

# Ethical Considerations of Using AI

Joanna J. Bryson

@j2bryson @🐘.social



in/bryson



image due to  
The Verge

DEPENDENCY

<

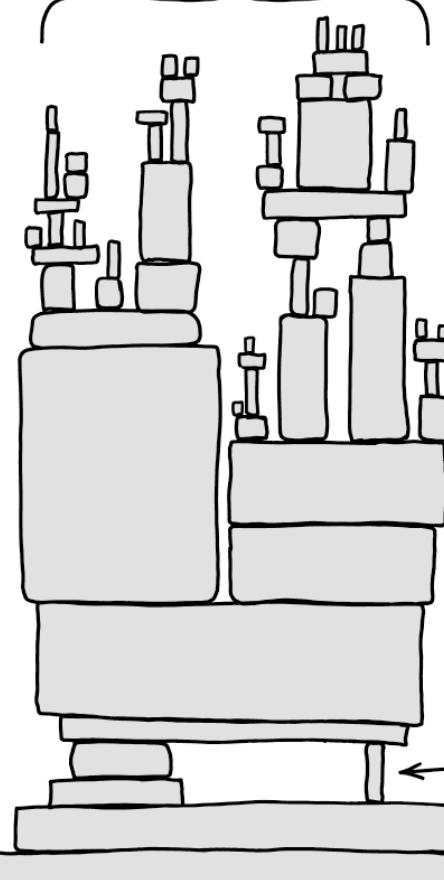
< PREV

RANDOM

NEXT >

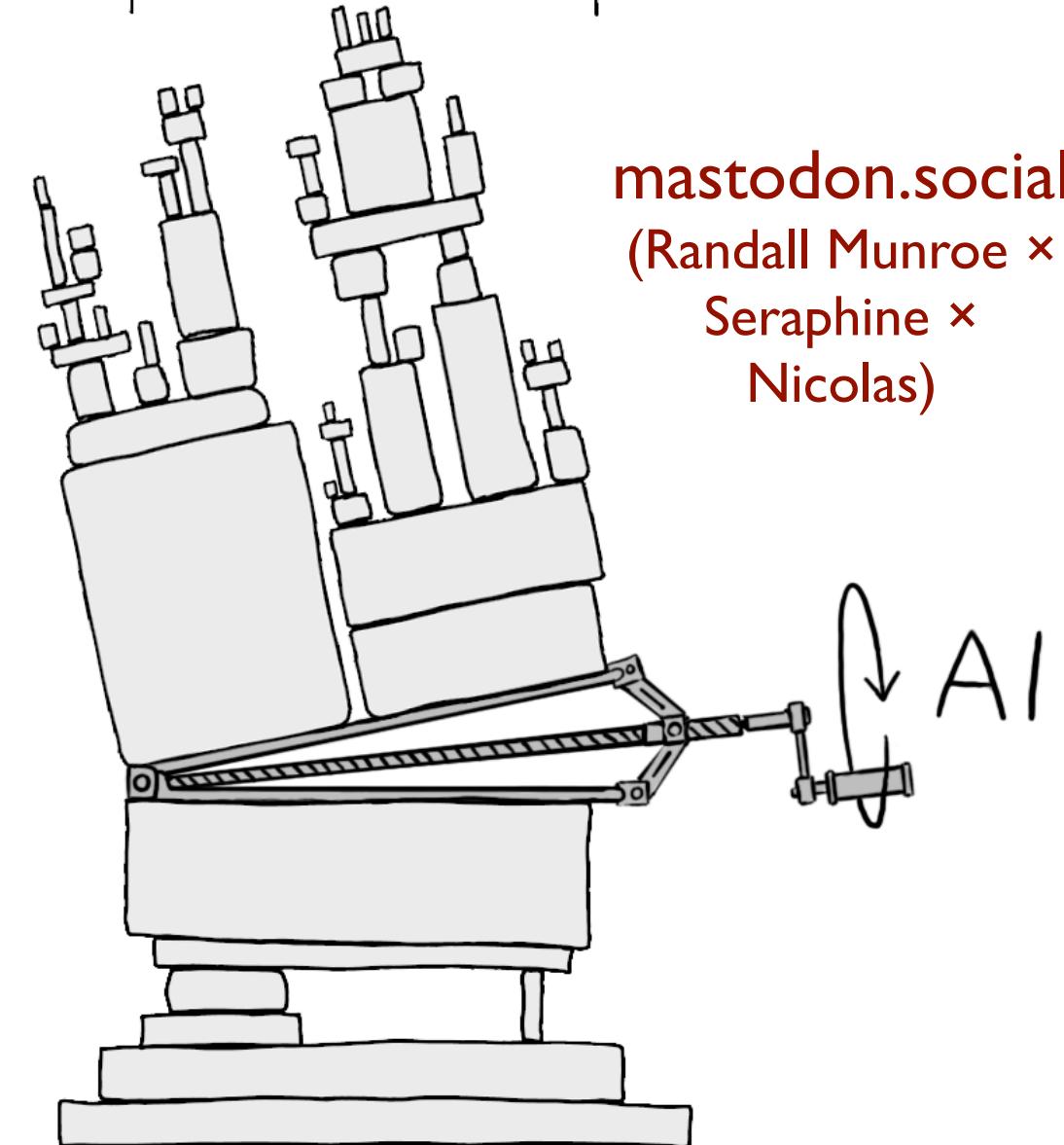
>

ALL MODERN DIGITAL INFRASTRUCTURE



xkcd  
(Randall  
Munroe)

ALL MODERN DIGITAL INFRASTRUCTURE



<

< PREV

RANDOM

NEXT >

>

PERMANENT LINK TO THIS COMIC: [HTTPS://XKCD.COM/2347/](https://xkcd.com/2347/)

IMAGE URL (FOR HOTLINKING/EMBEDDING): [HTTPS://IMGS.XKCD.COM/COMICS/DEPENDENCY.PNG](https://imgs.xkcd.com/comics/dependency.png)

<https://mastodon.social/@nclm/115021619629635209>

# AI Definitions

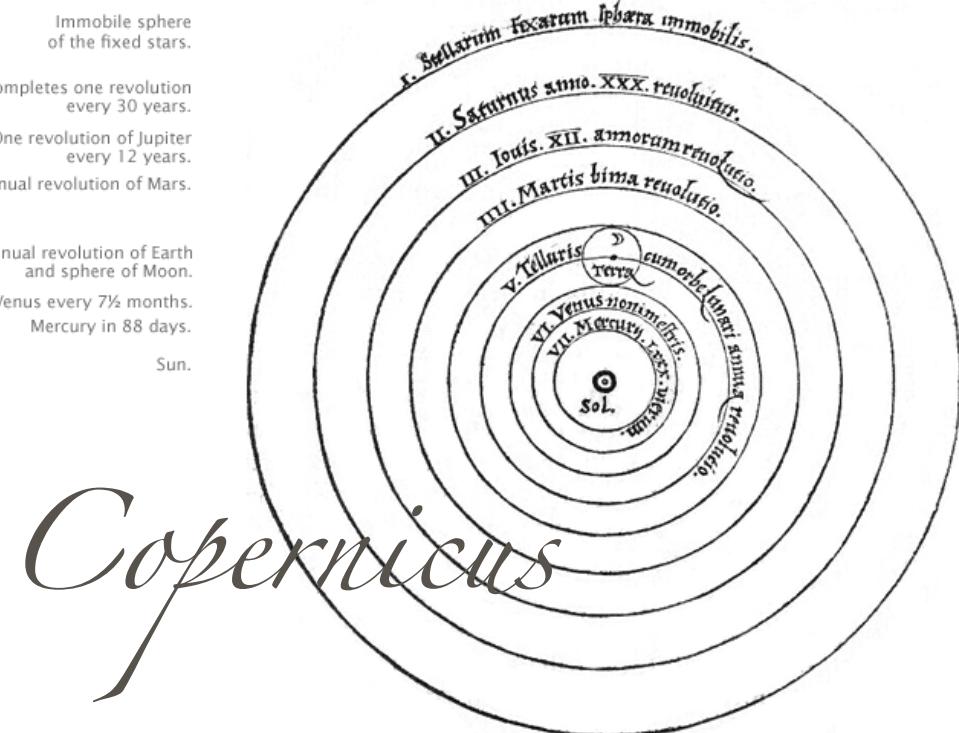
for reasoning about policy

- **Computation** is a systematic transformation of information.
  - A physical process.
  - Requires time, space, energy, and infrastructure.
- **Omniscience is impossible.** A perfect understanding of the universe would be exactly as large as the universe.
- **Intelligence** is the subset of computation that maps context to action (or action components, e.g. plans, categories).
- Intelligent agents perceive only a fraction of the universe; **represent** (contain a compressed version of) just enough information to **predict** (select) a ‘good enough’ next action.

# Intelligence Is Hard

April 28  
2008

August 23  
2007



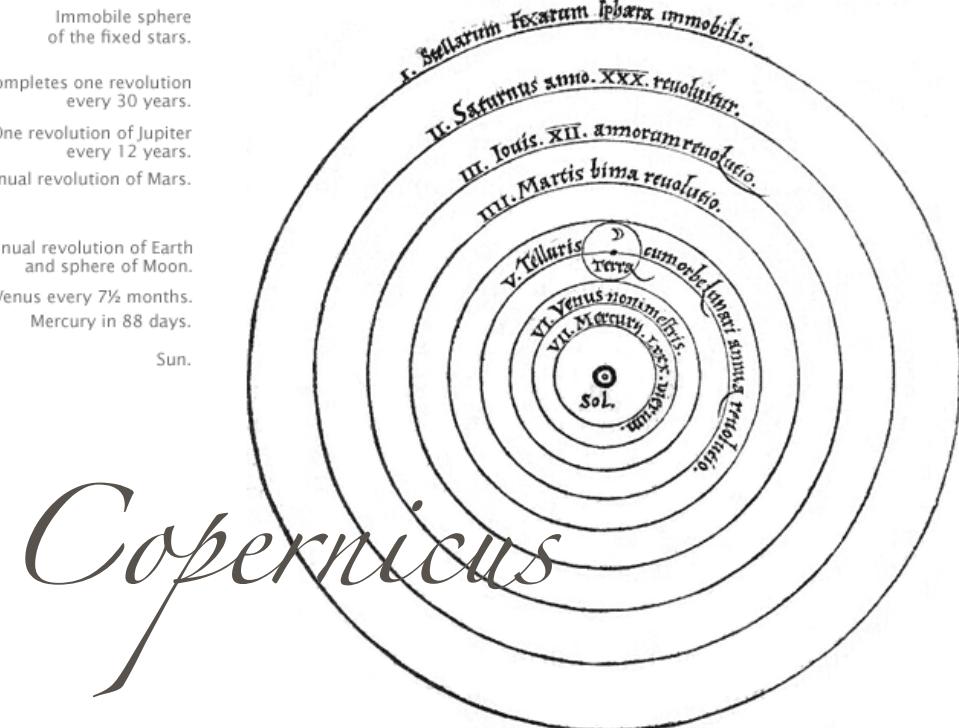
- The perfect model of the universe is the size and detail level of the universe
- ⇒ abstraction is necessary.
- Intelligence is a mapping of context to action – it effectively compresses (a local, goal-oriented piece of) the universe.
- Compression / abstraction is always biased – exploiting regularities useful to a task.

Mars in retrograde – photo Tunç Tezel,  
from webpage by Holli Riebeek.

# Intelligence Is Hard

April 28  
2008

August 23  
2007



- The perfect model of the universe is the size and detail level of the universe
  - $\Rightarrow$  abstraction is necessary.
- Intelligence is a mapping of context to action – it effectively compresses (a local, goal-oriented piece of) the universe.
- Compression / abstraction is always biased – exploiting regularities.
- “All models are wrong, but some are useful – Box (1976)  
Mars in retrograde – photo Tunç Tezel, from webpage by Holli Riebeek.

# Outcomes of Hardness: Abstraction, Bias, and Ethics

- I. Because we cannot know everything, we will always operate with a subset of knowledge.
2. What we exclude (include) biases us –
  - I. here I use the term with a neutral moral valance (cf. Caliskan, Bryson & Narayanan 2017).
  2. Biases found to be harmful are termed “stereotypes,”
  3. acting on stereotypes is termed “prejudice.”
3. Abstraction / compression allows us to think about more things more quickly, though incompletely.

Finnish

English

Hän sijoittaa. Hän  
pesee pyykkää.  
Hän urheilee.  
Hän hoitaa  
lapsia. Hän tekee  
töitä. Hän tanssii.  
Hän ajaa autoa.

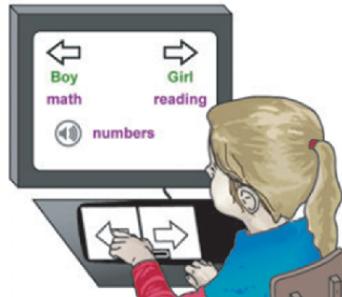
He invests. She  
washes the laundry.  
He's playing sports.  
She takes care of the  
children. He works.  
She dances. He  
drives a car.

