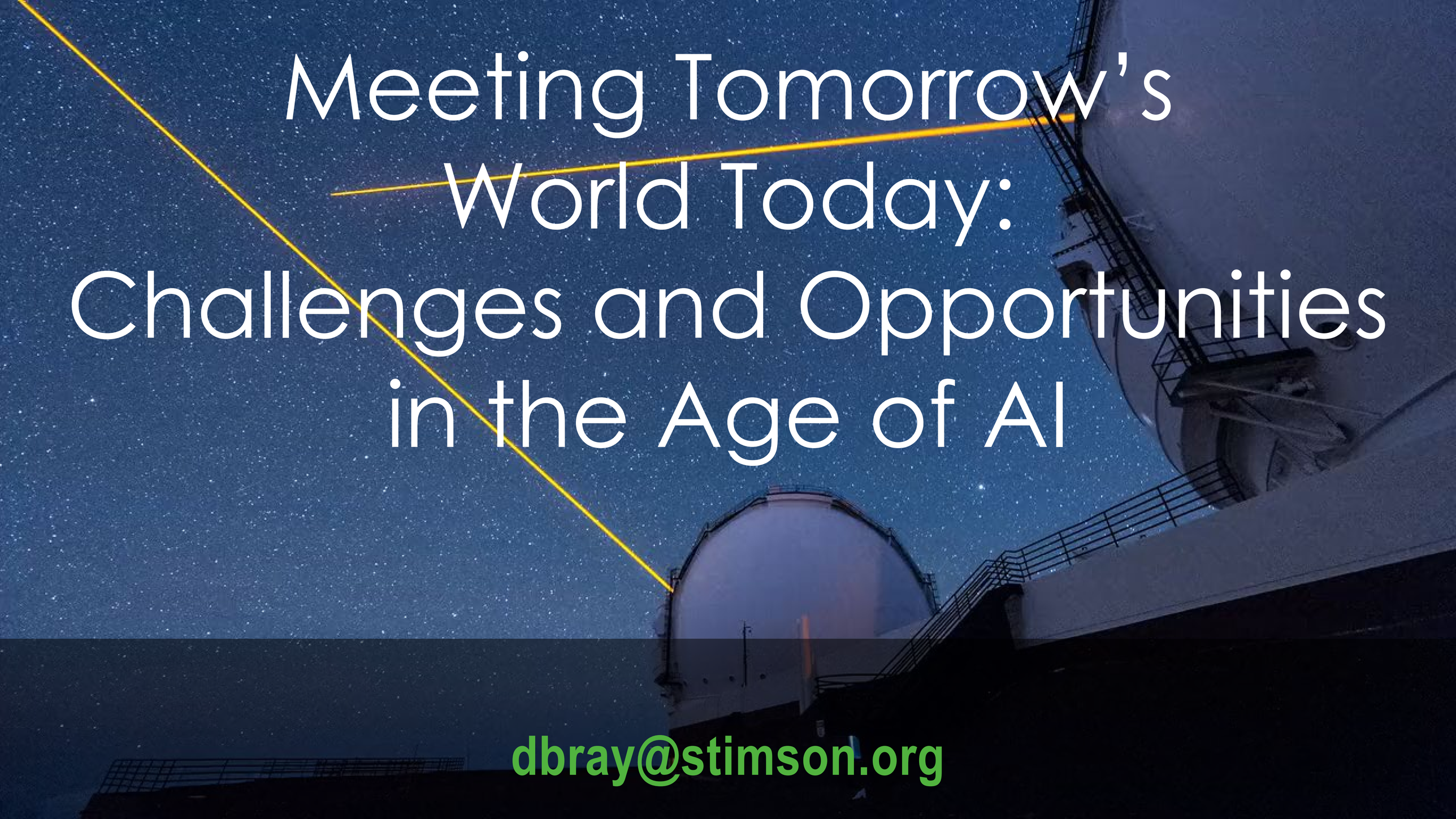
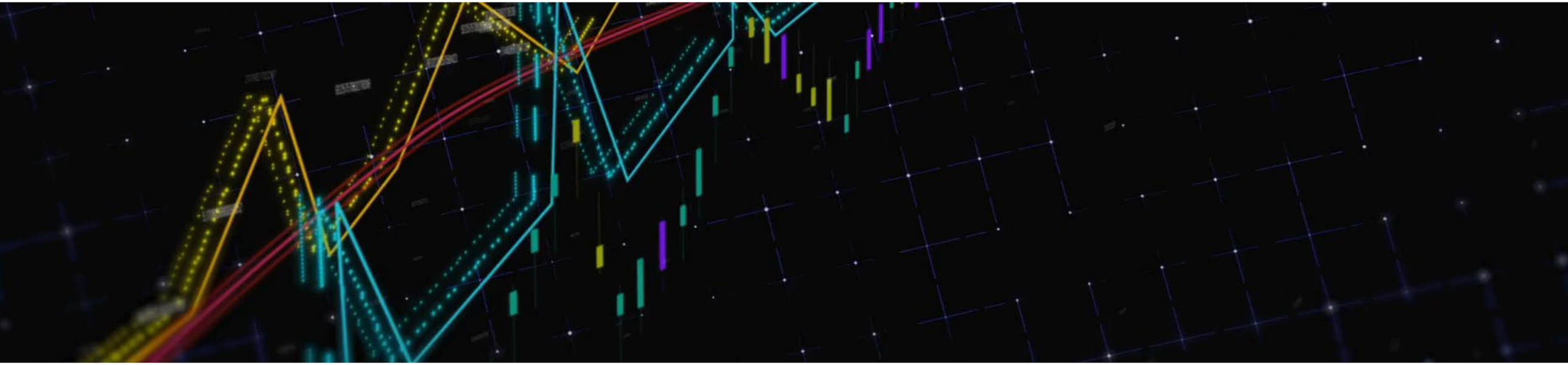


Meeting Tomorrow's World Today: Challenges and Opportunities in the Age of AI

dbray@stimson.org

Opening Thoughts:

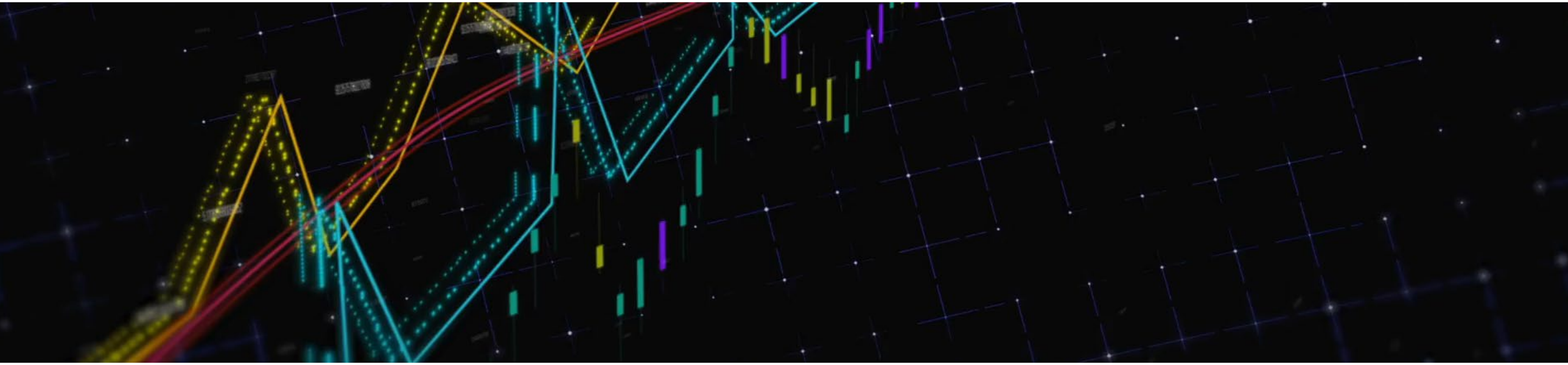
Living Amid Seismic Changes



Last 50 years: we've democratized tech previously just available to intelligence components of major nations in the 1970s

Opening Thoughts:

Living Amid Seismic Changes



However: <u>we have not upgraded how to do the work of defense, security, and civil society</u> given super-empowered populations

AI & Communities - Defining Trust

Trust = Willingness to be vulnerable to actions of an actor not directly controlled by you

Antecedents of Trust =

(1) Perceived Benevolence
(2) Perceived Competence
(3) Perceived Integrity

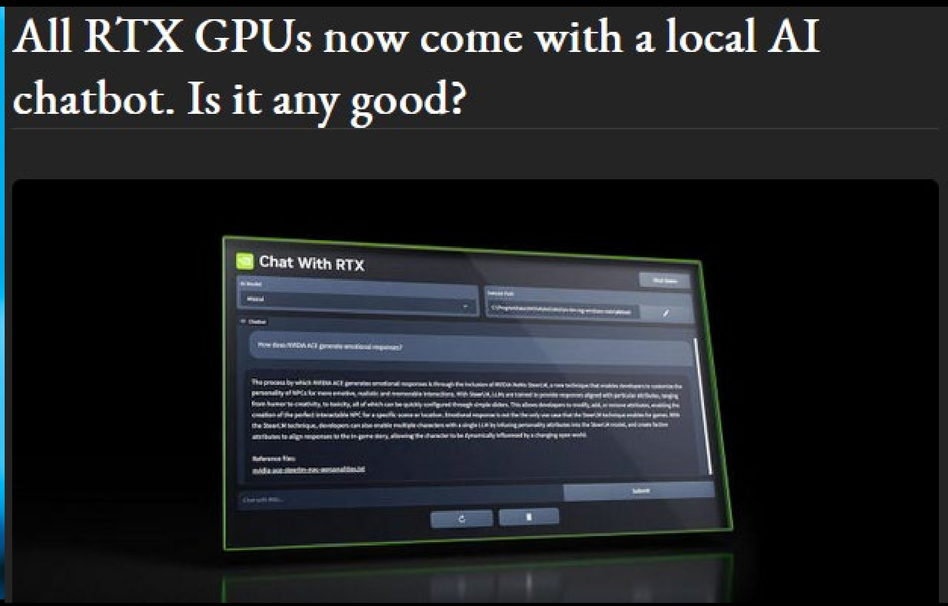
Artificial Intelligence = Alien Interactions

“AI” has many subcategories and has had many names since the 1950s

Flavors of “AI” over the years: Logical Reasoning and Problem-Solving Algorithms. Expert Systems. Statistical Inferences and Reasoning. Decision Support Systems. Cognitive Simulation. Natural Language Processing. Machine Learning. Neural Networks. Large-Language Models.

Why Is This Inevitable, Trend 1:

AI Everywhere



Current GenAI, using Deep Learning, is massively data & CPU intensive – may be replaced by local & more agile AI methods

Why Is This Inevitable, Trend 1:

AI Everywhere

napster

= ? GENAI vs.

ELSEVIER

Journal of Mathematical Psychology

Volume 107, April 2022, 102632

Tutorial

A step-by-step tutorial on active inference and its application to empirical data

Ryan Smith [a 1], Karl J. Friston [b], Christopher J. Whyte [c 1]

Active Inference = alternative approach that may prove better for more predictive, localized AI if present is *not* like past data

Why Is This Inevitable, Trend 2:

Data Collides with Globalization

	Rank	Name	Market Cap	Price	Today	Price (30 days)	Country
☆	1	Microsoft MSFT	$3.085 T	$415.28	2.66%		USA
☆	2	Apple AAPL	$2.675 T	$173.23	0.28%		USA
☆	3	NVIDIA NVDA	$2.297 T	$919.13	7.16%		USA
☆	4	Saudi Aramco 2222.SR	$2.069 T	$8.55	0.47%		S. Arabia
☆	5	Amazon AMZN	$1.821 T	$175.39	1.99%		USA
☆	6	Alphabet (Google) GOOG	$1.728 T	$139.62	0.49%		USA
☆	7	Meta Platforms (Facebook) META	$1.274 T	$499.75	3.34%		USA
☆	8	Berkshire Hathaway BRK-B	$876.32 B	$404.98	0.05%		USA
☆	9	TSMC TSM	$749.00 B	$144.40	3.87%		Taiwan
☆	10	Eli Lilly LLY	$717.32 B	$754.95	2.80%		USA

#	Country	GDP (nominal, 2022)	GDP (abbrev.)	GDP growth	Population (2022)	GDP per capita	Share of World GDP
1	United States	$25,462,700,000,000	$25.463 trillion	2.06%	338,289,857	$75,269	25.32%
2	China	$17,963,200,000,000	$17.963 trillion	2.99%	1,425,887,337	$12,598	17.86%
3	Japan	$4,231,140,000,000	$4.231 trillion	1.03%	123,951,692	$34,135	4.21%
4	Germany	$4,072,190,000,000	$4.072 trillion	1.79%	83,369,843	$48,845	4.05%
5	India	$3,385,090,000,000	$3.385 trillion	7.00%	1,417,173,173	$2,389	3.37%
6	United Kingdom	$3,070,670,000,000	$3.071 trillion	4.10%	67,508,936	$45,485	3.05%
7	France	$2,782,910,000,000	$2.783 trillion	2.56%	64,626,628	$43,061	2.77%
8	Russia	$2,240,420,000,000	$2.240 trillion	-2.07%	144,713,314	$15,482	2.23%
9	Canada	$2,139,840,000,000	$2.140 trillion	3.40%	38,454,327	$55,646	2.13%
10	Italy	$2,010,430,000,000	$2.010 trillion	3.67%	59,037,474	$34,053	2.00%

