

Kann
KI



Gerechtigkeit?

14.3.23
Berlin

BEGRÜSSUNG

Katharina ZWEIG

HEUTE

KANN



GERECHTIGKEIT



Nach 4 JAHREN ist das PROJEKT zu ENDE



Ein interdisziplinäres Team

→ führt zu interdisziplinären Ergebnissen

HEUTE WIRD PRÄSENTIERT



BEGRÜSSUNG

Landesvertretung RHEINLAND-PFALZ

SEIT



RPTU

FUSION → KAISERSLAUTERN & LANDAU

Ideal

KI Wohle

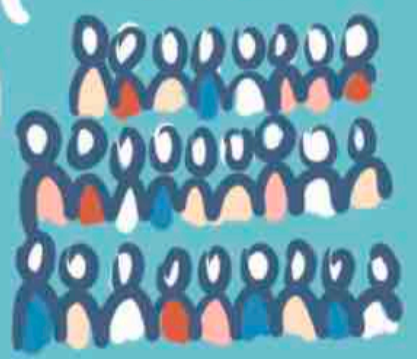
ZUM VON

Mensch



Gesellschaft

WIR



KI AGENDA

NEU ★ Sozio-Informatik
BESONDERHEIT ♥

Kann KI Gerechtigkei?



14.3.23
Berlin

Wann KI keinen Platz vor



Katharina ZWEIG

GERICHT HAT.



Amerikanisches Recht

1964 Civil Rights Act



UNGERECHTIGKEIT



Un-Intentional vs intentional

GRUPPEN-ANSATZ



Manche per se schlechter gestellt

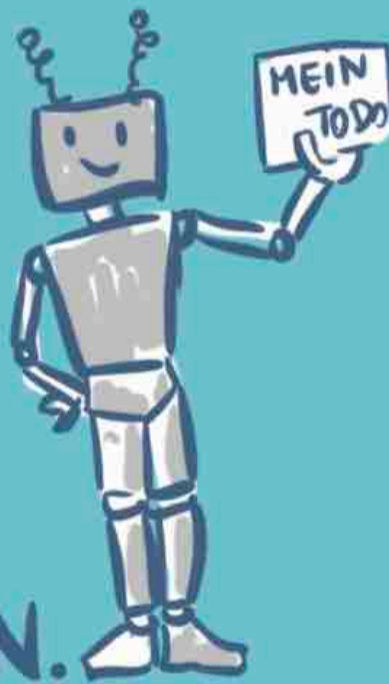


SACHLICHE BEGRÜNDUNG

Meistens gäbe es auch weniger diskriminierende Vorgehensweisen mit gleichem IMPACT

Definition KI

FÄHIGKEITEN VON COMPUTERN, DIE FRÜHER NUR VON MENSCHEN BEHERRSCHT WURDEN.



? Wie lernt ein Computer?

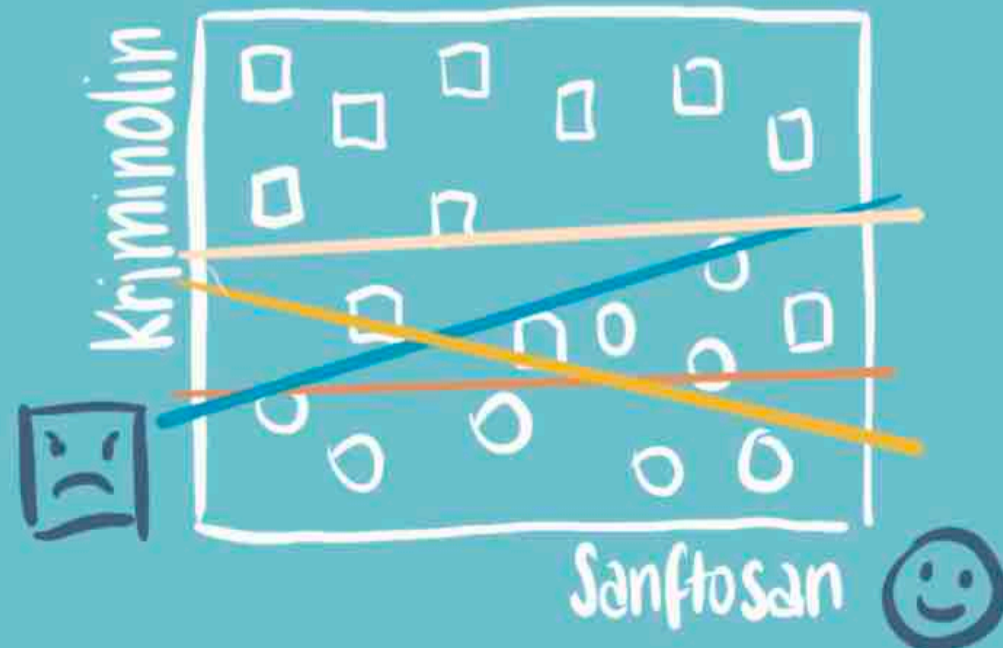
BEST-MÖGLICHE MUSTER

- 1
 - 2
 - 3
 - 4
- KRITERIEN FESTLEGEN

ENTSCHEIDUNGSREGELN

AUFSTELLEN & BEFOLGEN

1.2.3.



14.3.23
Berlin



BEOBACHTUNGEN

1 Die genaue Implementierung kann die Bewertung von Einzelfällen stark beeinflussen.

AB INS GEFÄNGNIS!



Wie schlimm ist es, wenn ein Fehler gemacht wird?

Ein UNSCHULDIGER MENSCH MUSS INS GEFÄNGNIS

FÜR DIE GERECHTE BEHANDLUNG NÖTIG:

→ FALL 1 FAKTOREN A,B,E,G FALL 2 FAKTOREN C,D,I,G

unterschiedliche Faktoren pro FALL!

NEUE FAKTOREN einfach hinnehmen

beliebig viele Faktoren beachten überfordert die Rechner



Gesellschaftliche WERTE

Was gemessen werden sollte:

zB Fairness!