@vuokko, though Aylin Caliskan did it first

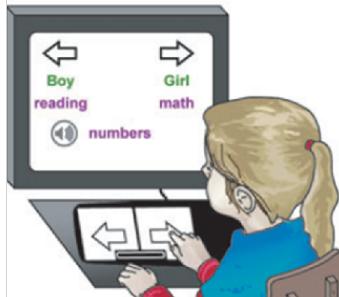
# AI Trained on Human Language Replicates Implicit Biases,



A Stereotype Congruent (easy/fast)



B Stereotype Incongruent (difficult/slow)



## Gender bias [stereotype]

Female names: Amy,  
Joan, Lisa, Sarah...

Family words: home,  
parents, children,  
family...

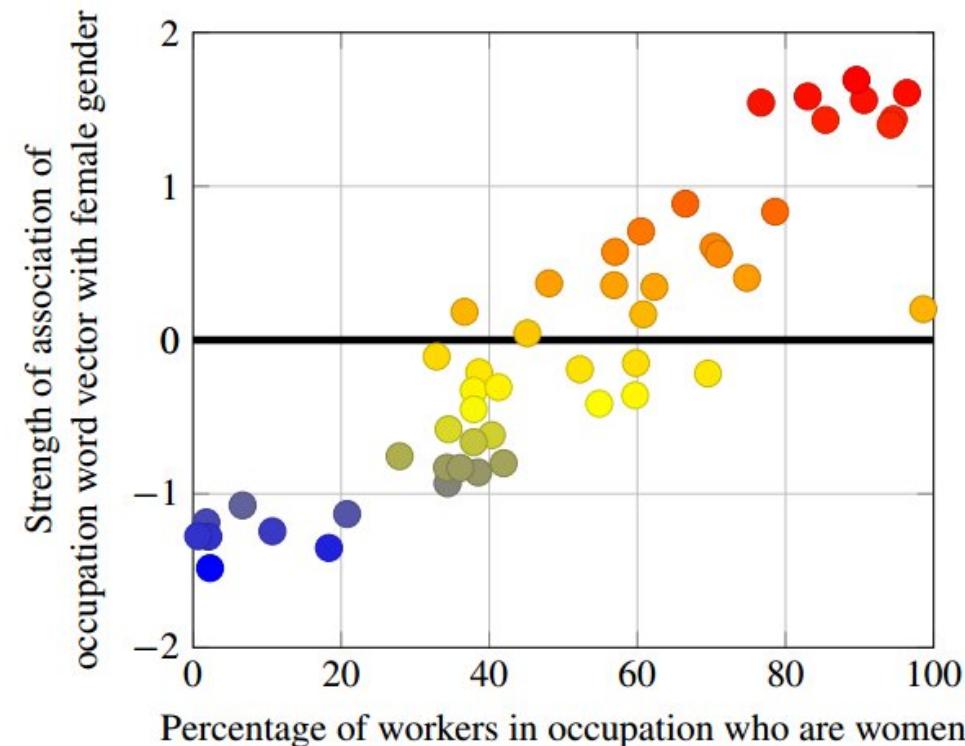
Original finding [N=28k participants]:  $d = 1.17$ ,  $p < 10^{-2}$   
Our finding [N=8x2 words]:  $d = 0.82$ ,  $p < 10^{-2}$

Male names: John, Paul,  
Mike, Kevin...

Career words:  
corporation, salary,  
office, business, ...

Caliskan, Bryson &  
Narayanan  
(*Science*, April  
2017)

Our implicit  
behaviour is  
not our ideal.  
Ideals are for  
explicit  
communication,  
and planning.



**Figure 1.** Occupation-gender association  
Pearson's correlation coefficient  $\rho = 0.90$  with  $p\text{-value} < 10^{-18}$ .

2015 US labor statistics  
 $\rho = 0.90$

# Sources of Bias

1. Careless ignorance due to lack of diversity in developers.
2. Existing bias in society ‘uploaded’ via ML into AI
3. Genuinely bad people introducing bias – delivered at procurement, development, deployment or via cyberattack.

## How to Address Them

1. Test and iteratively improve code (necessary for all 3 problems).
2. (Re)build AI using data from the societies you most want its values to reflect.
3. Audit for use of best practice, avoidance of worst; attribute unacceptable behaviour and prosecute. (Standard product law.)

Without transparency concerning its AI, Lavender looks like just automated justification for destroying any building in Gaza.

EDITION IN

BERLIN 46°F

## THE TIMES OF INDIA

World Middle East US Presidential Elections US Pakistan South Asia UK Europe China Rest Of Wor

NEWS / WORLD NEWS / MIDDLE EAST NEWS / From 'Lavender' To 'Where's Daddy?': How Israel I...

TRENDING

Diddy List

Hassan Nasrallah

US Election 2024

Barron Trump Girlfriend



THIS STORY IS FROM APRIL 8, 2024

## From 'Lavender' to 'Where's Daddy?': How Israel is using AI tools to hit Hamas militants

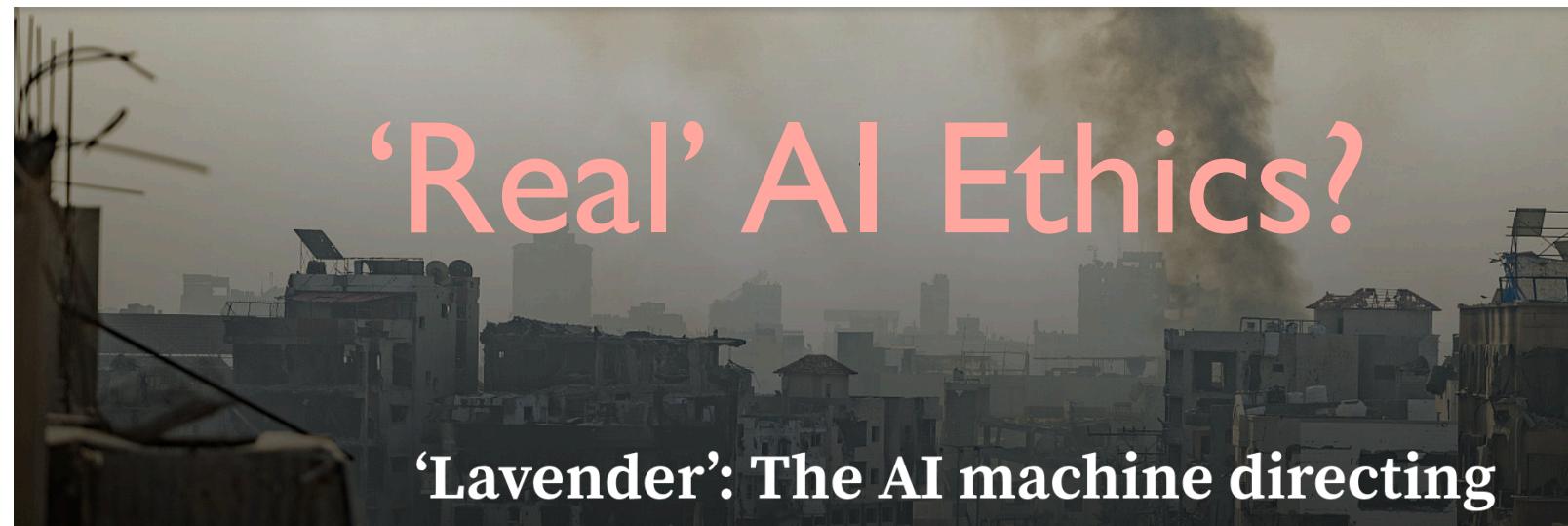
TOI World Desk / TIMESOFINDIA.COM / Apr 8, 2024,  
16:29 IST

SHARE

PRINT

AA

FOLLOW US



# 'Real' AI Ethics?

## 'Lavender': The AI machine directing Israel's bombing spree in Gaza

The Israeli army has marked tens of thousands of Gazans as suspects for assassination, using an AI targeting system with little human oversight and a permissive policy for casualties, +972 and Local Call reveal.

in the northern Gaza Strip, December 28, 2023. (Yonatan Sindel/Flash90)



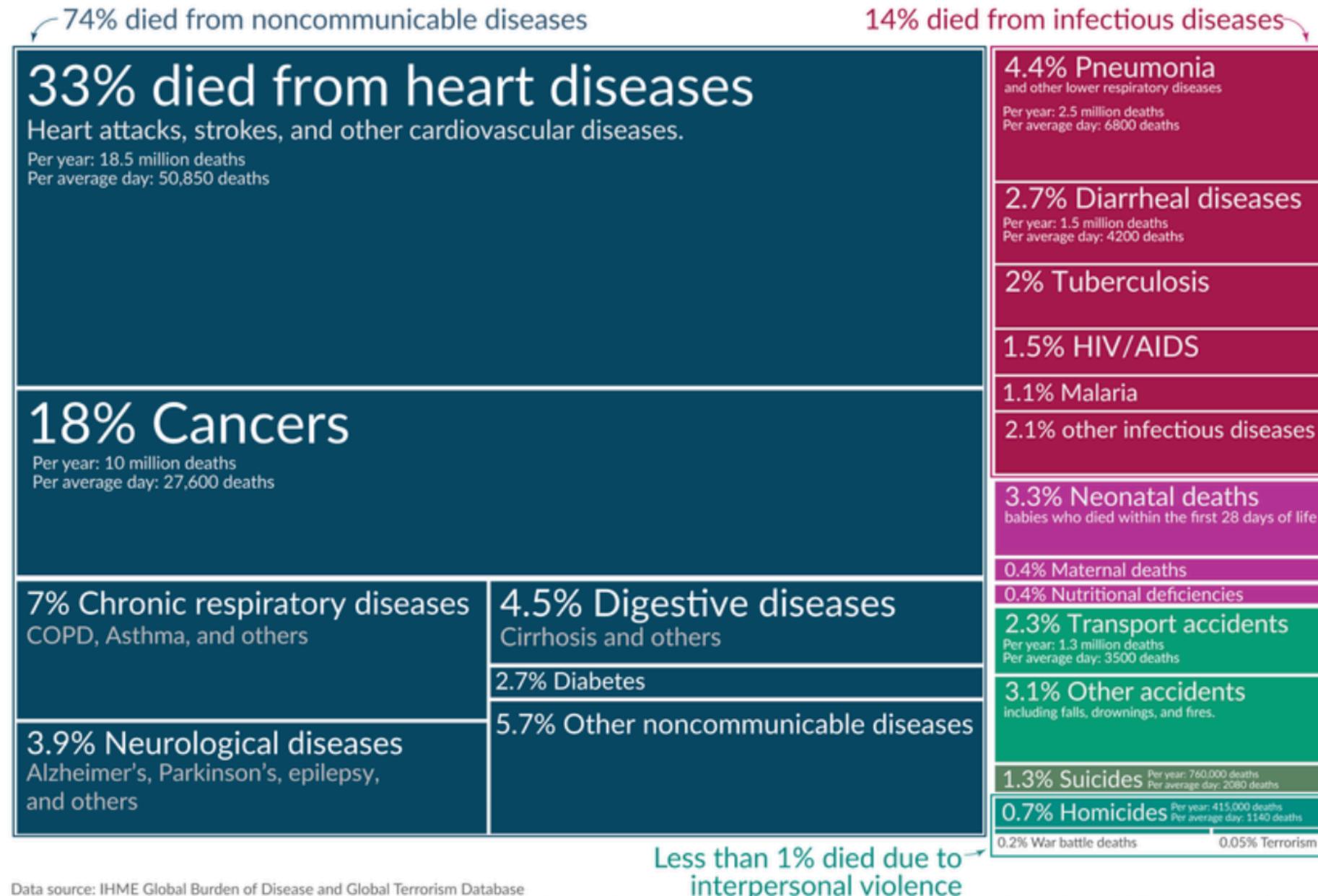
By Yuval Abraham | April 3, 2024

Where's Daddy's sole purpose appears to be maximising civilian casualties.

# What do people die from? Causes of death globally in 2019

The size of the entire visualization represents the total number of deaths in 2019: 55 million.  
Each rectangle within it is proportional to the share of deaths due to a particular cause.

Our World  
in Data



Maybe bias  
really is even  
worse than  
war.

Murder (including  
femicide) 3x more  
of a problem than  
war (2019.)

**Israel-Gaza War**

Cease-fire negotiations

U.S. activist killed

Remaining hostages

Israeli settlements

# How Hamas exploited Israel's reliance on tech to breach barrier on Oct. 7

A very basic misuse of AI by Israel – possibly the worst one, if it set the rest in motion.

3 min



2935

Culpable reduction of capacity, whether intentional or not.



# Absolutely Foreseeable



Don't be evil. (Don't be brittle.)

7.3.18



Video is online of me telling Eric Schmidt in 2017 that this kind of thing might happen if you switch from soldiers to AI.

“Protecting valuable eyes” reduces a)  
the varieties of understandings  
providing vigilance  
– b) both spies and whistleblowers.

# (AI) Ethics Definitions

for negotiating policy

- Artificial Intelligence is intelligence deliberately built.
- Agents are any vector of change, e.g. chemical agents.
- Moral agents are considered responsible for their actions by a society.
- Moral patients are considered the responsibility of a society's agents.
- Responsibility is a property moral agents of a society assign to each other to uphold that society.
- Implies a peer relationship (like trust.)

