



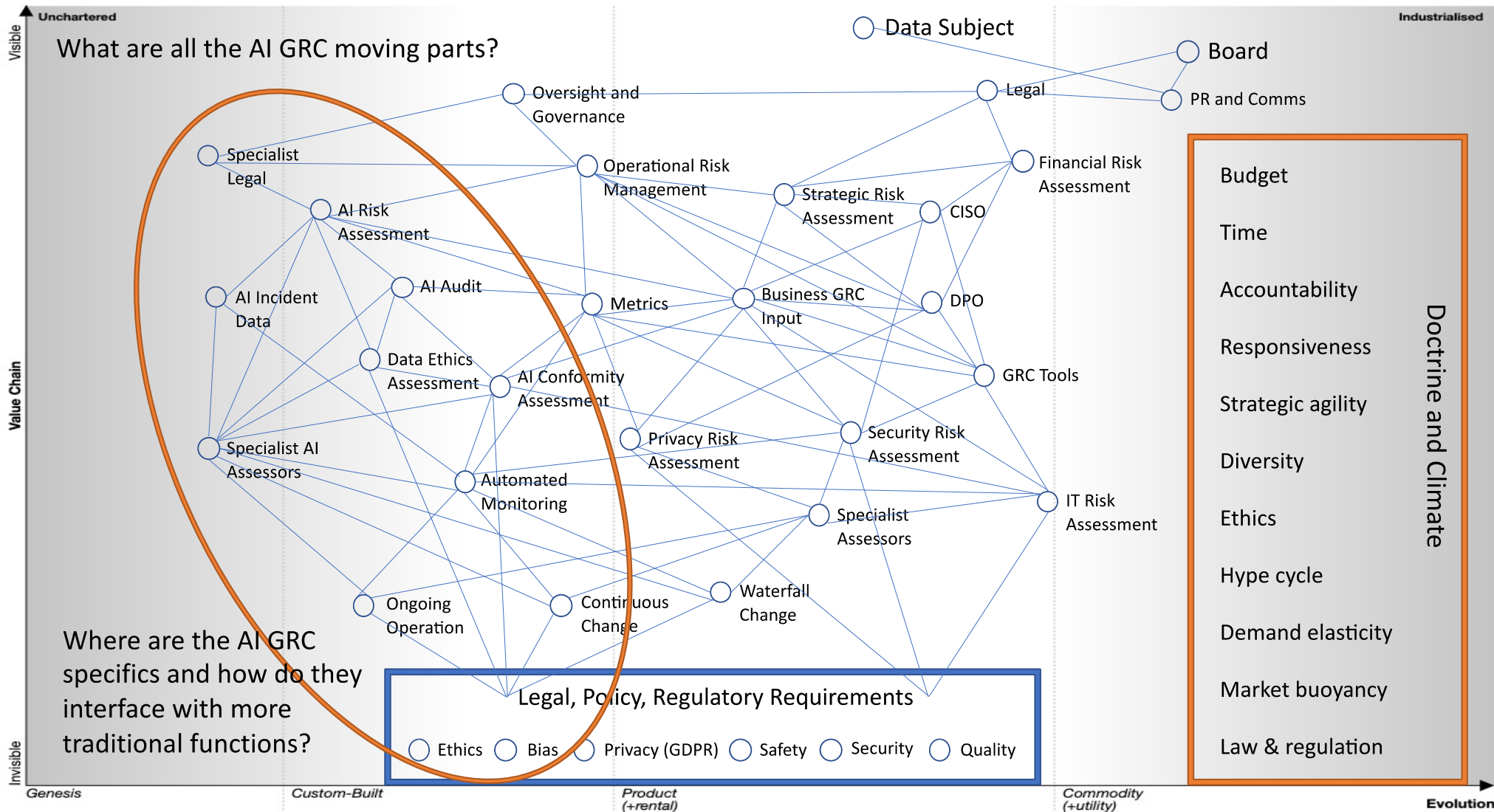
Making Space for Governance

Sarah Clarke - Owner Infospectives Ltd

Cybersecurity, Data Protection, Novel Technology GRC

Vendor Security Governance Guest Lecturer for Manchester University, IEEE
and World Ethical Data Foundation Contributor, Emeritus Fellow
ForHumanity

Lines of Communication



In a nutshell:

Accountability for risks must be **FORMALLY** assigned to decision-makers who have influence and means to effect change...

