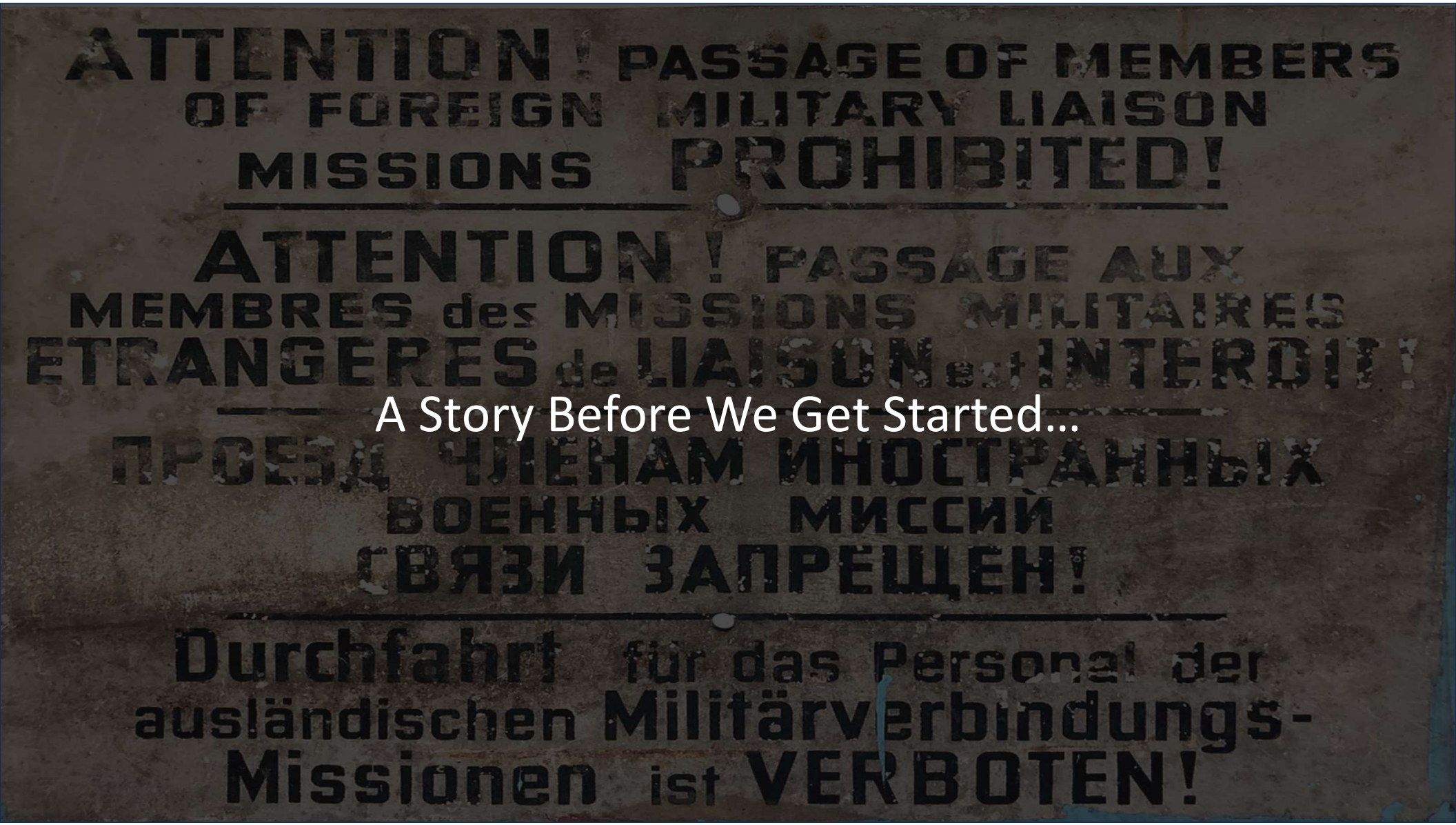
WILLIAM PATERSON UNIVERSITY

College of Science and Health

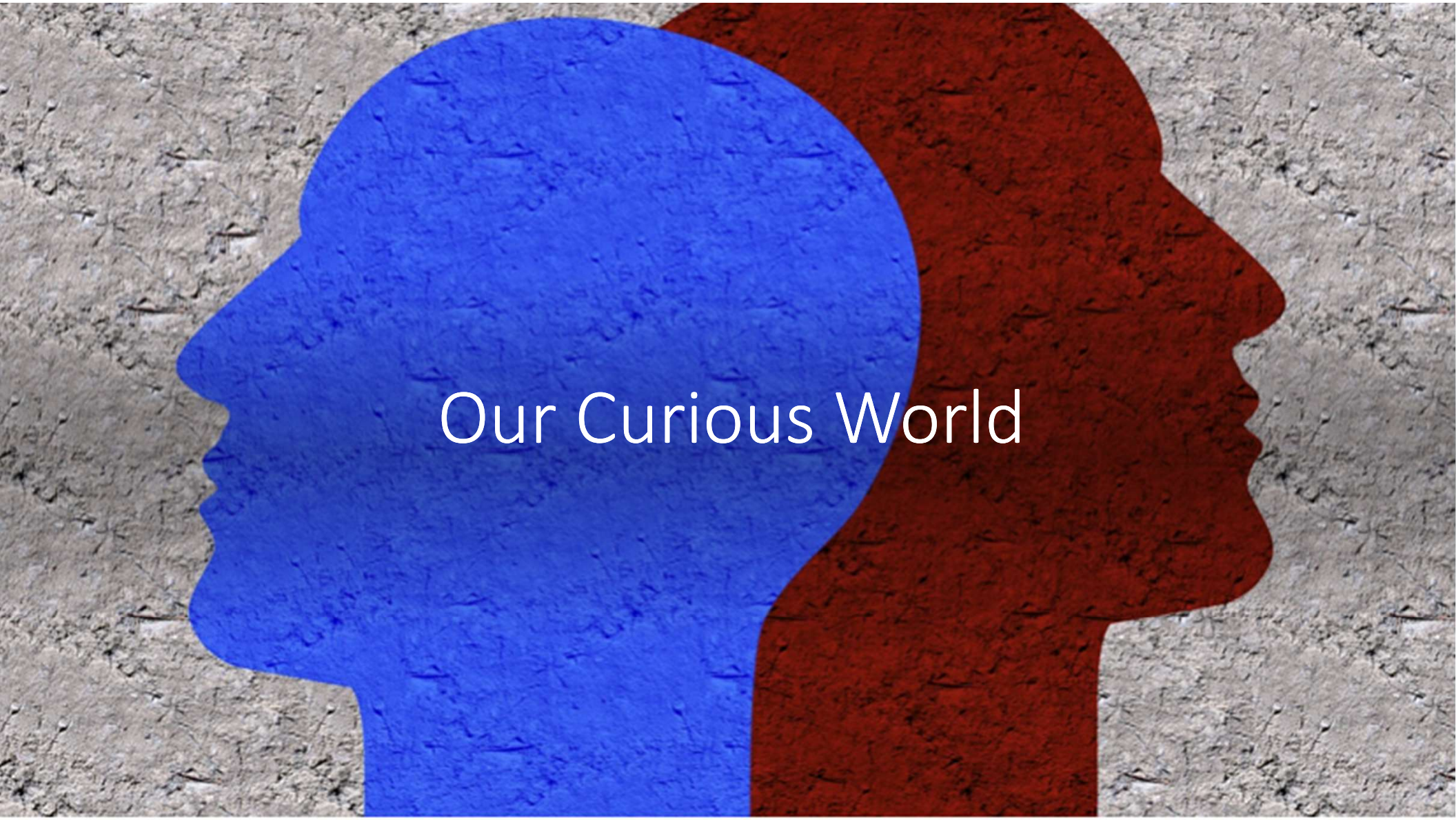
UPS Computer Information Lecture Series

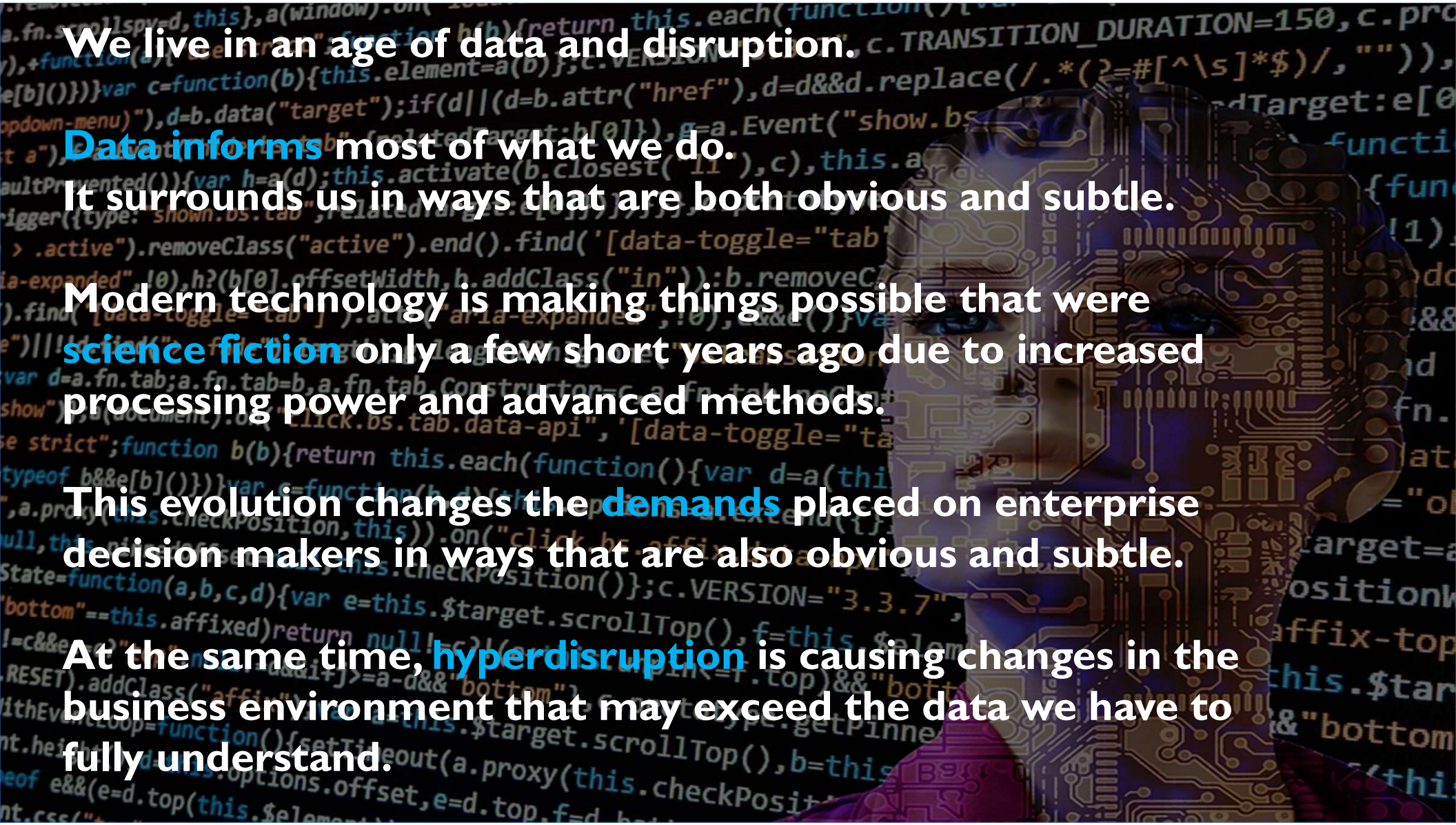
Data Science in the Context of Hyperdisruption:
Everything is Changing… at the SAME TIME!

Anthony Scriffignano, Ph.D.
Distinguished Fellow| The Stimson Center
Alfred Lee Loomis Innovation Council Member
April 2024



A Story Before We Get Started…


Our Curious World



We live in an age of data and disruption.

Data informs most of what we do.
It surrounds us in ways that are both obvious and subtle.

Modern technology is making things possible that were science fiction only a few short years ago due to increased processing power and advanced methods.