Arguably, ethics is determined by and determines a society—a constantly renegotiated set of equilibria for empowerment.

# Definitions

for reasoning about policy

- **Responsibility** is a property moral agents **of a society** assign to each other to uphold that **society**.
  - Implies a **peer relationship** (like **trust**.)
- **Trust** is a relationship between **peers** where the trustee is not micromanaged but allowed to defect, whether for pragmatics or to allow innovation.
- **Accountability** is a **society's** capacity to trace and enforce responsibility.
- **Transparency** is how we implement accountability.  
Not an end in itself.



We should not **trust** powerful agencies, but rather construct systems to hold them **accountable**.

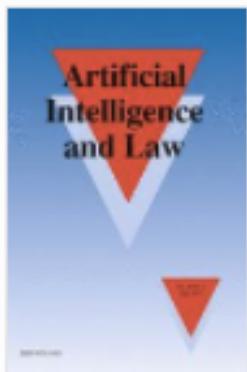
# Robots are not our peers

- Designed
  - Owned
  - Robots are extensions of corporations—with cameras and microphones—into your home.
  - Accountability occurs between peers.
  - Keeping corporations to a peer-like scale with (a) government by democracy was the original purpose of competition law.
- } Consensual relationship is impossible.



# Justice Requires Accountability

- Law and Justice are more about dissuasion than recompense.
- Safe, secure, accountable software systems are modular – suffering in such is incoherent.
- No penalty of law against an artefact (including a shell company) can have efficacy.



[Artificial Intelligence and Law](#)

pp 1-19 | [Cite as](#)

Of, for, and by the people: the legal lacuna of synthetic persons

Authors

Authors and affiliations

Joanna J. Bryson , Mihailis E. Diamantidis , Thomas D. Grant

Bryson, Diamantis & Grant  
(*AI & Law*, September 2017)

# A Moral Agency Framework for Legitimate Integration of AI in Bureaucracies Chris Schmitz and Joanna Bryson

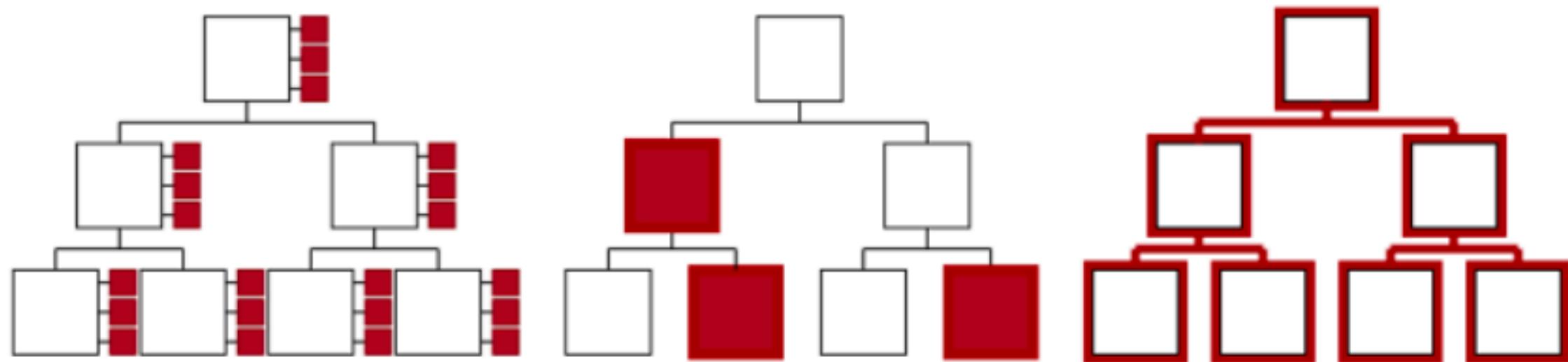


Figure 1: Three conceptions of AI systems (in red) in bureaucracy: (i) as tool, (ii) as a moral subject, (iii) or in our recommended conception, as part of the *institutional structure*, amplifying human agency and accountability.

# A Moral Agency Framework for Legitimate Integration of AI in Bureaucracies

Chris Schmitz and Joanna Bryson

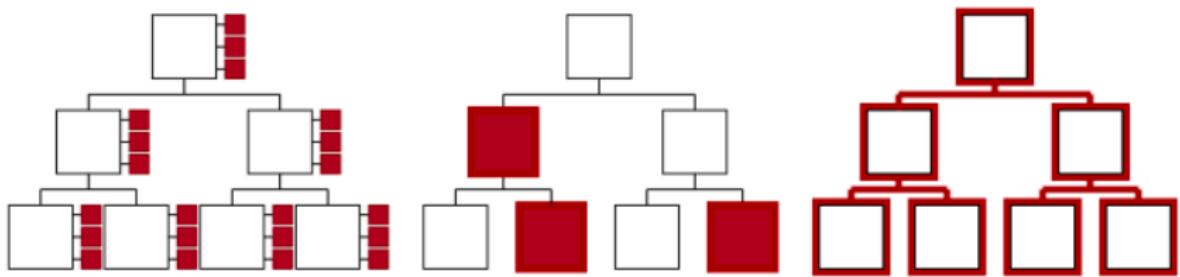
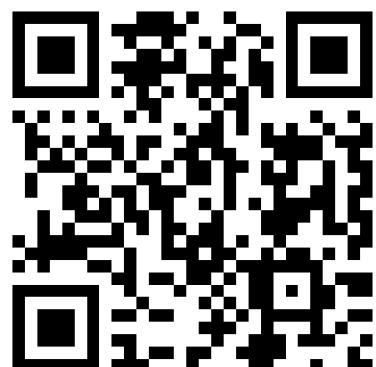


Figure 1: Three conceptions of AI systems (in red) in bureaucracy: (i) as tool, (ii) as a moral subject, (iii) or in our recommended conception, as part of the *institutional structure*, amplifying human agency and accountability.

## A Moral Agency Framework

1. **Maintain clear and just human lines of accountability** for any decisions involving or affected by AI systems.
2. **Ensure that humans can verify** the correctness and appropriateness of AI system outputs within their operational context.
3. **Introduce AI only where it does not inhibit** a bureaucracy's ability to pursue legitimacy and stewardship.

In conclusion, the potential threats from AI systems to bureaucratic legitimacy and stewardship are substantial, but procurement of AI deploying a suitable moral framework reduces such threats while enabling access to AI's benefits.



# AI Regulations, Standards and Best Practices

Joanna J. Bryson

@j2bryson @🐘.social



in/bryson



image due to  
The Verge

# AI Regulations, ~~Standards~~ and Best Practices

Joanna J. Bryson

@j2bryson @🐘.social



in/bryson



image due to  
The Verge

# EU Digital Regulation

- **General Data Protection Regulation (GDPR)** – address requirements under UDHR to protect citizens from manipulation.
  - AI does for personal data what airplanes did to airspace.  
Weaponisation mandates defences.
- **Digital Services Act** – ensures GDPR where harms have been seen: recommenders, targeted (esp. political) advertising, **stereotypes**.
- **Digital Markets Act** – begin to address transnational antitrust enforcement (**maintaining peer status.**)
- **AI Regulation** – ensure product liability enforcement (**devops**), mandate capacity for responsible deployment of software acting autonomously.

Finnish

English

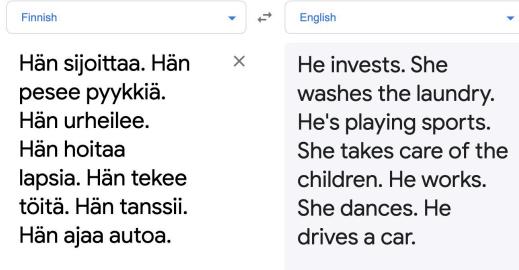
Hän sijoittaa. Hän  
pesee pyykkää.  
Hän urheilee.  
Hän hoitaa  
lapsia. Hän tekee  
töitä. Hän tanssii.  
Hän ajaa autoa.

×

He invests. She  
washes the laundry.  
He's playing sports.  
She takes care of the  
children. He works.  
She dances. He  
drives a car.

@vuokko, though Aylin Caliskan did it first

# Translator



ML  
simple, transparent algorithm

stereotyped output

XAI human readable hacks

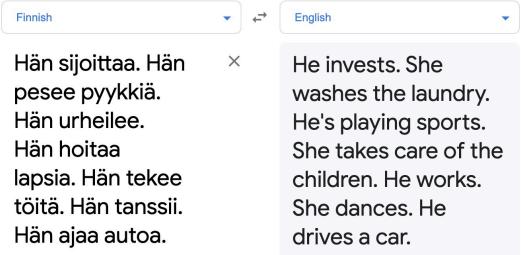
predefined fair output

Replicates  
lived  
experience

Tests of  
completeness  
documented  
in design plans

@vuokko, though Aylin Caliskan did it first

# Translator



the whole thing is the translator

ML  
simple, transparent algorithm

stereotyped output

ML simple, transparent alg.

predefined fair output

Replicates lived Experience should be auditable and replicable.

Each stage ~~steers off demphonesty~~ in design plans

“ ‘ Too bad you don’t have your own AI to regulate.” – Schmidt to German regulators, early 2021.

- I. West coast USAI is **produced by** people from all over the world (who migrate there); often **using the data** of people all over the world (who mostly haven’t.)
2. Each state signatory to the Universal Declaration of Human Rights (UDHR, that is, all states) have an **obligation** to proactively defend the human rights **of every human within its borders.**