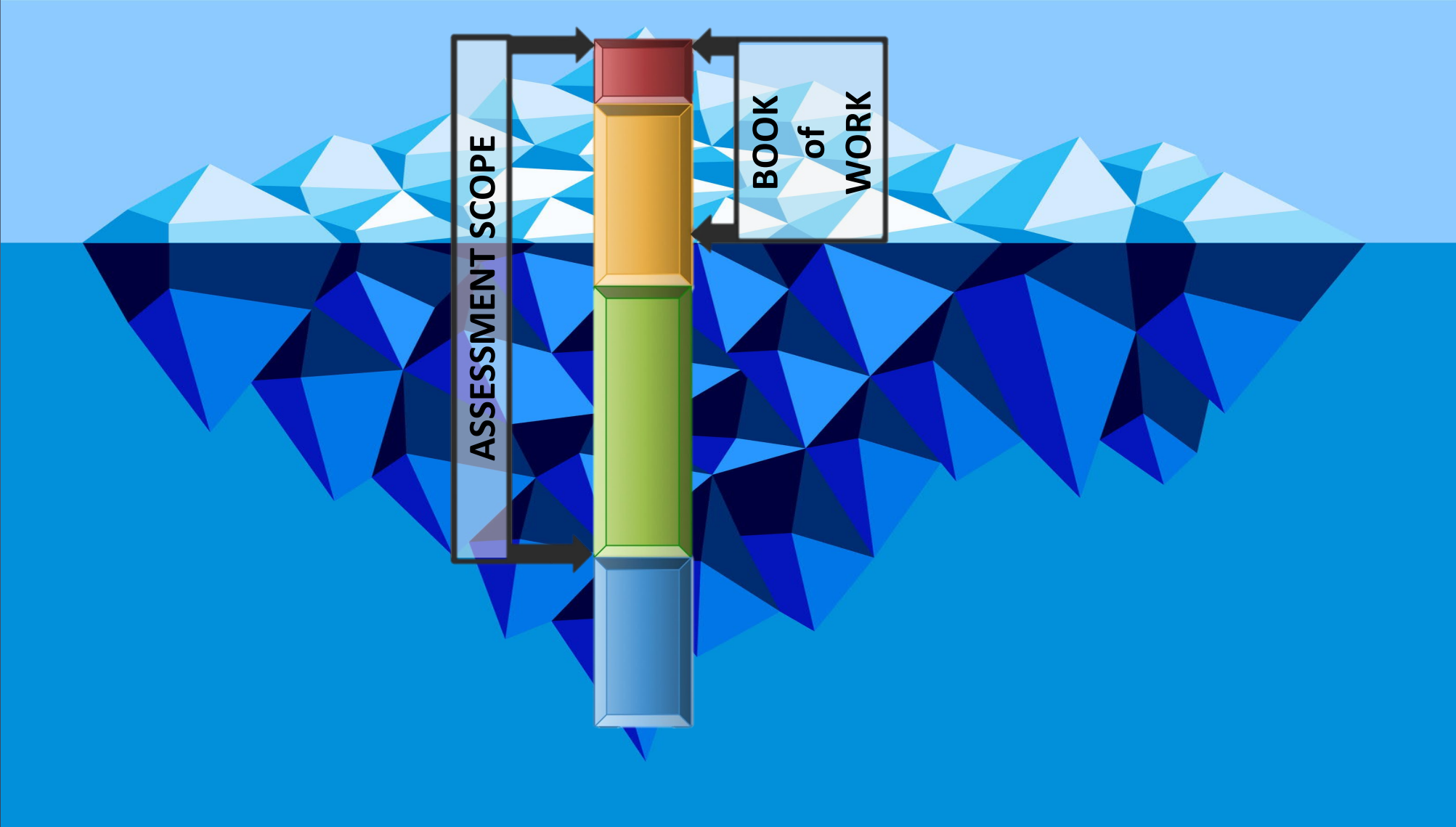
...decision-makers must **REMAIN** accountable until data/tech use for specified purposes ceases, or the role is formally handed over...