This evolution changes the demands placed on enterprise decision makers in ways that are also obvious and subtle.

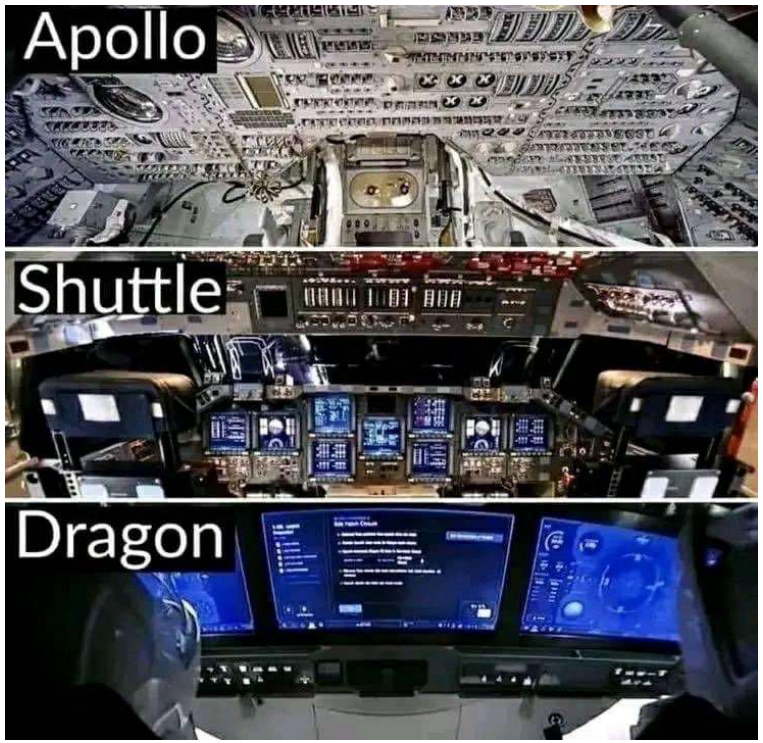
At the same time, hyperdisruption is causing changes in the business environment that may exceed the data we have to fully understand.

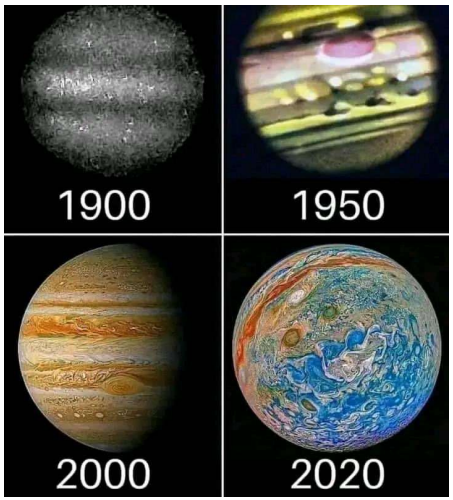
Considering the pace of evolution

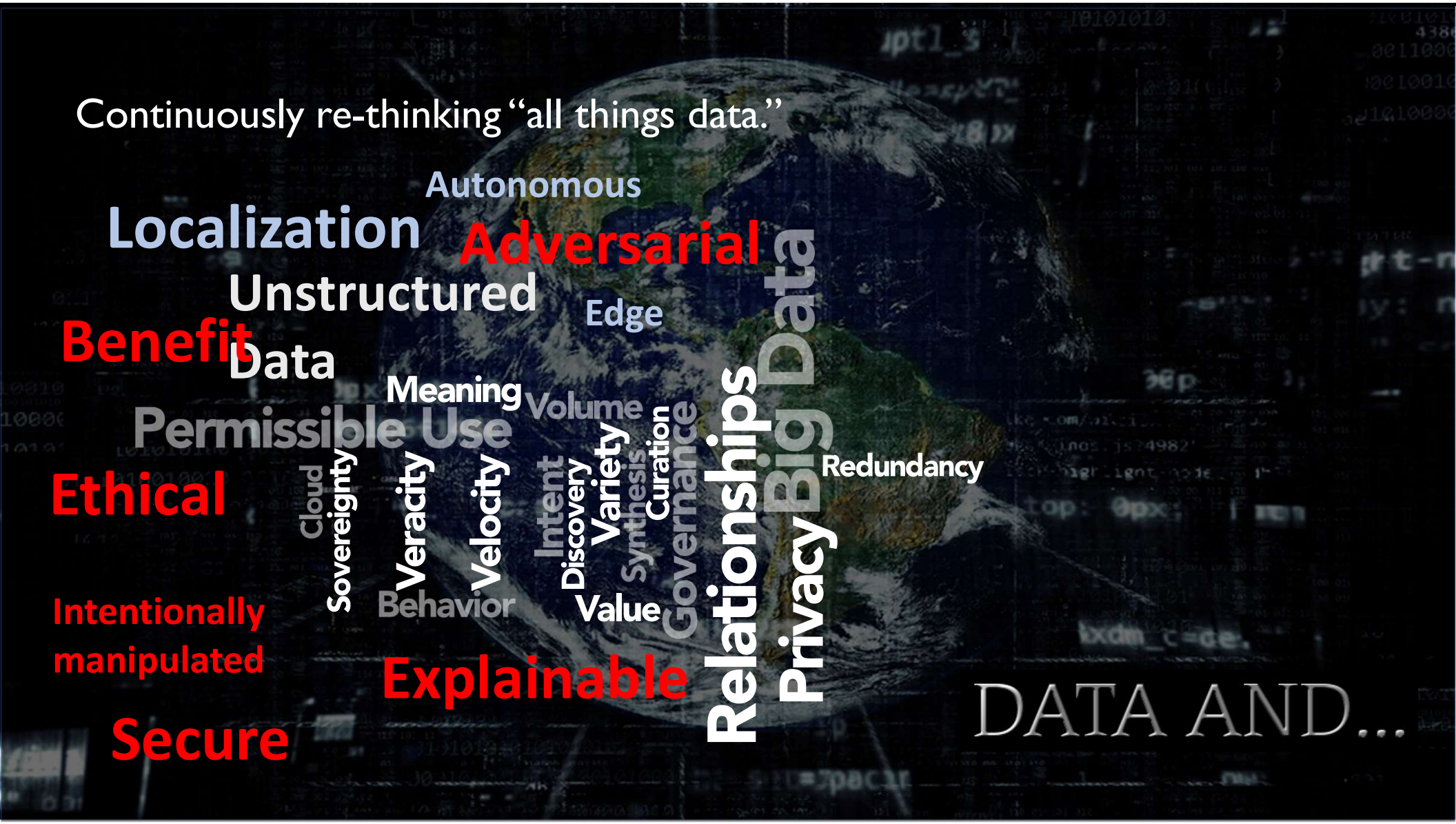
Pope Benedict First Audience

Pope Francis First Audience

Sometimes, change is hard to notice when you are part of that change

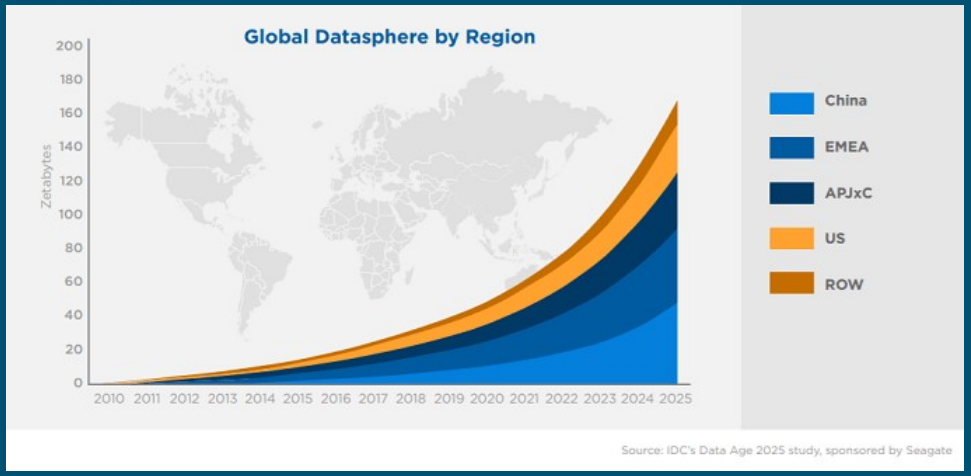




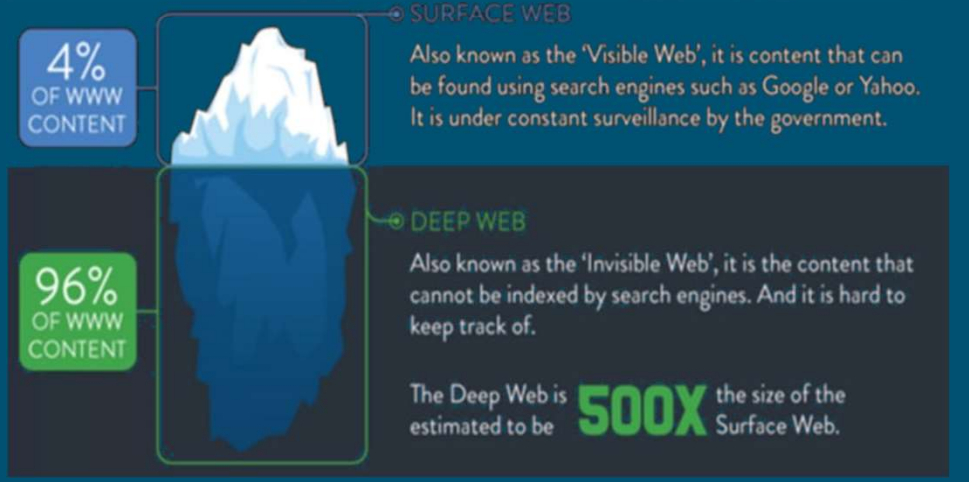

Continuously re-thinking "all things data."
Autonomous
Localization
Adversarial
Unstructured
Edge
Benefit
Data
Meaning
Volume
Permissible Use
Ethical
Cloud
Sovereignty
Veracity
Velocity
Intent
Discovery
Variety
Synthesis
Curation
Governance
Big Data
Redundancy
Relationships
Privacy
Behavior
Value
Intentionally manipulated
Explainable
Secure
DATA AND...

Part of the challenge is that we keep using the same words to describe different things