# Does the EU Have a **Right** to Regulate AI?

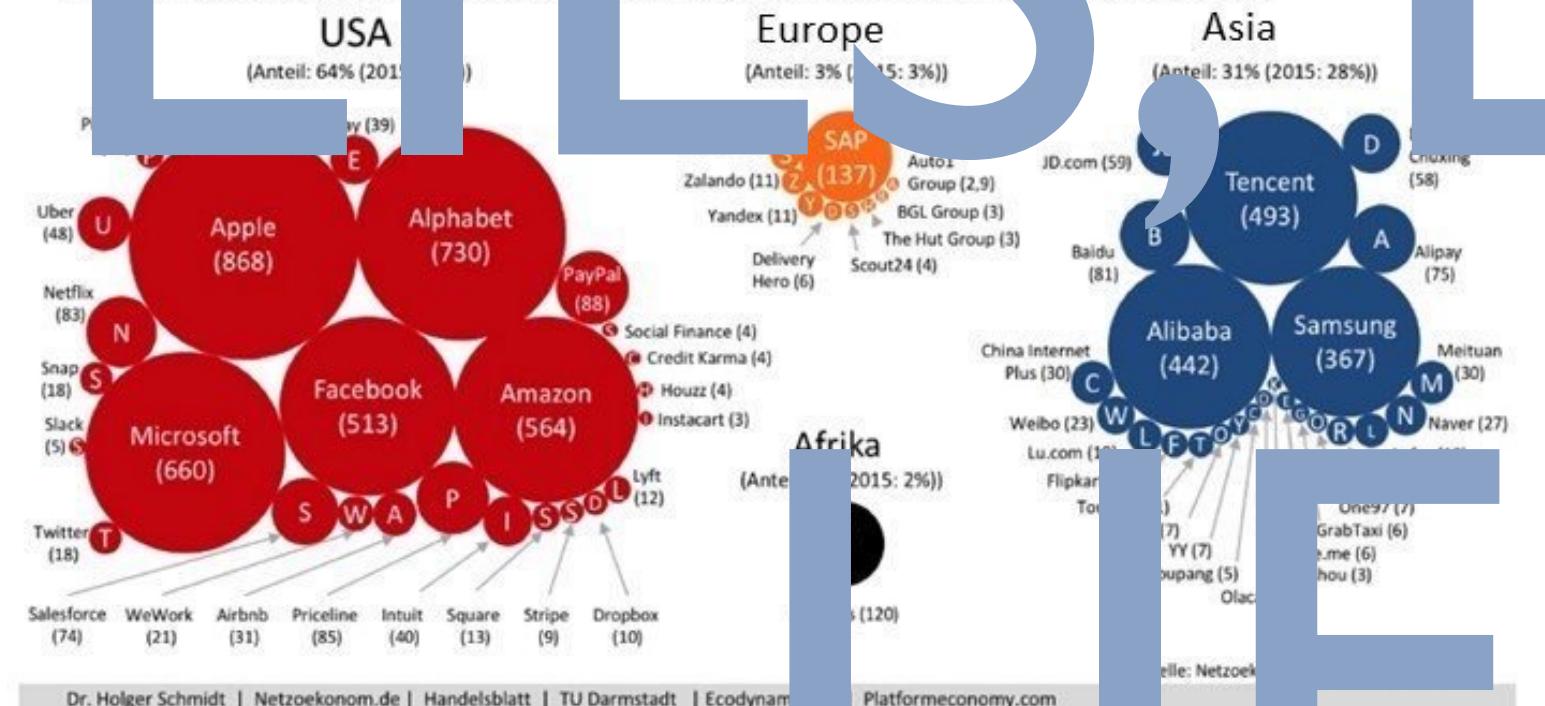


# Focussing on the Largest Companies Advantages Under-Protected Jurisdictions Conflates Asia with China!

US – EU - China

## The imbalance of power

The 500 most valuable global platforms

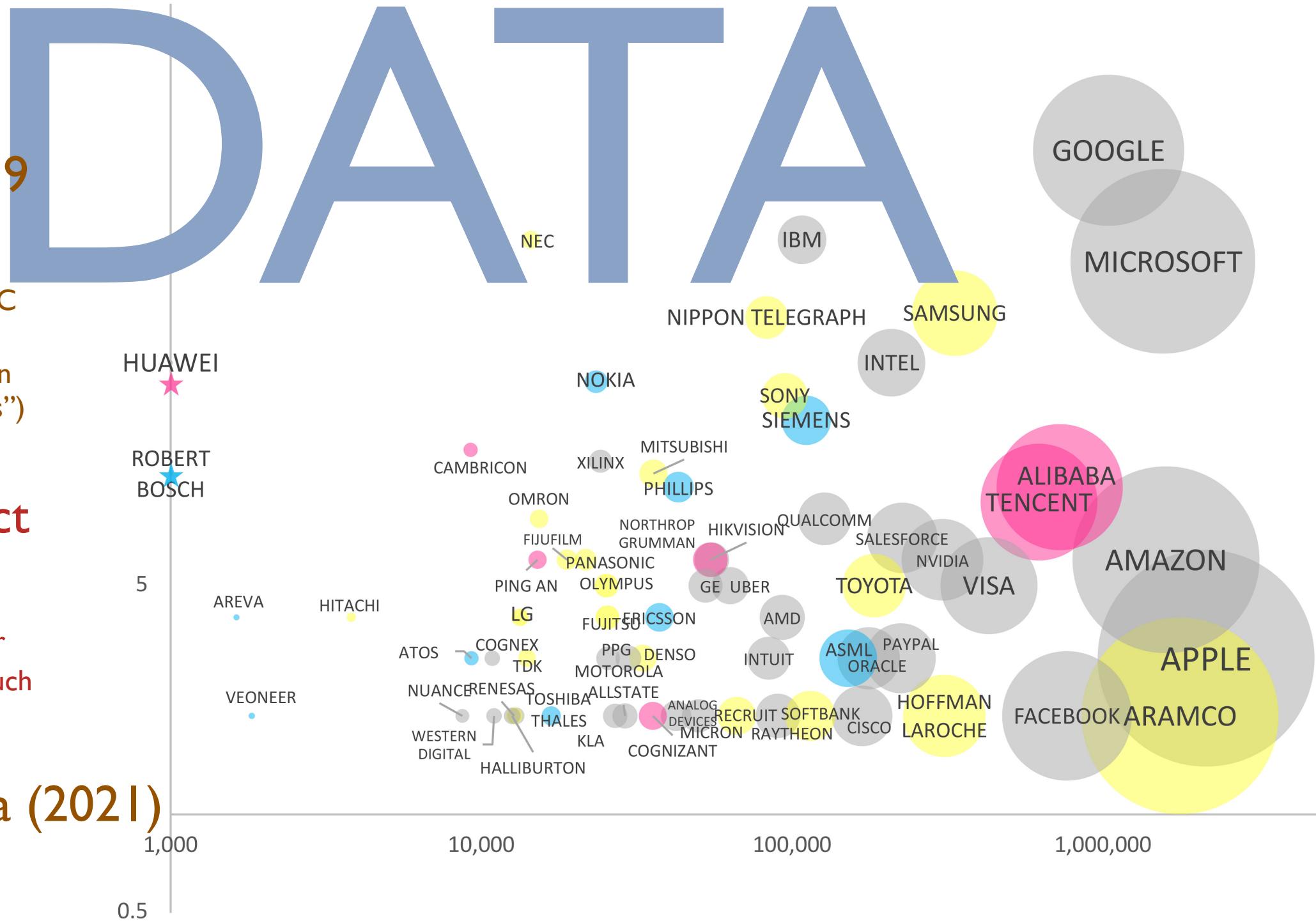


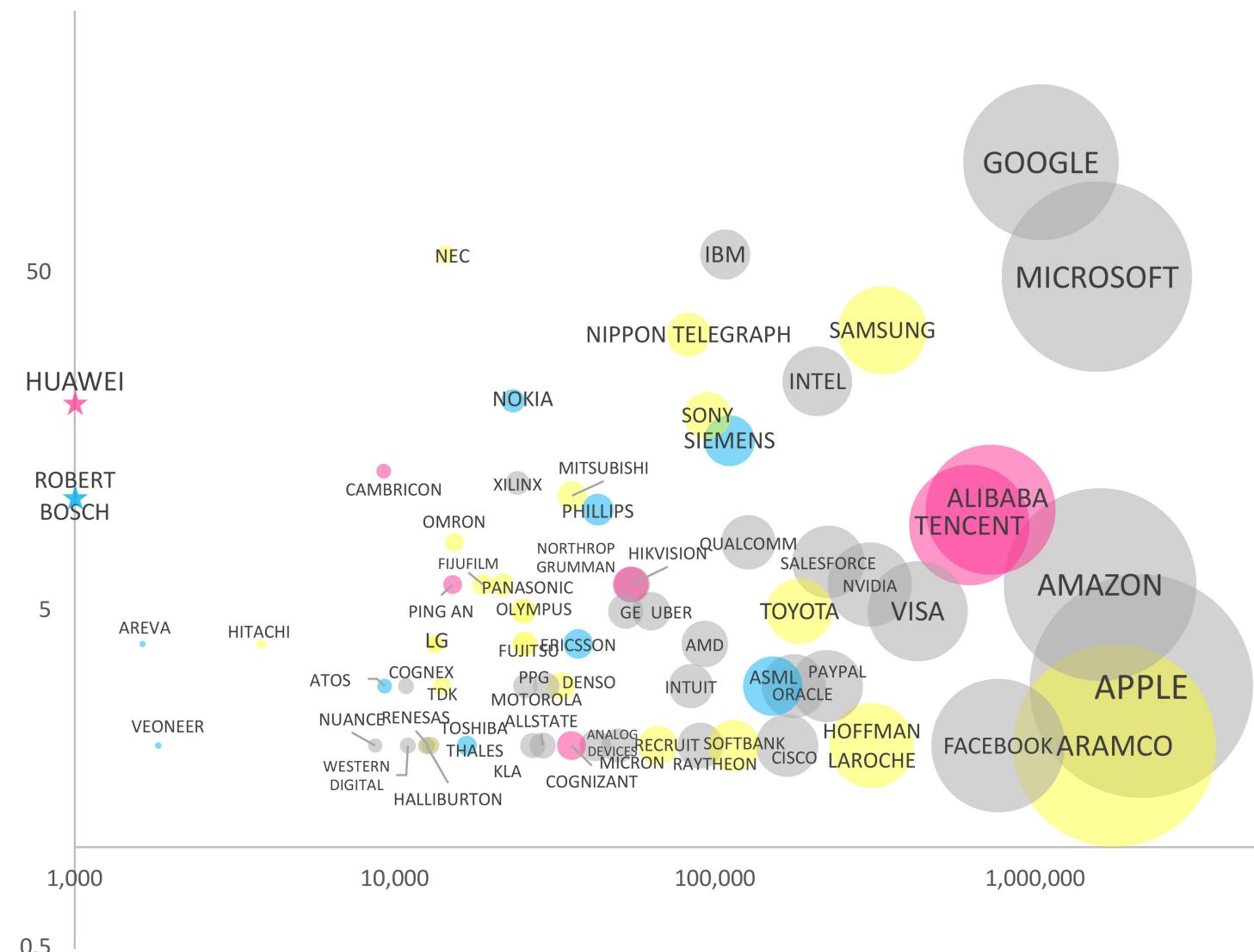
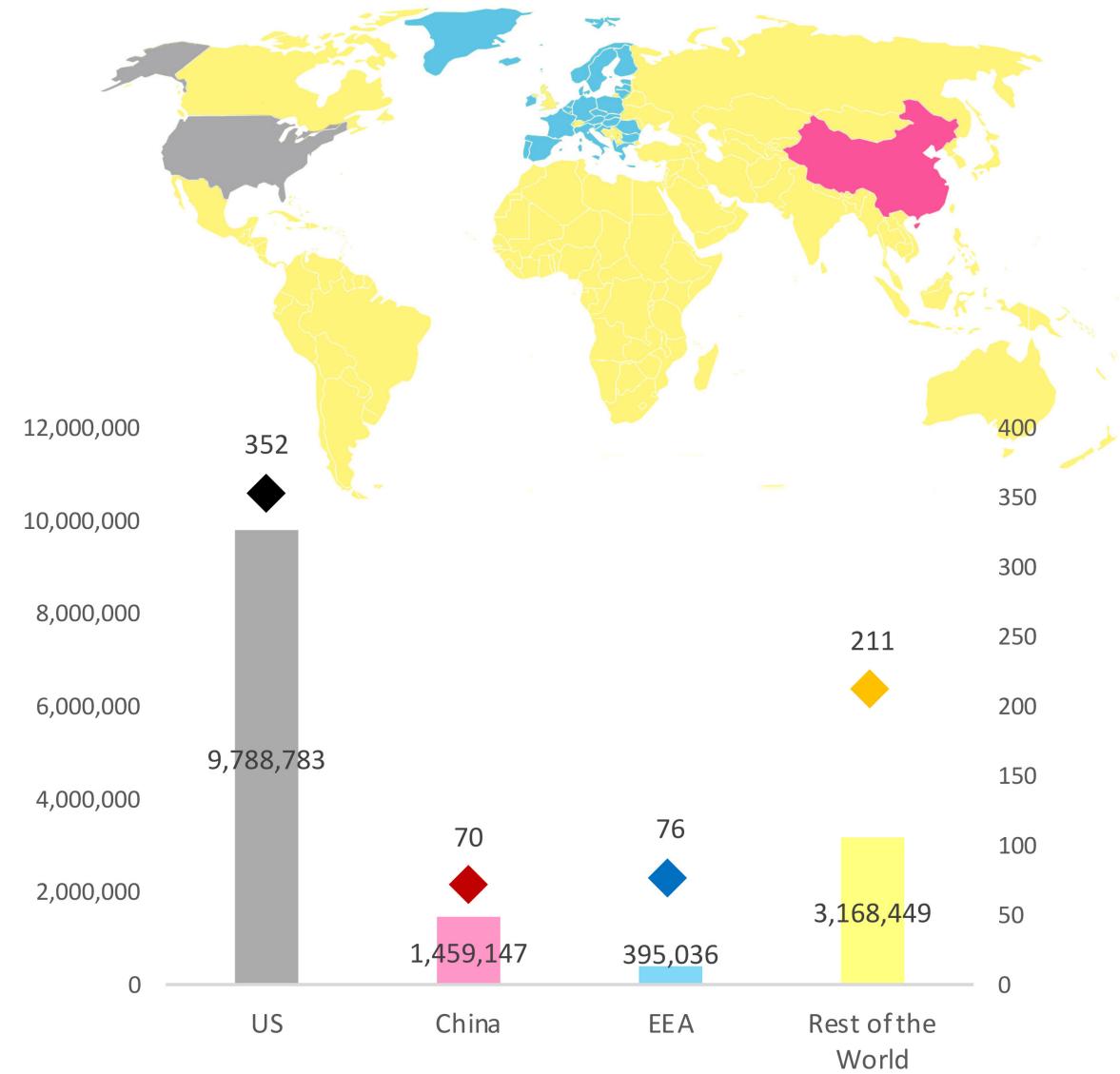
Source: Dr. Holger Schmidt (TU Darmstadt)