...**SPECIALISTS** are accountable for providing clear information about requirements, risks, and blockages...

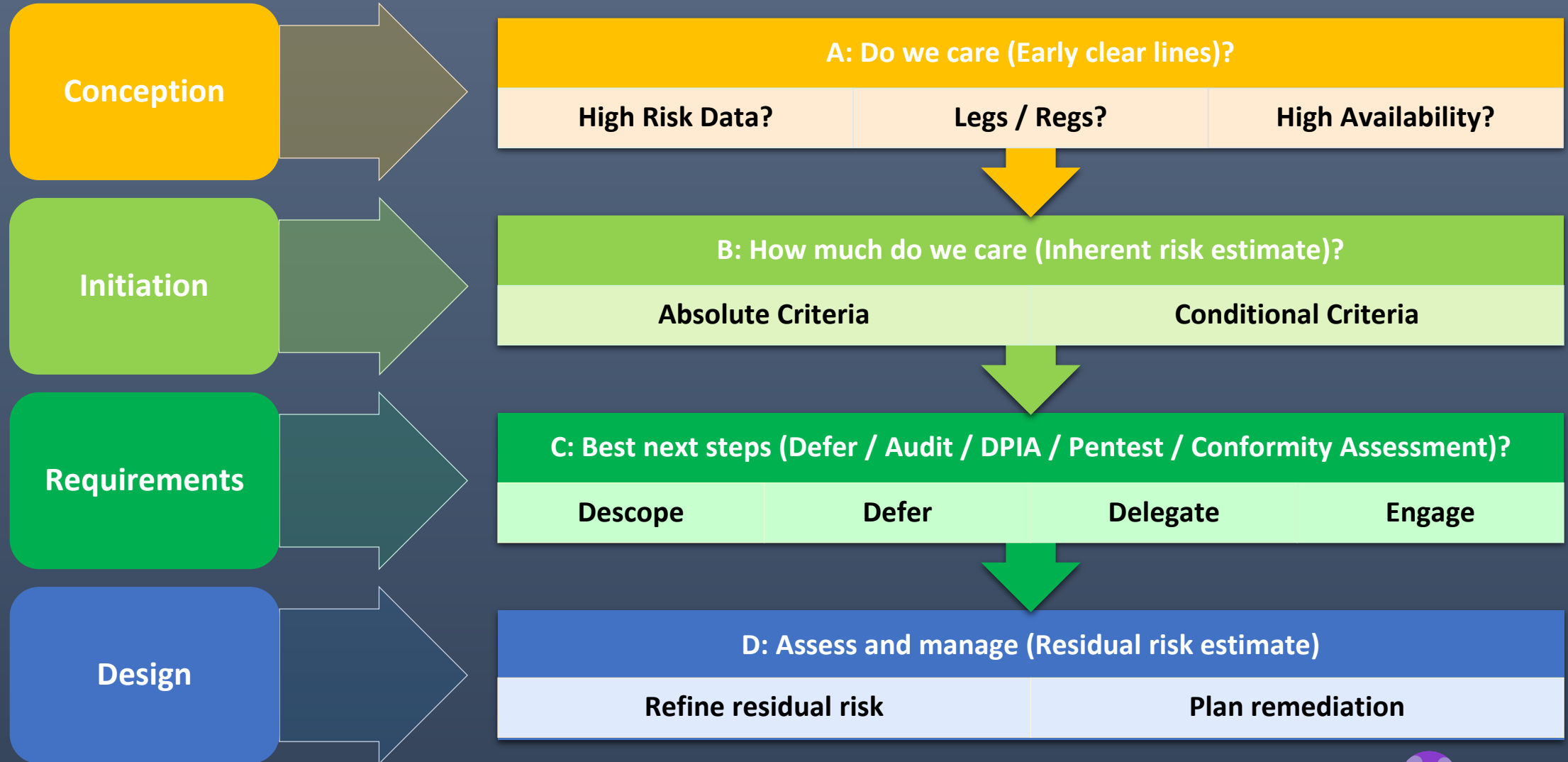
DECISION-MAKERS are accountable for providing sufficient time, money, and support to make that work...

