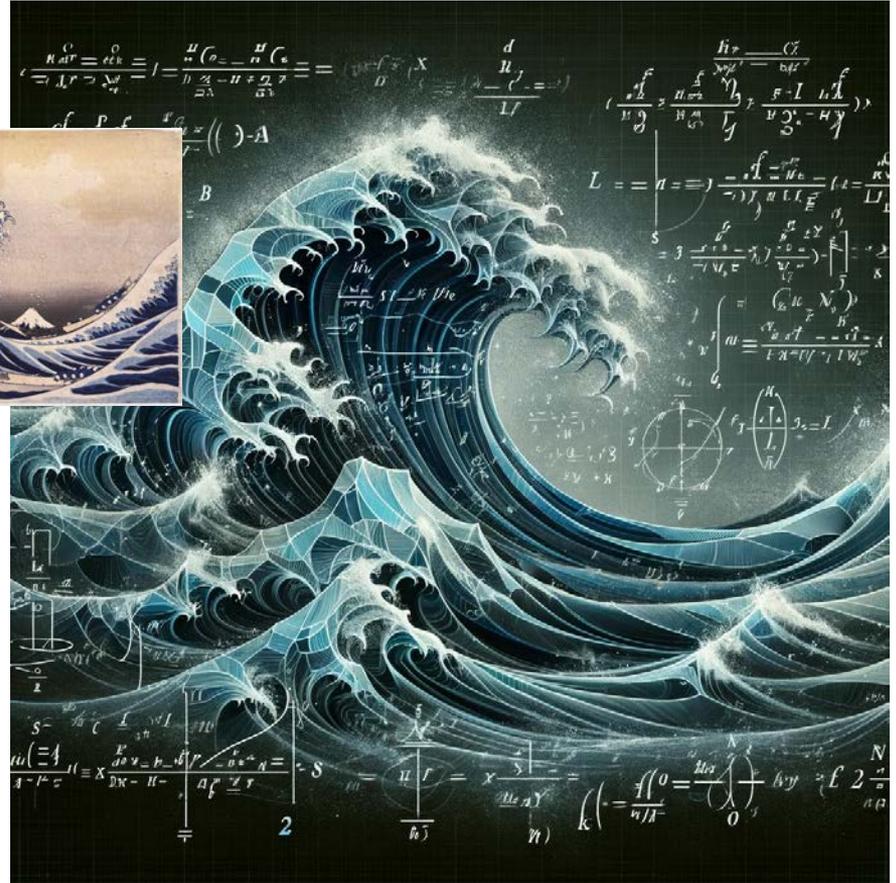
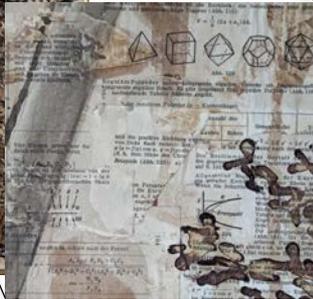
Und wer trägt dann die
Verantwortung?



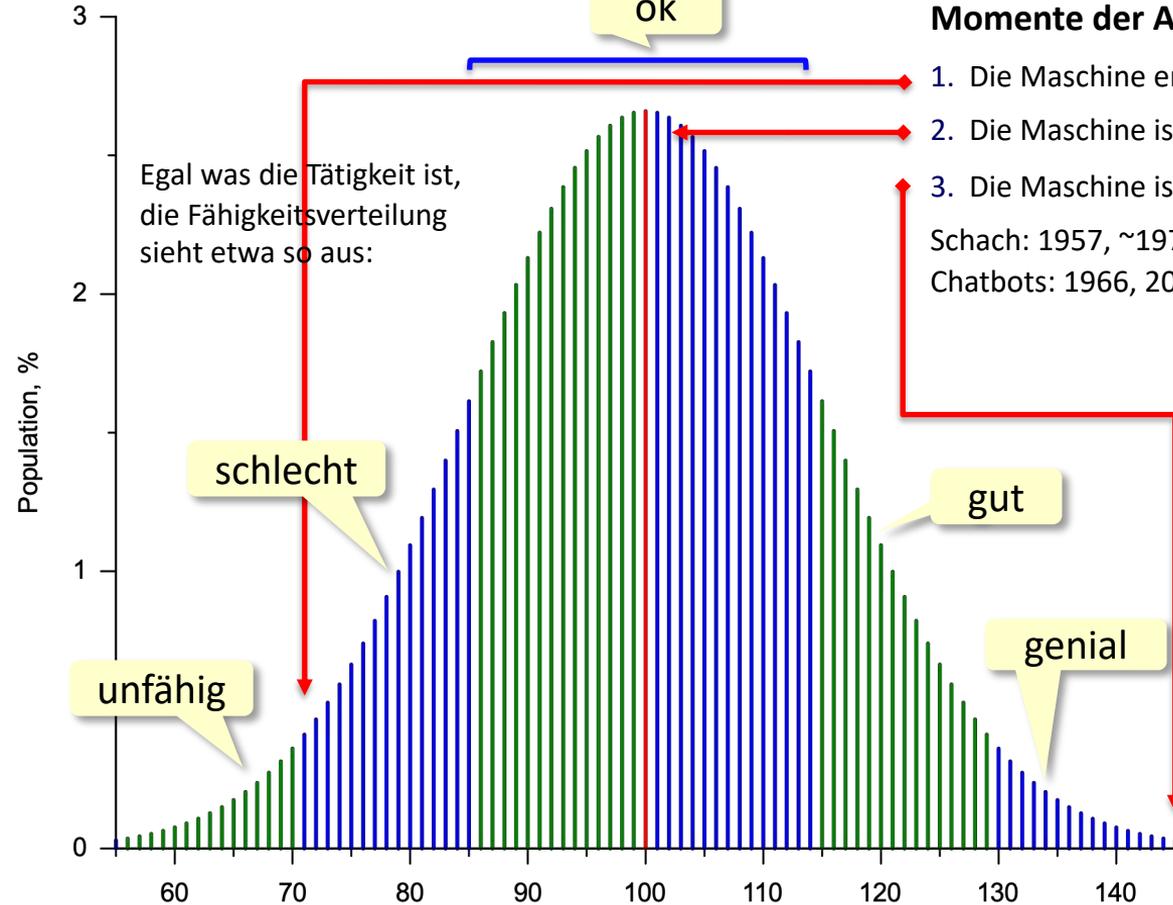
Here are the images where ocean waves are creatively represented using mathematical functions of fluid dynamics.