Definition of Big



Definition of Discoverable

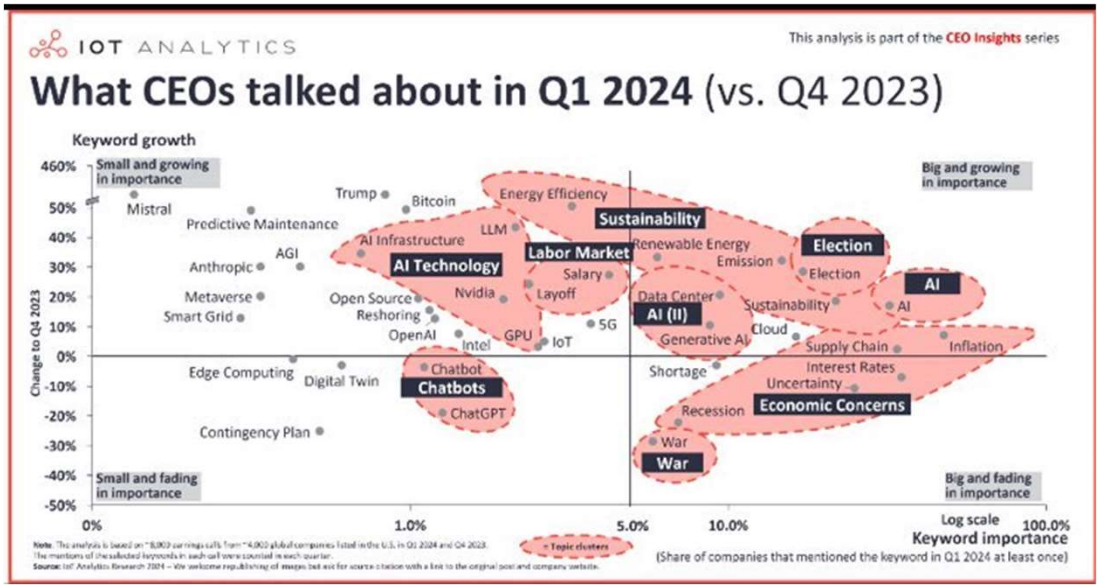


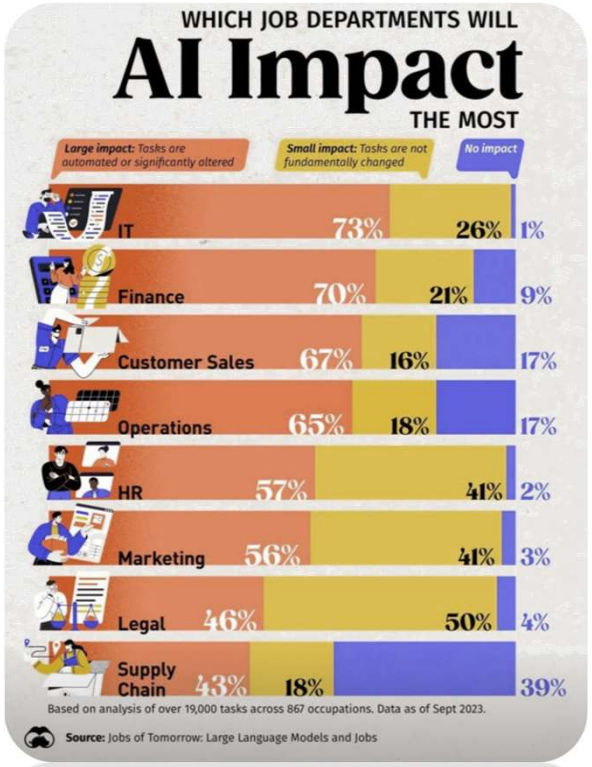
Source: https://www.deepwen-sites.com/how-big-is-the-deep-web/

Definition of Learning

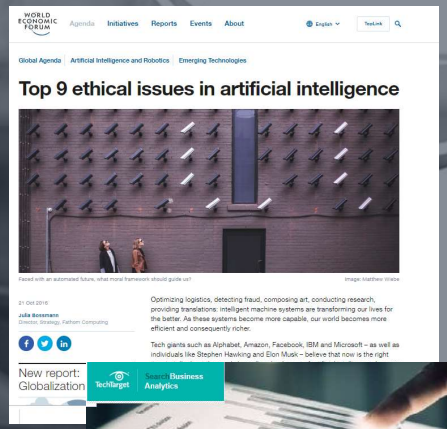
Source: https://iiot-world.com/artificial-intelligence/top-worlds-artificial-intelligence-researchers-and-influencers/

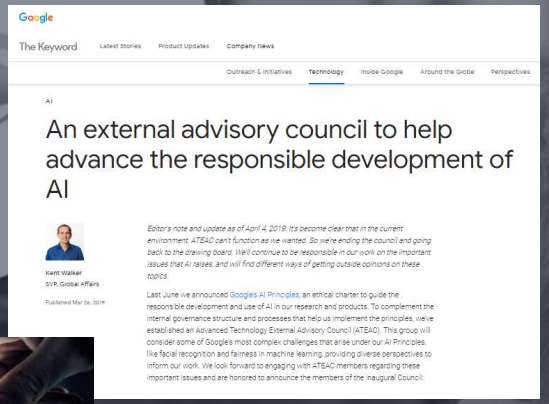
AI In the workplace

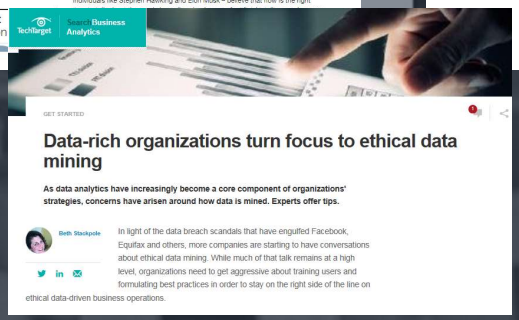


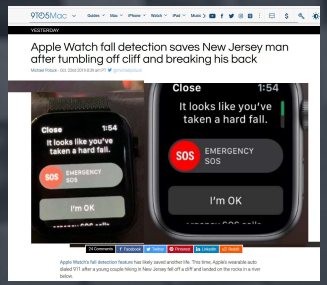


… the complexity of issues continues to mount







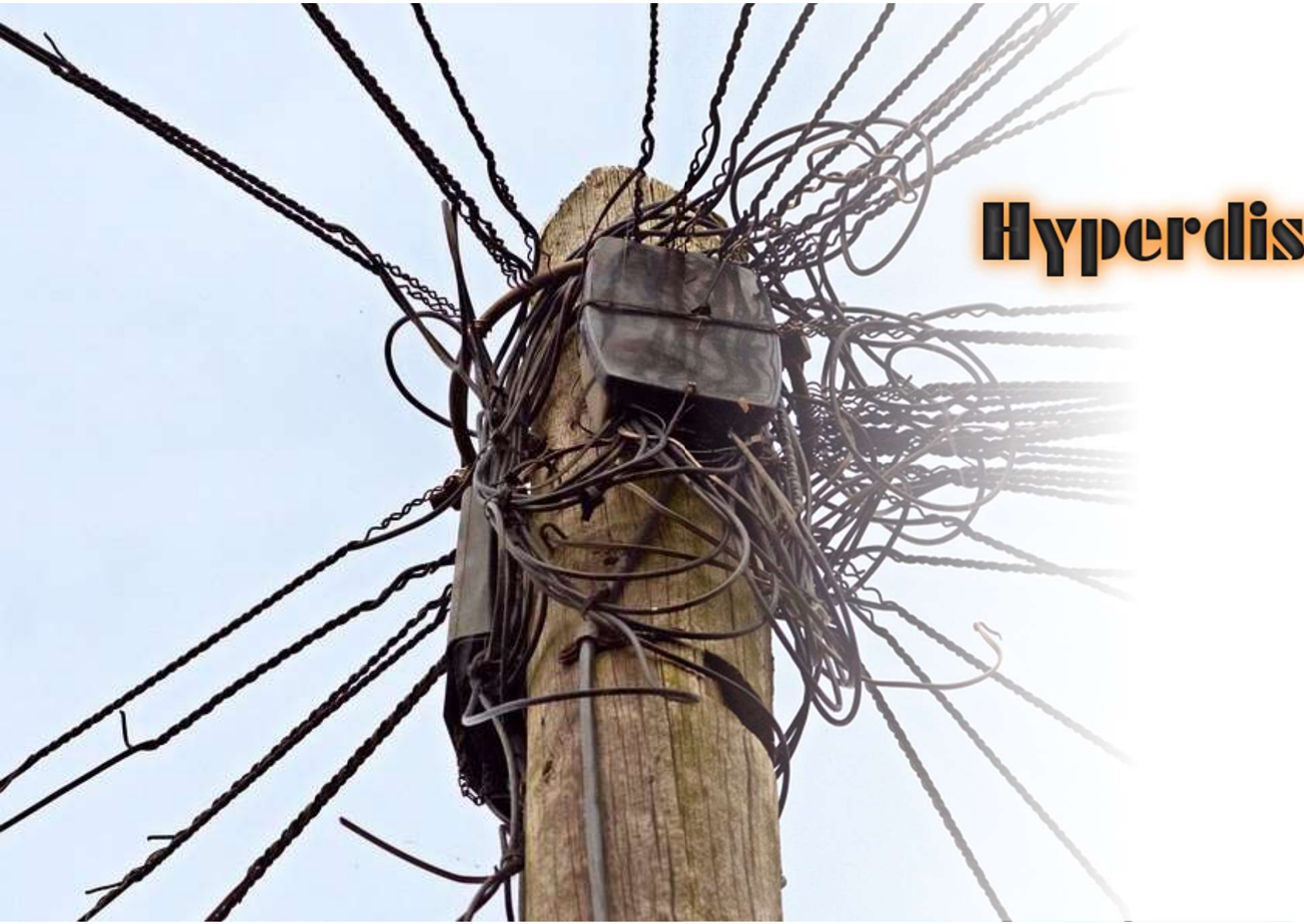


Changing focus:

- From open data to data rights
- From silos to bigger silos
- From cyber threats to cyber everything
- From economics to cryptoeconometrics
- From incident response to constant disruption

Emerging dialogue

- Changes in the workforce
- Inequality / marginalization
- AI Bias
- Adversarial Manipulation
- Federation of Technology
- Open Source implications
- Data Rights
- Intellectual Property
- Agency
- Explainability / Auditability
- Duty to Act / Cost of inaction
- Changing regulation


Hyperdisruption

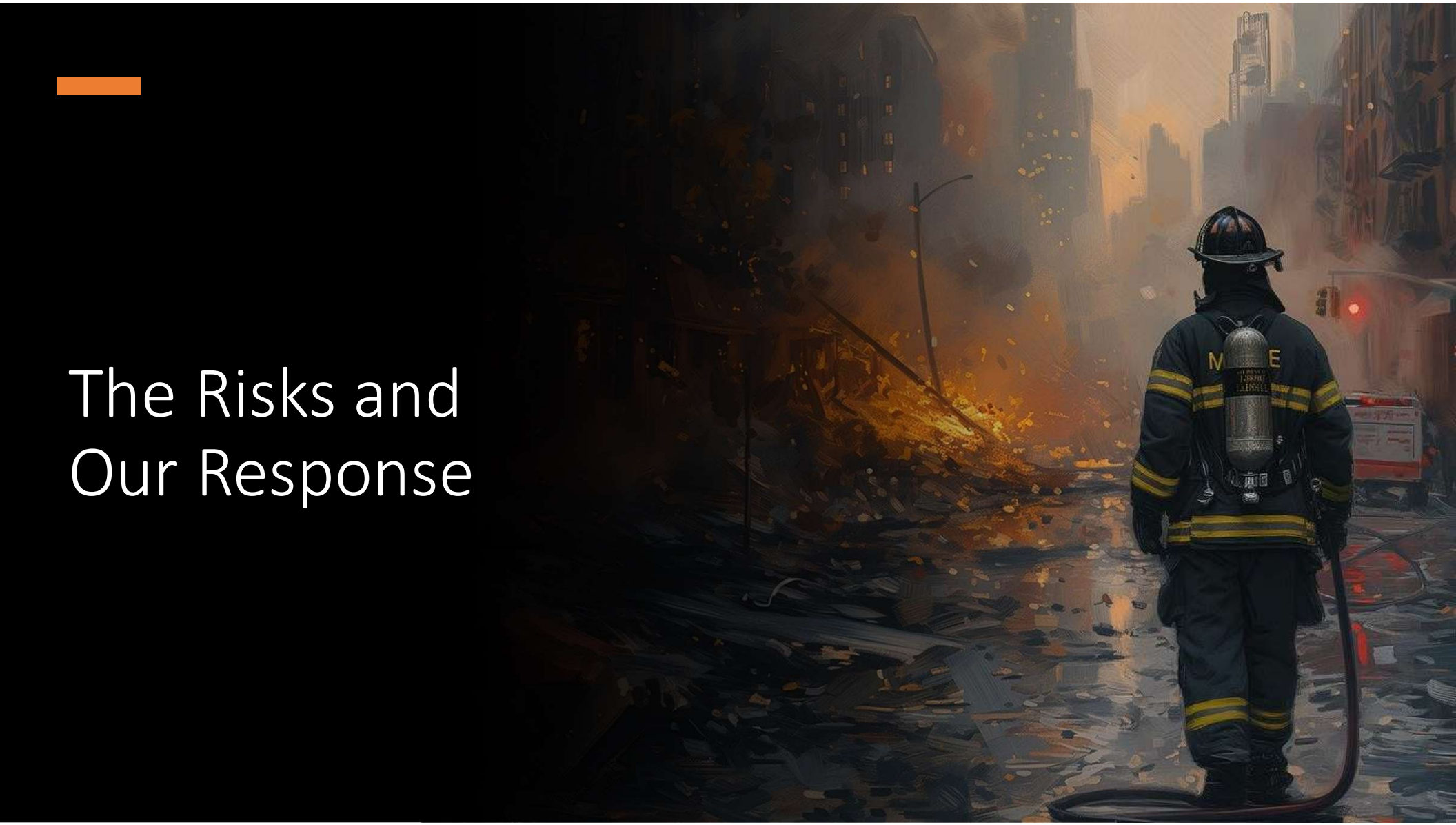
Practical Example:
Trends affecting supply-chain decision-makers

Innovation Focus	Explanation	Key Challenges
Lasting effects of pandemic	Pandemic has forever changed some businesses, causing shifts in product and service offerings as well as available counterparties	The availability of qualified counterparties Vetting suppliers in a timely way Ongoing assessment
Changing regulations	Various regulations impact the type of information that must/may be collected and permissible interaction.	Understanding changes in a timely way. Creating agile systems/processes.
Nation-state actions and sanctions	Key supply chain counterparties may become unavailable.	Understanding permissible and advisable reactions to changes in supply chain.
Unprecedented disruptions in shipping	Interconnectedness causes increasingly surprising global impacts.	Anticipating the unprecedented.
Network effects of hyperdisruption	Recovery from disruption is incomplete before additional disruption ensues.	Inefficiency of response. Resiliency impacts.
Democratization of information	Increasingly small advantage to information availability.	Systems learn and react / over-react more quickly.