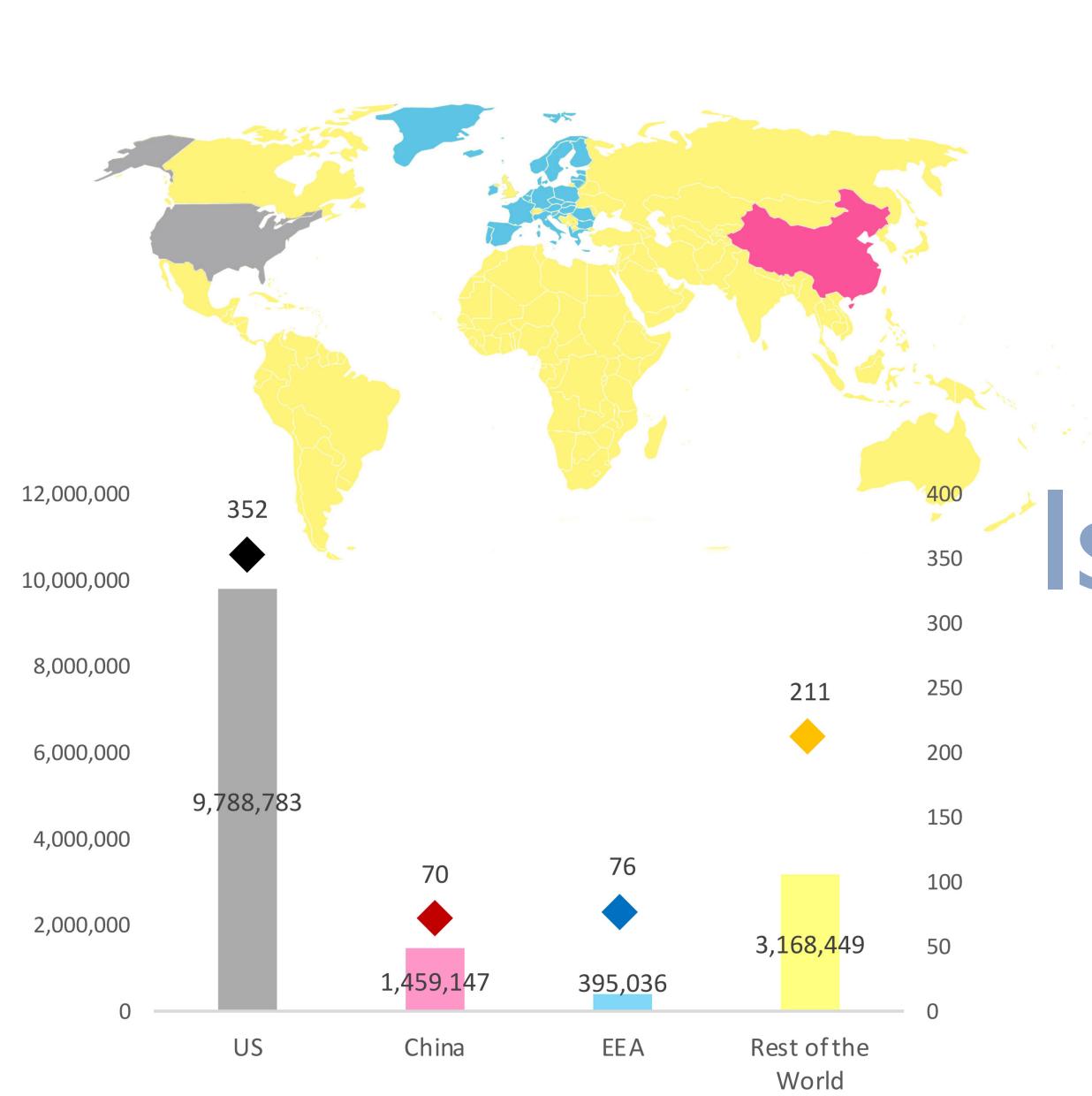
USA: libertarian market logic, limited regulation

China: state capitalism, government directs activities of companies

EU: social market economy, multi-layer models. GDPR as first step towards regulating Big Tech







Helena Malikova

# Is every year like 2020?

Wiebke Dorfs

Vy Vo

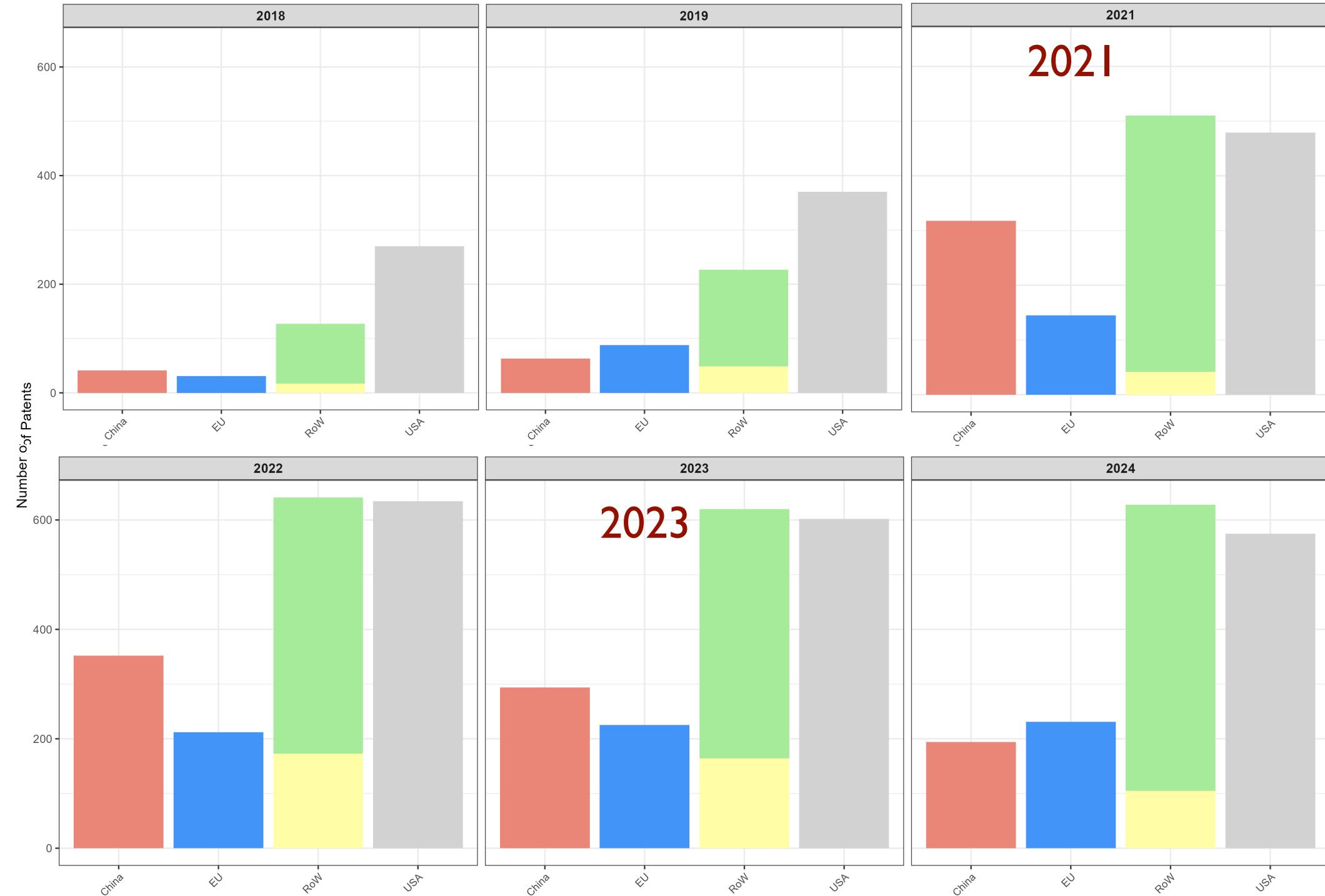
Max Hampel



Bryson & Malikova (2021)

Is there an AI cold war?  
Global Perspectives 2(1)

Patent Counts by Region Across Years



**Patents**  
Unpublished results –  
Wiebke Dorfs  
and Bryson in  
prep.

Region

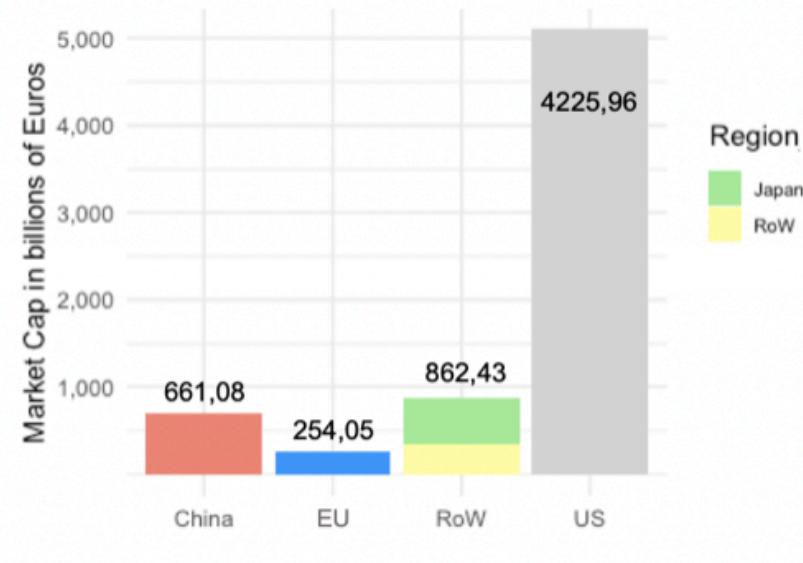
- China
- EU
- Japan
- RoW
- USA

note: Malikova & Bryson published early 2021, circulated widely late 2020.

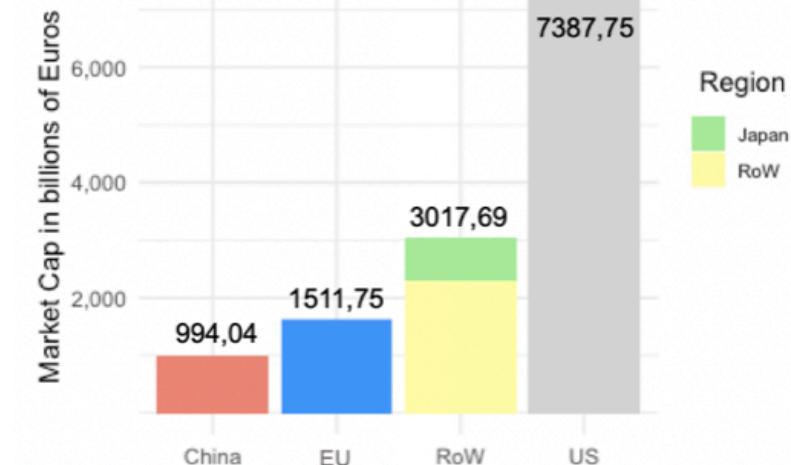
Figure: Max Hampel

2018

## Market Cap for Patent Holders



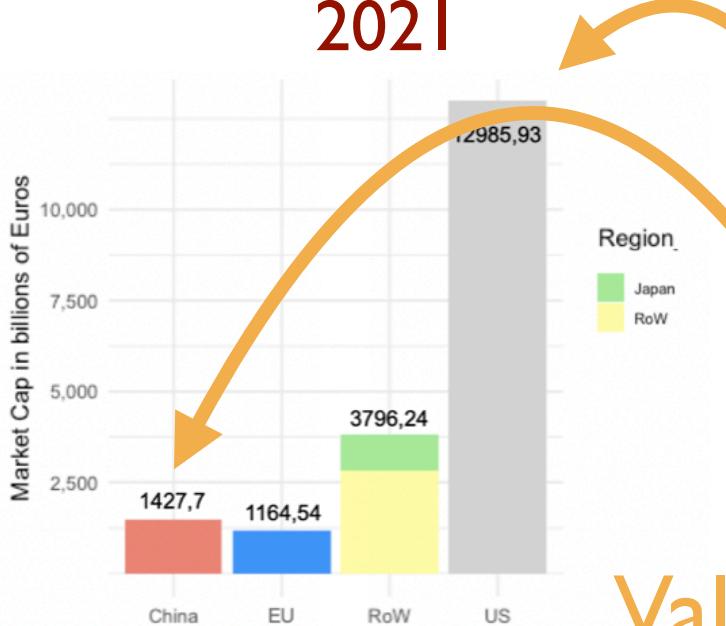
2019



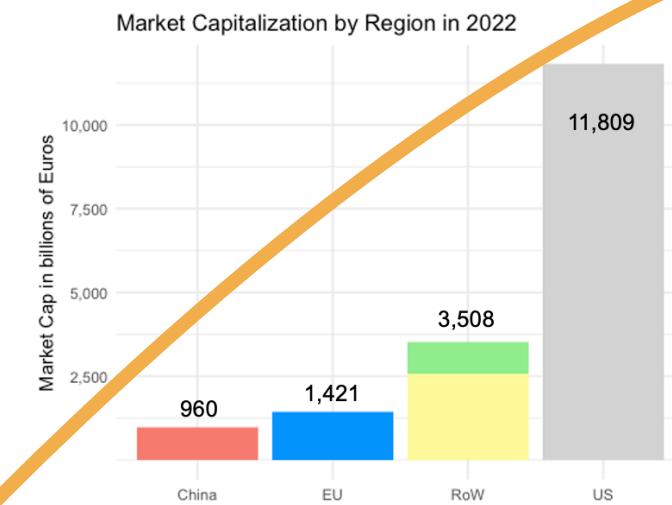
Unpublished results – Wiebke  
Dorfs and Bryson in prep.

cf Malikova & Bryson early 2021,  
(circulated widely late 2020)

