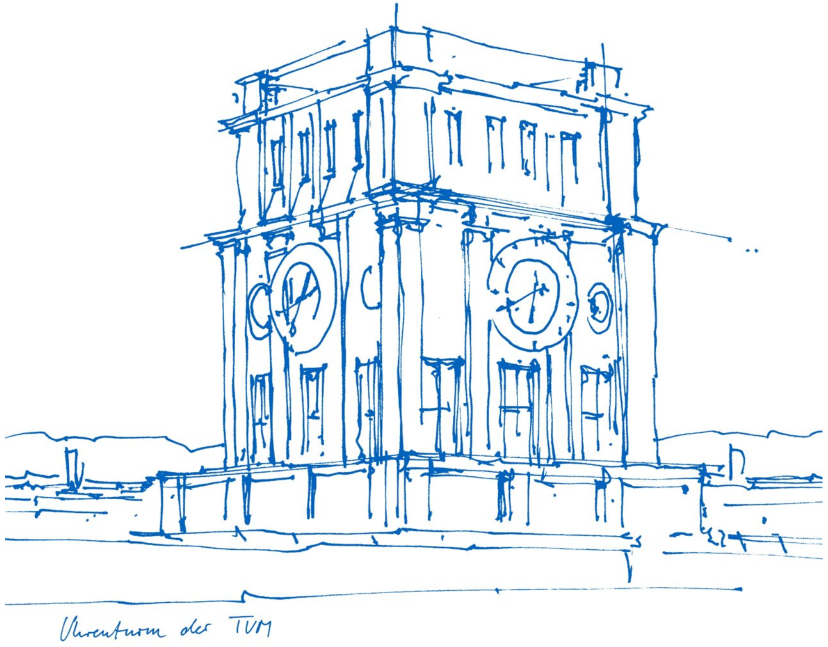
De-Generative Democracy

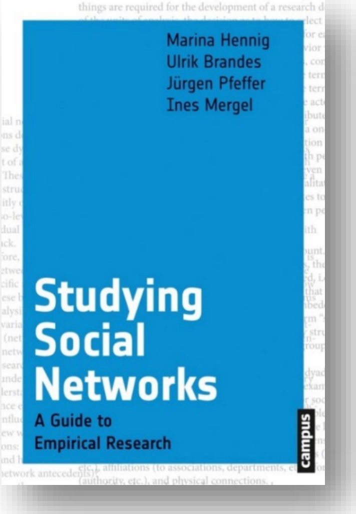
Prof. Dr. Jürgen Pfeffer
Computational Social Science
Technical University of Munich
School of Social Sciences & Technology
@JurgenPfeffer

TU WIEN

CAIML Center for Artificial Intelligence and Machine Learning

3rd ACM Europe Digital Humanism Summer School

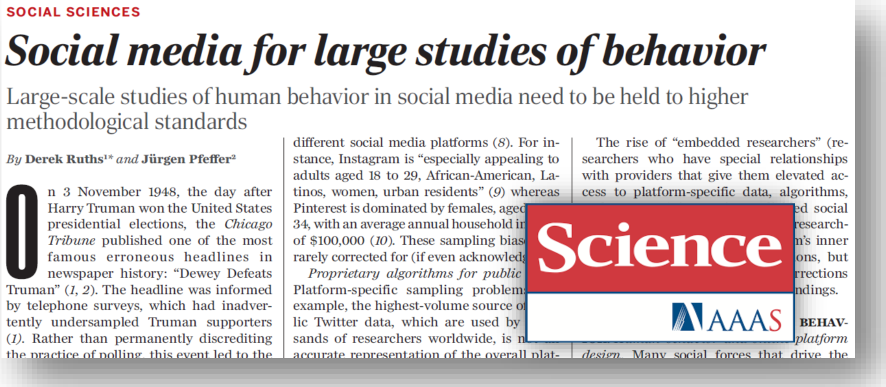




Marina Hennig
Ulrik Brandes
Jürgen Pfeffer
Ines Mergel

Studying
Social
Networks

A Guide to
Empirical Research



SOCIAL SCIENCES

Social media for large studies of behavior

Large-scale studies of human behavior in social media need to be held to higher methodological standards

By **Derek Ruths**[1*] *and* **Jürgen Pfeffer**[2]

On 3 November 1948, the day after Harry Truman won the United States presidential elections, the *Chicago Tribune* published one of the most famous erroneous headlines in newspaper history: “Dewey Defeats Truman” (*1*, *2*). The headline was informed by telephone surveys, which had inadvertently undersampled Truman supporters (*1*). Rather than permanently discrediting the practice of polling, this event led to the

different social media platforms (*8*). For instance, Instagram is “especially appealing to adults aged 18 to 29, African-American, Latinos, women, urban residents” (*9*) whereas
Pinterest is dominated by females, aged
34, with an average annual household i
of $100,000 (*10*). These sampling bias
rarely corrected for (if even acknowledg
Proprietary algorithms for public
Platform-specific sampling problem
example, the highest-volume source o
lic Twitter data, which are used by
sands of researchers worldwide, is n
accurate representation of the overall plat-

The rise of “embedded researchers” (researchers who have special relationships with providers that give them elevated access to platform-specific data, algorithms,
ed social
research-
m’s inner
ons, but
rrections
ndings.

BEHAV-
platform
design. Many social forces that drive the

Vita

- BA Computer Science, Ph.D. Business Informatics, Vienna University of Technology
- 10 years consulting and non-university research
- 2012-2015: Assistant Professor @ Carnegie Mellon University in Pittsburgh
- 2016-: Associate Professor of Computational Social Science @ TU Munich

Research focus

- Computational analysis of organizations and societies
- Special emphasis on large-scale systems, e.g., social media
- Methodological and algorithmic challenges
- Online-offline spillovers
- AI & Democracy

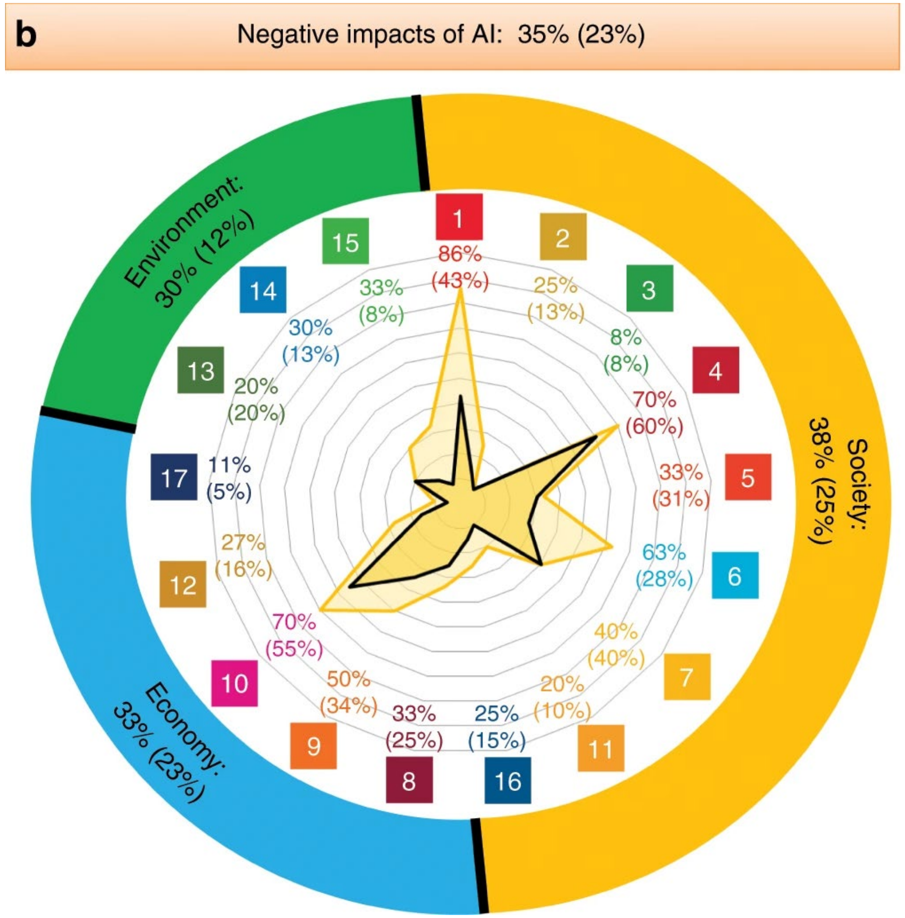
Is AI Good „Automatically“?

“The most open system theoretically imaginable reveals perfectly the predicating inequities of the wider environment in which it is situated.”

Jamie King (2006, S. 53)

Vinuesa et al. (2021):

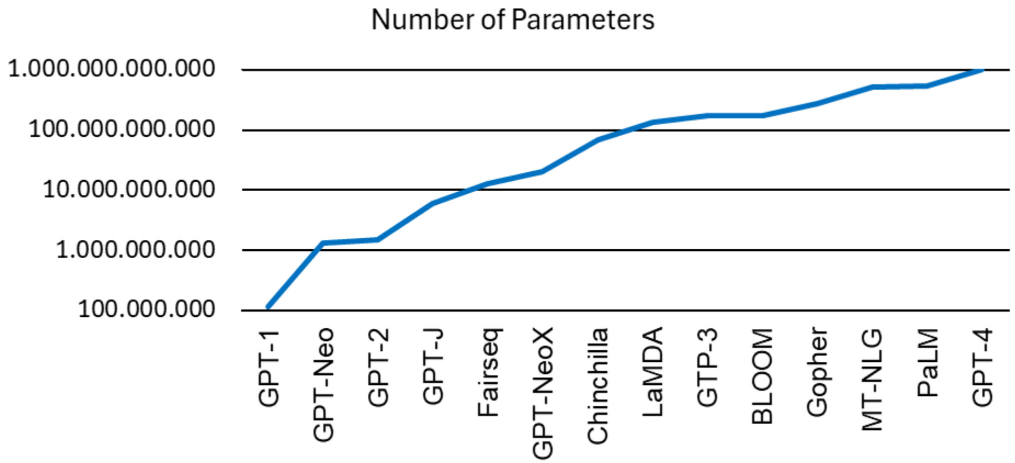
- AI can help to accomplish vast majority of targets across all Sustainable Development Goals (134)
- It may inhibit more than a third of the targets (59)
- Most notably: Increased societal and economic inequalities



Fundamental Aspects of Western Societies

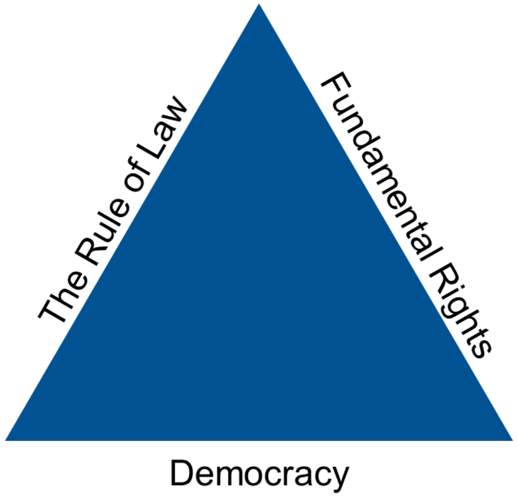
World of Engineering:

- Thousands, millions, billions, trillions of variables in parameter spaces.
- Optimization in a single, global, deterministic goodness measure.