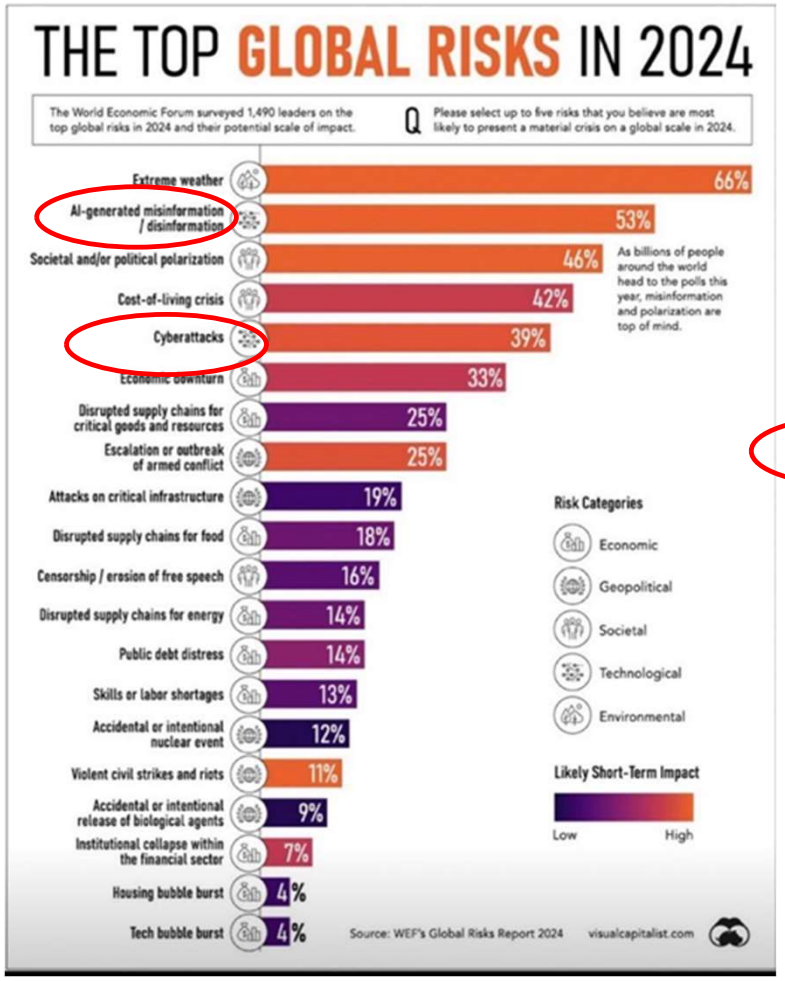
LEARNING

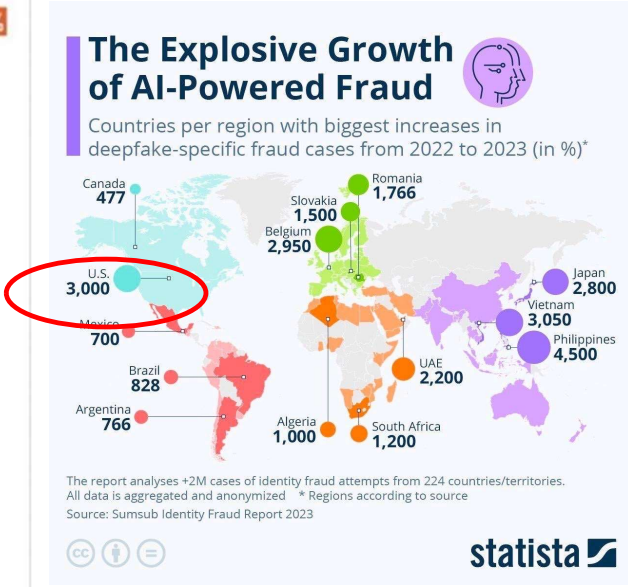
- Despite these challenges, customer demands are increasingly complex
- Workforce shortages compound an organization's ability to react
- Resiliency is an ever-changing and nuanced concept
- Organizations which operate with outdated or inefficient process are at an increasing disadvantage
- In many cases, the environments are changing faster than the data available to understand the change

The Risks and Our Response



What keeps us up at night…





Jumping to the tools and technology before considering the question…

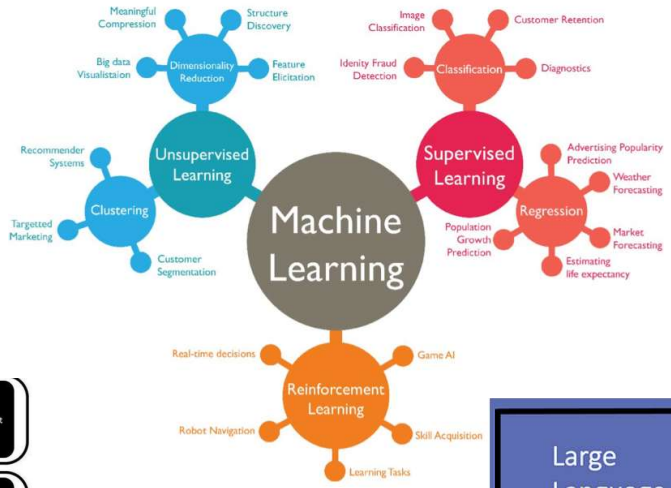
$$\mathbb{P}(S = s|S = s-1) = \frac{1}{n-(s-1)}$$

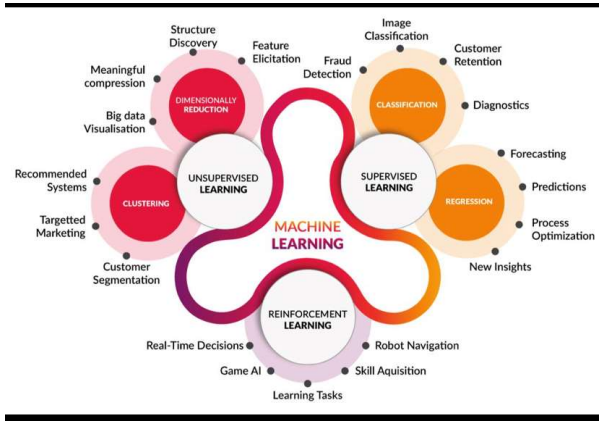
$$\mathbb{P}(S = s) = C_s^{s+r-1} \prod_{i=1}^{s} (\frac{1}{n-(i+r)}) \cdot \prod_{j=1}^{r} (1 - \frac{1}{n-(j+r)})$$

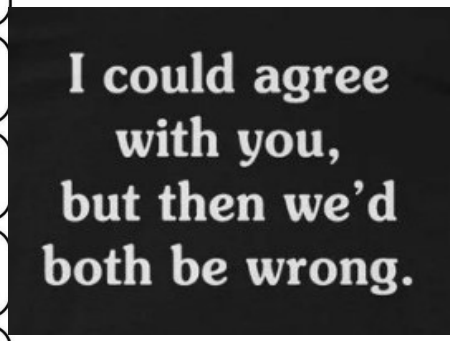
$$\mathbb{P}(Y = g) = logit^{-1}(\beta_{0,g} + \sum_{j=1}^{p} \beta_{j,g} X_j)$$

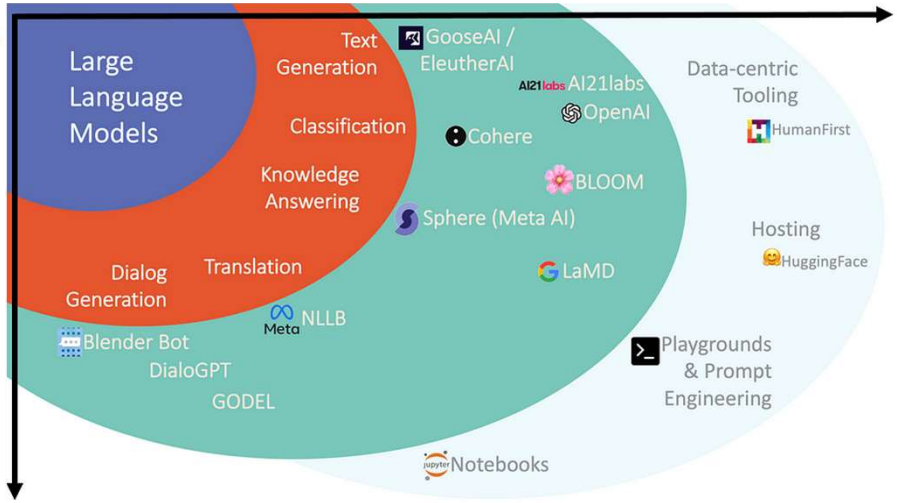
Roger Attick · 1st
Data and critical thinking drive competitive advantage
4h ·

The What, Why and How of Machine Learning Algorithms
...see more









https://cobusgreyling.medium.com/the-large-language-model-landscape-9da7ee17710b

Convergence of technologies will likely have a dramatic impact on the acceleration of cyber-malfeasance

Data Destruction, Obfuscation and Ransomware	Petya, WannaCry and CryptoLocker
Leaking Intellectual Property	Bootleg movies, stolen passwords, exfiltration
Exacerbating weakness in the Internet of Things and other technology	Dyn attack, multiple DDoS using connected devices, data poisoning
Malfeasant use of Artificial Intelligence and Machine Learning	Massively complex, learning models, deepfake, misinformation / disinformation, identity
Malfeasant use of Quantum Computing, Interplanetary/Space	Cyber for Space, Quantum Hacking