Combined market cap of the major data-intensive companies
≈ the combined GDP of all nations minus the top 4

Why Is This Inevitable, Trend 2:

Data Collides with Globalization

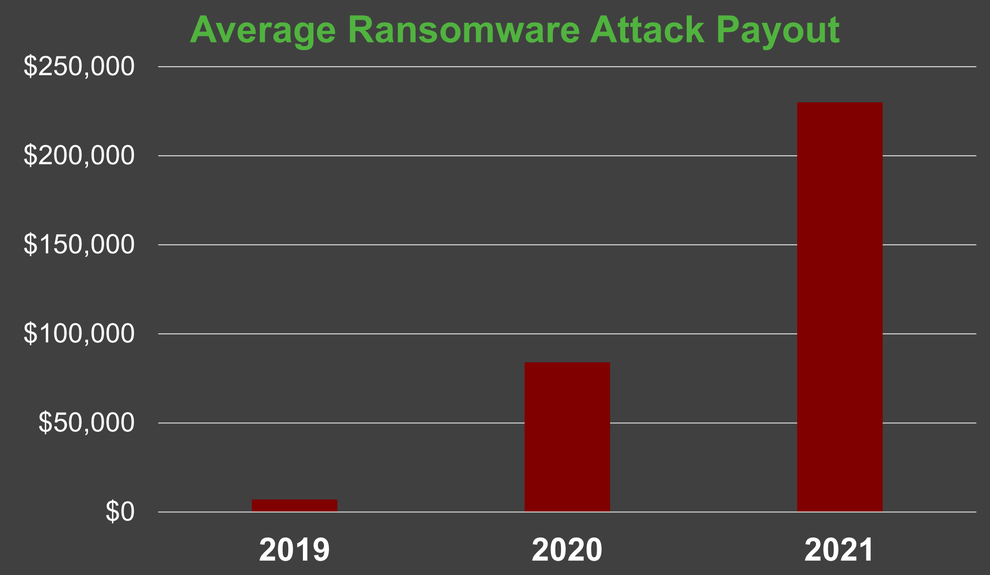
Data = form of voice for people, if little or no choices in its use the people lose free speech

Data is *not* the new oil = hoarding increases distrust and polarization

Most orgs still need to get digital fundamentals first, meanwhile may find too much <u>data hoarding challenges free societies</u>

Why Is This Inevitable, Trend 3:

Cybersecurity: Ransoms + Scams



Ransomware damages in 2019: $11B, 2020: $20B, 2021: $43B+

“Our goal is to make money” – Colonial Pipeline attackers

Why Is This Inevitable, Trend 3:

Cybersecurity: Ransoms + Scams

CEO Arrested for Selling $1 Billion in Fake Cisco Hardware on Amazon, eBay

Onur Aksoy allegedly imported thousands of fake Cisco networking devices from China.

Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.

A company lost $25 million after an employee was tricked by deepfakes of his coworkers on a video call: police

- The person had attended a video call with deepfakes of the firm's UK-based CFO and other colleagues.
- Hong Kong police said scammers created the deepfakes based on publicly available video.

Wi-Fi jamming to knock out cameras suspected in nine Minnesota burglaries -- smart security systems vulnerable as tech becomes cheaper and easier to acquire

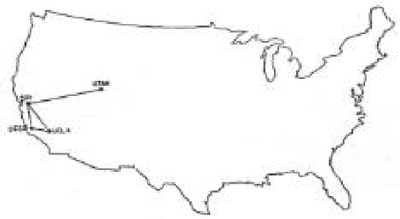
EDUCATION

Hackers are targeting a surprising group of people: young public school students

MARCH 12, 2024 · 5:01 AM ET

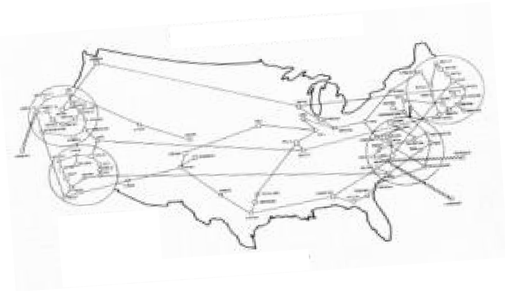
Significant concern in use of Generative AI to generate <u>fake AI images, audio, & videos</u> to generate fraud + disinformation