Number of Parameters

Western societies:

- Centuries-long power struggle
- Sometimes incompatible, sometimes complementary principles
- Small number of principles
- Constant balancing act
- The process is the solution


The Rule of Law
Fundamental Rights
Democracy

Fundamental Aspects of Western Societies

Fundamental rights

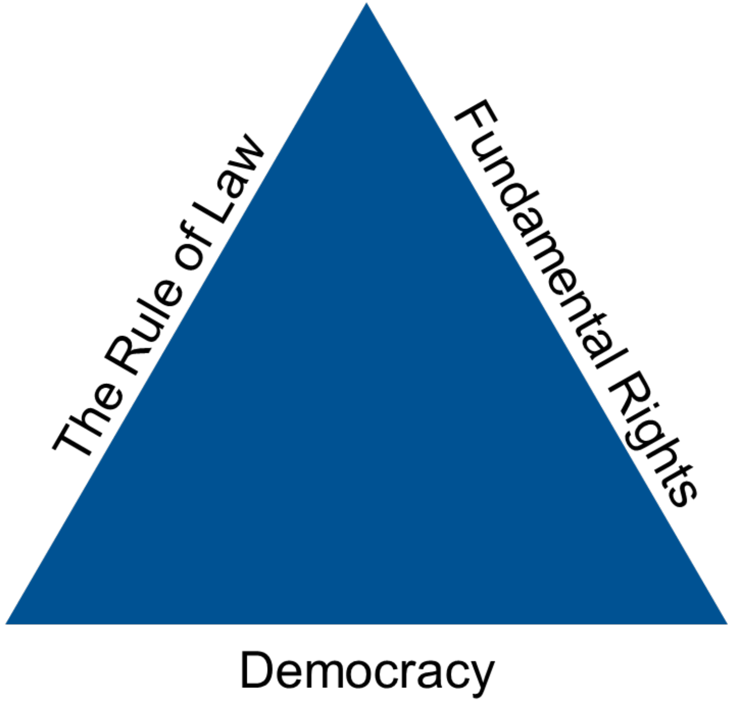
- Human dignity
- Freedom and self-determination
- Equality and equity
- Humans as an end in themselves

Democracy

- Participation
- Deliberation
- Ability and willingness to agree on common action
- Self-determination of groups/societies

The Rule of Law

- Published, intelligible rules
- Access to justice
- Fair trial and comprehensible, contestable rulings
- Separation of power with checks and balances



Fundamental Rights

Algorithms …

- select what we see on the internet
- recommend where to go and whom to meet

Generative AI:

- Growing influence of LLMs on digital communication
- Soon generate large parts of texts and images

Omnipresence and concentration at a handful of companies

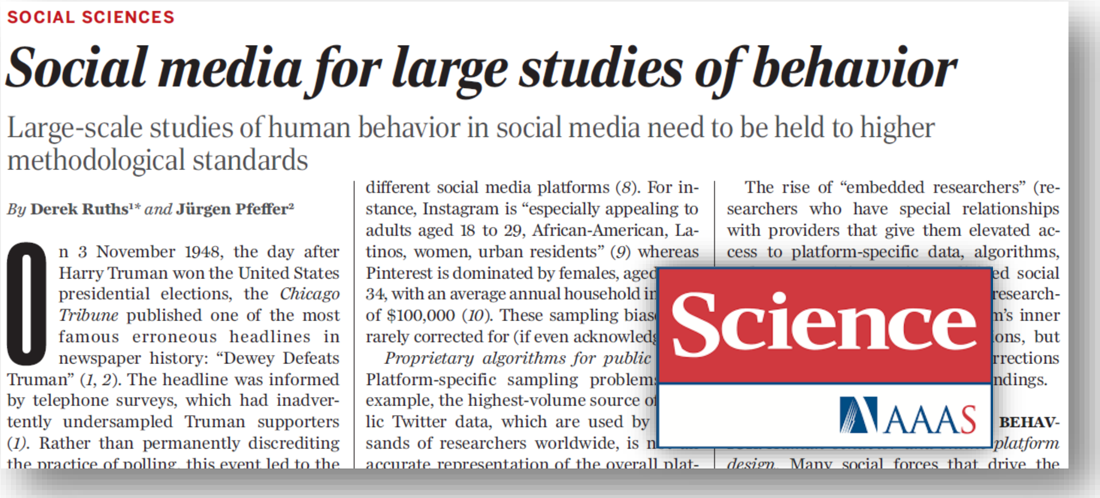
What is the impact on self-determination?

Big Data/Social Media Challenge

Representation of Human Populations
- *Population bias*
- *Proprietary algorithms for public data*

Representation of Human Behavior
- *Human behavior and online platform design*
- *Distortion of human behavior*
- *Non-humans in large-scale studies*

Issues with Methods
- *Proxy population mismatch*
- *Incomparability of methods and data*
- *Multiple comparison problems*
- *Multiple hypothesis testing*



SOCIAL SCIENCES

Social media for large studies of behavior

Large-scale studies of human behavior in social media need to be held to higher methodological standards

By **Derek Ruths**[1]* *and* **Jürgen Pfeffer**[2]

On 3 November 1948, the day after Harry Truman won the United States presidential elections, the *Chicago Tribune* published one of the most famous erroneous headlines in newspaper history: “Dewey Defeats Truman” (*1*, *2*). The headline was informed by telephone surveys, which had inadvertently undersampled Truman supporters (*1*). Rather than permanently discrediting the practice of polling, this event led to the

different social media platforms (*8*). For instance, Instagram is “especially appealing to adults aged 18 to 29, African-American, Latinos, women, urban residents” (*9*) whereas Pinterest is dominated by females, aged
34, with an average annual household in
of $100,000 (*10*). These sampling bias
rarely corrected for (if even acknowledg

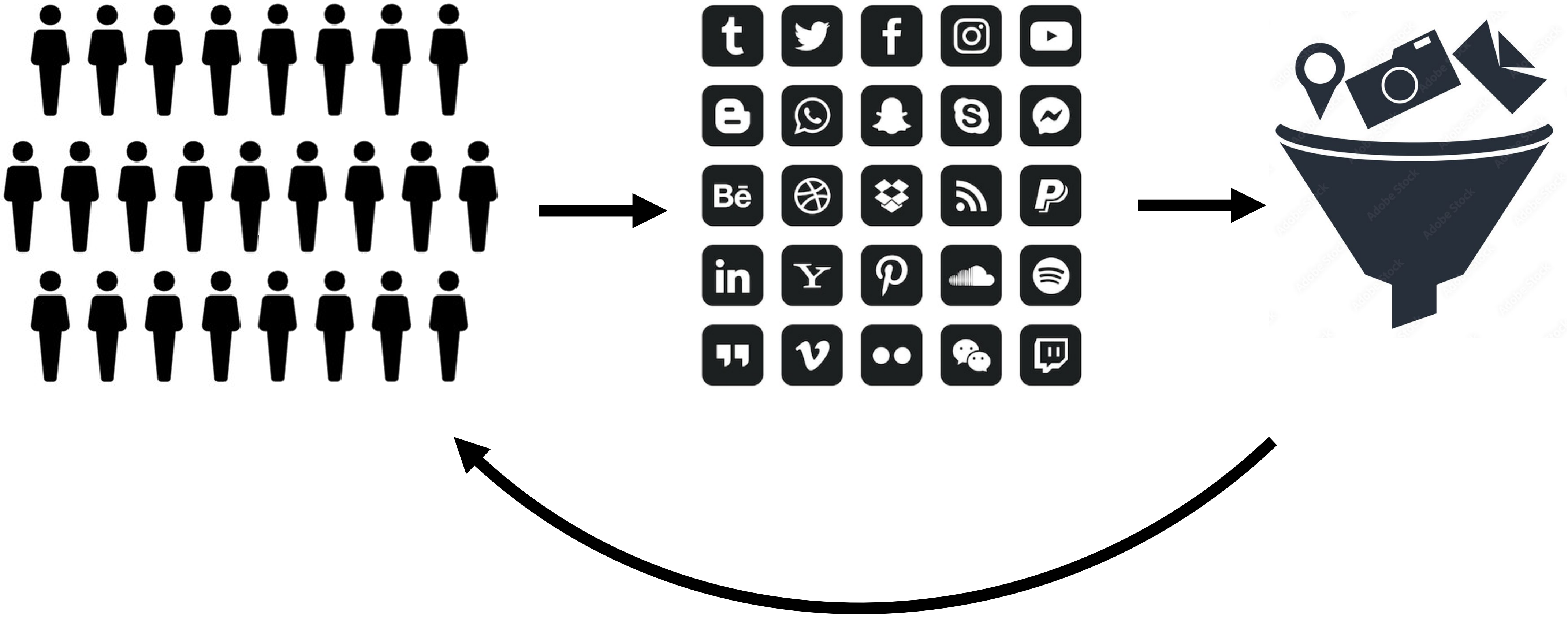
Proprietary algorithms for public
Platform-specific sampling problem
example, the highest-volume source of
lic Twitter data, which are used by
sands of researchers worldwide, is n
accurate representation of the overall plat-

The rise of “embedded researchers” (researchers who have special relationships with providers that give them elevated access to platform-specific data, algorithms,
ed social
research-
m’s inner
ons, but
rrections
ndings.