...because **NO-ONE** should be accountable for something that they can't influence, or don't understand.

Early Triage

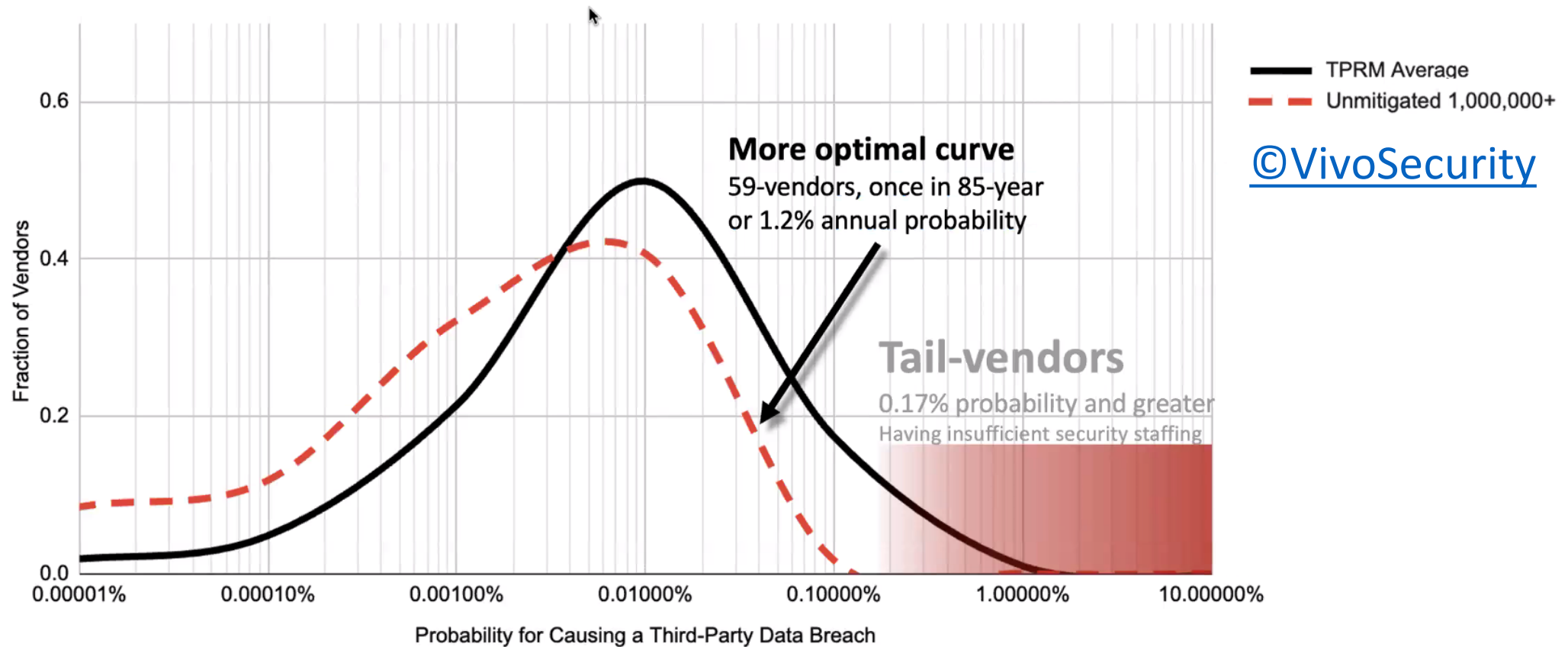


Where would you start?