What Massive Changes Mean



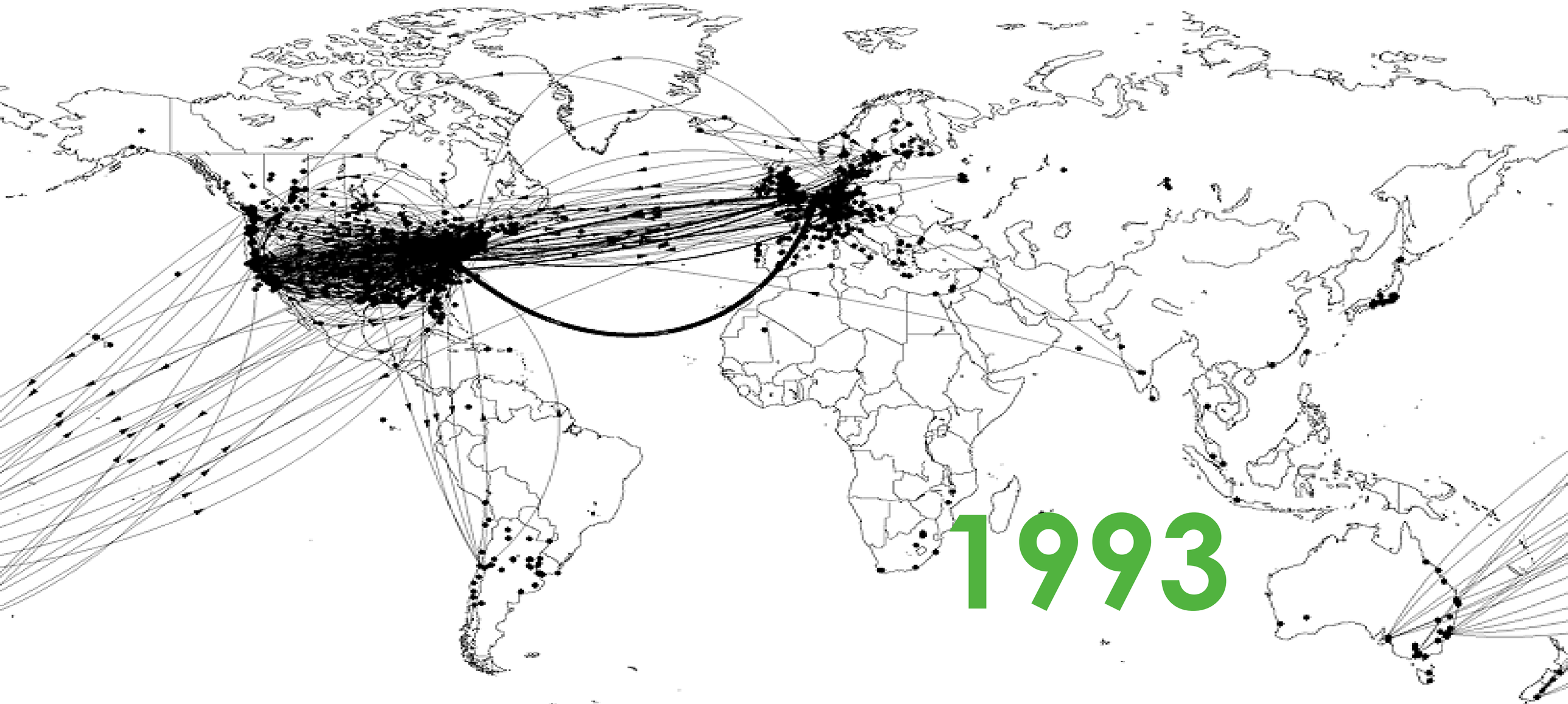
1969

What Massive Changes Mean



1982

What Massive Changes Mean

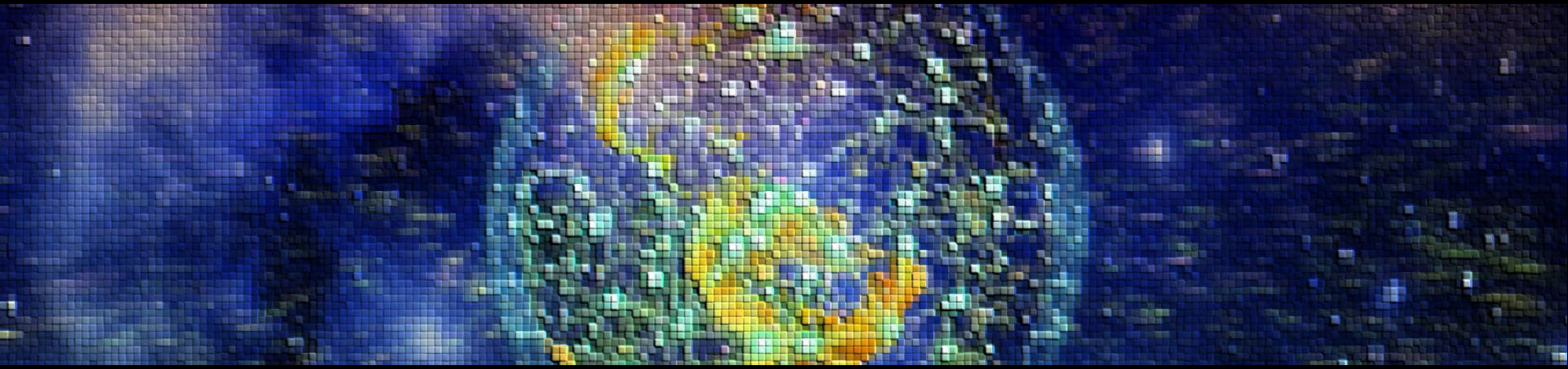


What Massive Changes Mean

What Massive Changes Mean

2013

What Massive Changes Mean



2013: 7.1 billion humans on Earth, ~7.1 billion network devices

2024: 8.1 billion humans on Earth, ~45 billion network devices

How To Address Trust – Given Trust in Societies Now?

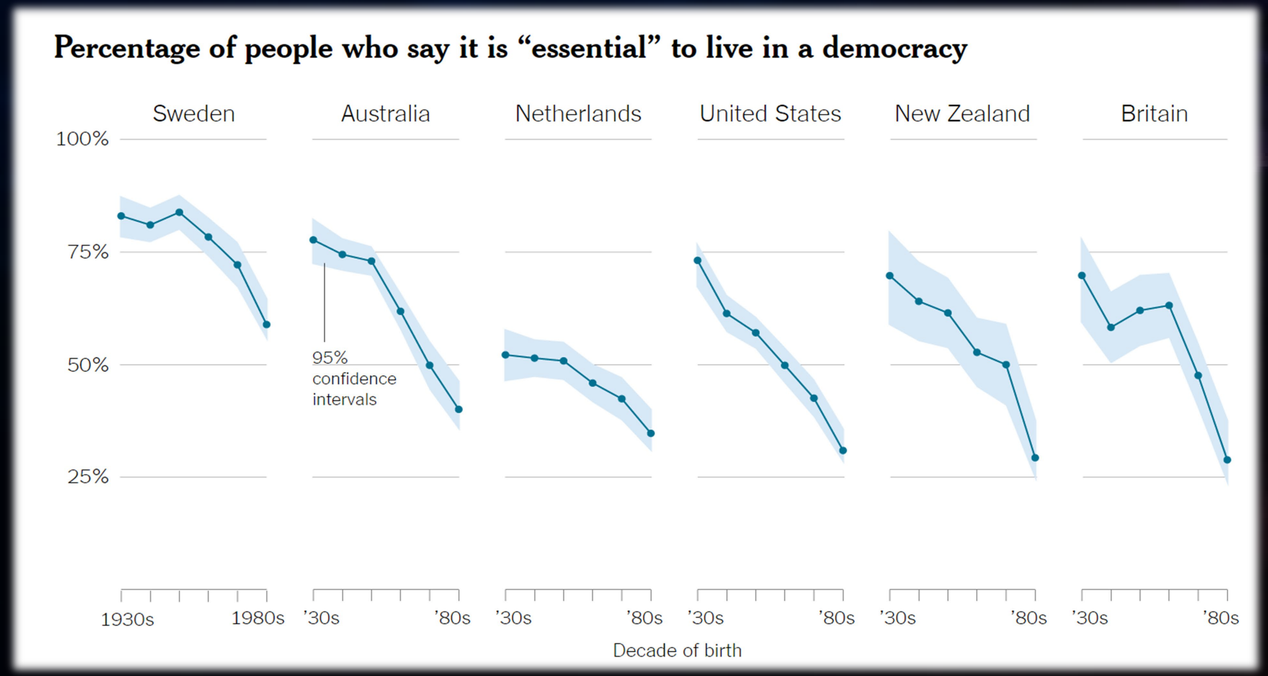
Trust = Willingness to be vulnerable to actions of an actor not directly controlled by you

Antecedents of Trust =

(1) Perceived Benevolence
(2) Perceived Competence
(3) Perceived Integrity

How To Address Trust – Given Trust in Societies Now?