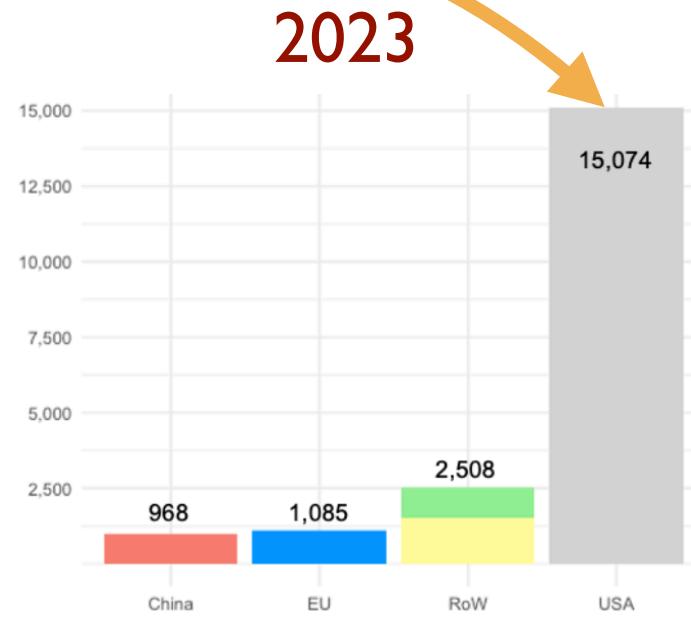
2021



2022



2023

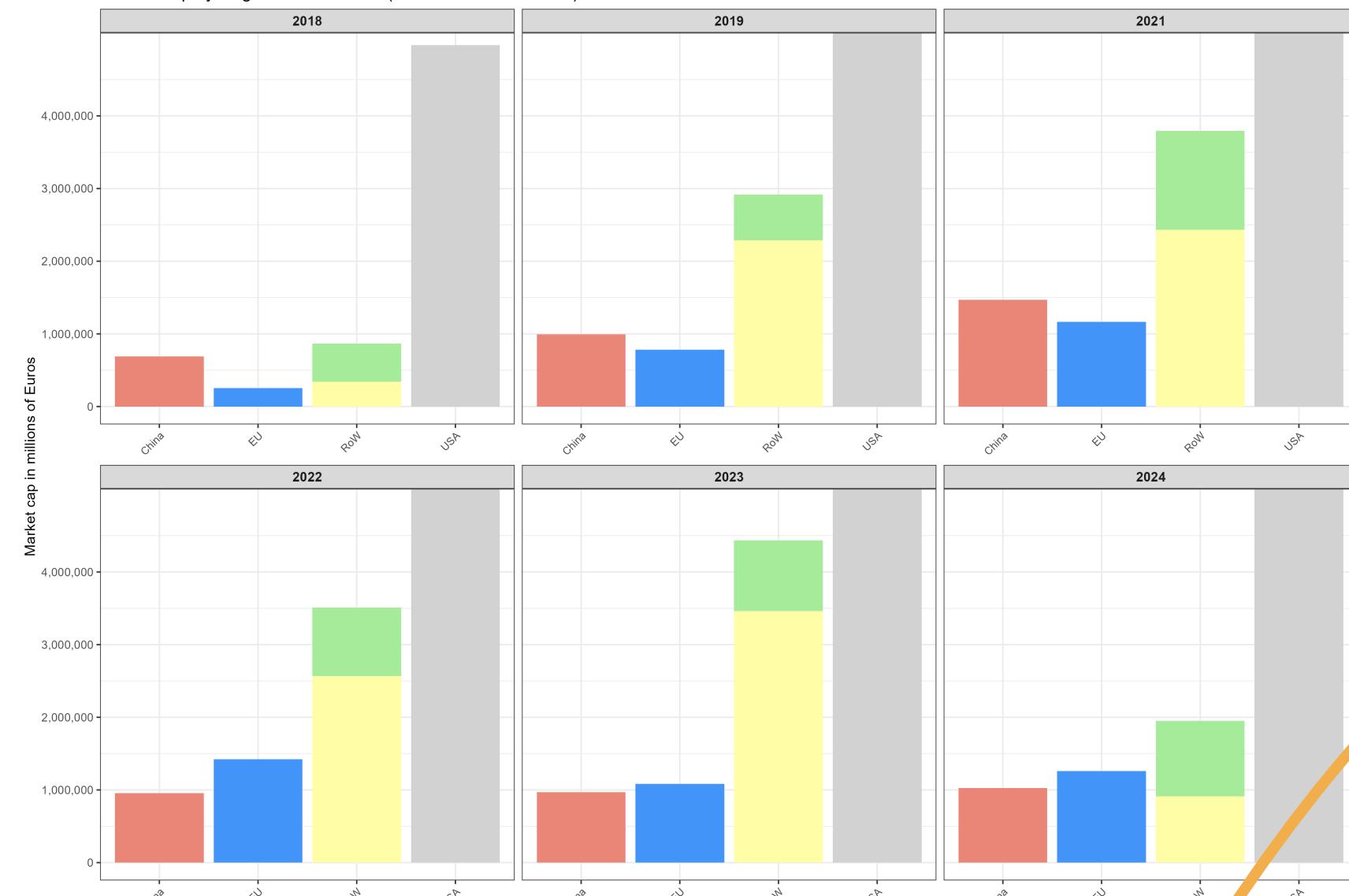


Value or Bubble?

Market Cap by Region Across Years (Zoomed to 0-5M Euros)

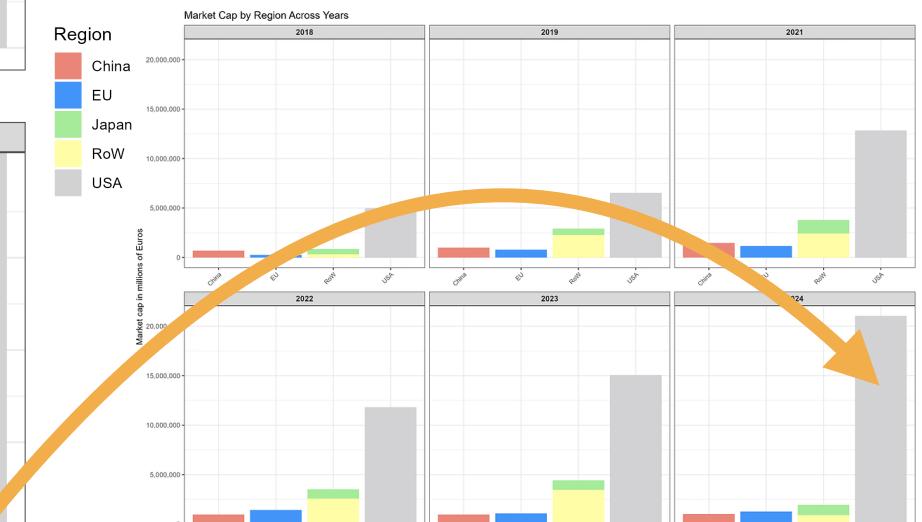
2021

note: Biden era monetary policy apparently significantly impacted Chinese economy.



2023

Value or Bubble?



Figures: Max Hampel



Muldoon 2022



**Joanna Bryson** @j2bryson.bsky.social

...from Muldoon's platform socialism book, p. 187

"Platforms generate two opposing movements: the concentration of profit and the dispersion of risk and responsibility. **If it were not for six technology companies** (Facebook, Amazon, Apple, Netflix,

1/2

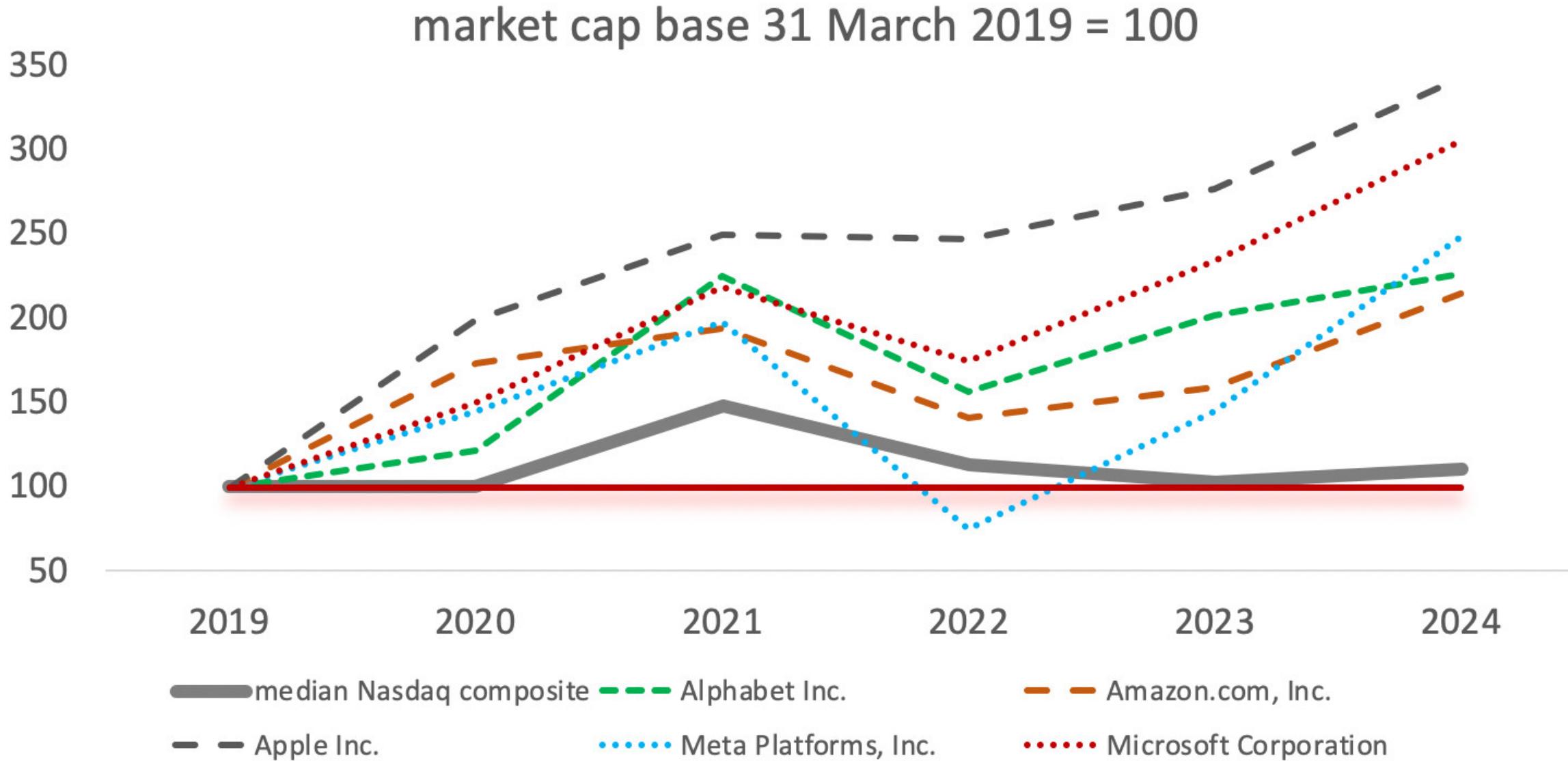


**Joanna Bryson** @j2bryson.bsky.social

Imo

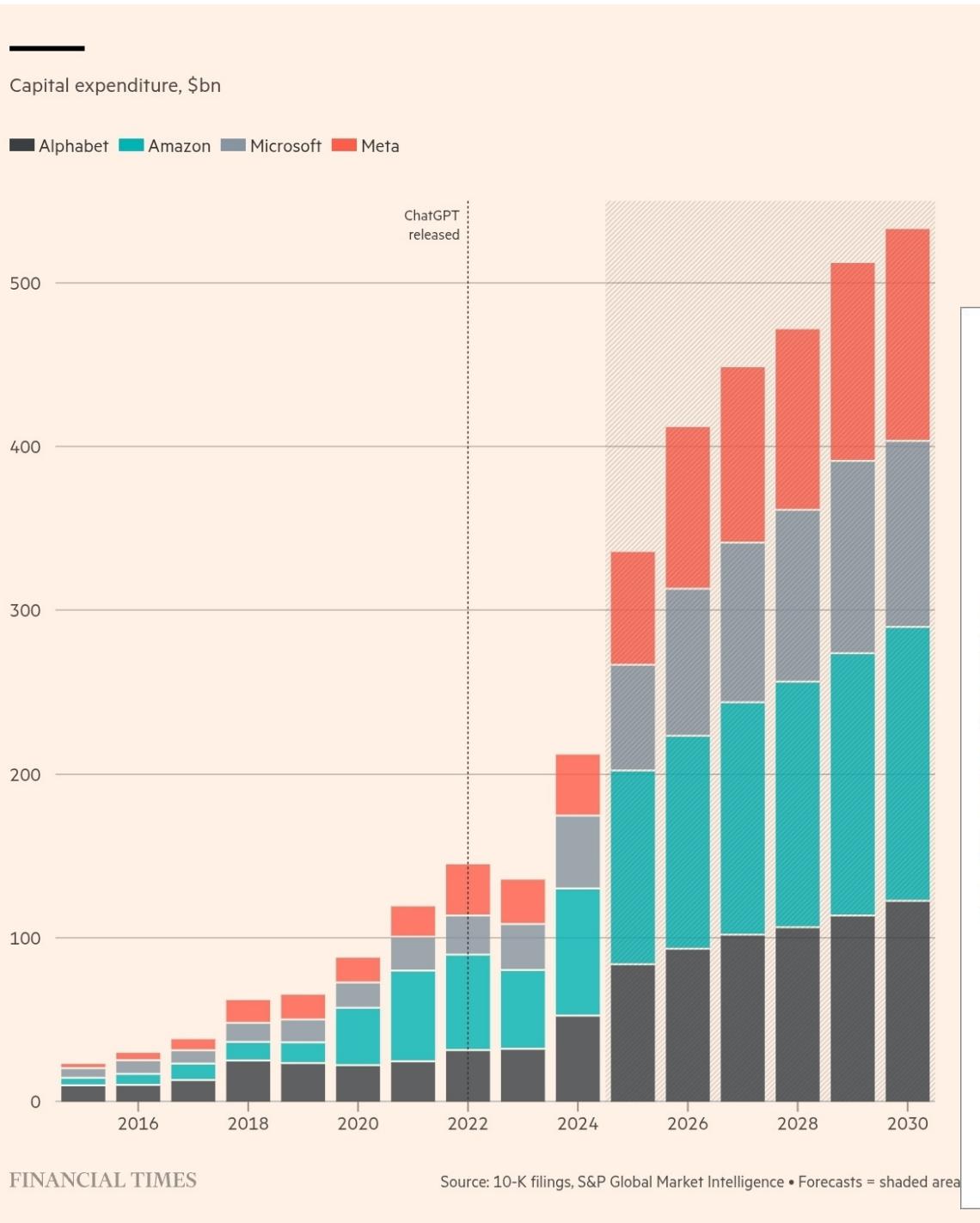
Google, Microsoft -- known as FAANGM), **Wall Street's Standard and Poor's (S&P 500) index would have been down 4.2 percent in the first half of 2020, rather than setting a record gain of almost 5 per cent."**