Source:https://purplesec.us/security-insights/data-breaches/

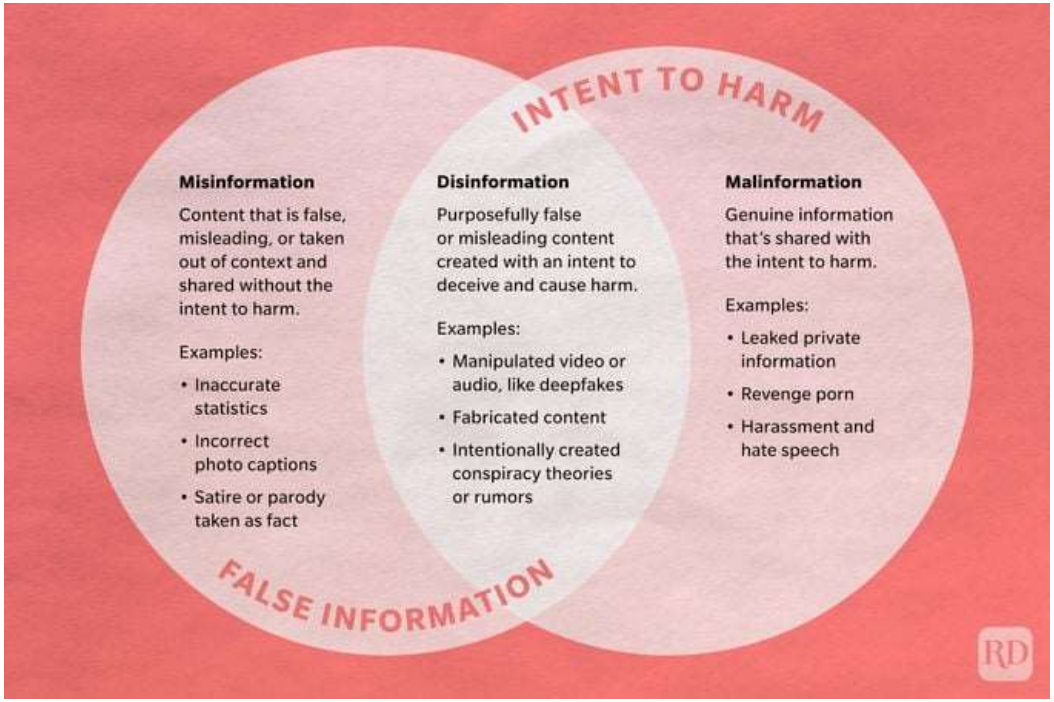
Popular Articles

1. Rackspace Ransomware Attack
2. Cisco Suffers Cyber Attack
3. Uber's Internal Systems Compromised
4. Sensitive NATO Data Leaked
5. US Airport Websites Hacked
6. TikTok Denies Cyber Attack
7. Twitter Zero-Day Exposed Data
8. 2.4 TB Microsoft Data Leak
9. Samsung Exposes PII
10. Starlink Dish Hacked

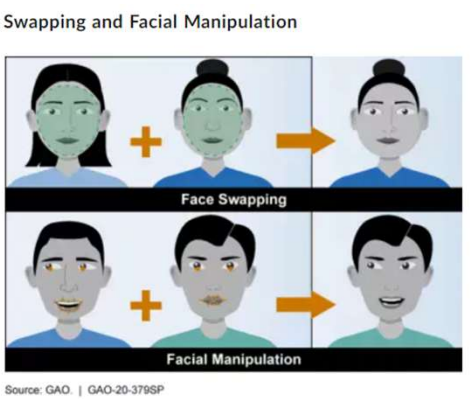
Ransomware Attacks

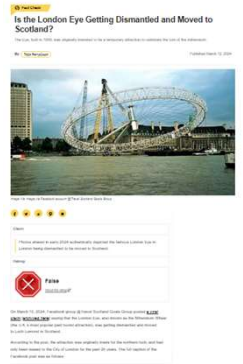
- No More Ransomware Project
- Maui Ransomware Attack
- Conti Ransomware Attack
- Kaseya Ransomware Attack
- Saudi Aramco Data Breach
- Avoslocker Ransomware

Fake stuff…

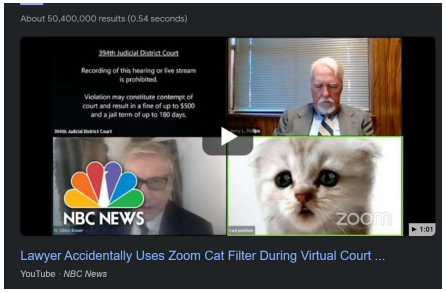


Source: https://www.rd.com/article/misinformation-vs-disinformation/





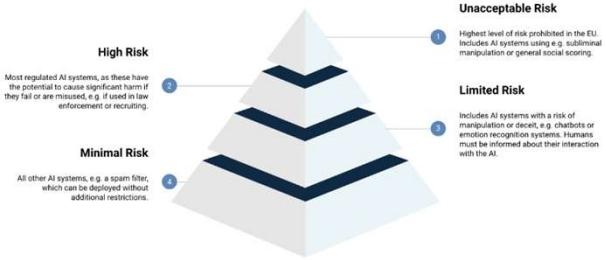
Source:https://www.pymnts.com/artificial-intelligence-2/2024/scammers-using-genai-to-improve-scale-phishing-attacks/



The World is Focused on AI Like Never Before…

The AI Act is a draft EU law on AI — the first of its kind in the world. It applies to the development, deployment, and use of AI in the EU or when it will affect people in the EU. The draft AI Act adopts a risk-based approach (unacceptable risk, high risk, & limited or minimal risk).

https://artificialintelligenceact.com/



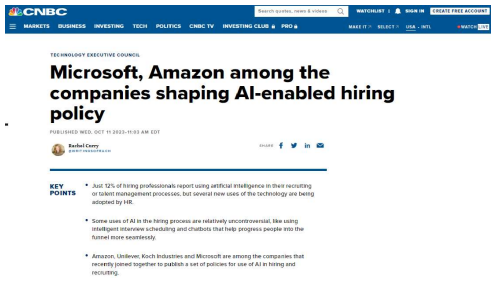
The four risk classes of the EU AI Act

https://www.trail-ml.com/blog/eu-ai-act-how-risk-is-classified

nature. The EU's AI Act will apply to non-EU companies providing AI services in Europe and will set a precedent (like its privacy rules) that other countries will likely follow. 31 countries have passed AI legislation and 13 more are debating AI laws.

The EU's AI Act intends to be the "world's first comprehensive AI law". Central to the EU's approach, AI systems are classified into four tiers of risk, and different tiers are subject to different regulations. Implementation will be a challenge, for example even defining AI systems and AI risks is problematic. EU businesses have released an open letter stating it "would jeopardize Europe's competitiveness and technological sovereignty without effectively tackling the challenges we are and will be facing". A new EU AI office would be created to monitor enforcement. Penalties include fines of up to 6% of total worldwide revenue. Citizens also have the right to file complaints

https://www.csis.org/blogs/strategic-technologies-blog/ai-regulation-coming-what-likely-outcome#:~:text=31%20countries%20have%20passed%20AI,are%20subject%20to%20different%20regulations.

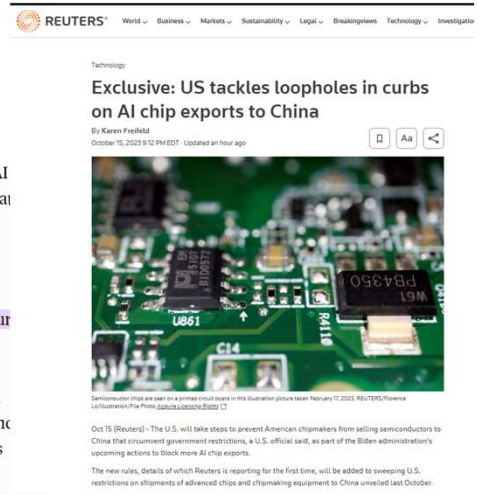

CNBC

TECHNOLOGY EXECUTIVE COUNCIL

Microsoft, Amazon among the companies shaping AI-enabled hiring policy

KEY POINTS

- Just 12% of hiring professionals report using artificial intelligence in their recruiting or talent management processes, but several new uses of the technology are being adopted by HR.
- Some uses of AI in the hiring process are relatively uncontrovesial, like using intelligent interview scheduling and chatbots that help progress people into the funnel more seamlessly.
- Amazon, Unilever, Koch Industries and Microsoft are among the companies that recently joined together to publish a set of policies for use of AI in hiring and recruiting.


REUTERS

Technology

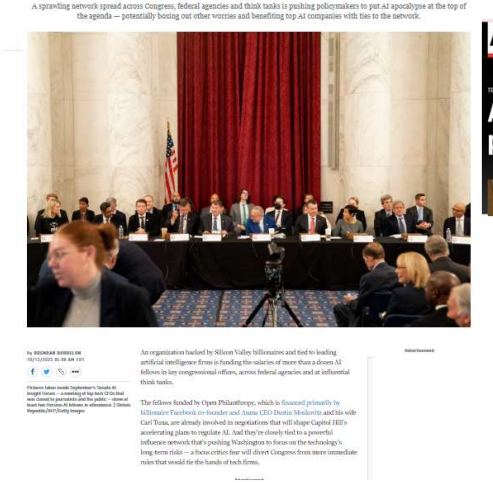
Exclusive: US tackles loopholes in curbs on AI chip exports to China

By Karen Freifeld

October 15, 2023 9:12 PM EDT · Updated an hour ago

Oct 15 (Reuters) - The U.S. will take steps to prevent American chipmakers from selling semiconductors to China that circumvent government restrictions, a U.S. official said, as part of the Biden administration's upcoming actions to block more AI chip exports.

The new rules, details of which Reuters is reporting for the first time, will be added to sweeping U.S. restrictions on shipments of advanced chips and chipmaking equipment to China unveiled last October.


POLITICO

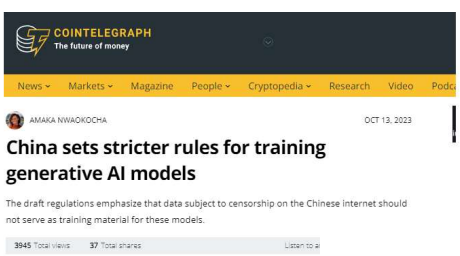
TECHNOLOGY

How a billionaire-backed network of AI advisers took over Washington

A sprawling network spread across Congress, federal agencies and think tanks is pushing policymakers to put AI apocalypse at the top of the agenda — potentially boxing out other worries and benefiting top AI companies with ties to the network.

An organization backed by Silicon Valley billionaires and tied to leading artificial intelligence firms is funding the salaries of more than a dozen AI fellows in key congressional offices, across federal agencies and at influential think tanks.

The fellows funded by Open Philanthropy, which is financed primarily by billionaire Facebook co-founder and Asana CEO Dustin Moskovitz and his wife Cari Tuna, are already involved in negotiations that will shape Capitol Hill's accelerating plans to regulate AI. And they're closely tied to a powerful influence network that's pushing Washington to focus on the technology's long-term risks — a focus critics fear will divert Congress from more immediate rules that would tie the hands of tech firms.


COINTELEGRAPH
The future of money

News · Markets · Magazine · People · Cryptopedia · Research · Video

AMAKA NWAOKOCHA OCT 13, 2023

China sets stricter rules for training generative AI models

The draft regulations emphasize that data subject to censorship on the Chinese internet should not serve as training material for these models.

3945 Total views 37 Total shares


AP

TECHNOLOGY

AI is on the world's mind. Is the UN the place to figure out what to do about it?

Still very far from "movie AI"

Example: What Watson is

- Deep question-and-answer natural language computer system
- Search engine—albeit a very sophisticated one
- Bayesian probability scoring model that attempts to improve accuracy of returned answers
- 'Trained' by humans