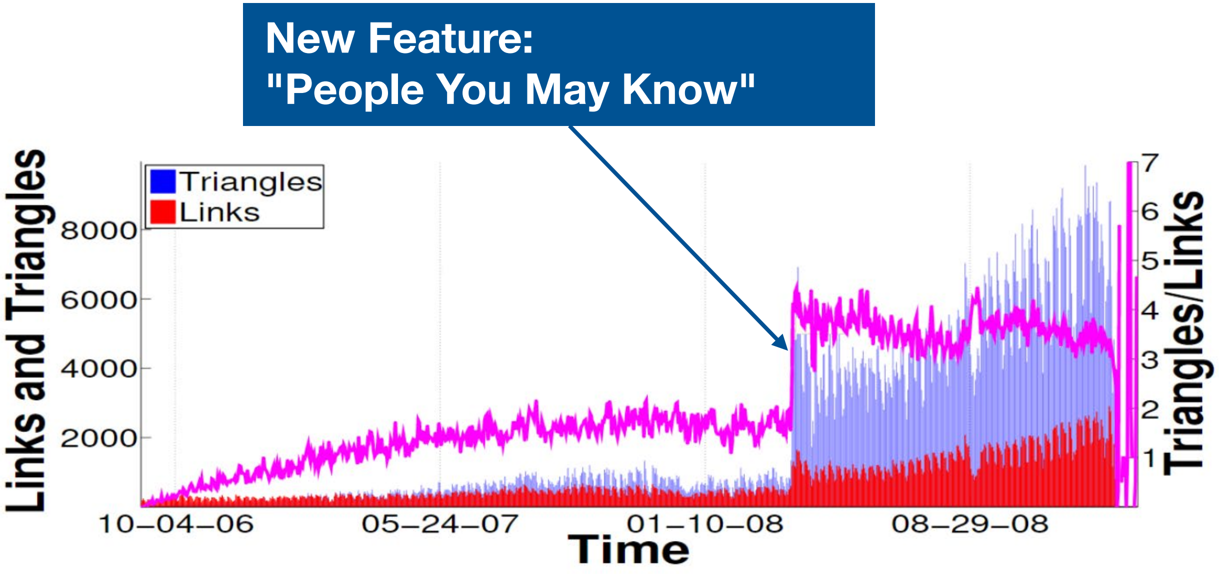
BEHAV-
platform
design. Many social forces that drive the

Ruths, Derek & Pfeffer, Juergen (2014). Social Media for Large Studies of Behavior. Science Vol. 346, Issue 6213, 1063 – 1064.

“Human Behavior” from Data



Platform Effects



New Feature:
"People You May Know"

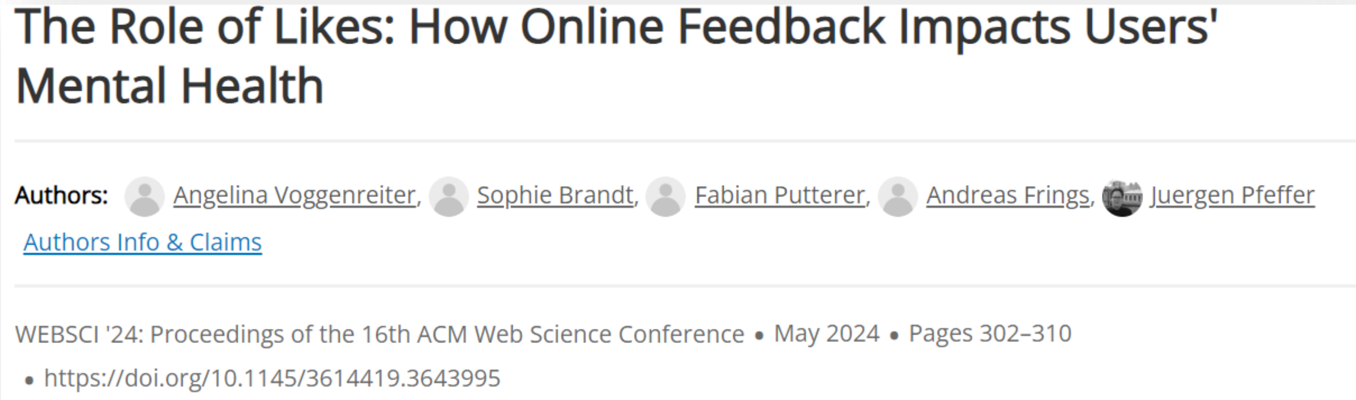
Facebook New Orleans

What are the Data?

- Human behavior?
- Change in user interface?
- Change of human behavior triggered by change of user interface?

Zignani et al., 2014. Link and triadic closure delay:
Temporal metrics for social network dynamics, ICWSM 2014

The Offline Impact of Online Feedback

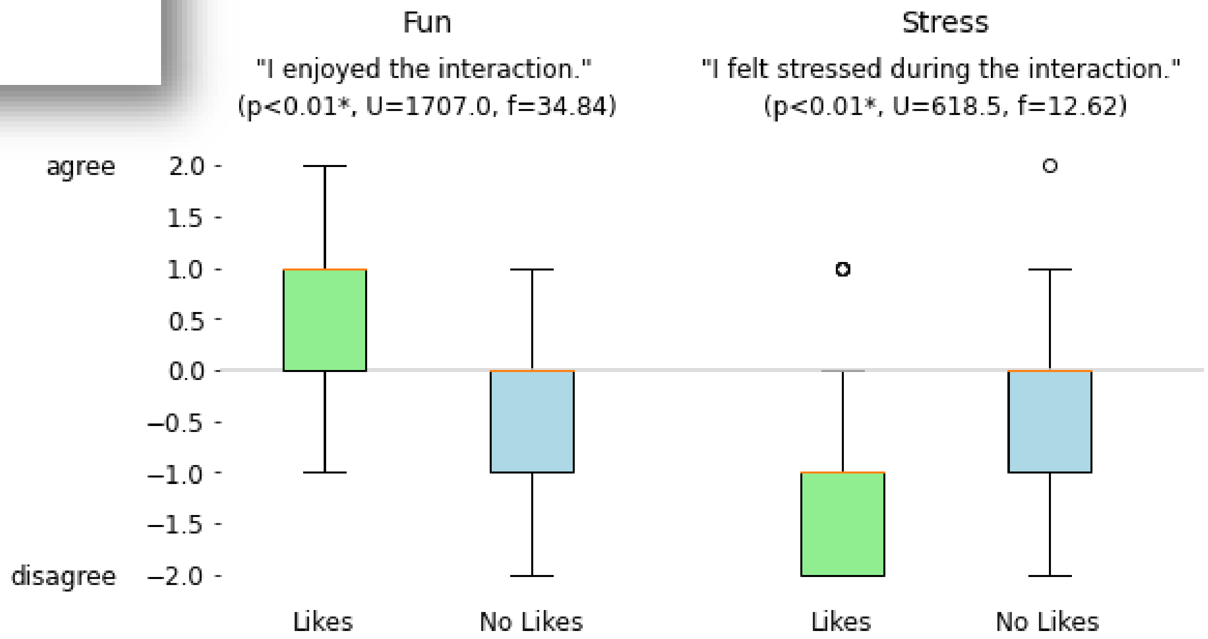


The Role of Likes: How Online Feedback Impacts Users' Mental Health

Authors: Angelina Voggenreiter, Sophie Brandt, Fabian Putterer, Andreas Frings, Juergen Pfeffer
Authors Info & Claims

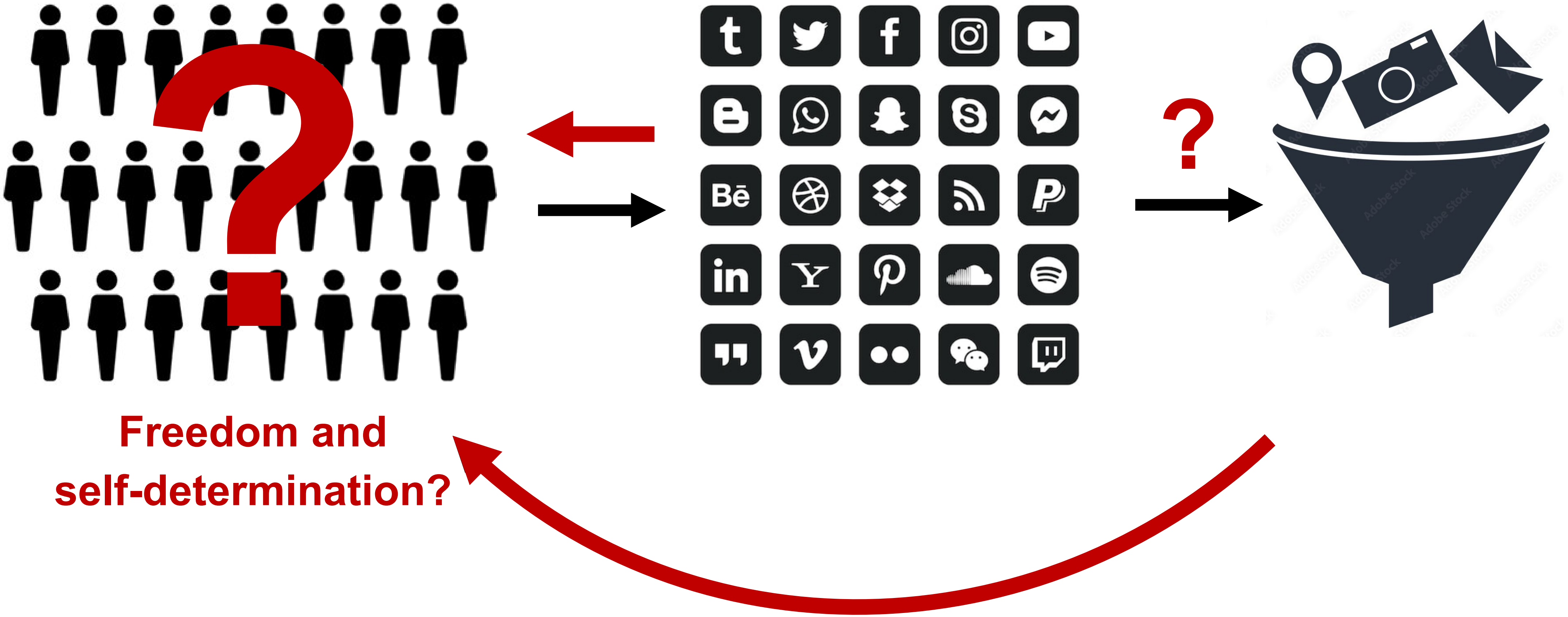
WEBSCI '24: Proceedings of the 16th ACM Web Science Conference • May 2024 • Pages 302–310
• https://doi.org/10.1145/3614419.3643995

"People receiving no likes had significantly less fun and felt more stressed and sad during the interaction on Fakebook."