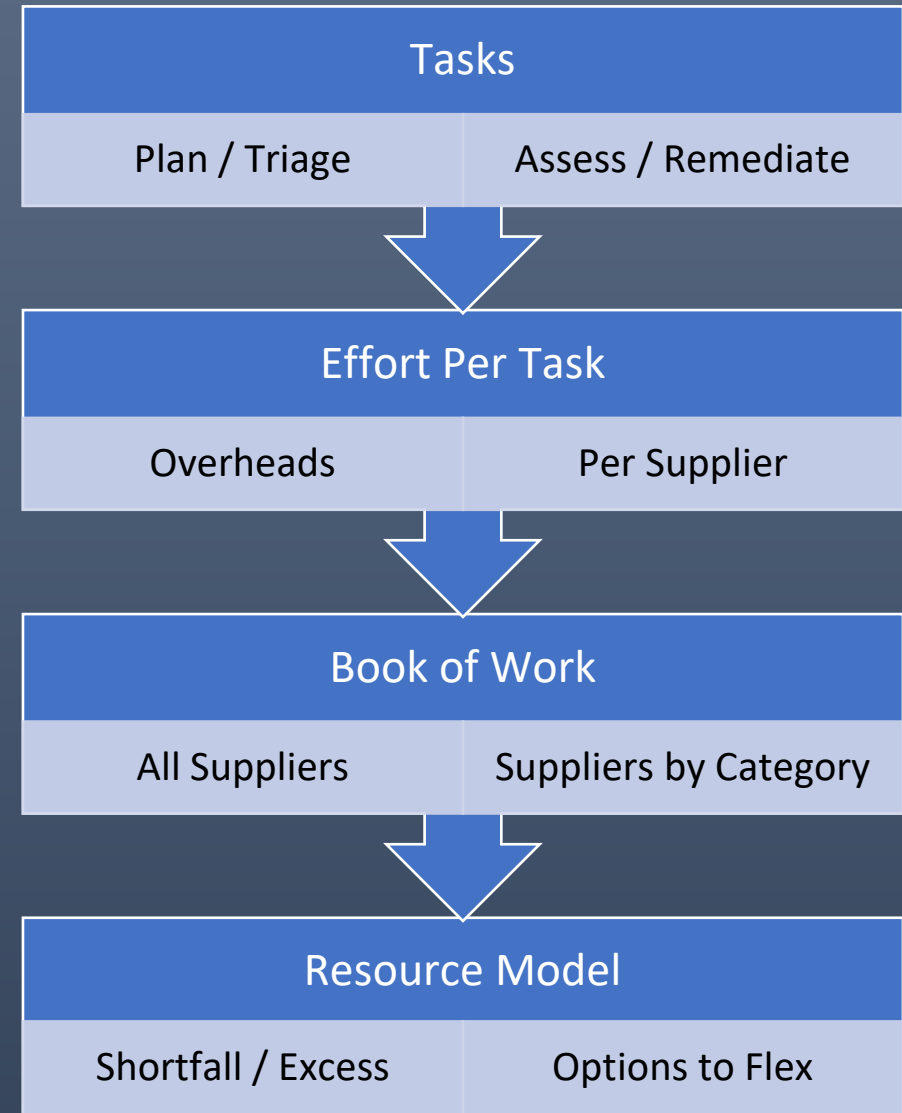
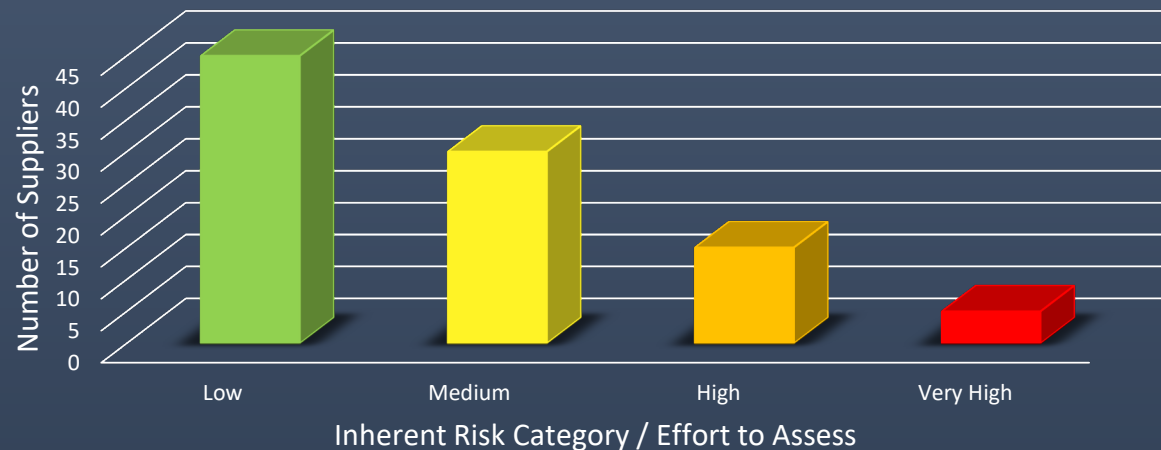
A TPRM program that mitigates risk from tail-vendors