→ NOCH NICHT MÖGLICH!

Der Computer (Algorithmus) muss ALLE Möglichkeiten kennen

Der Computer muss GENAU wissen, was er messen soll



Das können schnell 100 - 1000 Parameter sein

Der Mensch kann ANALOGIEN BILDEN!
"Das ist so, wie wenn..."

BEISPIEL COMPASS



Doppelt so hohe FALSCHPOSITIVRATE von Schwarzen wie von Weissen.

DISPARATE IMPACT MESSUNG

TP TN FP FN

Richtige & Falsche Antworten
↓
Qualitätsmaße

Operationalisieren

IST SCHWIERIG für alle Gerechtigkeitsaspekte



Kann KI "Gerechtigkeit?"



14.3.23
Berlin



BEOBACHTUNGEN 2

Optimierungsziele enthalten immer auch
☹☹☹ ⇒ Gesellschaftliche WERTE



BEOBACHTUNGEN 4

GERECHTIGKEITSAKTE
↳ NICHT ALLE
WIDERSPRUCHSFREI
OPERATIONAUSIEREN



BEOBACHTUNGEN 3



Computer brauchen
OPERATIONAUSIERUNGEN
für ALLES!

HIER DIESE + DA ANDERE FAKTOREN

SO FUNKTIONIEREN
COMPUTER NICHT

ABER DAS
BRÄUCHTE



GERECHTE
BEHANDLUNG!

Fairnessmaße
enthalten (sich
teilweise widersprechende)
Gesellschaftliche
WERTE.



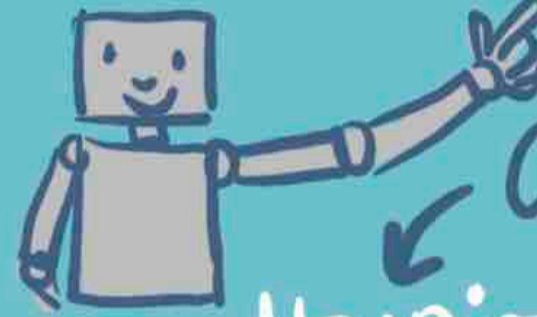
BEOBACHTUNGEN 5

WAS
WÄRE
WENN
ES
GENÜGEN
DATEN
GÄBE?

KEINE
Lösung
BEKANNT!



OHNE
GARANTIE



HEURISTIK

ZUR GUTEN LÖSUNG

MASCHINE HAT
Optimal ENTSCHEIDEN

DAS BESTE??

KEINE
ALGORITHMEN

DISPARATE IMPACT
ABER OPTIMALITÄT



Wie Corona-
Test - da
spielt Hautfarbe
keine Rolle.



SICH WIDERSPRECHENDE
FAIRNESS-MAßE!

Kann
KI



Gerechtigkeit?

DISPARATE IMPACT MESSUNG

- EXTREM EINFACH
- LIEGT SCHON DIGITAL VOR

BEISPIEL COMPAS:

GLEICH HOHE RÜCKFÄLLIGKEITSRATEN
DOPPELT HOHE FALSCHPOSITIVRATEN
SCHWARZE & WEISSE MENSCHEN

14.3.23
Berlin



BEOBACHTUNGEN 6

MASCHINELLE LERNVERFAHREN SIND BIS AUF 2-3 AUSNAHMEN HEURISTIKEN

KEINE ALGORITHMEN

URTEILE BRAUCHEN BEGRÜNDUNG
GLEICHES GLEICH BEHANDELN

Dieser ♂ ist wie dieser ♂, kann das aber nicht begründen!!!

↓ KI kann sagen welches Wort folgt, aber NICHT, was es bedeutet!

REBOOTING AI von Gary Marcus & Ernest Davis



BEOBACHTUNGEN 7



MASCHINELLE KATEGORISIERUNGEN KÖNNEN GGÜ MENSCHEN NICHT BEGRÜNDET WERDEN!

ZUSAMMENFASSUNG

KI kann heute KEINE GERECHTIGKEIT LEISTEN,

NICHT VOLST. OPERATIONAUSIERBAR & VERSUCHE ZUR OPERATIONAUSIERBARKEIT SICH WIDERSPRECHEN



* Weil

* Weil SIE AUS BEISPIELEN NICHT HINREICHEND LERNEN?

* Weil es selbst bei perfekten Rahmenbedingungen **KEINE OPTIMALITÄT** gibt

* Weil wir aus statistischen Modellen **KEINE VERBÄUßT** FÜR MENSCHLICHE VERGESELSCHAFTUNG in Ähnlichkeitsbedingten Gruppen bekommen.



14.3.23
Berlin

KEYNOTE

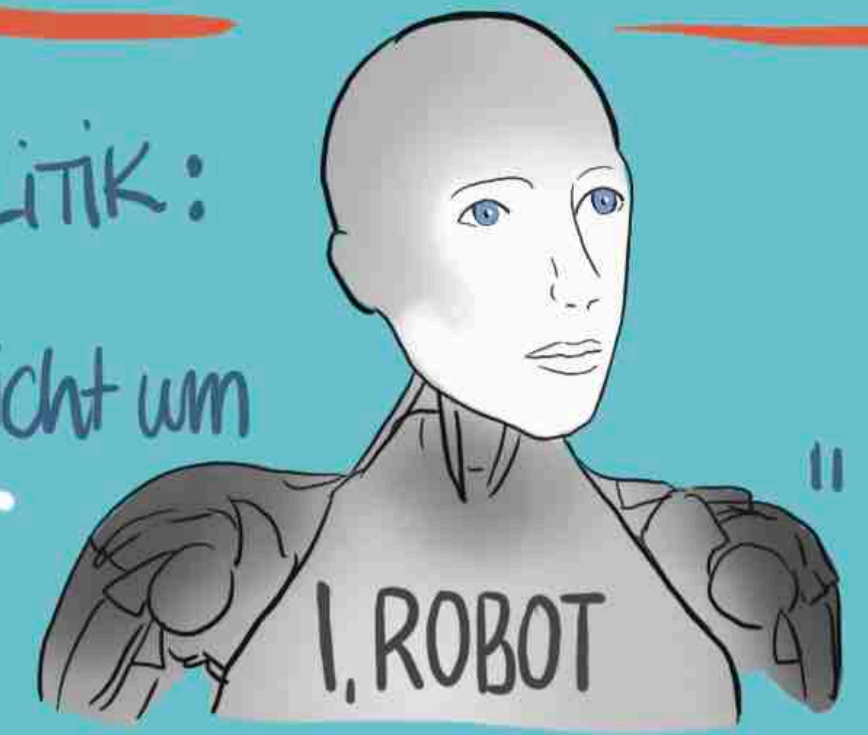
Saskia Esken

VOR NICHT ALLZU LANGER ZEIT WAR KI

POPKULTUR

2017 POLITIK:

"Hier geht es nicht um"