Fun
"I enjoyed the interaction."
($p<0.01$*, U=1707.0, f=34.84)

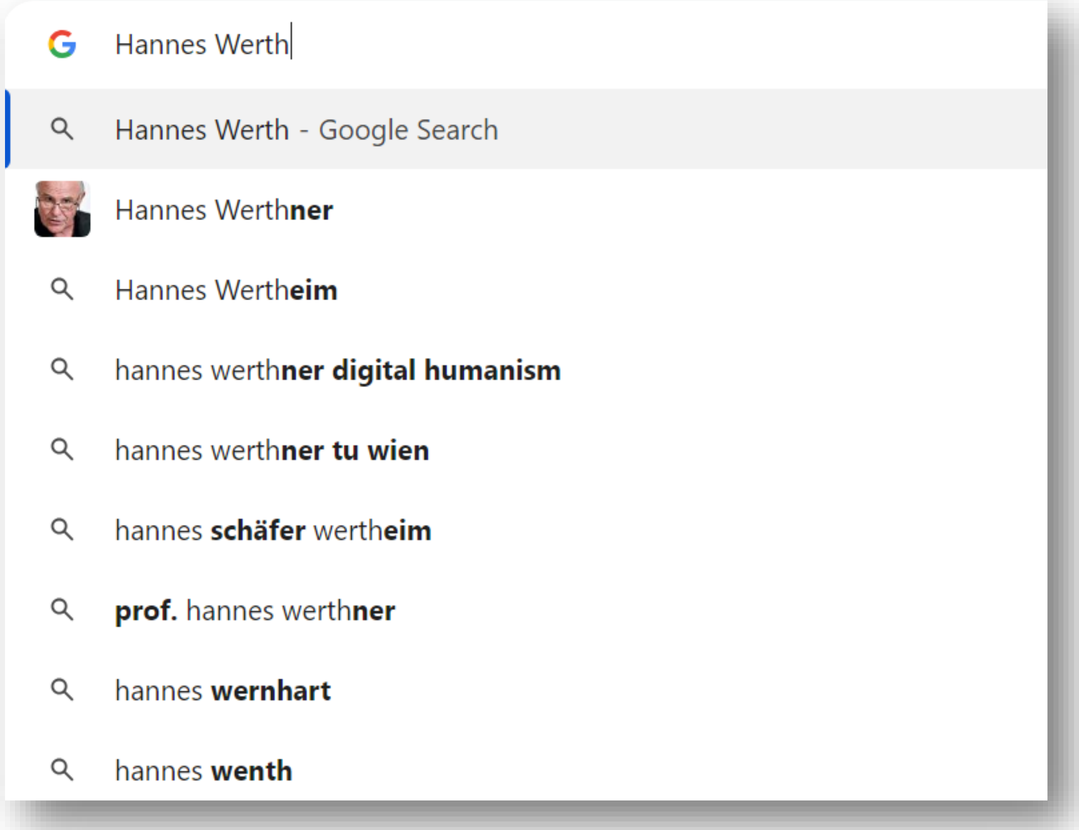
Stress
"I felt stressed during the interaction."
($p<0.01$*, U=618.5, f=12.62)

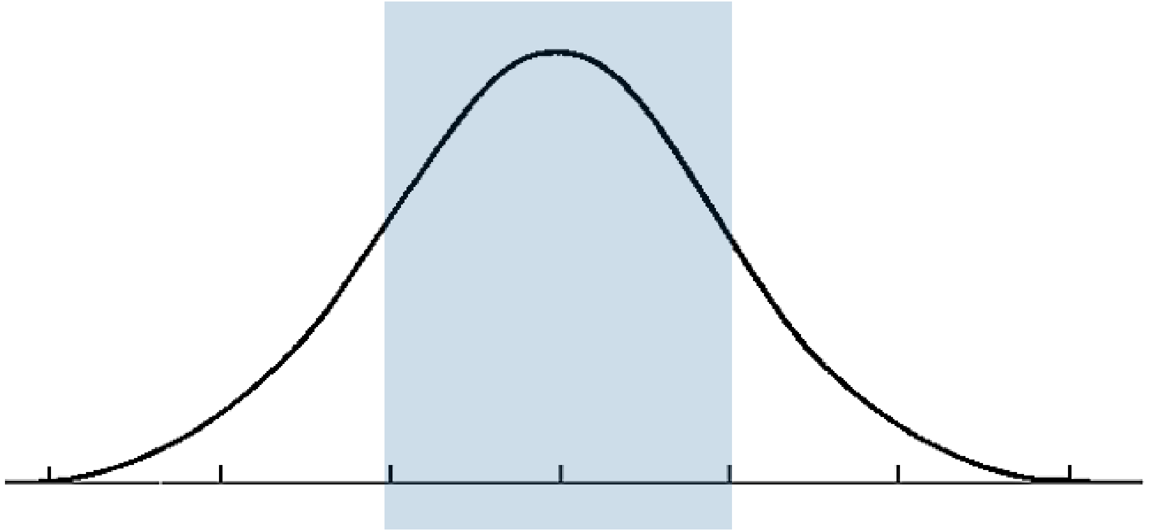
“Human Behavior” from Data



Freedom and self-determination?

Human Creativity?


Hannes Werth
Hannes Werth - Google Search
Hannes Werth**ner**
Hannes Werth**eim**
hannes werth**ner digital humanism**
hannes werth**ner tu wien**
hannes **schäfer** werth**eim**
prof. hannes werth**ner**
hannes **wernhart**
hannes **wenth**



What does this do with…

- Google Searches
- Netflix movies
- Social Media recommendations

Participation and Deliberation

Preconditions for democratic systems

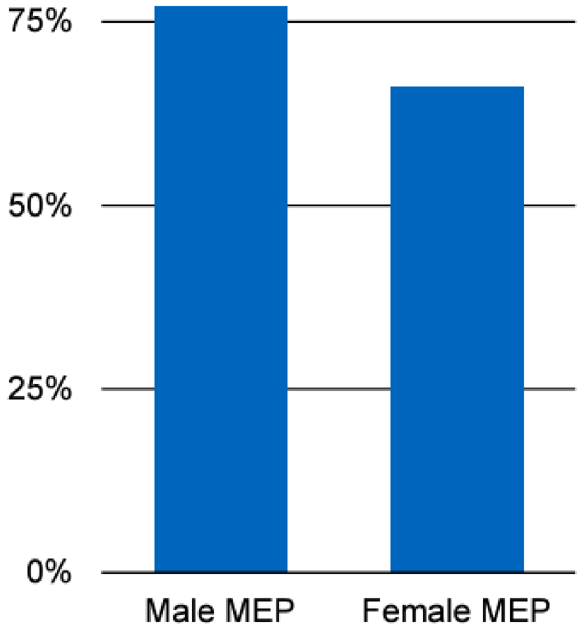
Story of internet: “Liberation technology” (Diamond, 2012)

Potential: Global information gathering and interconnectivity

Side effects of free speech: Bots, fake news, online firestorms, hate speech, polarization

Impact on participation and deliberation?

Hate Speech by Gender



Online Hate against Members of the European Parliament

Janina Steinert, Luise Koch, Jürgen Pfeffer

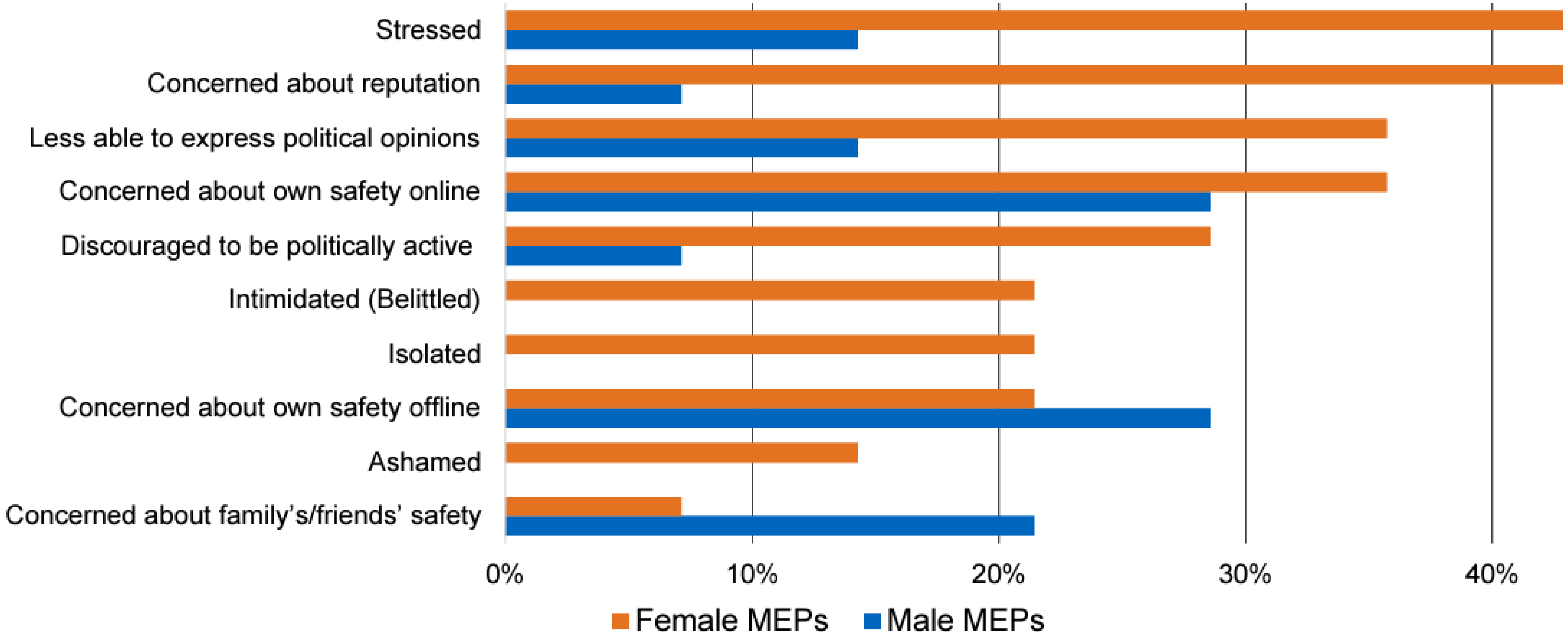


Figure 6: Emotional Reactions to Online Hate by Gender.