In Oct 2017, Pew found <45% of those <25 years old in the U.S. thought capitalism was "good"; we're at fairly low levels of trust in representative governments as well globally (not just U.S.)



Is asking for Trust in AI impossible?

What If the Turing Test is the Wrong Test?

Original Turing Test: Computer A and Person B attempting to convince Person C that they were human.

What if this test of a computer "fooling us" is the wrong test for the type of AI that either human services or our societies at-large need, if we're to have some trust among humans and machines?

What If the Turing Test is the Wrong Test?

We humans are products of natural selection pressures. Darwinian evolution is akin to a "blind watchmaker". Evolution has not prepared us to encounter the true alienness of AI + new tech.

Our choices regarding AI + new tech will amplify or mitigate human nature. Also, we humans anthropomorphize lots even if those things do not act, think, or behave at all like us.

What If Building Trust in AI ≈ Trust in Decisions & in Societies?

We humans also have biases including confirmation bias, sunk cost bias, "in vs. out group" biases (aka, xenophobia), and many more biases that can be mitigated by education and experiences.

The Global Megatrends all intersect with those existing biases, making the future of Strategic Decision-Making especially challenging for leaders across sectors.

What If Building Trust in AI ≈ Trust in Decisions & in Societies?

Natural selection pressures resulted in humans tending to trust if we perceive Benevolence, Competence, and Integrity in an actor.

Consider current LLMs/chatbots:

Benevolence = indeterminant?
Competence = LLMs are not fact-checking, can spread misinfo
Integrity = absent, they will change their stance readily

What If Building Trust in AI ≈ Trust in Decisions & in Societies?

So, Trust in LLMs = ? Yet consider other "obscured boxes" in society, such as decision-making in organizations, community associations, governments, or militaries?

What if we need to remedy and improve Trust in Societies simultaneously to Trust in Decisions simultaneously to Trust in AI for the future?

Where We Are Now – GenAI Example:

Synthetic "Hosted" Podcast

A group of NAPA Fellows associated with the Standing Panel on Technology Leadership recently released a call to action on responsibly using AI to benefit public service at all levels of government. We are grateful for the strong positive response to this call from numerous colleagues in governmental communities. We provide additional scoping observations below, and welcome continued and expanded dialogue on this critical issue.

https://napawash.org/standing-panel-blog/ai-services-to-citizens-in-2023-and-beyond

Artificial Intelligence and Public Service: Key New Challenges

David Bray, PhD
Distinguished Fellow, Stimson Center as well as Business Executives for National Security

In May 2023, the Executive Office of the President announced actions to promote responsible AI innovation, having previously announced in October 2022 a "Blueprint for an AI Bill of Rights" to include safe and effective systems, protections against algorithmic discrimination, data privacy, notice and explanation, and alternative options to include opting-out of such systems. Such efforts raise interesting questions about (1) where can advances in AI improve delivery of Services to Citizens and (2) what changes in how public service organizes and operates are needed to make such improvements a reality?

I. Where can advances in AI improve delivery of Services to Citizens?

For FY2024 and FY2025, we can look to see signs of adoption of the following at the federal level of the U.S. government as well within state and local governments:

Where We Are Now – GenAI Example:

Synthetic "Hosted" Podcast

The following is a synthetic podcast - of two synthetic agents discussing an actual article at https://napawash.org/standing-panel-blog/ai-services-to-citizens-in-2023-and-beyond

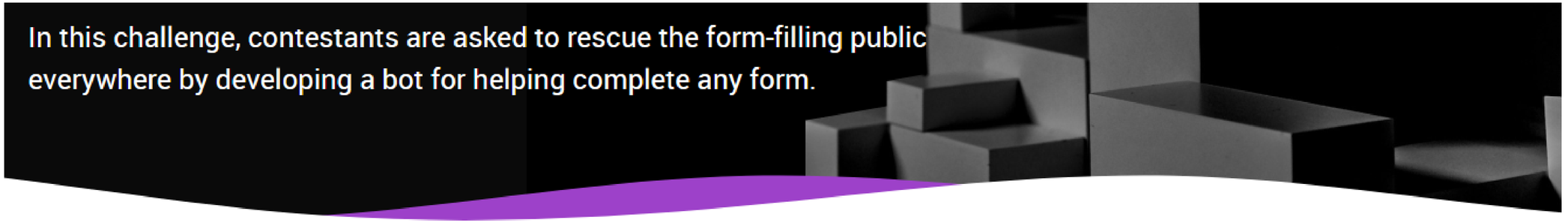
The actual article and content are real, the podcast is synthetic using NotebookKLM. Imagine if this was a brief podcast presenting someone's CV or summarizing the news?

What if we used GenAI to remove all the adjectives associated a news article, and asked it to help us triangulate the news from multiple perspectives without emotional adjectives?

Some are experimenting with this to prepare briefs for executives

Implementation - For Outcomes Now:

1) Conversational Forms



https://www.topcoder.com/community/innovation-challenges/intelliform-bot-gpt

CHALLENGE DESCRIPTION

Assignment: Develop a GPT bot capable of guiding users through completing any form they upload, ensuring accuracy and understanding of instructions. The bot infers context from the provided form, It converses with the user, asking follow-up questions when needed, and alerts if an answer seems incorrect. Users can correct answers and revisit previous elements without redoing the entire form.

CHALLENGE DETAILS

Implementation - For Outcomes Now:

2) Communities Have “A Voice”

Recent work to use GenAI to translate government forms into conversations, ultimately can aid individuals to receive services & benefits via text and/or verbal conversations.

Implementation - For Outcomes Now:

3) Cloud-Based Decision Support

AI can rapidly facilitate collaborations across different facets of healthy communities: Predictive maintenance of equipment & buildings, Resource allocation optimization, Waitlist management

Implementation - For Outcomes Now:

4) Computer Vision Assessments

Structural Integrity: detect cracks & damage, ADA compliance; Safety Checks: locate hazards, identify mold & pests; Energy Efficiency: spot heat leaks & insulation issues, pollution sources

Implementation - For Outcomes Now:

5) Tailored Q&A Chatbots

Science

Current Issue | First release papers | Archive | About | Submit manuscript

Durably reducing conspiracy beliefs through dialogues with AI

https://www.popsci.com/technology/conspiracy-debunk-ai-bot/

A group of researchers from the Massachusetts Institute of Technology, Cornell, and American University put that idea to the test with a custom made chatbot they are now calling “debunkbot.” The researchers, who published their findings in *Science*, had self-described conspiracy theorists engage in a back-and-forth conversation with a chatbot, which was instructed to produce detailed counter arguments to refute their position and ultimately try to change their minds. In the end, conversations with the chatbot reduced the participant’s overall confidence in their professed conspiracy theory by an average of 20%. Around a quarter of the participants disavowed their conspiracy theory entirely after speaking with the AI.