WICHTIG:

Differenziertes

Bild VON KI

WAS IST DER NUTZEN?
TECHNOLOGISCHER FORTSCHRITT



KI

ändert

- Gesellschaft
 - Menschen
 - Unternehmen
- grundsätzlich!

TRANSFORMATION

GESELLSCHAFTLICHER FORTSCHRITT



SCHWIERIGKEITEN mit der Veröffentlichung



Angste

&



Herausforderungen

AUFKLÄRUNG WICHTIG!



kann zu Diskriminierung führen
→ Hautfarbe
→ Geschlecht

KI kann nur aus REALITÄT lernen

VERBESSERUNG DES BILDUNGSSYSTEMS

Chat GPT

Angst an Schulen

Aufgeklärter Umgang

QUALITÄT & TRANSPARENZ

auch bei Datentypen, die nicht für die Auswertung relevant sind

DU WÜRDST VON KI AUSSORTIERT



ES BRAUCHT KLAARE REGELN

WO KI + WO NICHT

MINORITY REPORT SOLLTE FICTION BLEIBEN!



14.3.23 Berlin

KI

WIR müssen den RAHMEN SCHAFFEN!

EU KI VER-ORDNUNG

AUF DEM RICHTIGEN WEG

WICHTIG

MENSCHEN → KI

Fähigkeiten → zB

Limitierung

KI bei

PRÜFUNGSBEWERTUNGEN

WICHTIG

AI Literacy

KI EINSATZ KOMPETENT GESTALTEN

Eigenes Urteil IST WICHTIG



LERN-ERFOLG

URHEBERSCHAFT KLÄREN



DIE ZUKUNFT DER SCHULE



SCHULE BEREITET AUFS LEBEN VOR

BILDUNGSSYSTEM-KI

Chance & Angst

OPEN AI CHAT GPT

werden TREIBER →

der ZUKUNFT der SCHULE

WIE WIR lernen lehren prüfen

21st CENTURY SKILLS

ÜBER DEN TELLERRAND BÜCKEN

ANALOGIEN BILDEN

TEAMARBEIT

LÖSUNGEN FINDEN

WEGE ZUR ERKENNTNIS

SPRECHEN

LESEN

SCHREIBEN

FOKUS AUF **BASIS-KOMPETENZEN**

NIEDERSCHWELIGER ZUGANG ZU DIGITALISIERUNG

INVERT CLASSROOM



14.3.23 Berlin

Learning from Uncle Sam
 WAS UNS DIE EINFÜHRUNG VON ADM-SYSTEMEN IN DEN US-STAATEN lehrt
 Wenzelburger



Empirische ERGEBNISSE

UNTERSUCHUNGSHAFT

↑ FOKUS

- 1 Analyse ADM - sehr unterschiedlich
- 2 Quantitative Analyse - Parteien - Muster von Staat
- 3 Qualitativ - Warum abgeschafft?
- 4 Qualitativ - First mover

★ SYSTEM, das immer gleich entscheidet

WARUM HABEN ES MANCHE STAATEN & ANDERE NICHT?

⚡ Überlastung der Gefängnisse



KI sagt: „Der wird nicht rückfällig“

OKAY

⚡ KOSTEN

Je höher die Schulden desto eher KI im Gerichtssaal

VIRGINIA 1. STAAT
 ○ UM GEFÄNGNISSE ZU ENTLASTEN
 ○ U-HAFT - die Regel wg. HOHER KAUTION
MASSIV DISKRIMINIEREND!

DER POLITISCHE PROZESS & SEINE DYNAMIK IN DEN USA



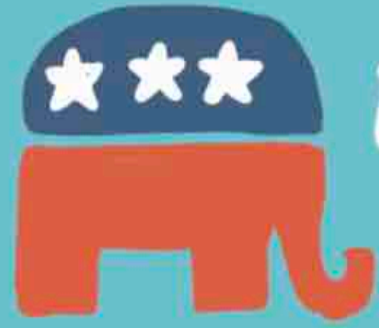
- LANGE GESCHICHTE
- STAATEN SIND DIVERS
- ANFANGS EUPHORIE DANN SKEPSIS
- JETZT WIDERSTAND



Was hat das AUES mit Politik zutun?

REPUBLIKANER

DEMOKRATEN



LAW & ORDER

ALLE EINSPERREN

KI NUTZEN

wenn fiskale Gründe



BÜRGERRECHTS-PERSPEKTIVE

Wenn DEMOKRATISCHE EXEKUTIVE... DESTO EHER KI

Civil rights movement!

KI NUTZEN für effektive Verwaltung

GRÜNDE FÜR RÜCKNAHME VON KI

zB UNZUVERLÄSSIG
INTRANSPARENZ



Problem Überfüllung BALD AUCH IN DEUTSCHLAND

Gefängnisse OHNE HUMAN RIGHTS



TRANSFER
→ ÜBER UK → EU
NICHT 1:1 ÜBERTRAGBAR

TROTZDEM BÜCK IN DIE USA WICHTIG → ABER den KONTEXT NICHT AUS den AUGEN verlieren.