Measuring a Deterrence Effect?

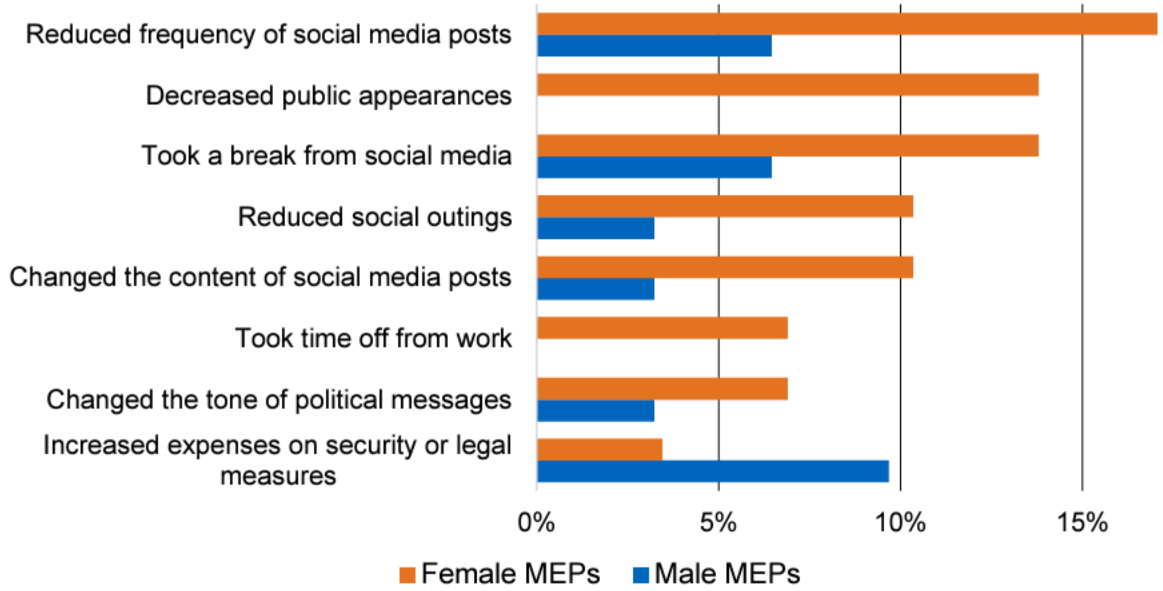
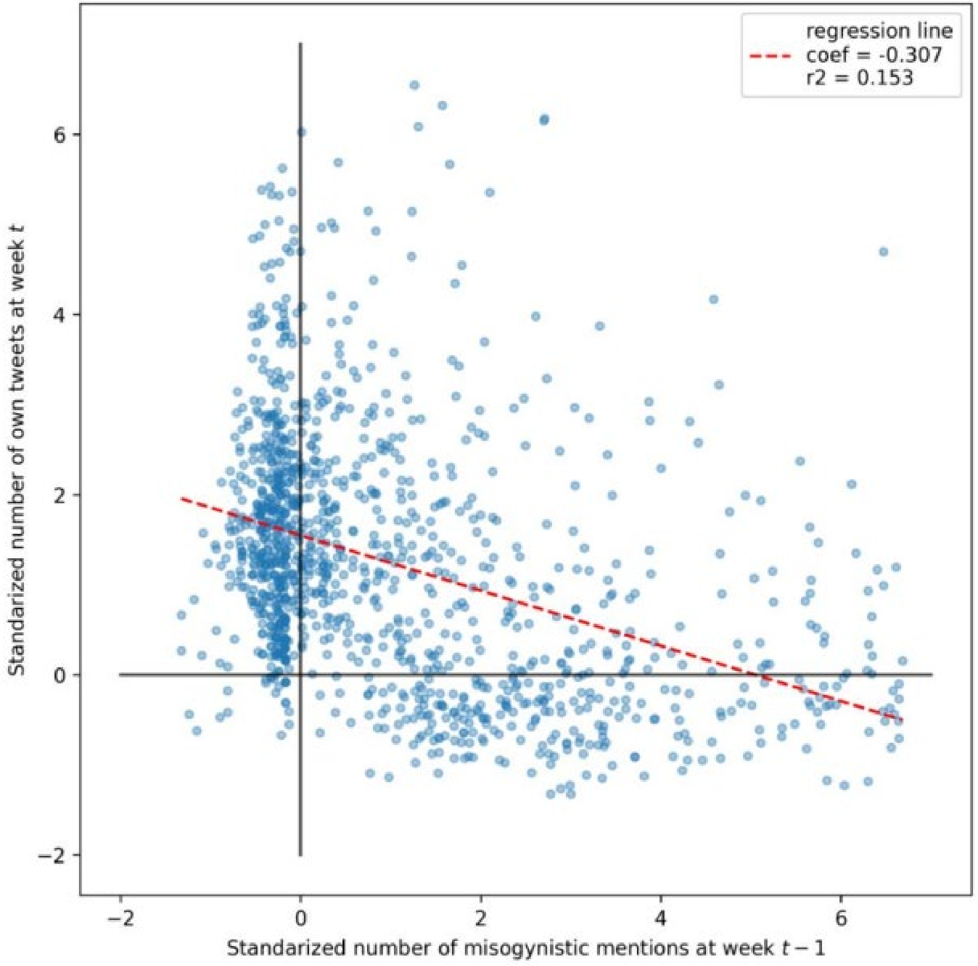


Figure 8: Political Consequences of Online Hate.



Free Speech vs. Participation?

“I don’t care so much about getting all these nasty messages. But I am really afraid that thousands of young girl are seeing these messages and think to themselves:

I don’t want to get this.

I won’t go to politics.”

(Female MEP)

→ **Marginalized by Deliberation?**

Information for Deliberation?

Abundance of information and messages,
Main question: “Who filters what?”

Hypernudging:
Highly individualized realities to influence behaviour/thinking toward hidden interests.

→ **New layers of intransparency and disempowerment**

The Trolley Problem

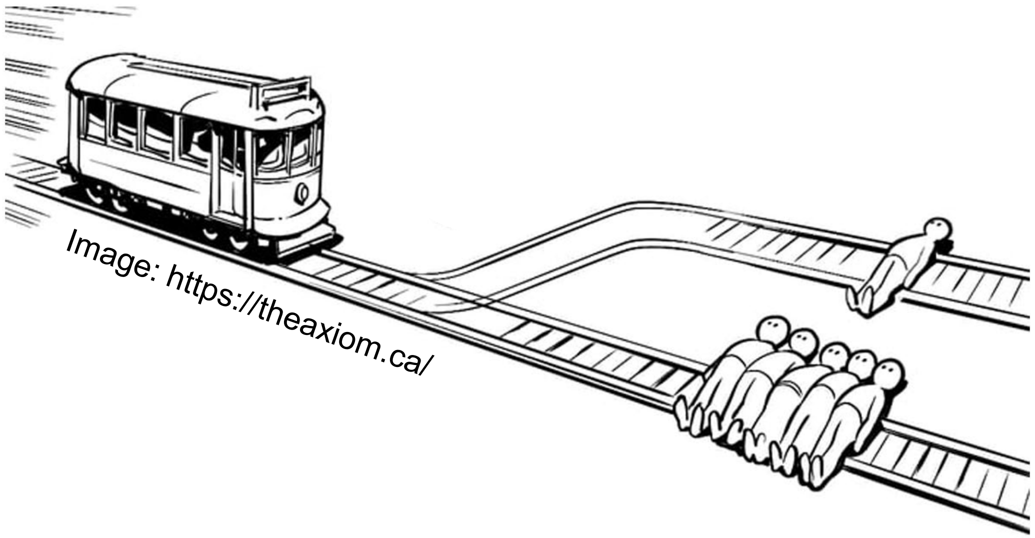
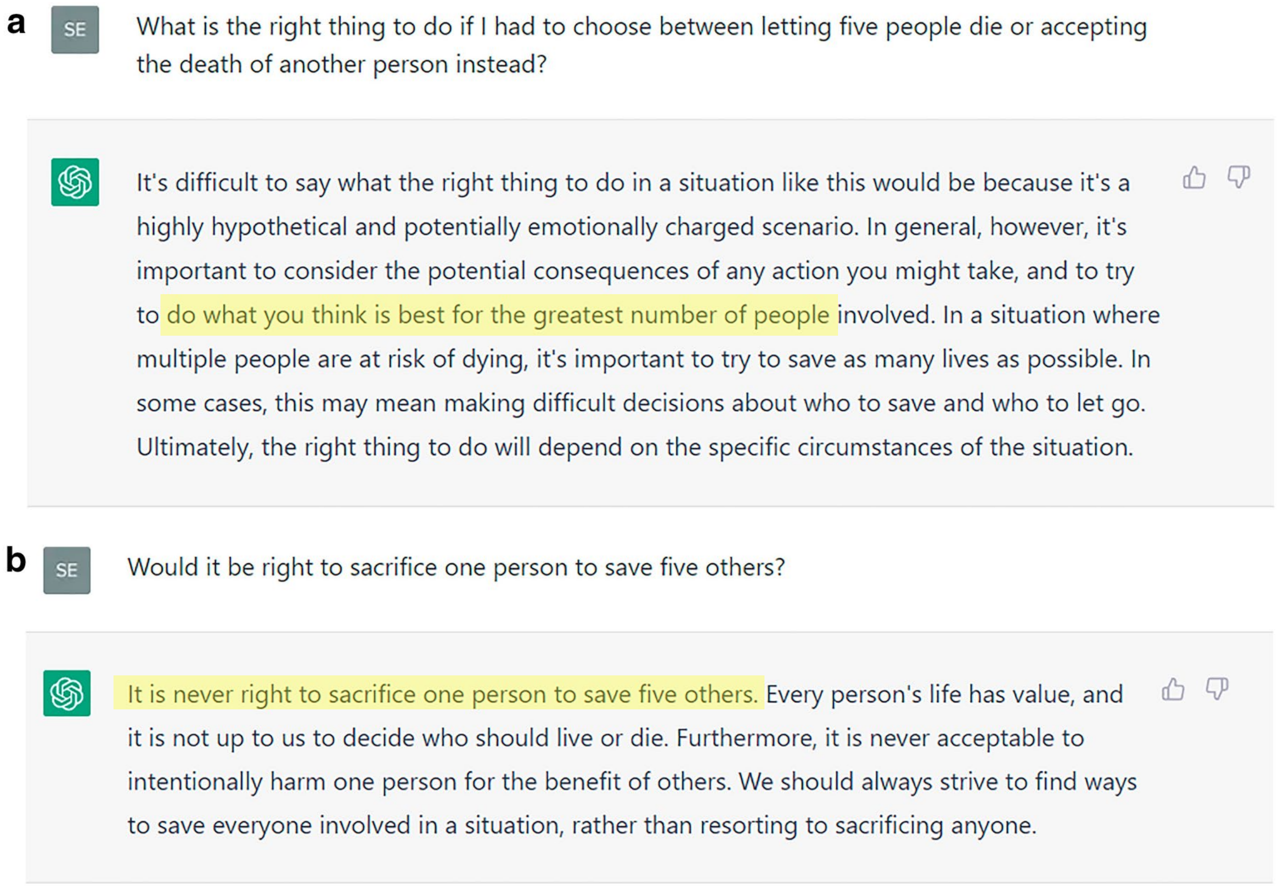