Rule (strategy): Don't expose 1M+ records through vendors over 0.1%-annual probability
Eliminates: 10-vendors (69-vendors to 59-vendors)
Cumulative-probability: **3%** (once in 34-years) to **1.2%** (once in 85-years)

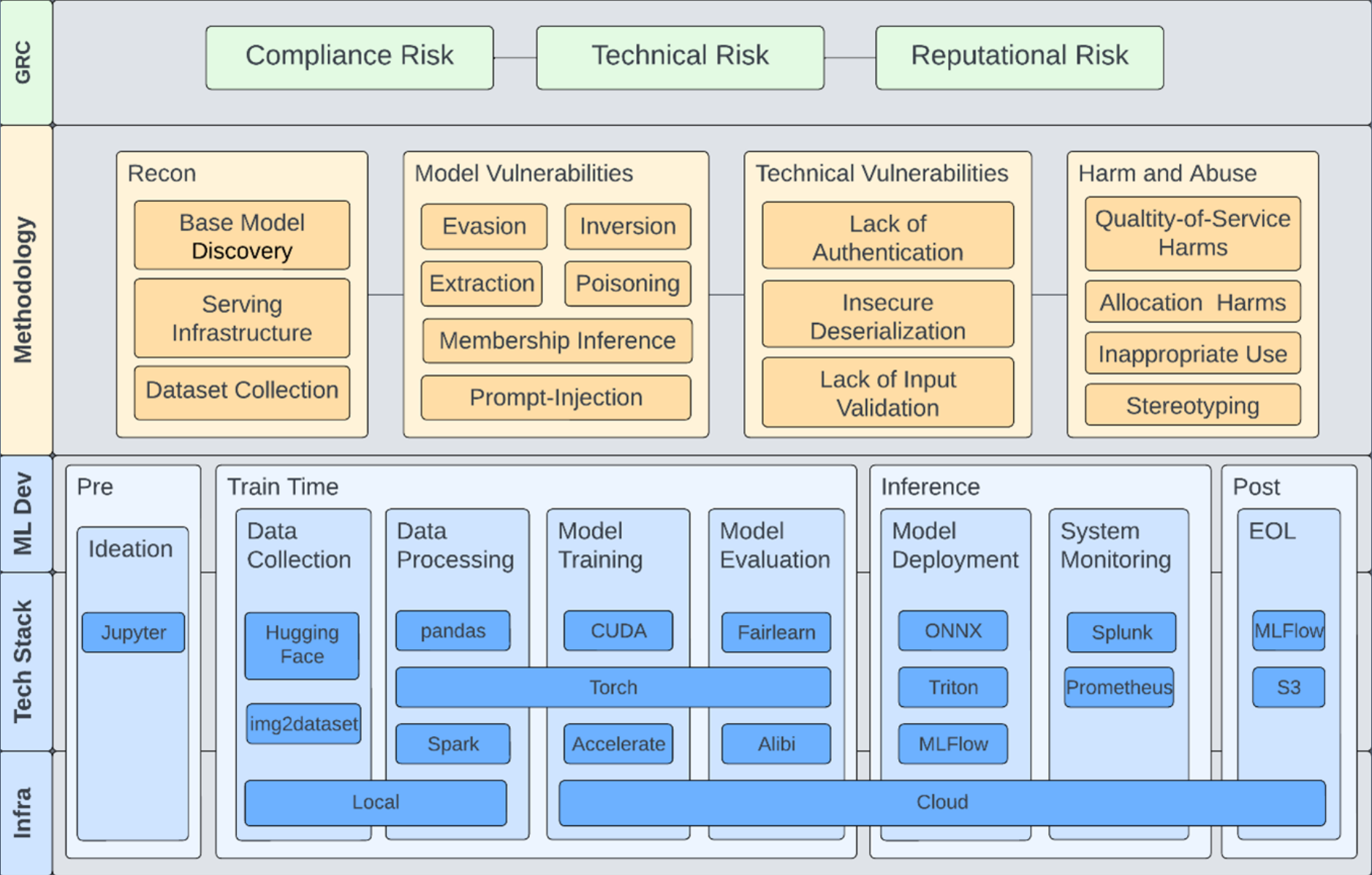


Resource modelling

Activity	Per Supplier / Overhead
Triage	Per Supplier
Assessment	Per Supplier
Remediation / Retesting	Per Supplier
Contract negotiation	Per Supplier
Governance data collation	Per Supplier
Remediation data collation	Per Supplier
Regular on-going governance	Per Supplier
Risk management	Per Supplier
Metrics and reporting	Overhead
Process development and planning	Overhead
Training	Overhead
Stakeholder management	Overhead



Defining uncertainty



Source: [NVIDIA Red Teaming Introduction, June 2023](#)

AI LIFECYCLE

Acquire and
Prepare Data

Train Model

Model
Evaluation

Model
Deployment

Model
Monitoring

AI AUDIT / ASSESSMENT – OVERSIGHT / DOCUMENTATION / PROCESS - REVIEW

Compliant	Compliant	Compliant	Compliant	Compliant
Non-Compliant	Non-Compliant	Non-Compliant	Non-Compliant	Non-Compliant
Unknown	Unknown	Unknown	Unknown	Unknown

AI AUDIT / ASSESSMENT – SYSTEM / CODE / DATA - TECHNICAL TESTING

Compliant	Compliant	Compliant	Compliant	Compliant
Non-Compliant	Non-Compliant	Non-Compliant	Non-Compliant	Non-Compliant
Unknown	Unknown	Unknown	Unknown	Unknown