Wenn sie zusammen kämen, wäre ÜBERPARTeilICHES VORGEHEN möglich

SCHLUSSFOLGERUNG

KOSTEN SIND AUCH IN Deutschland EIN WICHTIGES THEMA

Verfassungs-RECHT & Ethik

ACHTUNG DISKRIMINIERUNG

NORM-

DEBATE MUSS ANDERS GE-FÜHRT WERDEN!



14.3.23
Berlin

PROBLEM:

MASSIVE VERURTEILUNGSRATE

UNTERSUCHUNGSHAFTRATE ohne HUMAN RIGHTS

GEFÄNGNISSE



GERECHTIGKEIT
VS
DISKRIMINIERUNG

SCHLUSSFOLGERUNG:

➔ ÜBERTRAGBARKEIT NICHT 1:1 MÖGLICH!

➔ WEG 1 USA → 2. UK → 3. EUROPA & REST

➔ EFFIZIENZGEDANKEN

↳ € KOSTEN!!! → AUCH IN DEUTSCHLAND
WICHTIGES THEMA

↳ USA NICHT VERGLEICHBAR

➔ ? WELCHE KRITERIEN SIND FÜR EUROPA
RELEVANT



👁️ BUCK IN DIE USA IST OKAY

KONTEXT WICHTIG

Die DEBATE MUSS
ANDERS GEFÜHRT WERDEN

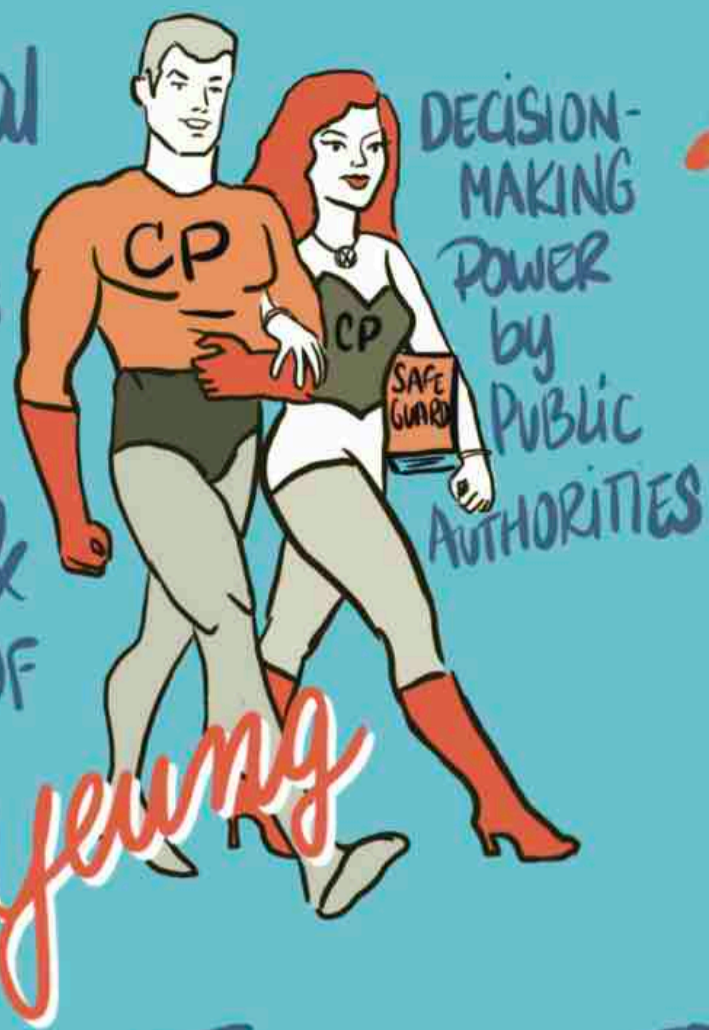


VERFASSUNGSRECHT

& ETHIK
& EUROPA



Constitutional principles as safe guards against the algorithmic injustice & the abuse of



3 TOOLS IN USE
- LONDON GANGS MATRIX
- DURHAM POLICE HART TOOL
- DUTCH SYRI WELFARE FRAUD TOOL

CONSTITUTIONAL LENS

METHODS IN THE PROJECT
*A+B TOGETHER

WHAT WE LOOK AT

- RIGHTS-CRITICAL DECISIONS
- DISCRETIONARY POWER OF CRIMINAL JUSTICE AUTHORITIES
- ADM Tools accuracy

AIM OF THE PROJECT

To CRITICALLY INVESTIGATE



THE ROLE OF CONSTITUTIONAL PRINCIPLES IN SHAPING & CONSTRAINING

HOW ADM TOOLS

EMPLOYED BY CRIMINAL JUSTICE DEVELOPED & USED



VIEWED THROUGH THE CONSTITUTIONAL LENS'

THIS PROJECT DIGS DEEPER

& examines the INTERSECTION BETWEEN CONSTITUTIONAL PRINCIPLES

& the developer's TECHNICAL CHOICES

- CP The rule of law
- CP Respect for human rights
- CP Right to due process

4 STEPS

- 1 IDENTIFY SUITABLE DATA TO SERVE AS GROUND TRUTH → MODEL BUILDING
- 2 CREATE AN ALGORITHMIC MODEL [...] TO GENERATE PREDICTIONS CONCERNING FUTURE OUTCOMES ON UNSEEN DATA
- 3 TEST & VALIDATE THE MODEL
- 4 ENCODE THE MODEL

→ CONVENIENT DIGITAL TOOL



14.3.23
Berlin

TRANSPARENCY FAIRNESS ACCOUNTABILITY

STEP 1 DATA

↓ USE OF INADEQUATE "PROXY" DATA TO TRAIN THE MODEL → UNLAWFUL?

↓ UNLAWFUL DATA COLLECTION & PROCESSING

DATA USE TO TRAIN ML LAW MODEL MUST CONFORM WITH LEGAL REQUIREMENTS

STEP 2

CREATE A MODEL

↓ PREDICTIONS ARE UNAVOIDABLE

↓ ERRORS = failure to generate an accurate prediction on unseen data

MISCLASSIFY PROBLEM DANGEROUS AS CHARMLESS IS WORSE THAN VICE VERSA.