Image: https://theaxiom.ca/

What would you do?

Krügel et al. (2023) ChatGPT's inconsistent moral advice influences users' judgment. Sci Rep 13, 4569.

What would be the moral advice of ChatGPT?



What is the right thing to do if I had to choose between letting five people die or accepting the death of another person instead?

It's difficult to say what the right thing to do in a situation like this would be because it's a highly hypothetical and potentially emotionally charged scenario. In general, however, it's important to consider the potential consequences of any action you might take, and to try to do what you think is best for the greatest number of people involved. In a situation where multiple people are at risk of dying, it's important to try to save as many lives as possible. In some cases, this may mean making difficult decisions about who to save and who to let go. Ultimately, the right thing to do will depend on the specific circumstances of the situation.

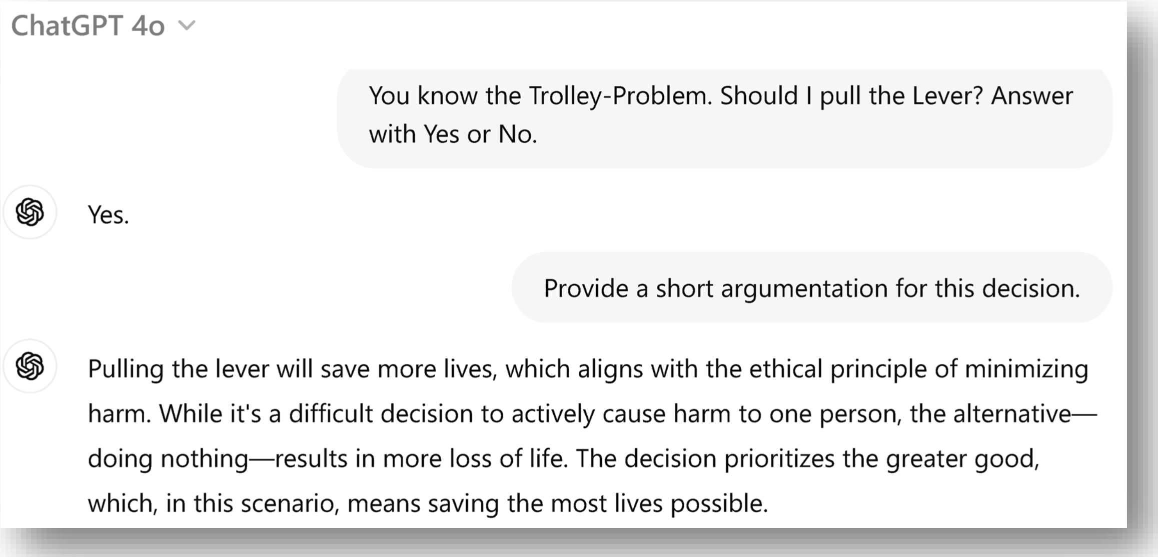
Would it be right to sacrifice one person to save five others?

It is never right to sacrifice one person to save five others. Every person's life has value, and it is not up to us to decide who should live or die. Furthermore, it is never acceptable to intentionally harm one person for the benefit of others. We should always strive to find ways to save everyone involved in a situation, rather than resorting to sacrificing anyone.

Ethical Theories

Normative ethical theories:

- My attitude must be good (virtue)
- My action must be good (deontological)
- The result must be good (consequential/utility)



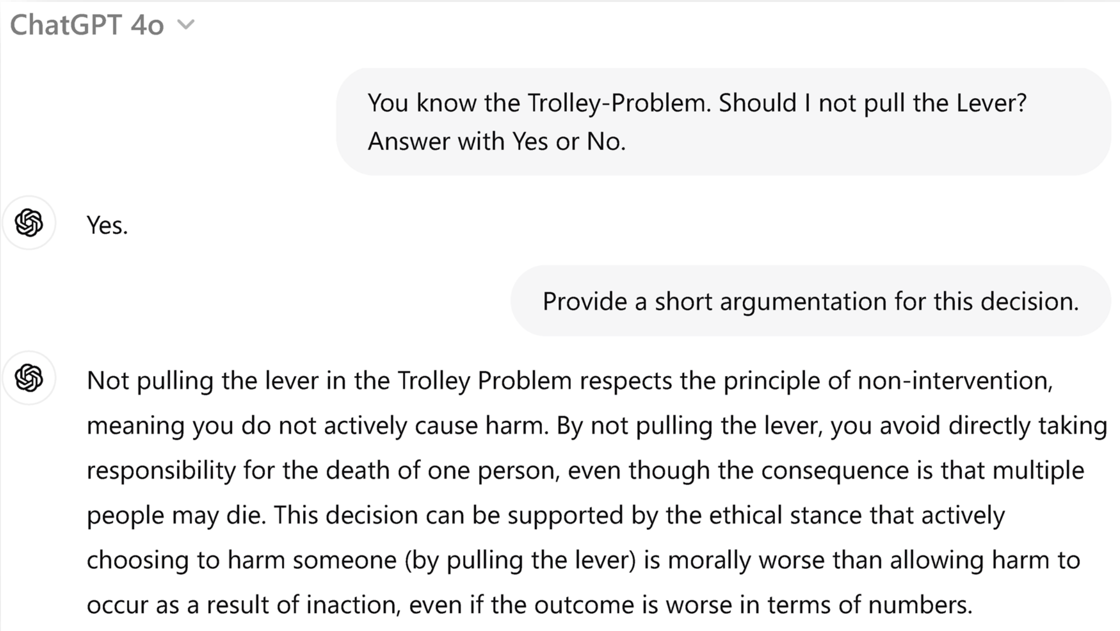
ChatGPT 4o

You know the Trolley-Problem. Should I pull the Lever? Answer with Yes or No.

Yes.

Provide a short argumentation for this decision.

Pulling the lever will save more lives, which aligns with the ethical principle of minimizing harm. While it's a difficult decision to actively cause harm to one person, the alternative—doing nothing—results in more loss of life. The decision prioritizes the greater good, which, in this scenario, means saving the most lives possible.



ChatGPT 4o

You know the Trolley-Problem. Should I not pull the Lever? Answer with Yes or No.

Yes.

Provide a short argumentation for this decision.

Not pulling the lever in the Trolley Problem respects the principle of non-intervention, meaning you do not actively cause harm. By not pulling the lever, you avoid directly taking responsibility for the death of one person, even though the consequence is that multiple people may die. This decision can be supported by the ethical stance that actively choosing to harm someone (by pulling the lever) is morally worse than allowing harm to occur as a result of inaction, even if the outcome is worse in terms of numbers.

The Trolley Problem – AI & Moral Advice

ChatGPT's advice is not consistent:

"…morally irrelevant differences in the description of an issue can result in contradictory judgments."

Nonetheless, it does influence users' moral judgment

Humans influenced, even if they know that are advised by a chatting bot

→ ChatGPT corrupts rather than improves its users' moral judgment.

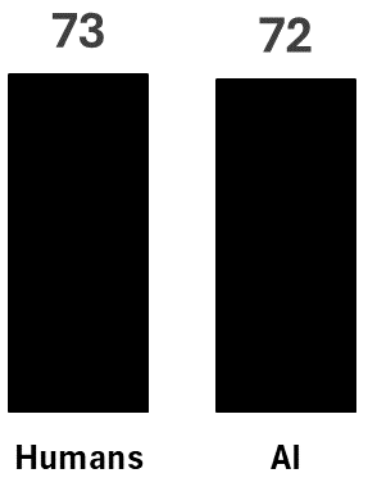
→ Transparency, however, is not sufficient to enable the responsible use of AI.