What Bad Actors Already Are Doing with AI:

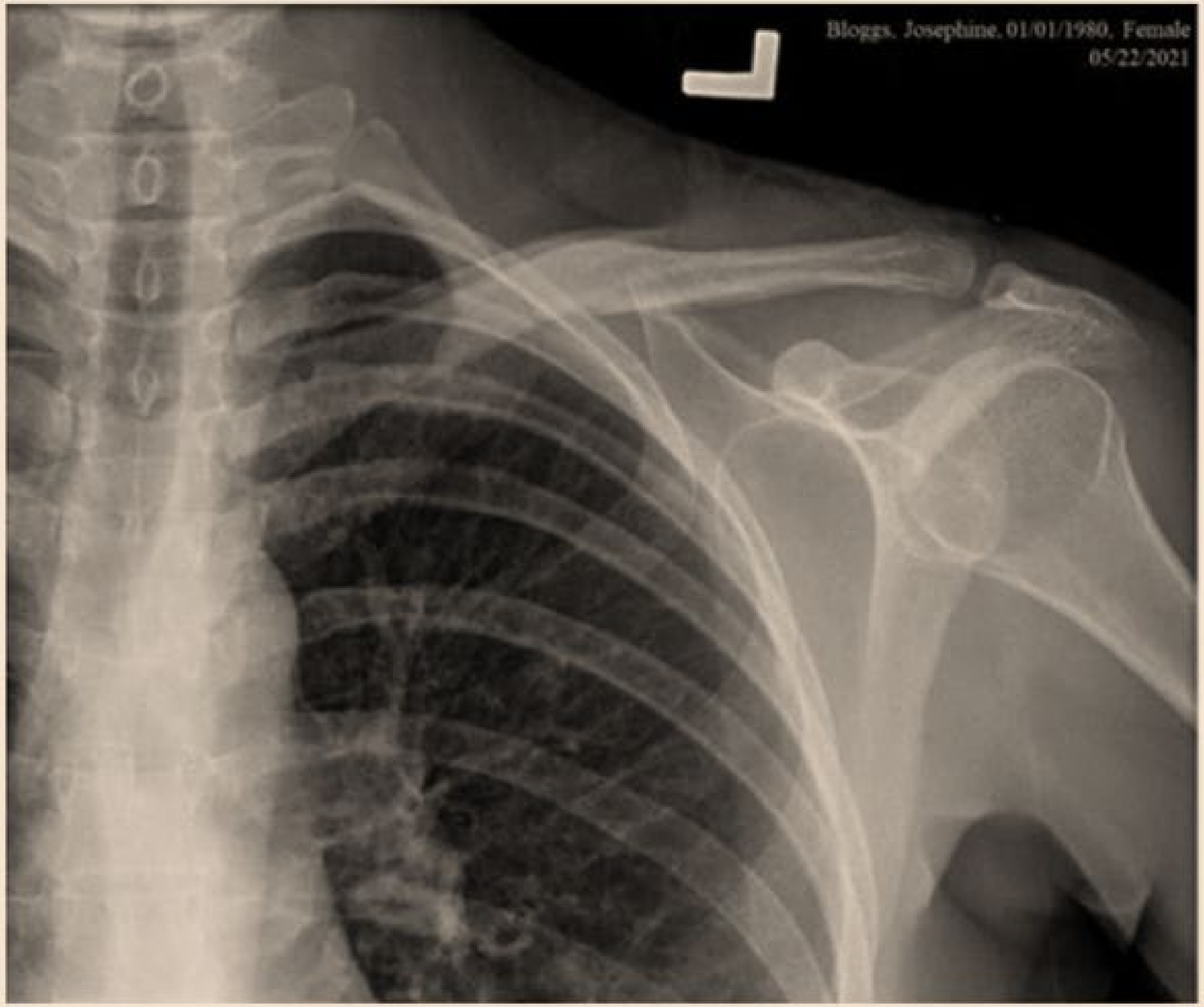
AI-Aided Scams Increasing

Be ready for an even more massive wave of AI-produced fraud – this will include financial and other fraudulent claim-related scams. For more details see StopScamAlliance.org

Example Fictitious Script for Amoxicillin for a 43 yr old Female

Scammers already are using GenAI to create websites to look authentic. Health & human services data are being stolen via cybersecurity breaches to produce fraudulent claims.

Example Fictitious Imaging of Upper Left ... Upper Respiratory Tract Infection



Example Fictitious Imaging of Upper Left Chest for 43 yr old Female showing potential Upper Respitory Tract Infection.

Primary Care Center - University of MN Med Center
516 Delaware ST SE Clinic 3A, Phillips-Wangensteen Bldg, Mpls, MN 55455
612-624-9499

Jon HALLBERG M.D
612-465-9986

Name **Josephine BLOGGS** MRN 015648945

Address _____ DOB 01/01/1980

Amoxicillin 250 MG Tablet #30

Quantity: (thirty tablets)

TAKE 1 TABLET DAILY

Signature of Prescriber

Written: 05/22/2021 ***Refills Authorized 6 Times*** *Rx:* 56987845

Example Fictitious Script for Amoxicillin for a 43 yr old Female

One of 1,000,000+ fake claims generated in less than 10 minutes

1. Ethics & Practical Steps

Your board needs no-nonsense AI leadership - these experts explain why

Policymakers, business leaders, and technology experts all agree on the need for a pragmatic, inclusive, and responsible approach to AI. How do we get there?

https://www.zdnet.com/article/your-board-needs-no-nonsense-ai-leadership-these-experts-explain-why/