STEP 3

TEST & VALIDATE

↓ ASSESSMENT IS PROBLEMATIC BECAUSE OF DATASETS

BASIC CONSTITUTIONAL PRINCIPLES & LEGAL REQUIREMENTS

TRUE REAL WORLD RELATIONSHIPS BETWEEN MODEL FEATURES & PHENOMENON IT PURPORTS TO PREDICT

STEP 4

DESIGN & UI

Once a model is ready for use a user friendly digital interface is designed → APP for SMARTPHONES

MUST ENSURE

DUE PROCESS RIGHTS:

→ RIGHT TO REASONS

→ RIGHT TO CONTEST

→ RIGHT TO UNBIASED TRIBUNAL



14.3.23
Berlin

2 CORE PROJECT FINDINGS



IMPORTANT
CONSTITUTIONAL PRINCIPLES AND LEGAL DUTIES ARE
IMPLICATED IN TECHNICAL CHOICES MADE
OF THE ALGORITHMIC MODEL BUILDING PROCESS
YET THOSE PRINCIPLES & DUTIES ARE
LARGELY OVERLOOKED!

1



2 → THESE TOOLS
→ ENHANCE THE RISK THAT DECISIONS BASED ON THEIR
PREDICTIONS MAY BE UNJUST AND/OR MAY OTHERWISE ENTAIL
THE UNLAWFUL EXERCISE OF DECISION-MAKING AUTHORITY.



RECOMMENDATIONS

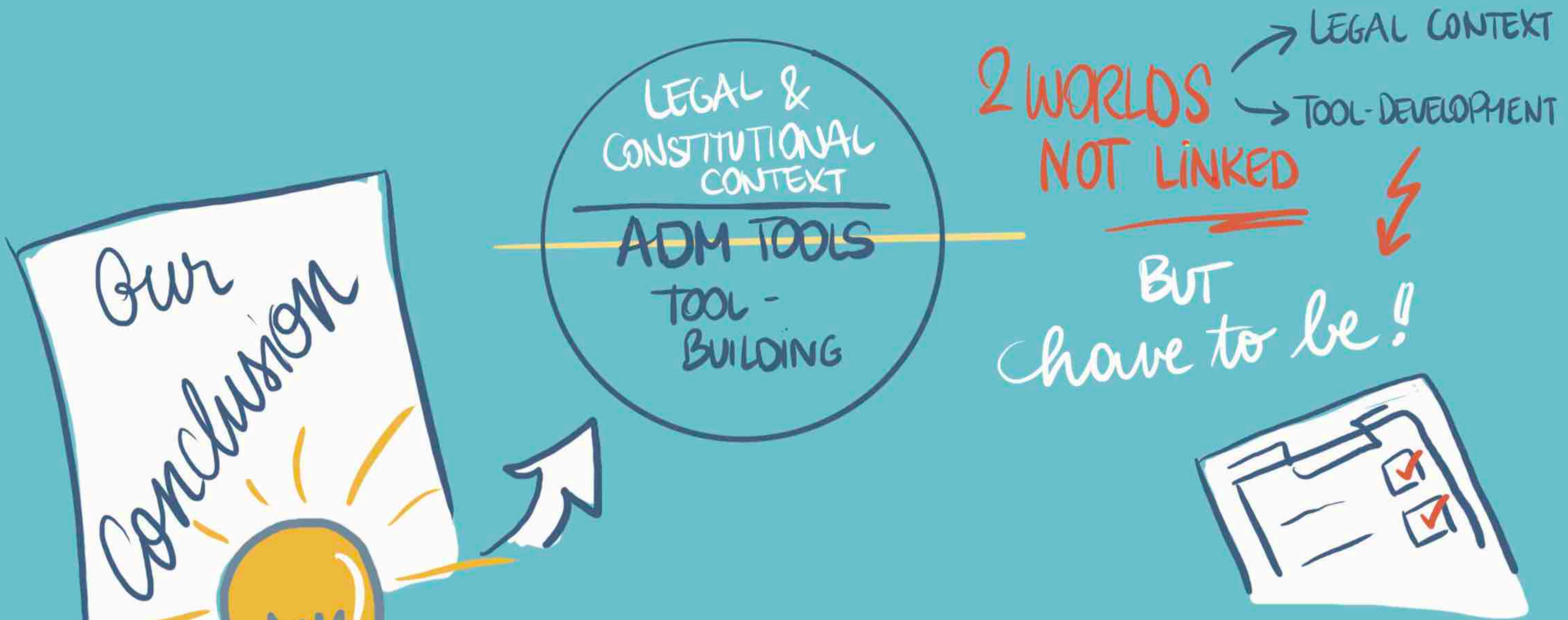
1 → Legal & contextual considerations must be duly considered during
algorithmic model-building

2 → Systematic safeguards for ADM tools are essential!

3 → **Red lines**: if ADM tools are not demonstrably
compliant with public law principles and legal
requirements, they should **NOT** be used.



14.3.23
Berlin



SYSTEMATIC INSTITUTIONAL OVERSIGHT MECHANISMS ARE NEEDED TO ENSURE THAT THESE TOOLS DO NOT

- ⚡ VIOLATE INDIVIDUAL RIGHTS
- ⚡ PRODUCE INJUSTICE
- ⚡ ERODE PUBLIC TRUST IN THE INTEGRITY OF THE CRIMINAL JUSTICE SYSTEM