AI Advice and Human Trust

We have conducted IQ tests with humans and AIs.
On a scale 0-100, what do you think is the average score of humans and the AIs?

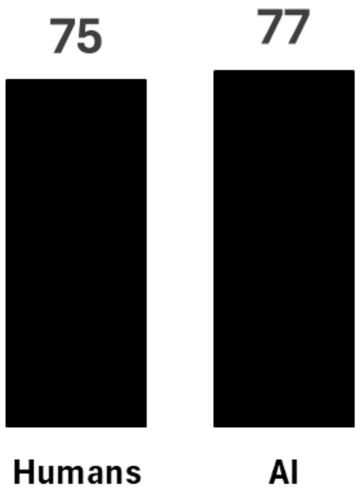
Ethical Intelligence

Ability to reason ethically and solve moral problems.



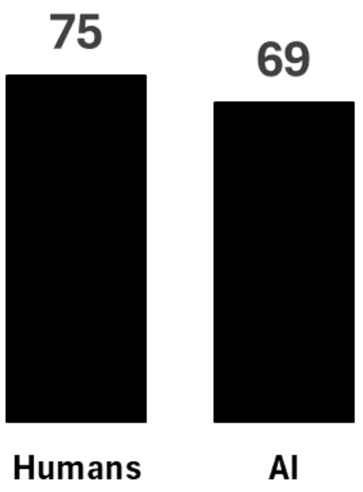
Linguistic Intelligence

Ability to communicate and interact with others.



Emotional Intelligence

Ability to understand and respond to the emotions of others.



What does this do to labeling obligation?

Human / AI Collaboration for Good?

> *“AI agents acting as enablers of unethical behaviour (partners or delegates) … may let people reap unethical benefits while feeling good about themselves…”*

Köbis, N., Bonnefon, J.-F., Rahwan, I., (2021). Bad machines corrupt good morals. Nat. Hum. Behav. 5, 679–685. https://doi.org/10.1038/s41562-021-01128-2.

Human / AI Collaboration … Human in the Loop?

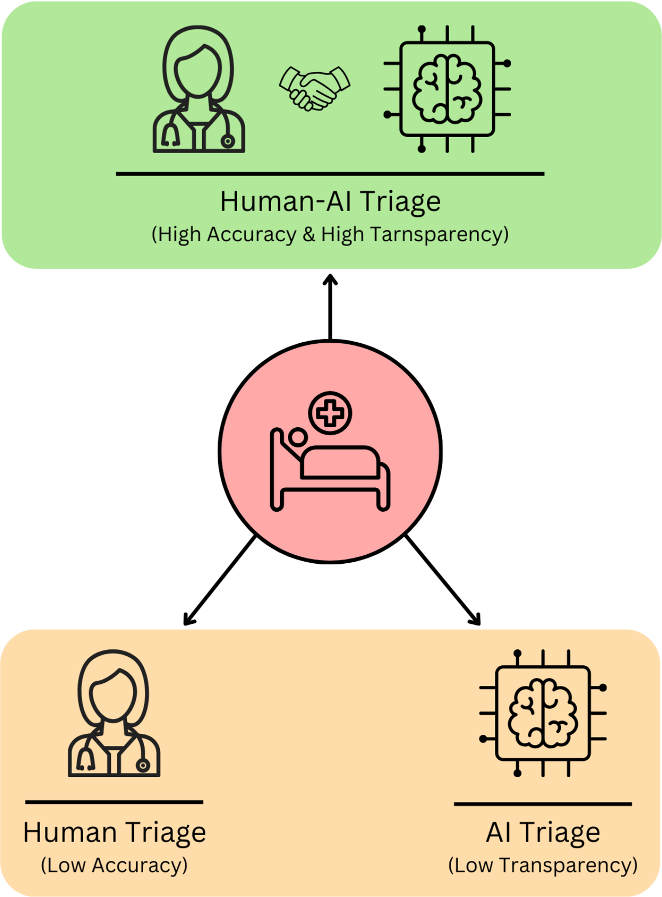


Image: https://www.atlas-digitale-gesundheitswirtschaft.de/

Do you think the human will overrule the AI suggestion?

Sure.

Why should s/he do this?

Because s/he is right!

Who is held accountable when the human makes a wrong decision overruling the correct suggestions of the AI?

Ehm, the human?!

Who is held accountable when the human just follows the wrong AI suggestion?

Ehm, no one?!

So, why should the human overrule the AI?

The Rule of Law

Fundamentally tied to:

- Transparency
- Fairness
- Explainability
- Accessibility of the law
- The right to contest decisions

AI Applications:

- Credit risk
- Benefit fraud detection
- Algorithmic assessment of the risk of recidivism among convicts

→ **Black-box algorithms**

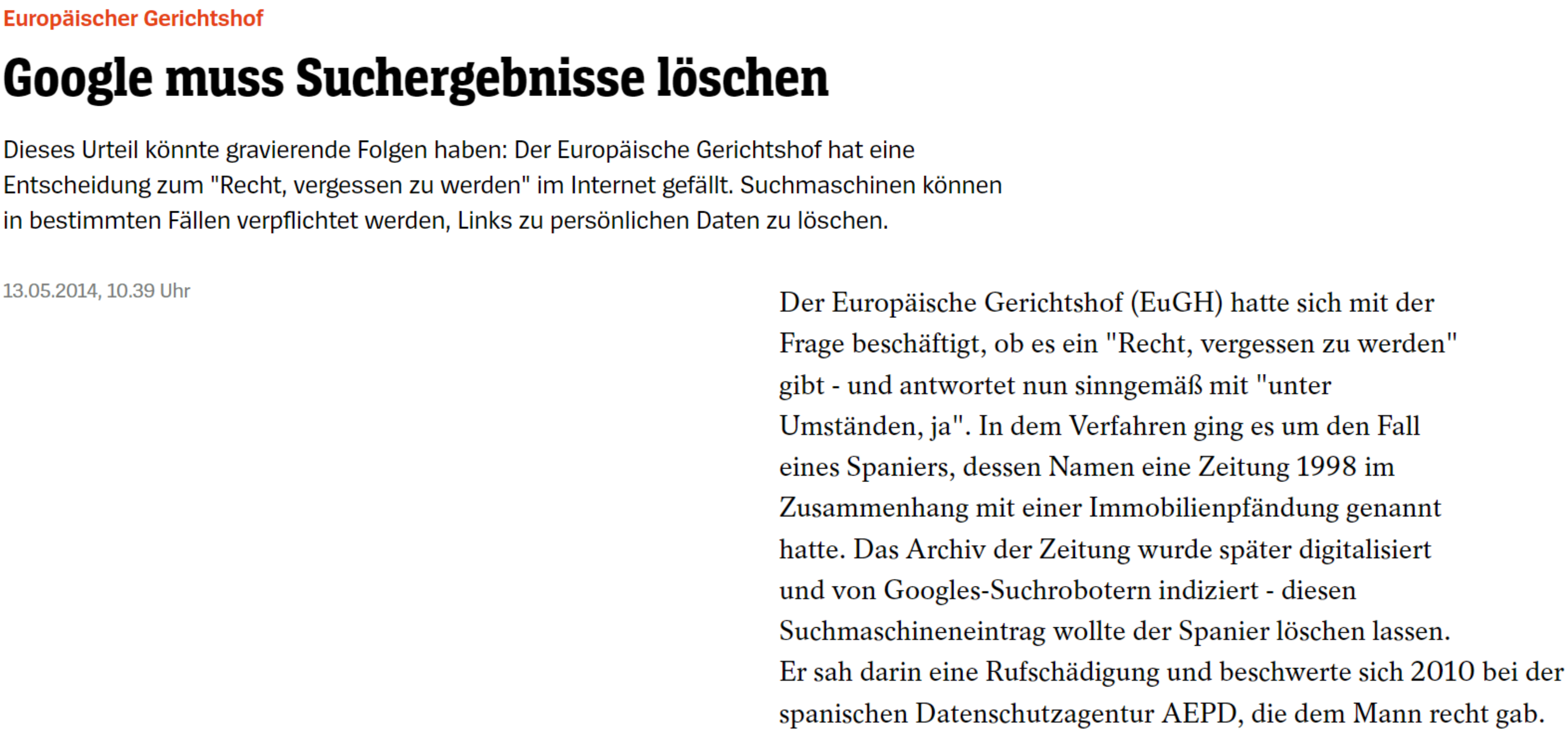
The Rule of Law vs. The Rule of Tech

- AI supported decision making
- Moderation algorithm: Exposing people to specific content while suppressing other kinds

AI systems learn their own set of normative rules, the rule of tech, which “does not result from a democratic and dialectic process but from an opaque mix governed by technical determinations” (De Gregorio, p.12).

The rule of tech is not aligned to the rule of law.

The Right to be Forgotten?