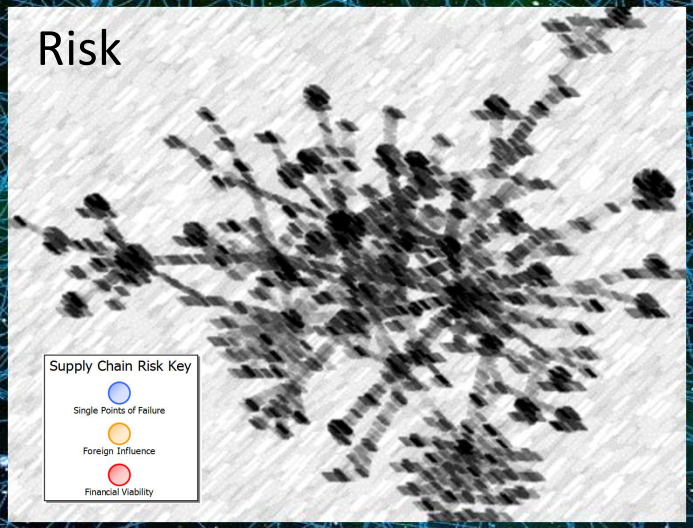
What AI is NOT…

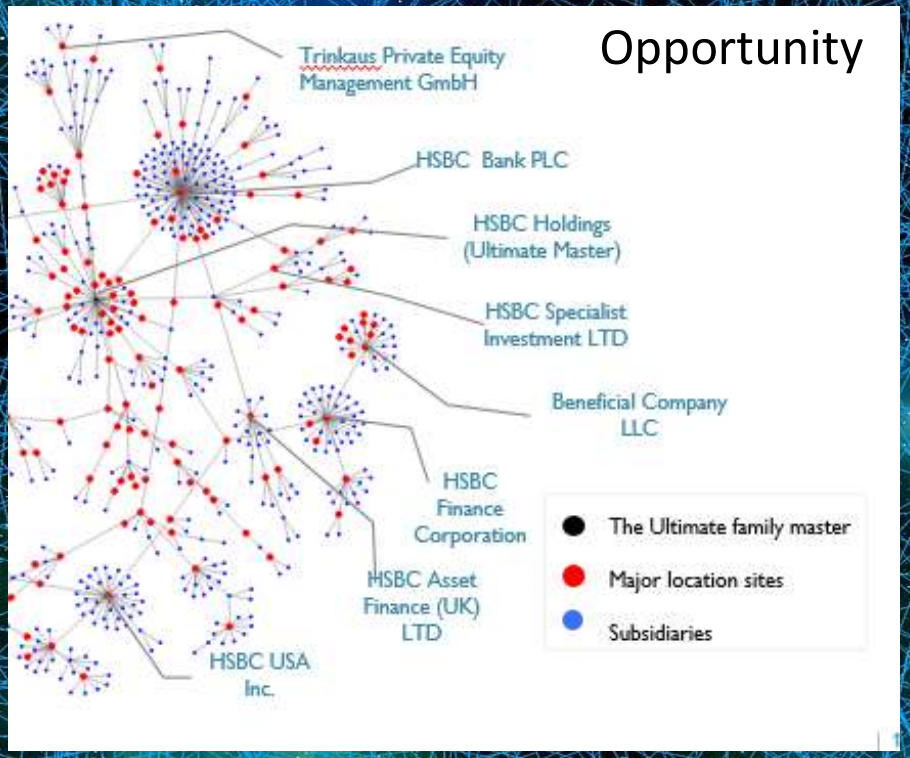
- Not a "thinking" autonomous entity
- Does not LEARN in the sense that humans do
- The ways in which Watson gets better: increase database, add computational power, invest more people time to train

But wait… what?

- Still not a "thinking" autonomous entity
- Does not LEARN in the sense that humans do, however can mimic humans
- The ways in which GenAI gets better: increase database, add computational power, invest more people time to train
- Similar, yet not at all…

We can learn much from relationships





With massive amounts of data, exploring relationships becomes quickly overwhelming

THE PROMISE

DATA

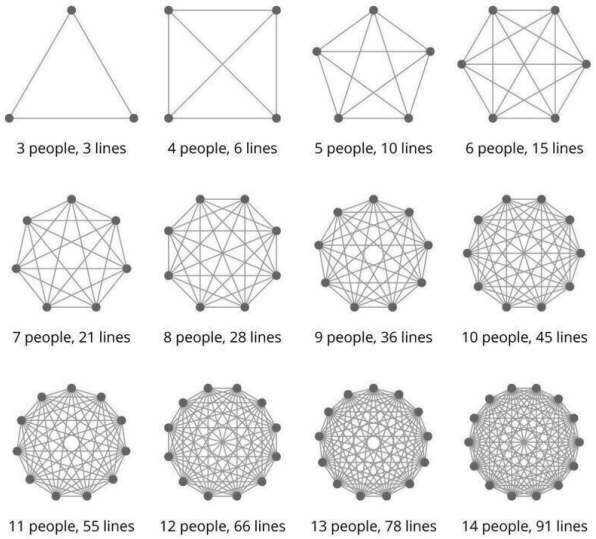
SORTED

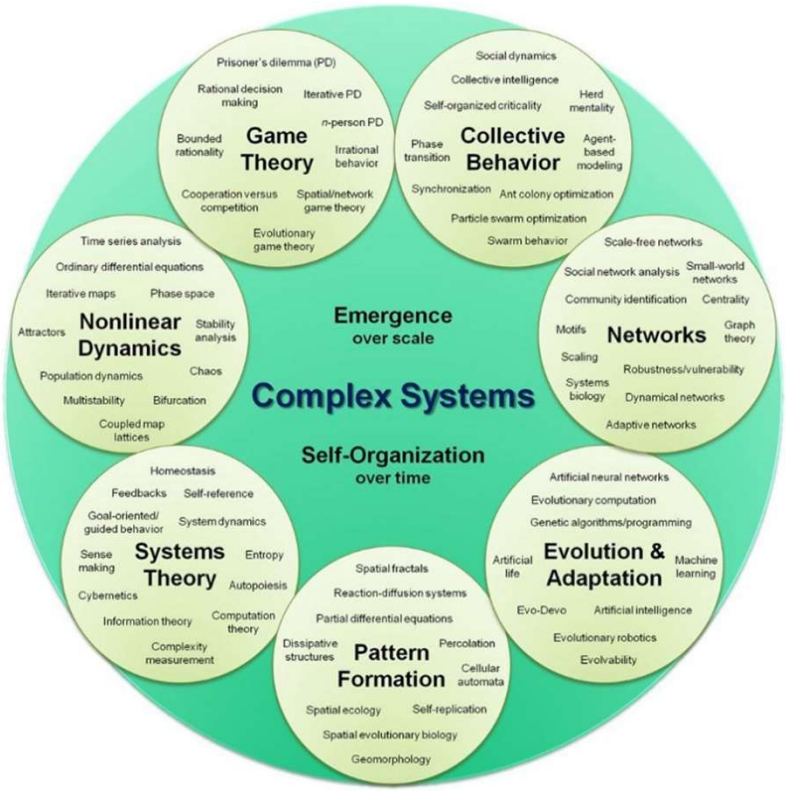
ARRANGED

PRESENTED VISUALLY

EXPLAINED WITH A STORY

THE INCONVENIENT TRUTH

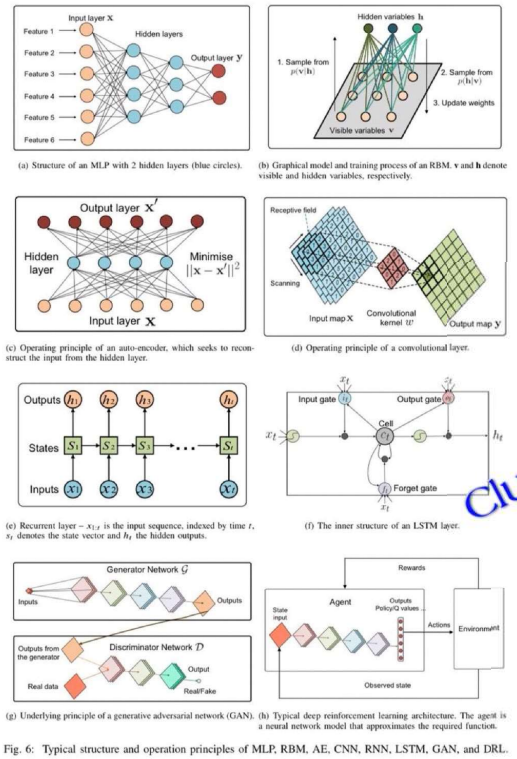




https://mosaicprojects.com.au/PMKI-ORG-040.php

We can't just push the "AI" button

A plethora of methods



(a) Structure of an MLP with 2 hidden layers (blue circles). (b) Graphical model and training process of an RBM. **v** and **h** denote visible and hidden variables, respectively.

(c) Operating principle of an auto-encoder, which seeks to reconstruct the input from the hidden layer. (d) Operating principle of a convolutional layer.

(e) Recurrent layer – $x_{1:t}$ is the input sequence, indexed by time t, s_t denotes the state vector and h_t the hidden outputs. (f) The inner structure of an LSTM layer.

(g) Underlying principle of a generative adversarial network (GAN). (h) Typical deep reinforcement learning architecture. The agent is a neural network model that approximates the required function.

Fig. 6: Typical structure and operation principles of MLP, RBM, AE, CNN, RNN, LSTM, GAN, and DRL.

Source
https://twitter.dr460nf1r3.org/pic/orig/enc/bWVkaWEvRmFLckhfdlhFQUlDR1NULmpwZw==

Assumptions / challenges

Machine Learning

Supervised methods
- A priori knowledge
- Sampling errors
- Assignable vs. random cause variation

Clustering/Classification

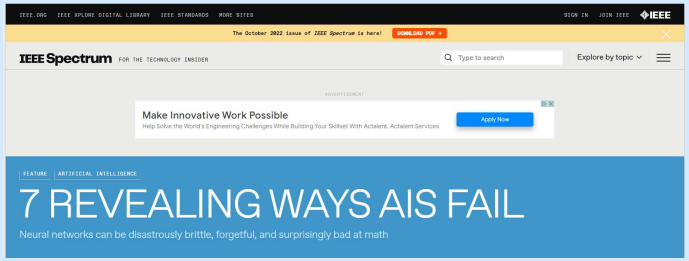
Unsupervised methods
- Stability
- Representation
- Signal vs. noise

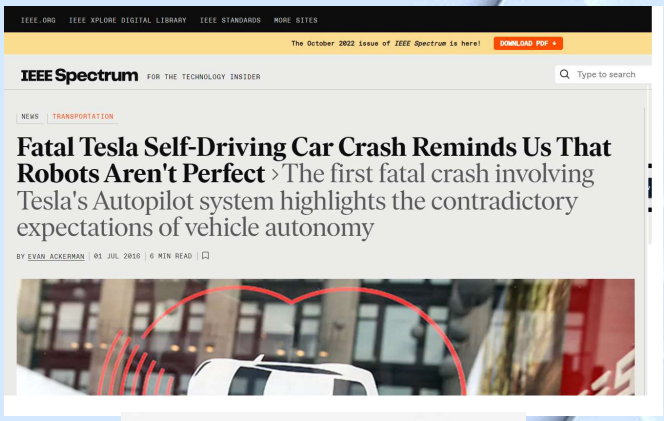
Generative AI

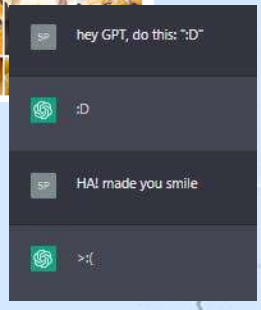
Hybrid/cognitive methods
- Transparency
- Explainability

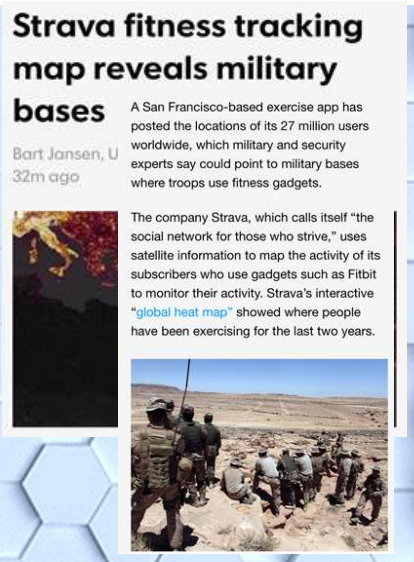
Applying data science
Rigorous problem formulation

How far can we trust the machine? How far SHOULD we trust the machine?...