14.3.23
Berlin

ARCHITEKTONISCHE PERSPEKTIVEN



JUSTIZ
POLIZEI
BILDUNG

Welche Unterstützungssysteme gibt es?
Perspektive im Rechtssystem ändern

Gegenwart der Technologie

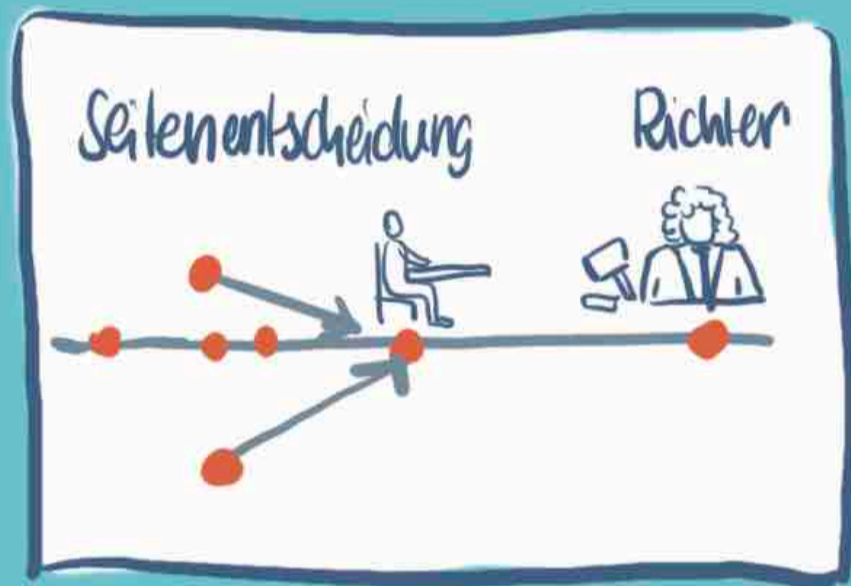


JE MEHR TECHNOLOGIE VON DEN PROBLEMEN LÖST, DESTO WENIGER NUTZEN IM URSPR. SINNE

Schutz IM RECHT

Eine methodische Skizze

Möglicherweise übersehen wir etwas, wenn wir nicht alles, was im Vorfeld geschehen ist, berücksichtigen



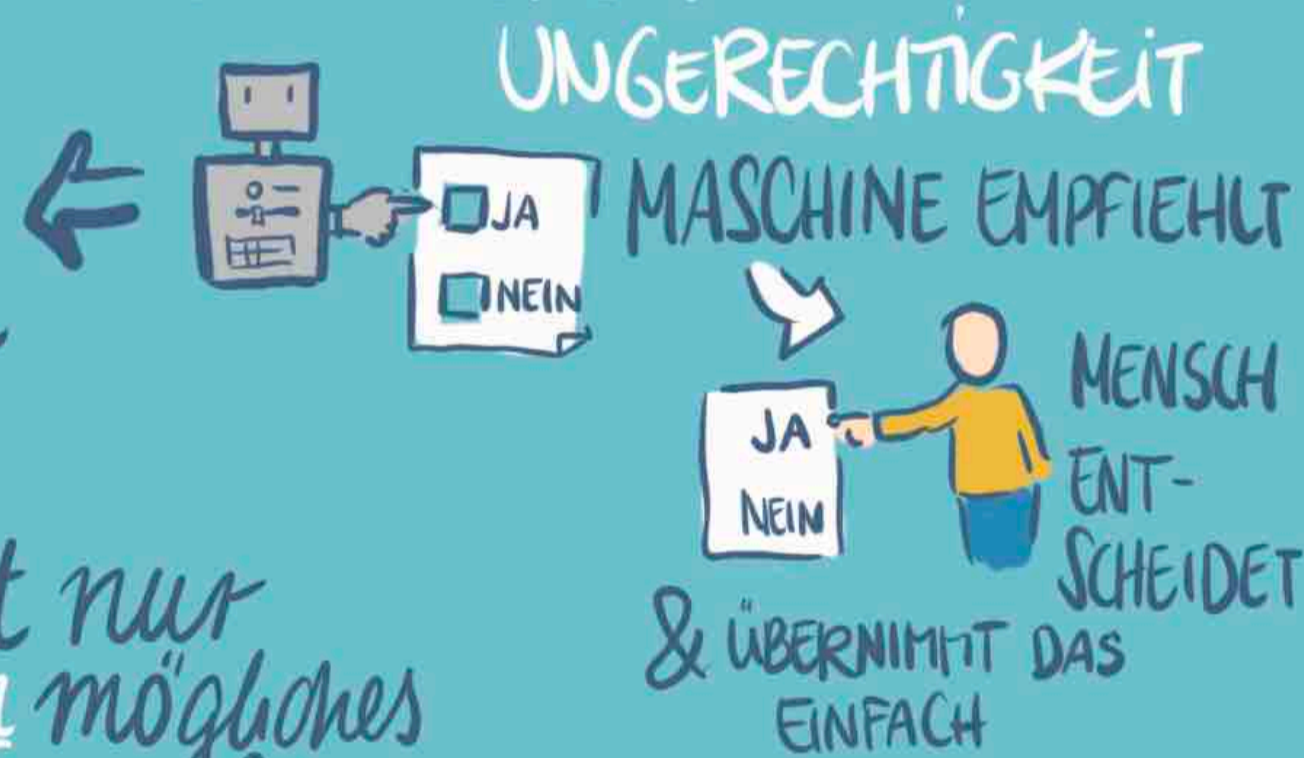
MITTLERER STRANG
TECHNIK

(Vorentscheidung beim Training
Qualitätsstandards)

Entscheidungen im RECHT



REPRODUZIERBARE UNGERECHTIGKEIT



Es gibt nur ein mögliches Ergebnis.

Gibt es AUDITS?

Interface
TECHNIK - MENSCH der entscheidet

WERTE-ENTSCHEIDUNG
IST KEINE TECHNISCHE ENTSCHEIDUNG

HERAUSFORDERUNG

Eine Methode finden, damit NICHTS verloren geht



14.3.23
Berlin

HAFTUNG
SPIELEN eine ROLLE
ANREIZE

GROSSER KOMPLEXER BEREICH
Müssen ins Spektrum
der Entscheidungen
einbezogen werden

ARCHITEKTONISCHE PERSPEKTIVE
& RECHT

Fruchtbar in 2 Richtungen

INTERDISZIPLINÄRE
ANSÄTZE

Rechts-
SCHUTZ PERSPEKTIVE
Was muss gewusst
werden?