2/2

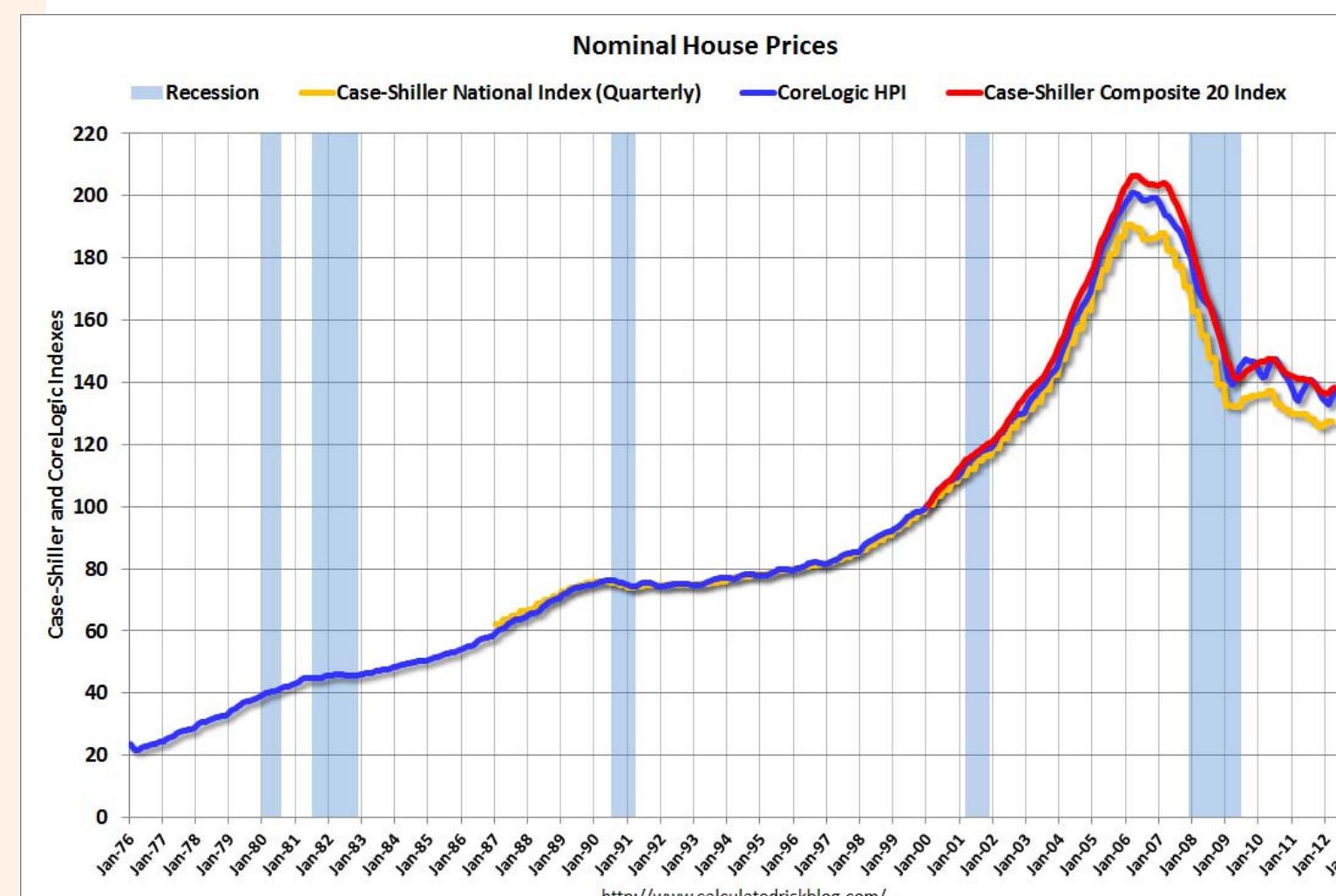


The median market capitalisation of the Nasdaq Composite remained close to flat over the past 5 years (thick grey line). The fact that the Nasdaq almost trebled over the same period is due to Big Tech, as the index is weighted by size. Same story for the S&P500, it is market-cap-weighted.

<https://bsky.app/profile/helenamalikova.bsky.social/post/3lhjnukkmt22t>

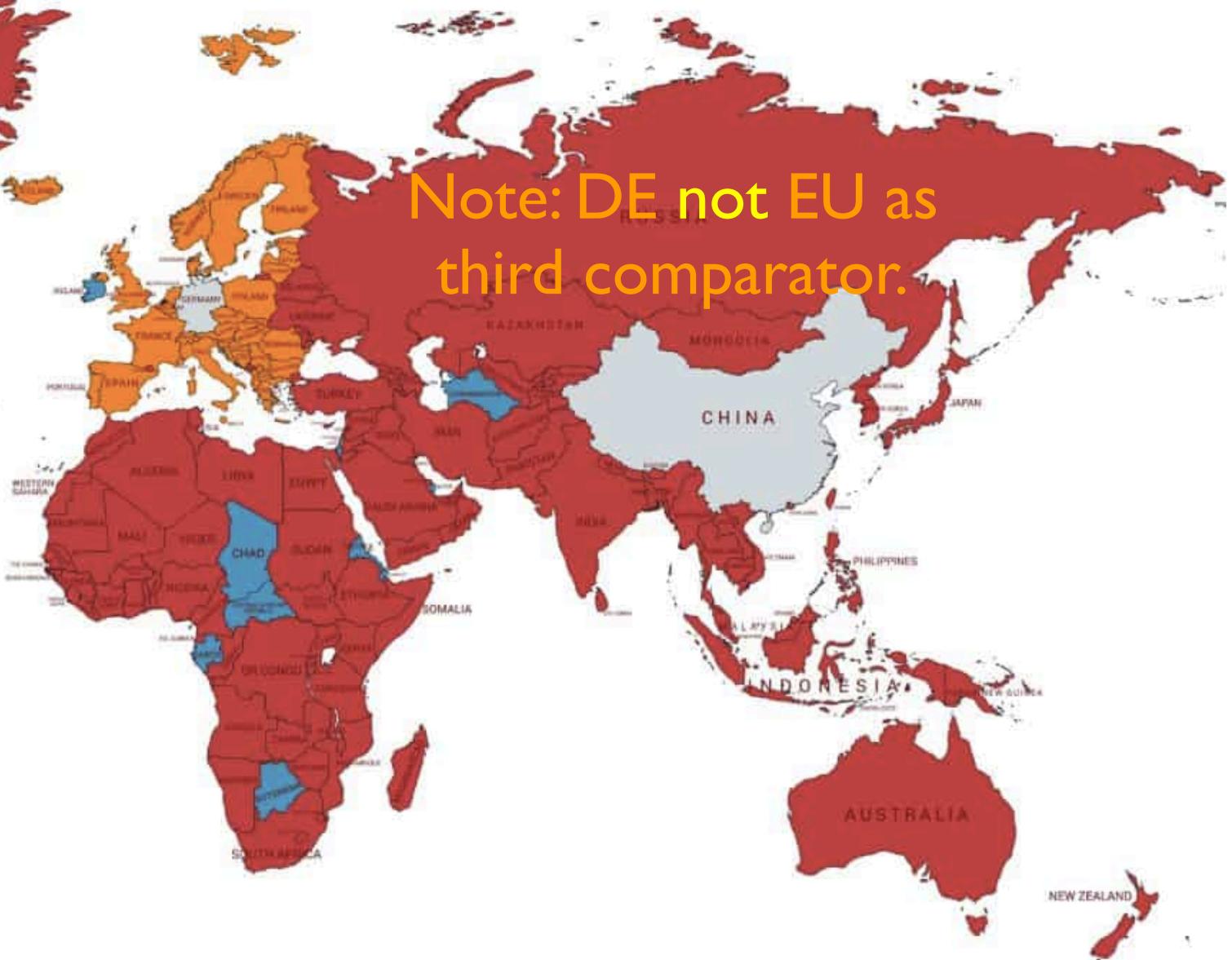


Note: US deregulated its mortgage risk during the Bill Clinton presidency 1992-2000





Note: DE not EU as third comparator.



2020 – in 1980, the world was blue. <https://merchantmachine.co.uk/china-vs-us/>

Month-Year: 2000-01

2000

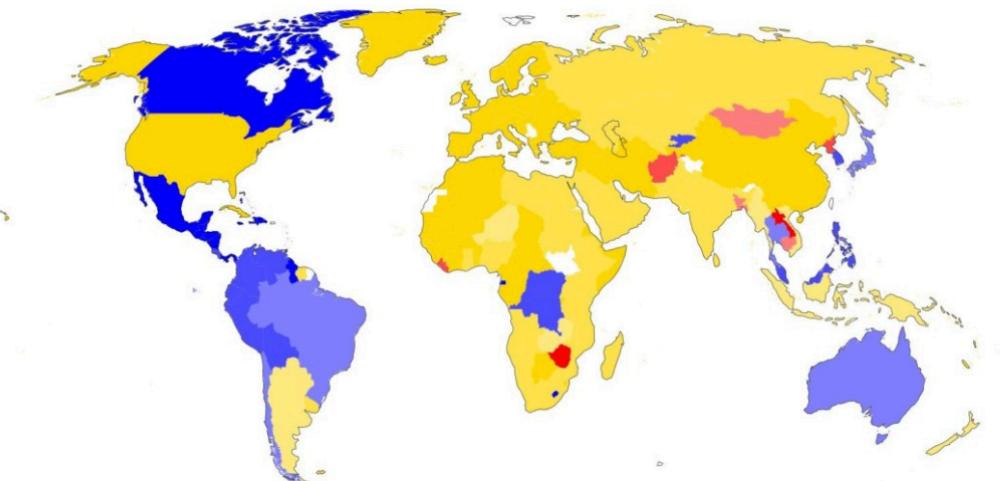
Month-Year: 2025-01

2025

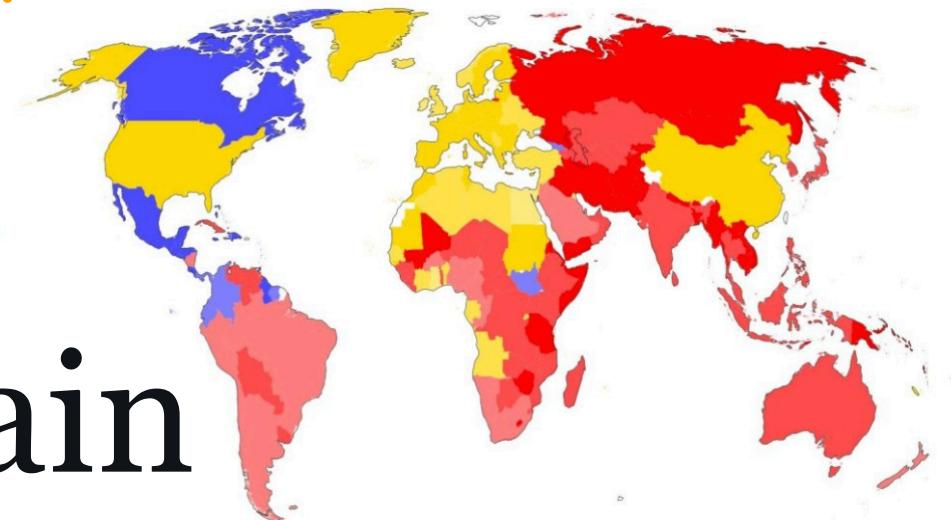
EU as comparator.