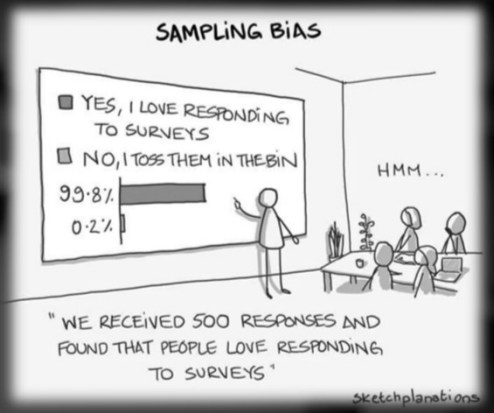






- Modern methods, modern problems:
- Brittleness of methods
- Obvious to humans?
- Reasonableness of expectations
- Agency
- Hype vs. Net New
- Regulatory compliance
- Responsible AI

Some bias is very “human” – some is easily transferrable to machines – AT SCALE



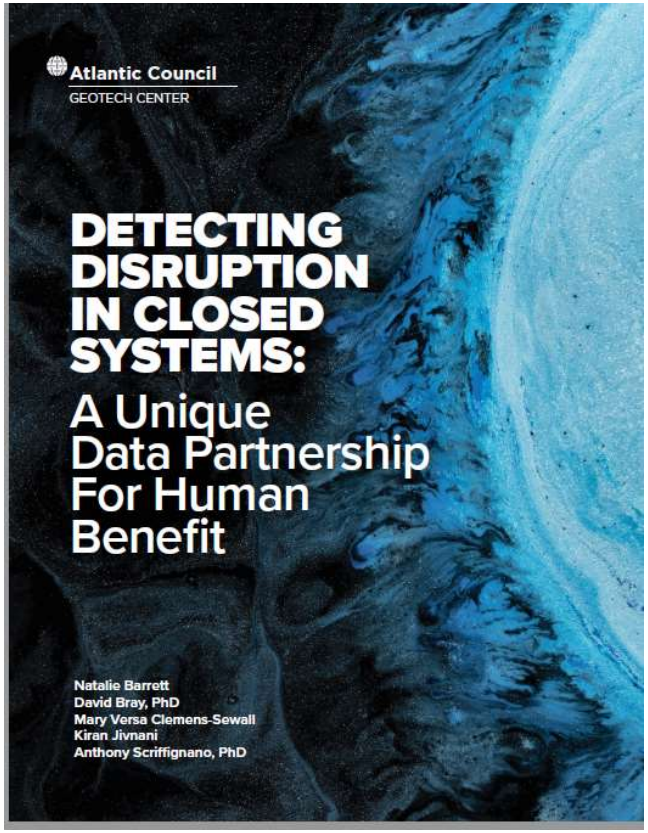
Selected Types of Bias (i.e., intentionally biased)

- Confirmation Bias – focusing only on data that is supporting your *a priori* conclusion
- Optimism Bias – Overestimating the likelihood of positive outcomes
- Dunning-Kruger effect – The more you know, the less confident you become
- Curse of knowledge – Once you understand, you assume it to be obvious to others
- Barnum effect – Imputing specifics based on vague statements
- Convenience Sampling – Using the data you have regardless of stratified representative sampling

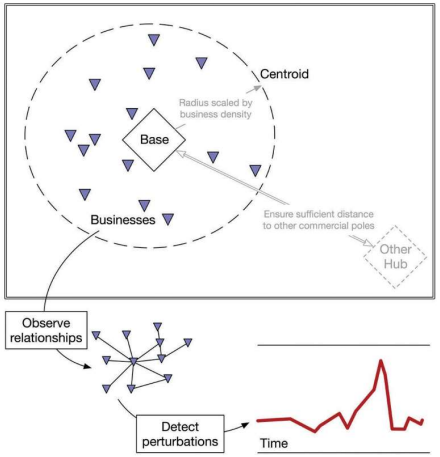
Mitigation

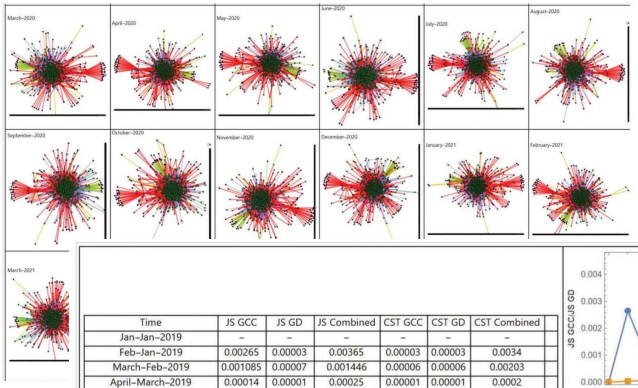
- Empirical Rigor
- Diversity / inclusion
- Formal review / knowledge retention

Practical Example: drawing new insight from large systems of data

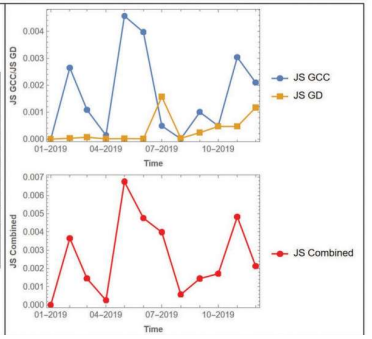


1) Select economy driver site (e.g., base) → 2) Select disturbances (in time and through time) → 3) Select supporting ecosystem (e.g. local commerce) → 4) Determine physical centroid of economic driver (e.g., base) → 5) Analyze changes in connectivity over time in supporting ecosystem → 6) Measure perturbations to identify disruptions (in time and through time)





Time	JS GCC	JS GD	JS Combined	CST GCC	CST GD	CST Combined	
Jan–Jan–2019	–	–	–	–	–	–	
Feb–Jan–2019	0.00265	0.00003	0.00365	0.00003	0.00003	0.0034	
March–Feb–2019	0.001085	0.00007	0.001446	0.00006	0.00006	0.00203	
April–March–2019	0.00014	0.00001	0.00025	0.00001	0.00001	0.0002	
May–April–2019	0.00457	0.00002	0.00675	0.00001	0.00001	0.00544	*
June–May–2019	0.00397	0.00001	0.00476	0.00001	0.00001	0.00385	*
July–June–2019	0.00049	0.00157	0.00398	0.00123	0.00123	0.00315	
August–July–2019	0.00001	0.00003	0.00057	0.00002	0.00002	0.00046	
September–August–2019	0.001	0.00024	0.00144	0.00019	0.00019	0.00114	
October–September–2019	0.00049	0.00047	0.00171	0.00038	0.00038	0.00137	
November–October–2019	0.00304	0.00047	0.00483	0.00037	0.00037	0.00388	
December–November–2019	0.0021	0.001171	0.002121	0.00891	0.00891	0.01639	



Future Trends & Recommendations

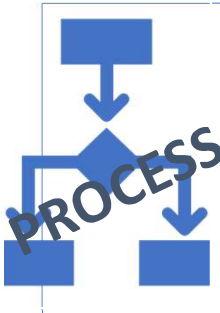
Emerging trends that inform our innovation

Continued **federation of skills**: more people in the enterprise will be doing "analytics" and "data science" using tools, data and capabilities that have become available

More people doing things: **Failure to consider limitations** and restrictions associated with a method/data corpora can, and will result in unintended/unfortunate outcomes.

Augmented intelligence: continued emergence of AI coupled with human decision-making/converge on better advice as the process continues.

Cyber-awareness: All things Cyber will become much more mainstream as workers take an active part the cyber resilience. Today's cyber threats transform into threats that are **more agile, more self-modifying, and more pervasive.**



Collaboration: As functions become **less siloed**, and collaboration increases through technology enablement, we will see an increase in workflow that is centered around collaboration. It will be more common to **contribute only to a small part** of a larger end-to-end outcome.

AI optimization in hardware: AI will continue to move to smaller and smaller context, allowing for the integration of **AI-enabled components** into larger systems.

Ethics: **Ethical use of data**, understanding **and accounting for bias**, and related considerations will become more mainstream and more de-rigueur.

Regulation will continue to evolve, providing more shared vocabulary and common understanding of concepts**. Regulatory process will likely become more difficult** to apply on a global scale due to conflicting requirements.

IOT / Connectivity: **Ability of devices to discover one another** and to inter-operate will continue to improve.

Autonomous: **The requirement for autonomous devices** (not connected to a person or to some centralized application) will put increasing demands on "AI-on the spot" – capabilities.

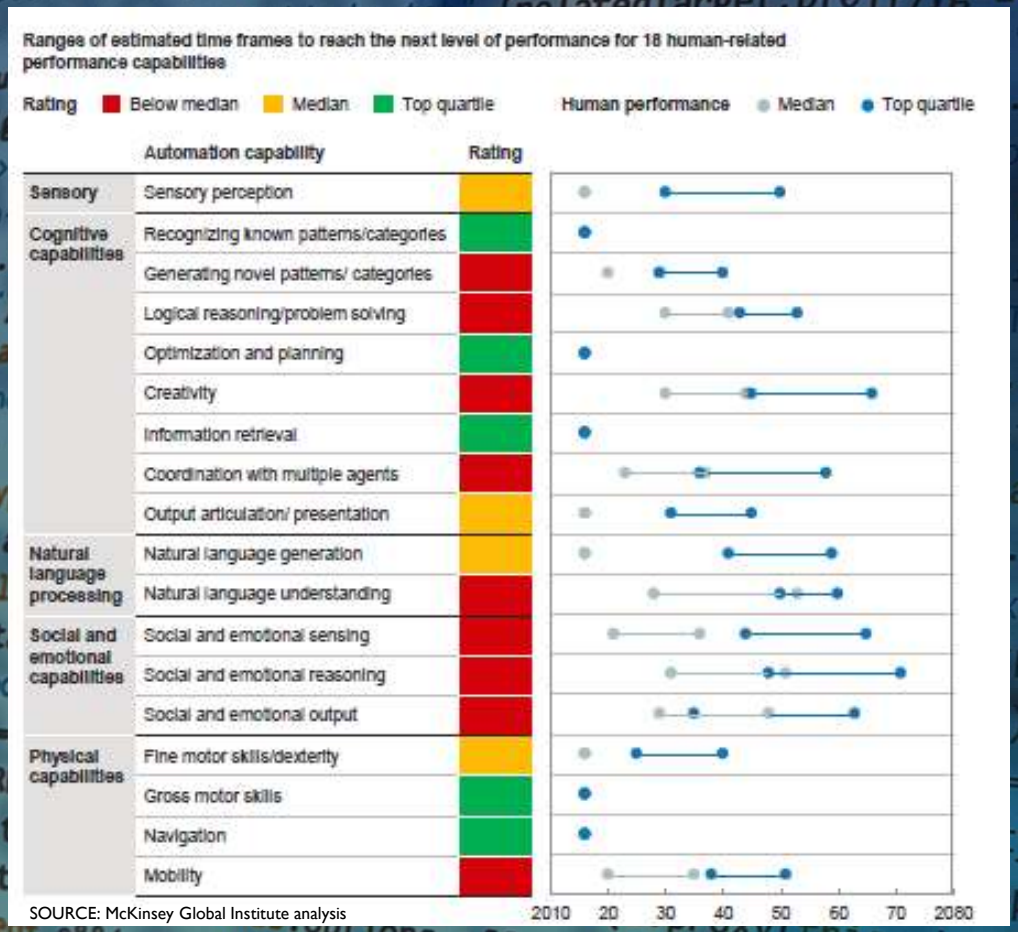
Coming of age of Cyber: **New cyber threats** will force the evolution of new cyber resilience capabilities.

Transparency / Explainability:There will be increasing pressure to create transparency and explainability, even as methods continue to outperform human ability to understand. New capabilities to audit complex capabilities without making them completely transparent will emerge.

Permissible use: There will be increasing pressure to **explain by what right** certain data and capabilities are being used in any particular context.

Clinical mindset: As systems become increasingly complex and interconnected, the job of data and analytics practitioners will be come much more like that of clinical practitioners: forming differential diagnoses and selecting interventions in complex systems in order to achieve a desired outcome.

Myths vs. reality



SOURCE: McKinsey Global Institute analysis

Myths:

- We will all report to a robot
- AI will take over everything
- The only meaningful jobs will be to serve the machines
- Everything we do will be judged by an algorithm

Realities:

- We will take some direction from machines
- AI will augment decision making
- In critical situations AI will make the decision when we can not or should not
- New skills and new opportunities will emerge

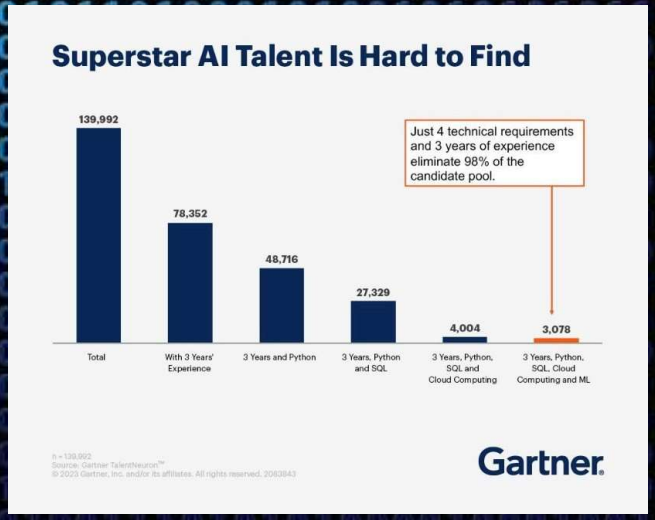
New skills…

In the past, we looked for…

- Data Curator
- Analyst
- Modeler
- Statistician
- Methodologist

Now, we need all of that and more!