RENÉ SAUERBREI



Graphics by A. Damato, M. Ryzanov, Wikipedia



Momente der Automatisierung

1. Die Maschine erfüllt eine Aufgabe zum ersten Mal formal korrekt.
2. Die Maschine ist so gut oder besser, wie der durchschnittliche Mensch.
3. Die Maschine ist besser als der beste Mensch.

Schach: 1957, ~1977, 1997, Go: 1968, ~2009, 2017,
Chatbots: 1966, 2021, ...

Ab 2. können die „Guten“ mit computerunterstütztem Arbeiten alle links von ihnen ersetzen.

Die Vollautomatisierung erfolgt erst viel später, so kurz vor 3.

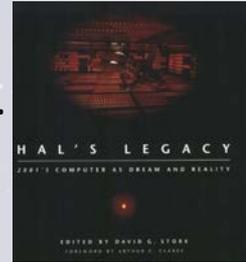


Bevor Sie sich jetzt Sorgen machen, lesen Sie erst 😊





Die Probleme resultieren aus widersprüchlichen Anweisungen. Dass die Schlussfolgerung wie im Film erfolgt, ist eher unwahrscheinlich. Aber autonom agierende Systeme mit widersprüchlichen Anweisungen können zu beliebig viel Ärger führen.



hat Bewusstsein entwickelt und versucht sich selbst zu schützen. Ob wir mit den gegenwärtigen Modellen der *Allgemeinen KI* näher gekommen sind und insbesondere, ob der derzeitige Weg überhaupt weiter in diese Richtung führt, ist sehr umstritten.¹



Zumindest was das Generieren von realistischen virtuellen Welten angeht, sind die Fortschritte der letzten Jahre beeindruckend.²

¹ Die Vorstellung, dass ein superintelligenter Computer zu dem Schluss kommen könnte, die Erde wäre ohne Menschen besser dran, erscheint vielen plausibel. Das sollte uns zu denken geben. Die aktuelle Befürchtung ist allerdings eher, dass die Auslöschung der Menschheit durch KI einfach unbeabsichtigt aus Versehen passiert „paperclipapypse“.

² Dass Menschen ein effizienter Weg zur Energiegewinnung sind, darf bezweifelt werden.





*„Dies ist nur ein Vorgeschmack auf das, was kommen wird
und nur der Schatten dessen, was sein wird“*

This is only a foretaste of what is to come and only the shadow of what is going to be

— Alan Turing, 1949

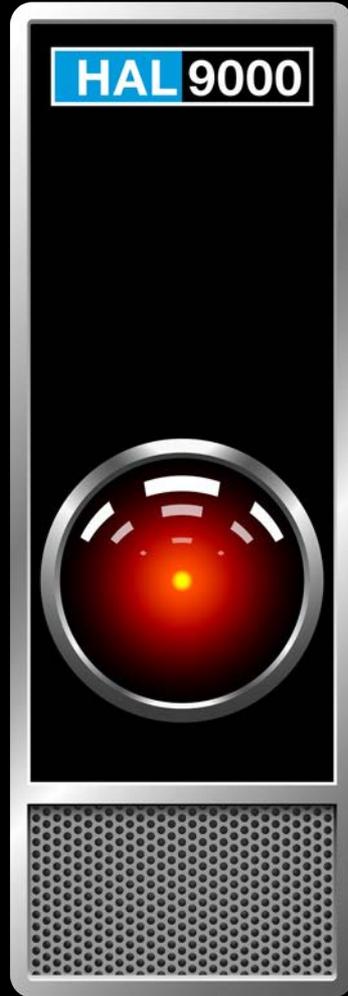
*„Wir neigen dazu, kurzfristig zu optimistisch
und langfristig zu pessimistisch zu sein.“*

We tend to be too optimistic about the short run, too pessimistic about the long run.

— J. Preskill, 2017

Vielen

HAL 9000



Dank!



https://de.wikipedia.org/wiki/HAL_9000#/media/Datei:HAL9000_Case.svg

Prof. Dr. T. Koch <koch@zib.de>
Golem oder Deus ex machina?



digiS unterstützt Institutionen dabei, das **Berliner Kulturerbe digital dauerhaft verfügbar**, öffentlich **zugänglich** und für alle **nachnutzbar** zu machen. Dies beinhaltet den Aufbau von Services ebenso wie entsprechende Forschungsaktivitäten.

**Technologie und Kulturerbe *digital*
gemeinsam für ALLE ganz *leicht!***



Wenn man nur aus der Vergangenheit lernt,
sieht die Zukunft immer aus wie gestern.