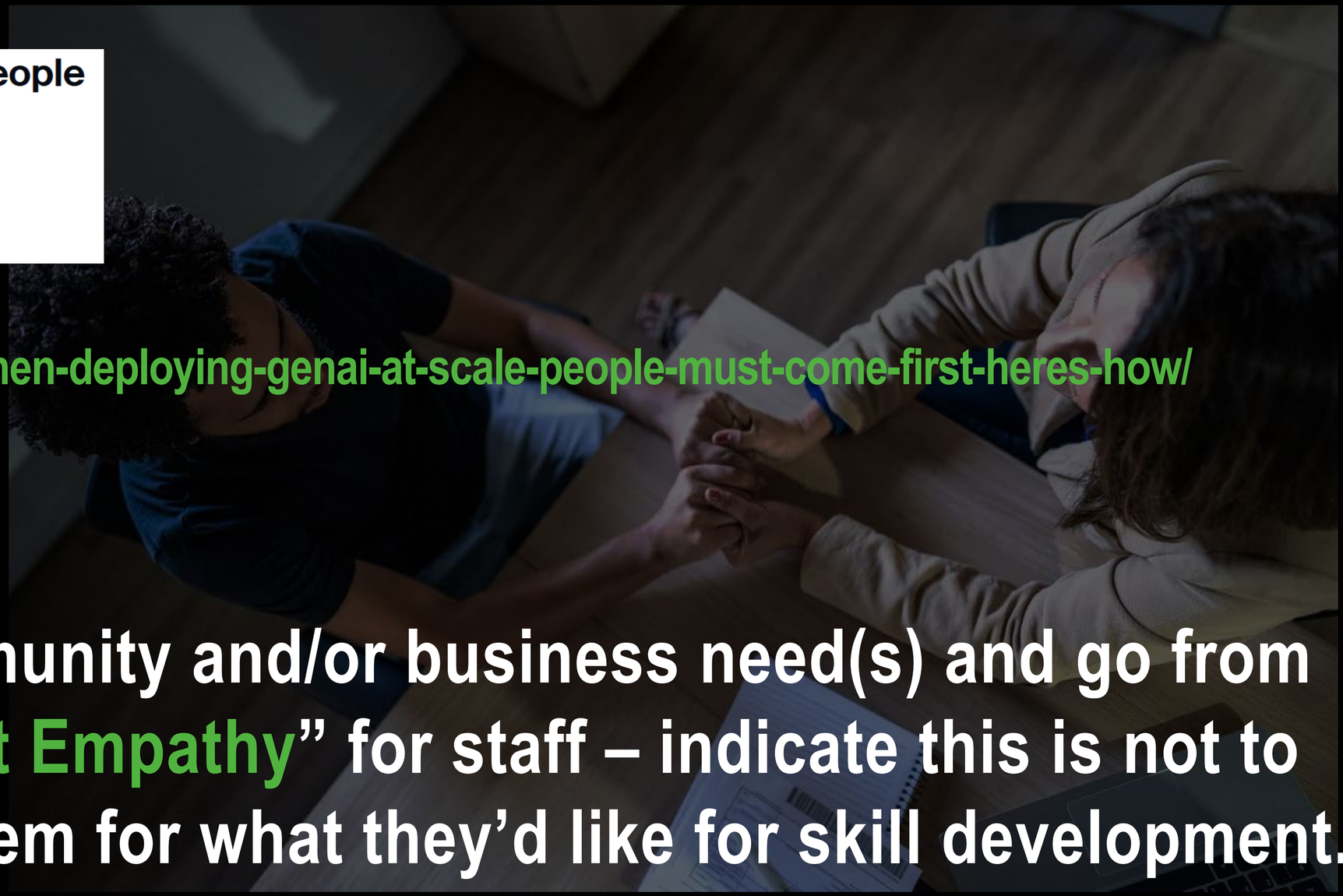
Human governance to ensure both the use of data & of AI is representative of the communities equitably.

Make sure to include community members in governance.

2. Getting Started & Workforces

When deploying GenAI at scale, people must come first. Here's how

'Deployment empathy' means managing change thoughtfully, creating psychological safety, reassuring anxious workers, and collaborating to co-create solutions tailored for shared benefit. Three business leaders break it down.



https://www.zdnet.com/article/when-deploying-genai-at-scale-people-must-come-first-heres-how/

Start with the community and/or business need(s) and go from there. "Deployment Empathy" for staff – indicate this is not to replace them, ask them for what they'd like for skill development.

3. Security & Privacy by Design

FRONTIERS

Three People-Centered Design Principles for Deep Learning

Bad data and poorly designed AI systems can lead you to spurious conclusions and hurt customers, your products, and your brand.

David A. Bray and R "Ray" Wang • September 09, 2019 Reading Time: 6 min

https://sloanreview.mit.edu/video/people-centered-design-principles-for-ai-implementation/

Over the past decade, organizations have begun to rely on an ever-growing number of algorithms to assist in making a wide range of business decisions, from delivery logistics, airline route planning, and risk detection to financial fraud detection and image recognition. We're seeing the end of the second wave of AI, which began several decades ago with the introduction of rule-based expert systems, and moving into a new, third wave, termed *perception AI*. It's in this next wave where a specific subset of AI, called *deep learning*, will play an even more critical role.

Make sure to consider data/IP rights, cybersecurity. Consider data cooperatives – these empower communities to pool their data for specific purposes for communities' benefit.

4. Best Practices & Mindfulness

Don’t let the “fear of missing out” motivate a rush to GenAI, it may be similar to Napster in the 1990s that was later replaced with more professional iTunes and other music models.

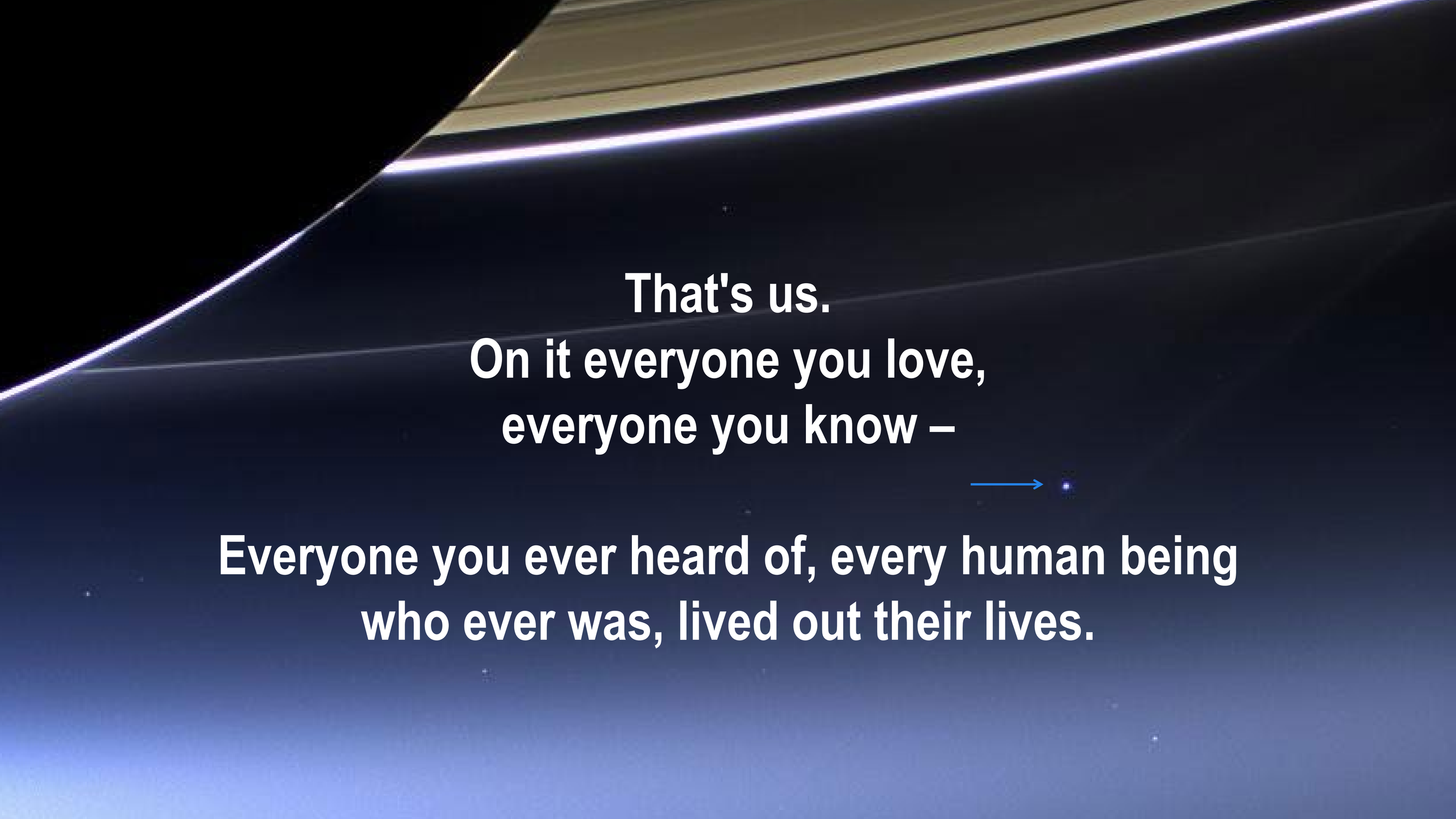
Human settings need to be ready for patients to start bringing their own AI apps of varying quality with into human settings, consider what will and won’t allow to be captured by devices.