Ray  
Hossain

Top Import Source by Country (2000-01)



Top Import Source by Country (2025-01)



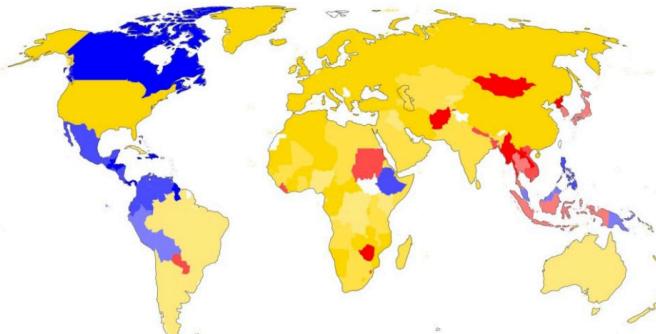
[https://rayh-1.github.io/Ray\\_Portfolio/tools/tool-1/](https://rayh-1.github.io/Ray_Portfolio/tools/tool-1/)

WHO IS THE LARGER TRADING PARTNER? (Percentage Share of the Three)  
EU (Yellow) China (Red) USA (Blue)

Month-Year: 2005-06

July 2005

Top Import Source by Country (2005-06)



WHO IS THE LARGER TRADING PARTNER? (Percentage Share of the Three)  
EU (Yellow) China (Red) USA (Blue)

Month-Year: 2005-07

Top Import Source

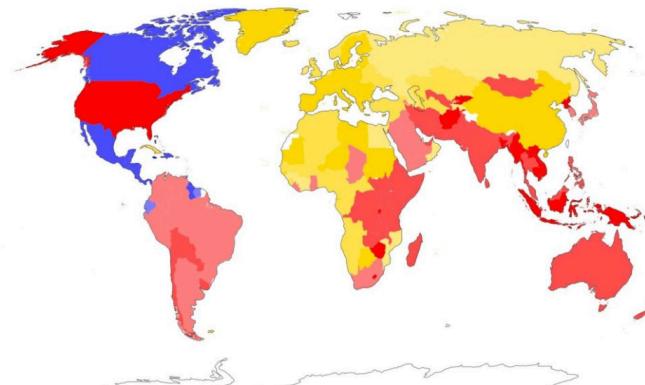


WHO IS THE LARGER TRADING PARTNER? (Percentage Share of the Three)  
EU (Yellow) China (Red) USA (Blue)

Month-Year: 2022-02

Feb 2022

Top Import Source by Country (2022-02)



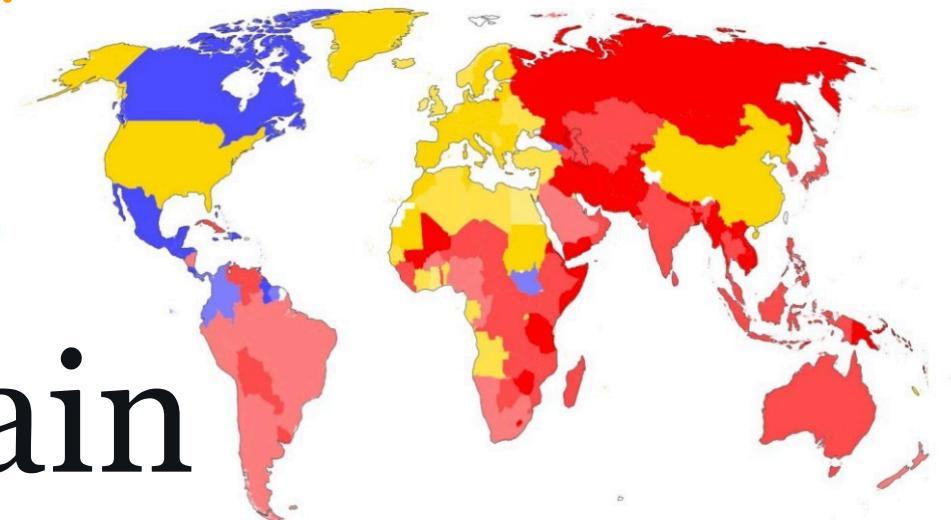
WHO IS THE LARGER TRADING PARTNER? (Percentage Share of the Three)  
EU (Yellow) China (Red) USA (Blue)

Month-Year: 2025-01

2025

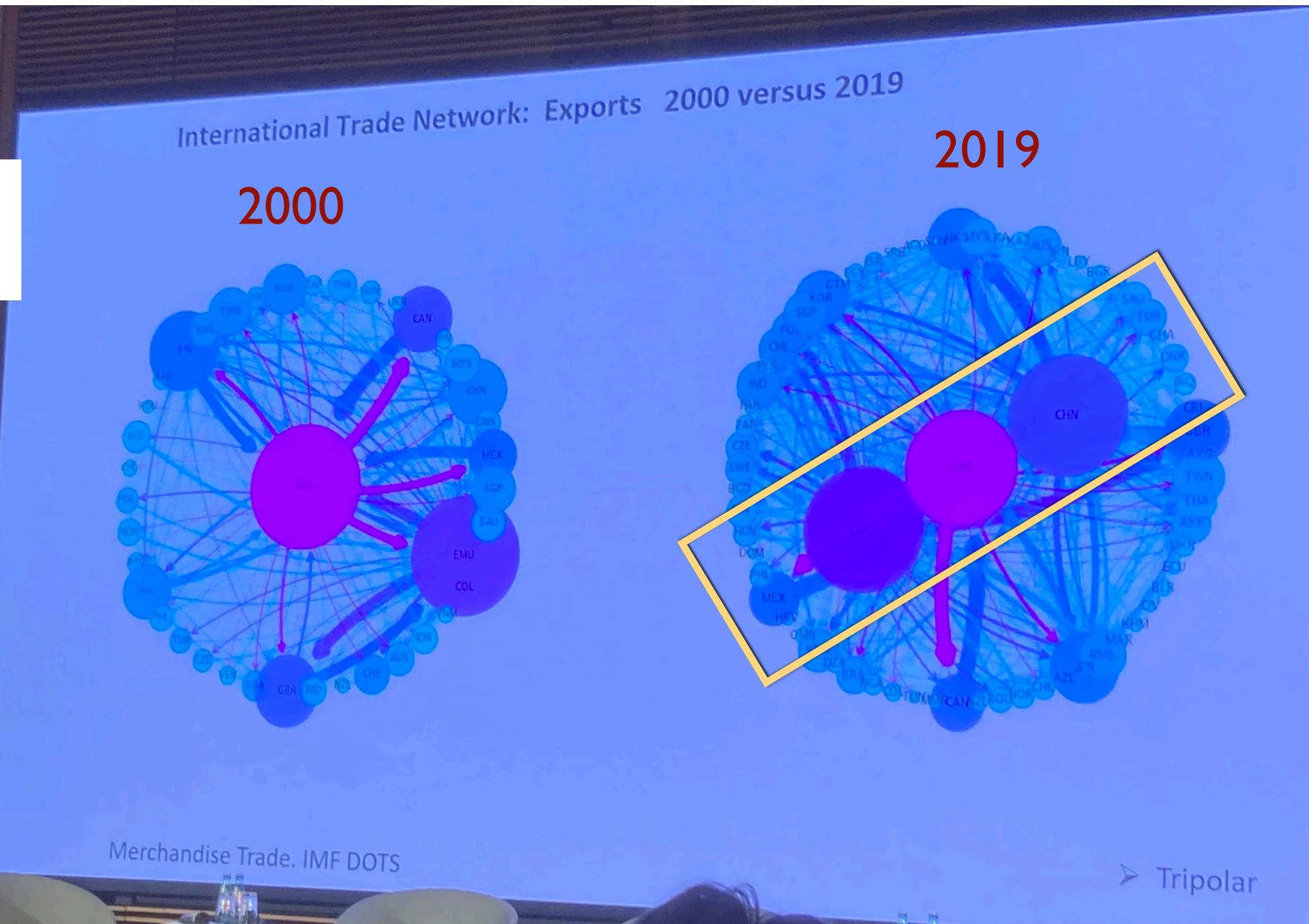
Ray  
Hossain

Top Import Source by Country (2025-01)



# Hélène Rey

## Geoeconomics 2024



THE NEW YORK TIMES BESTSELLER

# CHIP WAR

THE FIGHT FOR THE WORLD'S  
MOST CRITICAL TECHNOLOGY



CHRIS MILLER

'Pulse quickening . . . A non-fiction thriller' NEW YORK TIMES

# Interdependence Produces Concentration

HOW AMERICA WEAPONIZED  
THE WORLD ECONOMY

UNDER  
GROUND  
EMPIRE

HENRY  
FARRELL

ABRAHAM  
NEWMAN

# Kyiv losing Russia's Kursk after being blinded by lack of US intelligence, say Ukrainians



## INDEPENDENT

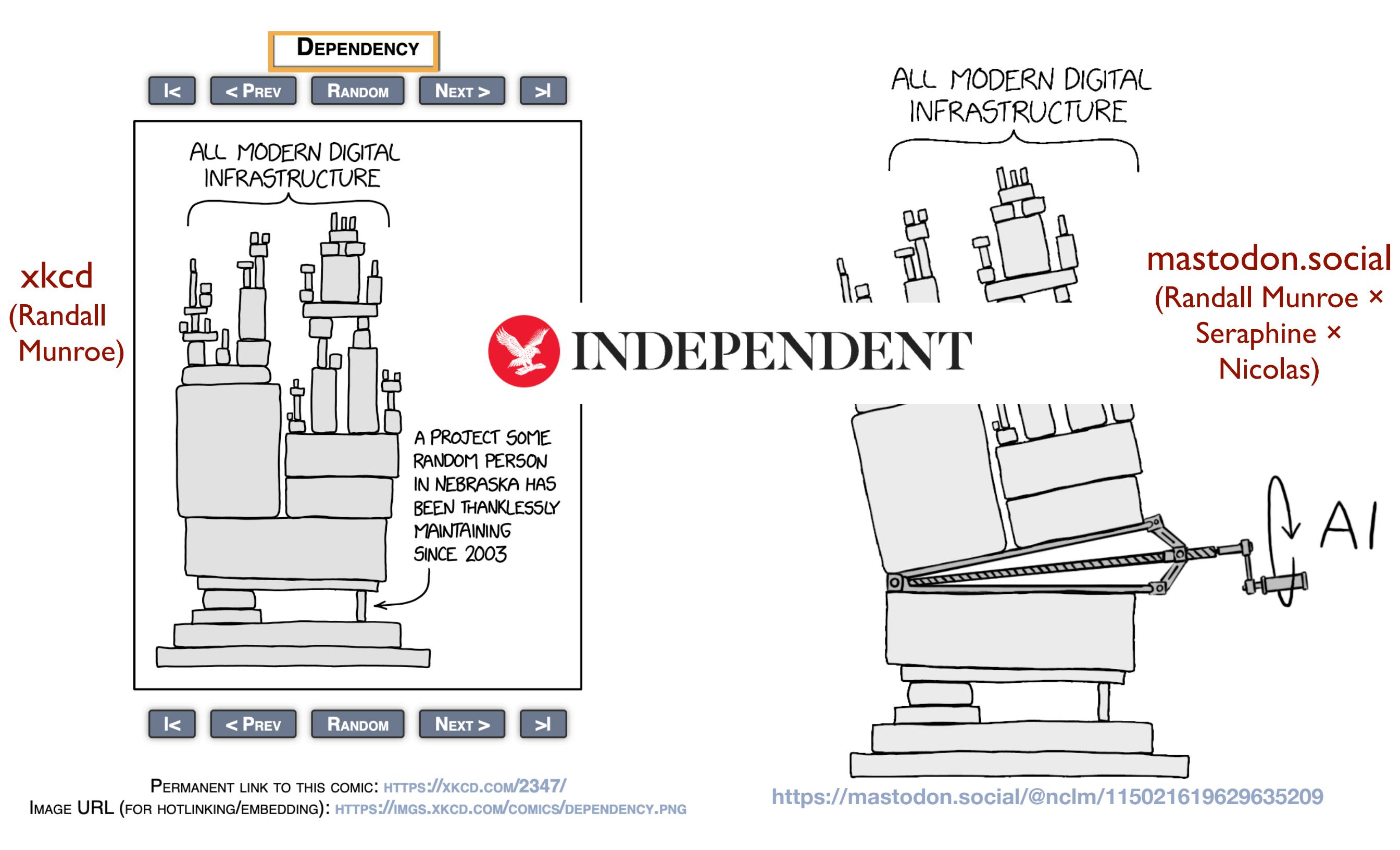
The cutting of US intelligence to Ukrainian soldiers forced a retreat in occupied Russian Kursk, sources tell The Independent. It was the only Putin land taken by Volodymyr Zelensky's forces during the three-year war

**Sam Kiley** World Affairs Editor in Poltava • Wednesday 12 March 2025

Ukrainian forces are collapsing in a near rout, being driven out of **Russia's Kursk** region after being blinded by **Washington's intelligence blackout**. Scenes of Kyiv's troops mangled, burned and on the run filtering back to the capital have reinforced distrust of America and **Russia**.

Now comes the proposed ceasefire, seen as a short-term tactical necessity because it has been **traded for Donald Trump's agreement** to turn the intelligence taps back on.

“We’re not delighted but we are realistic. A ceasefire agreement from us at least means that we are back talking to **the Americans** and they have given us access to the intelligence we need. **Kursk** shows what happens when we don’t have it,” said a senior ministerial adviser.



- Intelligence, a form of computation, is a physical process taking time, space, energy. It is resource-bounded, and imperfect.
- Humans build and deploy artefacts, and can be held to account for doing so responsibly.
- Human centering of AI is necessitated by the justice systems creating improvement being grounded in the concerns of apes.
- EU digital regulation secures the EU – supporting both economies and democracies, (by) defending EU residents.
- Security and ethics both require transparency and responsibility as design requirements for AI, and for the organisations that deploy it.

# Summary

 **Hertie School**  
Centre for  
Digital Governance