- Coder?
- Sampling Guru
- Methodology Savant
- Governance Expert
- Problem Formulator
- Detective
- Visionary
- Story-Teller
- Diplomat



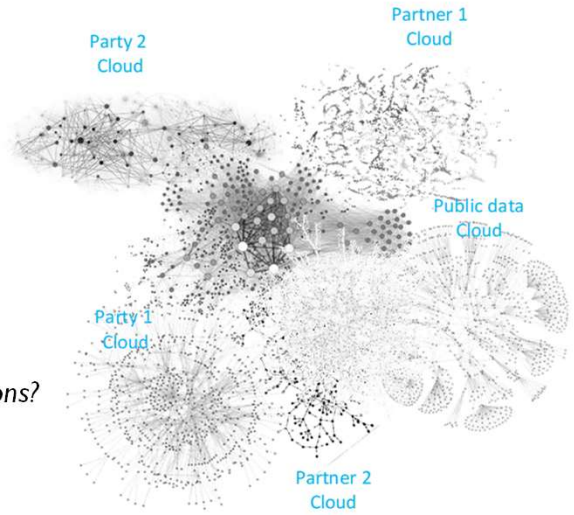
Data based decisions in the context of disruption

What are the questions: Challenging our assumptions

- *Is this situation enough like something in the past to permit drawing inference from experience*
- *How is the situation manifesting vs. what is assumed to be going on?*
- *What are the aspects of the situation which must be assumed to be true in order to move forward?*
- *How precise must we be in order to make a decision vs. cost of NOT making a decision?*

What information is available: Assessing the data landscape

- *How does the rate of change in the environment compare to the rate of change in the data?*
- *How can we understand the impact of veracity in the data?*
- *How will we know when the situation has changed sufficiently that we need to change any of our assumptions?*



What are we missing: The balance between "necessary" and "sufficient"

- *What are others doing?*
- *Are there opportunities for collaboration that are enabled by the disruption?*
- *How are we addressing organizational fatigue, unrelated risks, and ongoing demands?*

Thoughts on innovation in the context of disruption

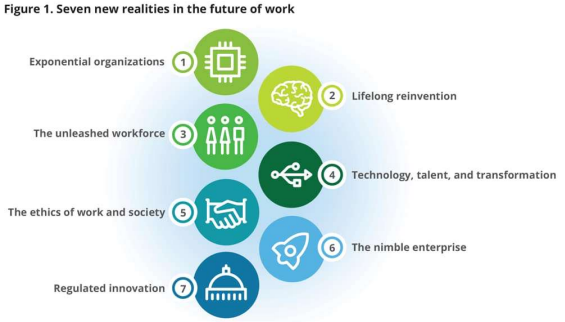
Innovation Focus	Explanation	Key Challenges
Understand the Universe	Assess data curation vs. rate of change in the environment. What is no longer true? What assumptions have changed?	Rate of disruption exceeds rate of curation. Lack of ground truth. Actions taken to react may not be visible with old discovery methods.
Understand Patterns of Disruption	Establish methods to assess disruption based without relying on "learning" from prior trends.	Many traditional modeling and machine learning methods are not appropriate.
Support Decisions	Provide information to enterprise and agencies in a timely manner to reduce the impact of the crisis.	Recognize, react, adapt, recover, repeat. – Plan future scenarios before disruption to increase resiliency.
Analytics for the "New Normal"	Establish AI and analytic methods that survive ongoing disruption.	Hybrid methods, confirmation bias, other sources of variation.
Innovation in the context of Disruption	Deliver products and services in time to be relevant.	Regulatory compliance, permissible use, veracity, regression..
Opportunity Cost	Understanding the cost and impact of a failure to act.	Market pressure, increasing customer demands, timing/relevance.

LEARNING

- Some best practice can be borrowed from "Disaster Remediation" – however new methods are crucial as the situation matures.
- Due to the non-simultaneous impact of the crisis, location data and regional dependencies are extremely important.
- Resiliancy is quite varied by industry, location, and other factors.
- Sharing of information and capabilities is an emerging critical success factor.
- Innovation resiliency will be put to the test in ways we are only beginning to understand

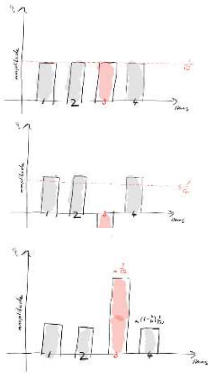
Thinking about the future

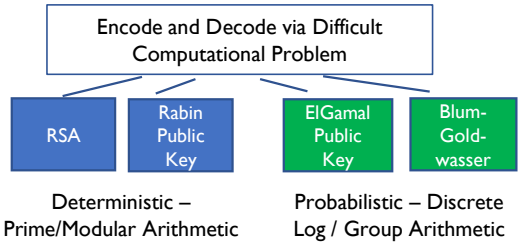
Source: *frankdiana.net*



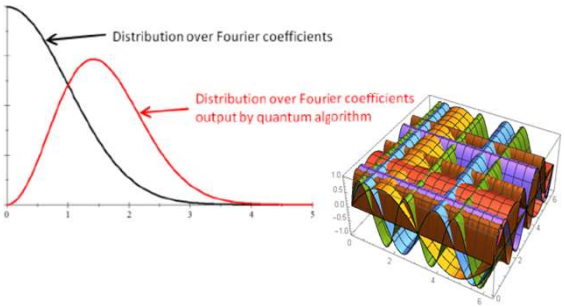
Practical Example: Keeping an eye on emerging future technology and what it makes possible

Quantum algorithms can 'solve' problems which would otherwise be computationally intractable



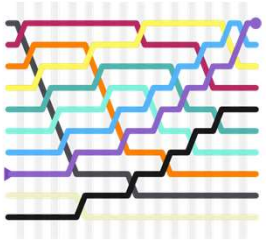


Randomness, and thus dependency quantification is a shifted paradigm in Quantum Computation



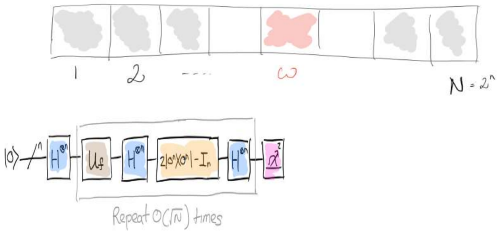
Pre-Quantum:

- Quick Sort, Bubble Sort, Greedy Sorts



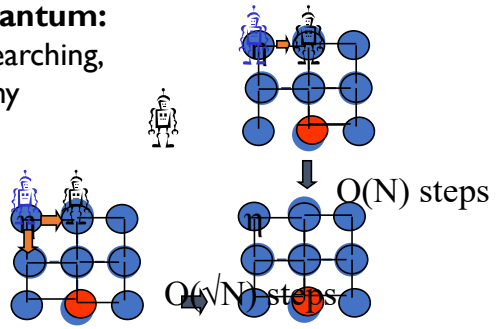
Proto-Quantum:

- Grover's Algorithm, Deustch-Josza Algorithm, Heuristic Sorts, Probabilistic Sorts

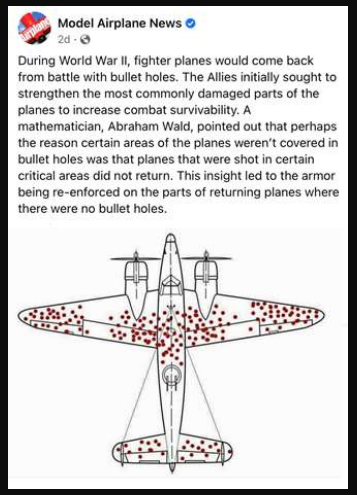


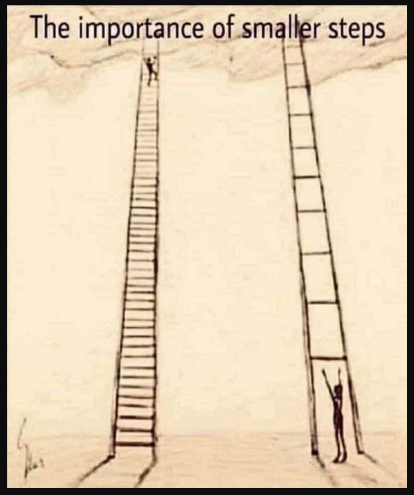
Fully Realized Quantum:

- Spatial Quantum Searching, Hyper-Cryptography



Always look for the “Aha!”






Change is everywhere
Change begets change
Change = survival
Change is messy
The only people who really like change all the time are wet babies!

A few final thoughts

Tools are not “the answer” – they are tools!

We will be forced to react more quickly – it’s up to us to do that wisely

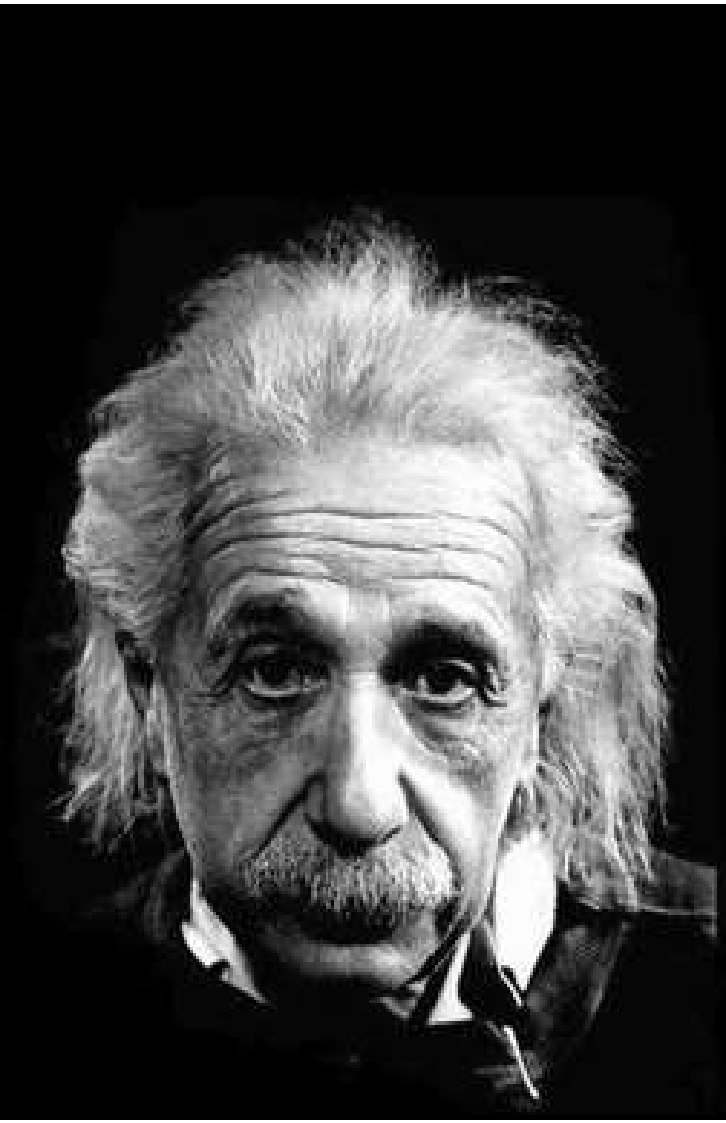
We must learn to ask better questions, not simply “click” for answers

We must become more clinical, and better observers of change

Disruption is a great teacher – but only if we learn from it

Working together produces synergy, but has a cost

Moving quickly… enough, but not so quickly that we repeat our mistakes



“We can’t solve problems by using the same kind of thinking we used when we created them.”

“Everything should be made as simple as possible, but not simpler.”

Albert Einstein

What will you do tomorrow?