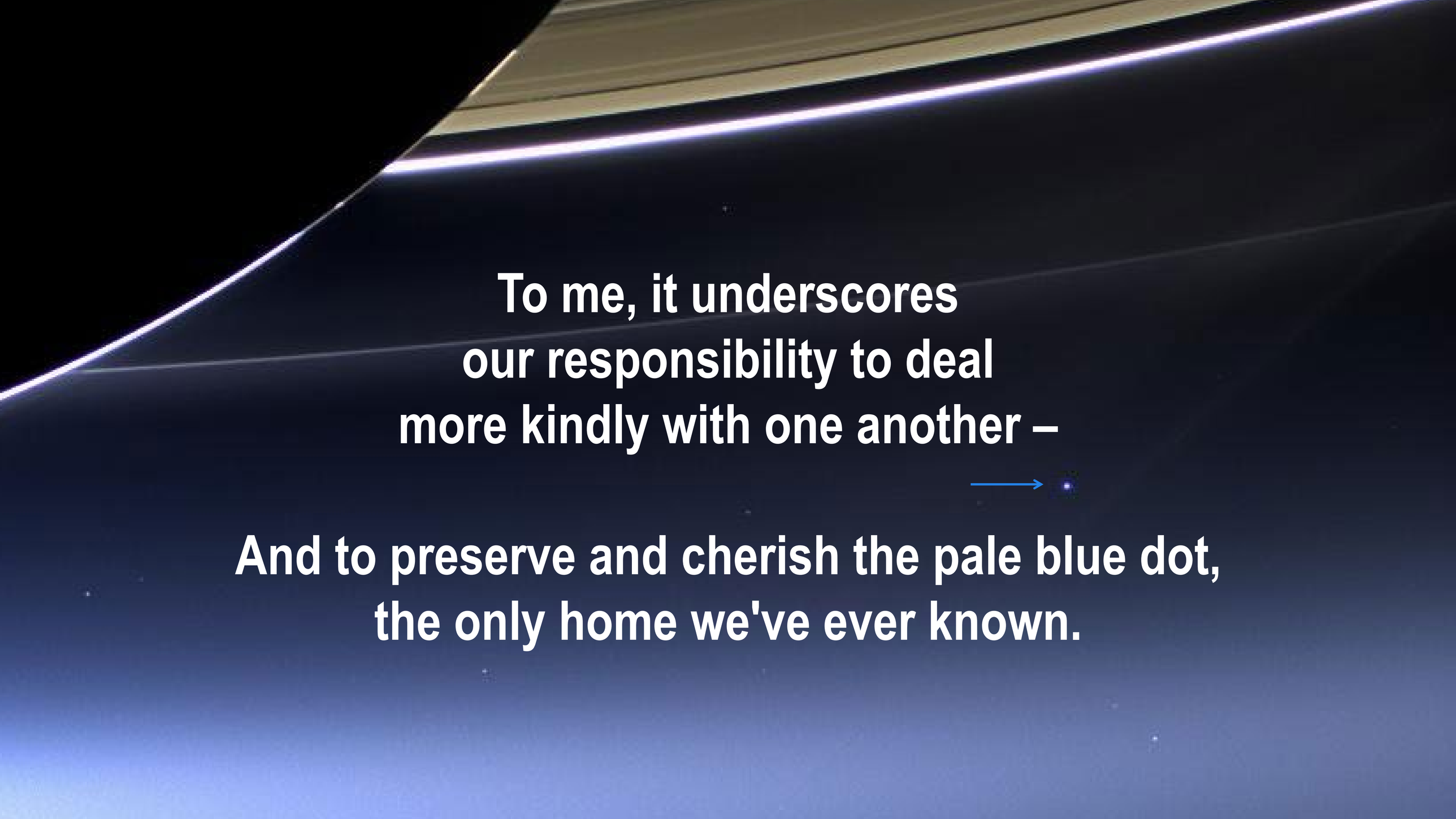
5. Regulation & Auditing

Housing & human services data will need even more security as lots of companies are striving to get this data to train AIs, at the same time this data is the voice of individuals & communities.

AI providers will need to start recording when an AI aids in outputs stored in a human record, to avoid the risks of AI reading in an AI output and **introducing model destabilization**.


Carl Sagan in 1994:
Look again at that dot.
That's here. That's home.


That's us.
On it everyone you love,
everyone you know –
Everyone you ever heard of, every human being
who ever was, lived out their lives.


To me, it underscores
our responsibility to deal
more kindly with one another –
And to preserve and cherish the pale blue dot,
the only home we've ever known.

Be Bold, Be Brave, Be Benevolent

Additional Questions?

2020 – MIT Sloan Mgmt Review article on People-Centered AI:
https://mitsloan.mit.edu/ideas-made-to-matter/5-steps-to-people-centered-artificial-intelligence

2023 – National Academy of Public Admin on AI & Public Service:
https://napawash.org/standing-panel-blog/a-call-to-action-the-future-of-artificial-intelligence-and-public-service

dbray@stimson.org