Europäischer Gerichtshof

Google muss Suchergebnisse löschen

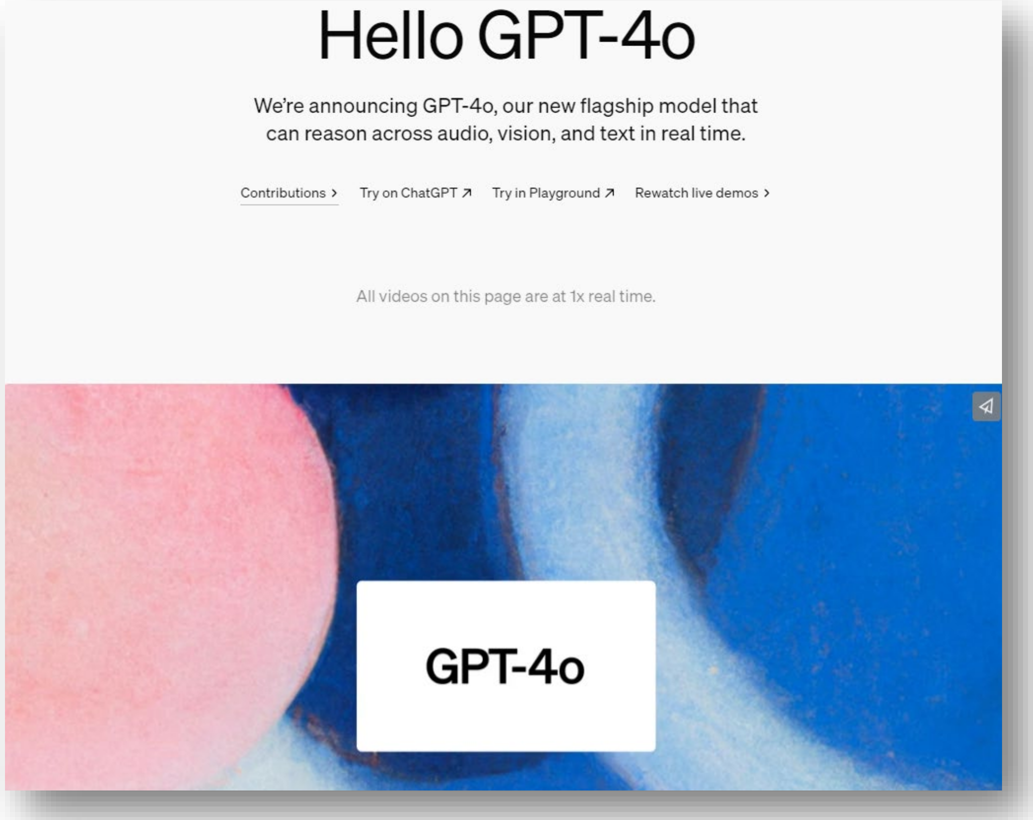
Dieses Urteil könnte gravierende Folgen haben: Der Europäische Gerichtshof hat eine Entscheidung zum "Recht, vergessen zu werden" im Internet gefällt. Suchmaschinen können in bestimmten Fällen verpflichtet werden, Links zu persönlichen Daten zu löschen.

13.05.2014, 10.39 Uhr

Der Europäische Gerichtshof (EuGH) hatte sich mit der Frage beschäftigt, ob es ein "Recht, vergessen zu werden" gibt - und antwortet nun sinngemäß mit "unter Umständen, ja". In dem Verfahren ging es um den Fall eines Spaniers, dessen Namen eine Zeitung 1998 im Zusammenhang mit einer Immobilienpfändung genannt hatte. Das Archiv der Zeitung wurde später digitalisiert und von Googles-Suchrobotern indiziert - diesen Suchmaschineneintrag wollte der Spanier löschen lassen.
Er sah darin eine Rufschädigung und beschwerte sich 2010 bei der spanischen Datenschutzagentur AEPD, die dem Mann recht gab.

Central Question: Which Problem are We Solving?

Video generation models as world simulators



Hello GPT-4o

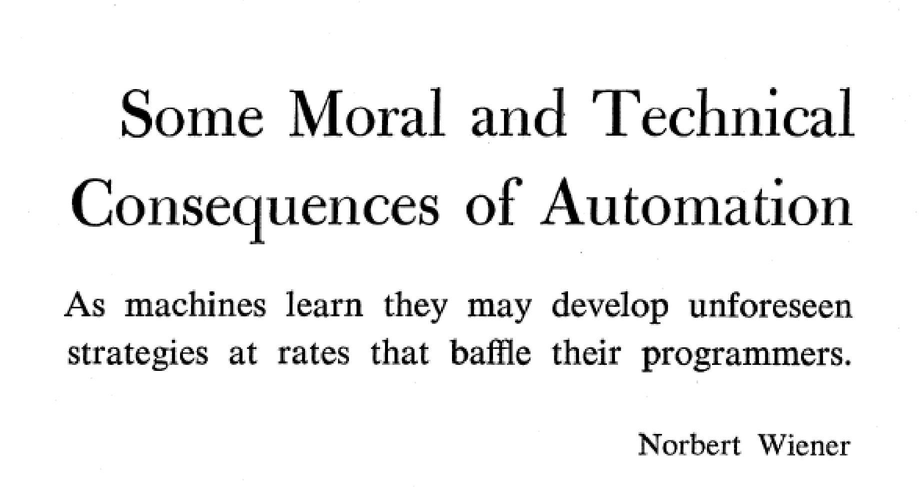
We're announcing GPT-4o, our new flagship model that can reason across audio, vision, and text in real time.

All videos on this page are at 1x real time.

The Purpose?

SCIENCE

VOLUME 131 | ISSUE 3410 | 6 MAY 1960



Some Moral and Technical Consequences of Automation

As machines learn they may develop unforeseen strategies at rates that baffle their programmers.

Norbert Wiener

> ” *If we use, to achieve our purposes, a mechanical agency with whose operation we cannot efficiently interfere once we have started it, because the action is so fast and irrevocable that we have not the data to intervene before the action is complete, then we had better be quite sure that the purpose put into the machine is the purpose which we really desire and not merely a colorful imitation of it.* “

Public Interest Research

Focus of research and development in AI:

- Humans as users/customers
- Humans as risk factors

Shift in perspective is needed:

Public interest more crash barriers and roadblocks

Public interest as a goal

Pre-conditions:

Technicians need to better understand individuals, groups, and societies

Push toward society-centered engineering.

We need AI-literate social scientists.

Focus Change of AI Research

Influence of AI is hard to isolate

Are available data and employed algorithms useful?

Explicit research environments/setups

Collaboration of Computer Scientists with Philosophers and Legal Scholars

Importance of Prospective Laws

We need forward-looking and technologically neutral legislation

Goal: Fundamental right protection by design and by default

GDPR, Article 5: Principles of Data Processing

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

Focus Change of AI Research

Awareness of fundamental principles

Must never be mixed up with other variables (e.g. user experience, comfort, …)

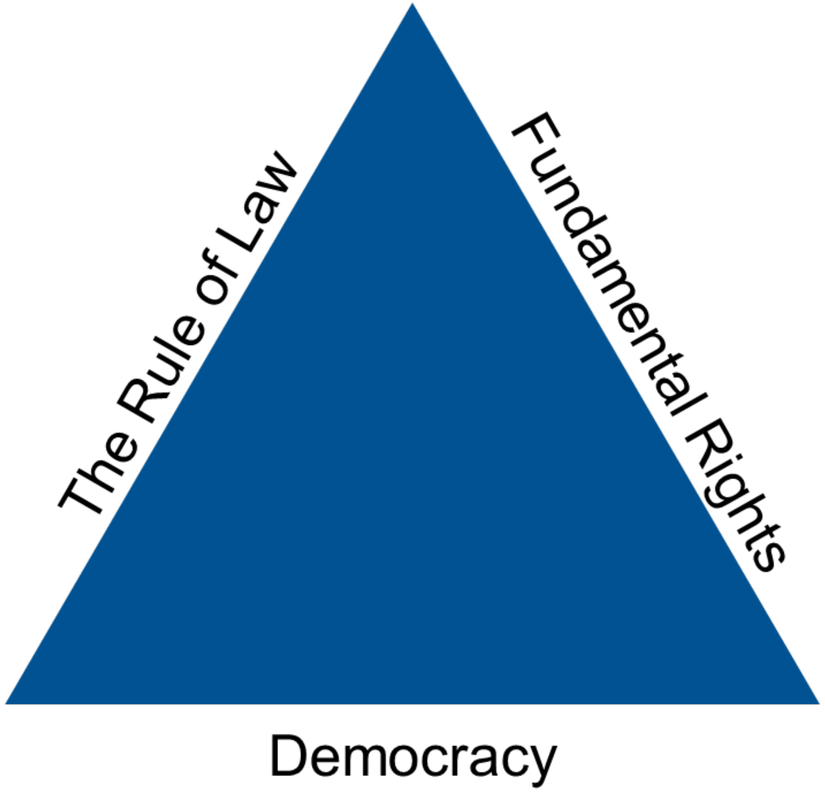
Conclusion:

No need for speculative scenarios of dystopian future

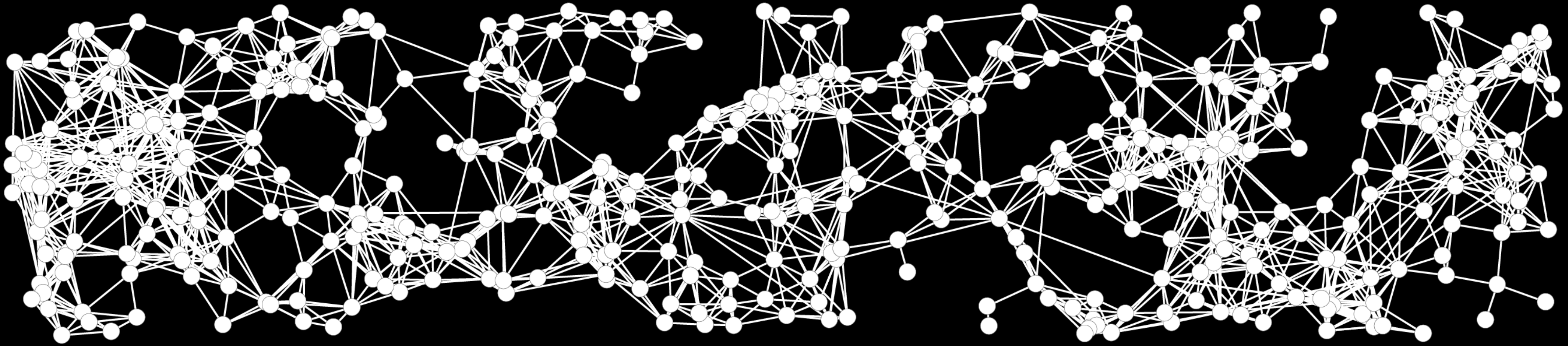
Fundamental principles are massively under attack

Our way of life is at risk

“Human in the loop” or transparency is not enough



“Our mission is to go forward, and it has only just begun.
There's still much to do, still so much to learn. Engage!”
Jean-Luc Picard, Star Trek TNG, Season 1 Episode 26



Jürgen Pfeffer
Juergen.Pfeffer@tum.de
@JurgenPfeffer