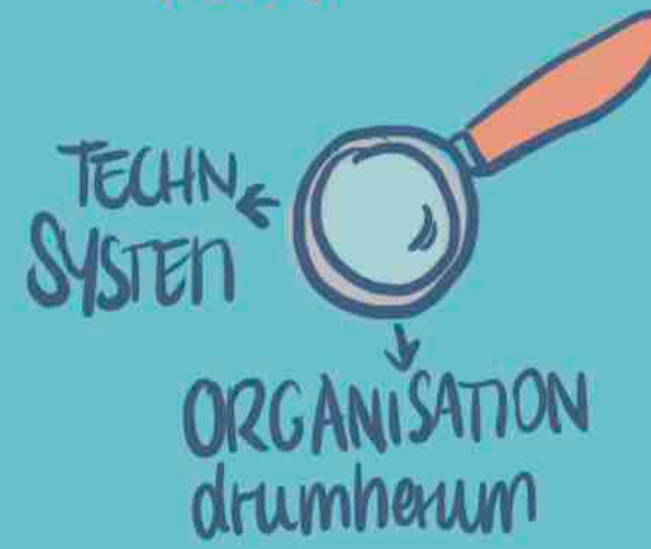
ZB.
Dokumentationspflicht
Evaluationspflicht
Qualitätsstandards
...

Welche
Plattformen
↓
Sozio-technische
Ökosysteme

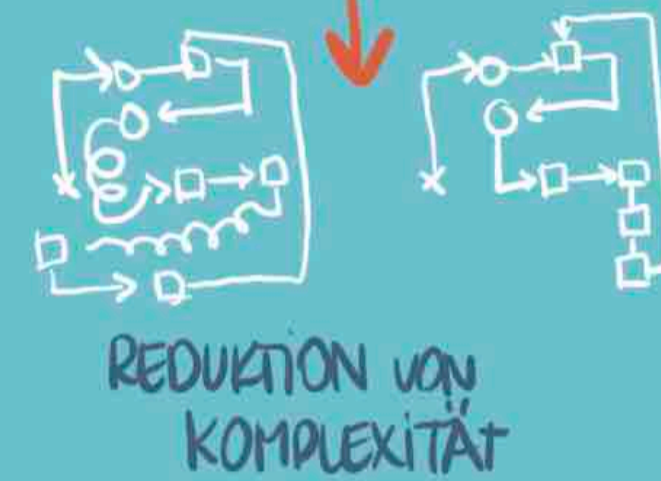


ZB. FLUGZEUGE
können autonom fliegen
aber das ist für die Fluggesellschaften
noch nicht tragbar.

UNABHÄNGIG
VON DER
TECHNIK



Perspektiven
der Informatik
verstehen
↓
zB Gängige
Systeme der
Informatik



Wir bauen Bilder ein
Sehr hilfreiches Tool
um interdisziplinär
zusammen zu arbeiten
zB Informatiker & Juristen
→ Wenn man
Technologien
in Entscheidung
integrieren
will

Werbung für interdisziplinäre
Exploration



14.3.23
Berlin



WIE REAGIEREN RICHTER*INNEN AUF KI?

Psychologische Sicht:



BSP. GOOGLE MAPS → Routenempfehlung

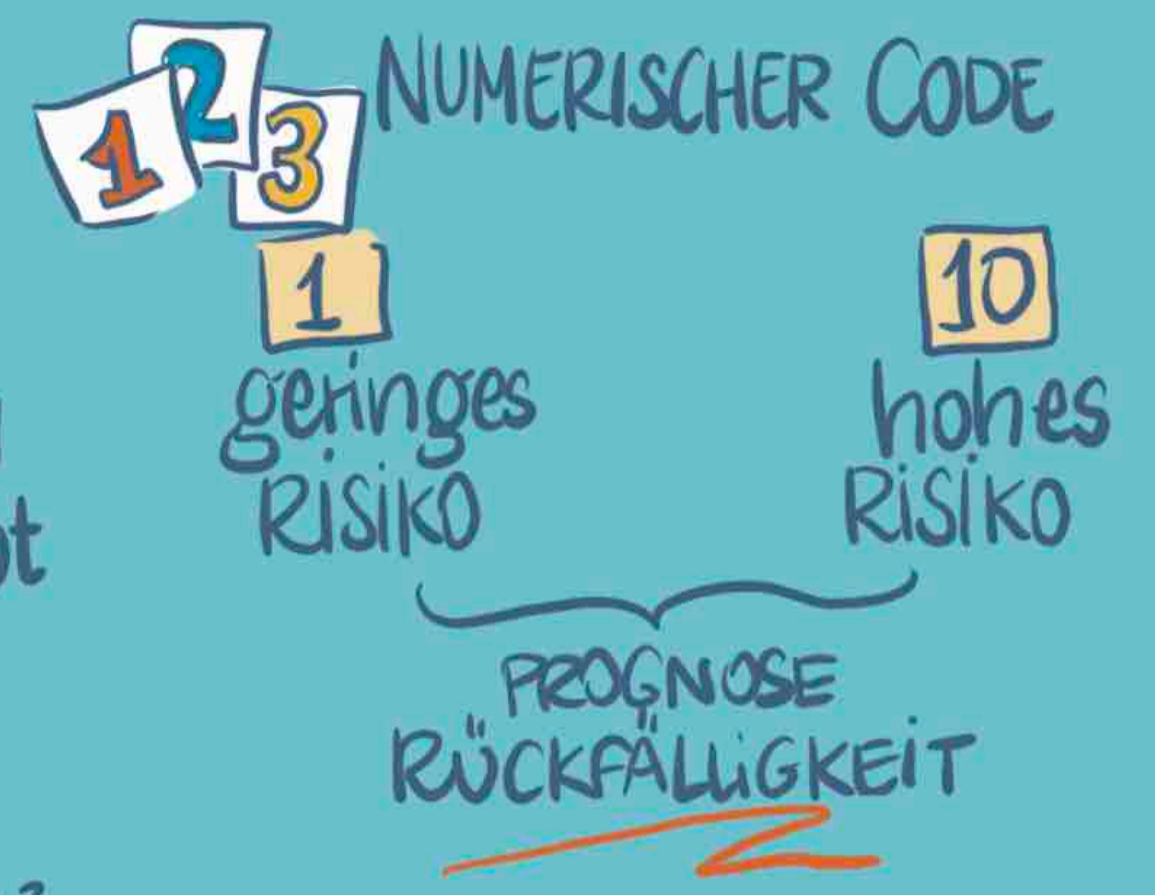


PROBLEME

- Empirische Forschung
- LAIENBETEILIGUNG kritisch reflektiert
- Allgemeine Bevölkerung statt RECHTSEXPERTEN

USA: COMPAS ALGORITHMUS
RICHTERLICHE ENTSCHEIDUNGEN
↳ BEWÄHRUNGSFRAGE

1. URTEIL auch vorläufig bleibt präsent
1. Urteil auch vorläufig bleibt präsent



REIHENFOLGE
FRÜH PRÄSENTIERT

ALGORITHMEN MENSCH

ÜBERSCHÄTZUNG DER KOMPETENZ & LEISTUNG VON KI

WILL KEINE KI
KI KANN MEINE ENTSCHEIDUNG ZU BEEINFLUSSEN

Verankerungsheuristik
Richter werden durch numerische Codes beeinflusst

z.B. Empfehlung 10 Jahre

WÄRE ANDERS OHNE ZAHLEN ??

Forschungsfrage: Wie wirken sich algorithmische Entscheidungen auf menschliche Entscheidung aus.



14.3.23
Berlin

DAS DESIGN DER STUDIE

1.  Keine Empfehlung

2.  Supervision Richter Empfehlung

3.  Algorithmus Empfehlung

Numerische Risikobewertung

1 → 10

3 FELD-STUDIEN

1. Allgemeinbevölkerung } wurden von Empf. beeinflusst
2. Rechtsreferendare } beeinflusst
3. Richter } weniger scharfe Urteile als 1.

10 Fiktive Kriminalfälle → ? Bewährung



3 VARIANTEN IN DER KOMPLEXITÄT

1. PRO ANGEKLAGTE
2. CONTRA ANGEKLAGTE
3. KOMPLEXE FÄLLE (in der Realität am häufigsten)

EFFEKTE

- Wie bewerten Sie das Risiko? 1-10
- zur Bewährung JA X NEIN □
- wie sicher sind Sie, dass das die RICHTIGE ENTSCHEIDUNG?



ERGEBNIS zeigt: LAIEN ↓ EXPERTEN } DEUTLICHE UNTERSCHIEDE



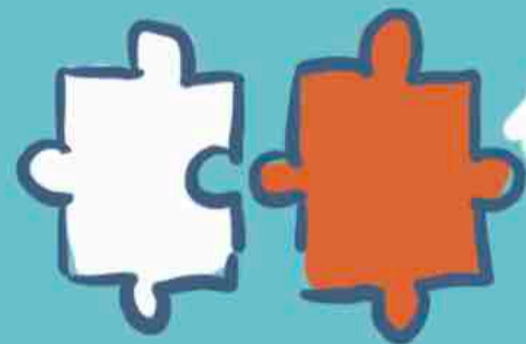
14.3.23
Berlin



FÜR
ZUKÜNFTIGE
FORSCHUNG



RECHTSEXPERTEN ALS PROBANDEN
WICHTIG: ENGE ZUSAMMENARBEIT



IMPLIKATION: In die Ausbildung zukünftiger Richter



VERTRAUEN IN DIE EIGENE ENTSCHEIDUNG HABEN
OHNE BEEINFLUSSUNG



SCHWIERIGKEIT ERFAHRENE RICHTER FINDEN

LAIEN MIT DUNNING-KRUGER-EFFEKT
& CSI-WISSEN!



14.3.23
Berlin

Chat GPT
GARY MARCUS FOR AI
 QUEST

TECH-OPTIMISTIC MEDIA
 RAPIDLY CHANGED ITS TUNE
RADICAL SHIFT

WHERE DOES ALL THIS LEAVE US?
 AI DOESN'T KNOW WHAT WORDS MEAN → INTELLIGENT PEOPLE WOULD FACT CHECK
 NEEDS DEBUGGING

IT TAKES A VILLAGE → TO RAISE AN AI
 THE AGI DEBATE

LOOKS INTELLIGENT → BUT IT IS NOT
 WE NEED: **POLICY & REGULATIONS**
 REGARDING JUSTICE → ? **IMPORTANT QUESTION**
 ARE THE AI'S LIMITATIONS? PEOPLE AWARE OF?

BRIEF HISTORY OF AI

1967 MARVIN MINSKY HUBRU
 Problems will be solved!

2012 GPUs were GOOD
 NOV 2012 changing THE WORLD
 SEE THE PROMISE IN

DEEP LEARNING
 NEW YORKER 1st CRITICAL VOICE I

2022 DEEP LEARNING IS HITTING THE WALL

MORE PROGRESS

DALE SOLVED AI PROBLEMS

PLOT

TWIST
 PROMISE 1
 DRIVERLESS CARS
 SCALED BACK

2 LARGE LANGUAGE MODELS ARE WORRIED ABOUT
 PEOPLE ABUSED THE SYSTEM

3 SKEPTICS CHANGED THEIR TUNE!
 Sam Altman, Elon Musk

2023 has been WILD

WAKE-UP CALL
 CHAT GPT IS NOT AI

NUMERATOR DENUMERATOR
 I know 3 jokes! BUT OUT OF HOW MANY JOKES IN THE WORLD?

Even a taller ladder won't get us to the moon

SUGGESTIONS:
 LOOK TOWARDS HUMAN COGNITION

ABSTRACTION
 Humans can do AI still struggles

REASONING ABOUT THE WORLD

COMPOSITION
 HUMANS UNDERSTAND eg SIMILAR PATTERN

FACTUALITY REliable? → ELON MUSK DIED IN A CAR ACCIDENT
 2018 → NEWS

TAKE IT WITH CAUTION!