Assurance and governance guidance is evolving



OWASP
AI security & privacy guide

Information Technology Laboratory

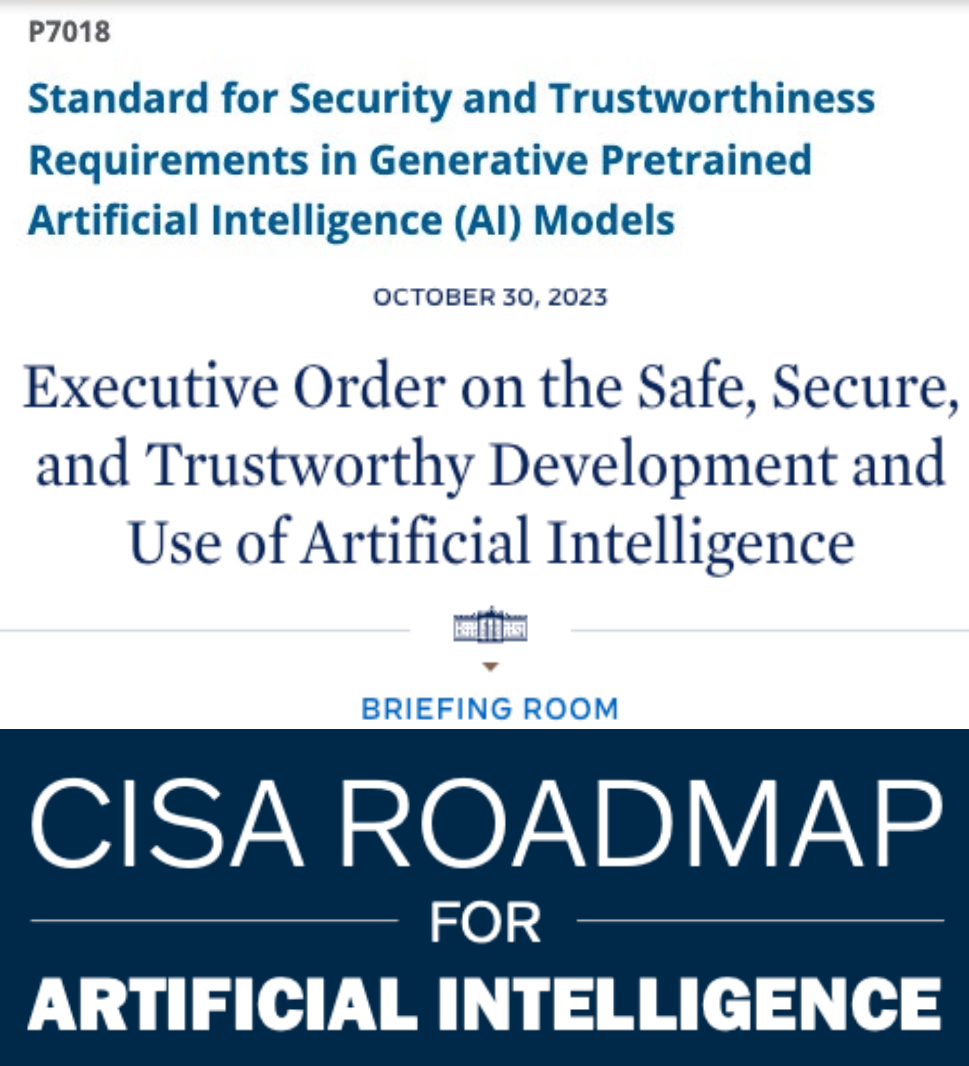
AI RISK MANAGEMENT FRAMEWORK

EU AI Regulation

Conformity Assessments in the EU AI Act: What You Need to Know

ISO/IEC 42001:2023
Information technology
Artificial intelligence
Management system

Status : **Published**



P7018

Standard for Security and Trustworthiness Requirements in Generative Pretrained Artificial Intelligence (AI) Models

OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

BRIEFING ROOM

CISA ROADMAP FOR ARTIFICIAL INTELLIGENCE



ARTIFICIAL INTELLIGENCE AND CYBERSECURITY RESEARCH

ENISA Research and Innovation Brief

JUNE 2023

Sources counter-clockwise: OWASP, NIST, HolisticAI, ISO/IEC, ENISA, CISA, The White House, IEEE,



Live the hiring rhetoric

AI the next big challenge for the digital skills gap, EU's Schmit says

By [Luca Bertuzzi](#) | [Euractiv.com](#) ⌚ Est. 4min 📅 18 Jul 2023 (updated: 📅 19 Jul 2023)

[Source: Euroactive 2023](#)

OpenAI recruiters are trying to lure Google AI employees with \$10 million pay packets, report says

[Source: Business Insider 2023](#)

How do you assess potential impacts?

Oversight and control team

(DIRECT STAKEHOLDER)

Who will troubleshoot, manage, operate, oversee or control the system during and after deployment? Who can discontinue the system?

E.g., Microsoft, consumer customer, enterprise customer, B2B, B2C

Privacy and Security

POTENTIAL HARM

Microsoft Privacy Standard

Refer to the standard for compliance.

Microsoft Security Policy

Refer to the policy for compliance.

Accountability

POTENTIAL HARM

Fit for purpose

What harms might this stakeholder experience if the system does not effectively solve the intended problem?

Transparency

POTENTIAL HARM

Significant intelligibility

What harms might this stakeholder experience if there is not enough information to make appropriate decisions about people, using the system's outputs?

Fairness

POTENTIAL HARM

Allocation

Could the system recommend the allocation of resources or opportunities to a stakeholder differently based on their demographic group(s)?

Reliability & Safety

POTENTIAL HARM

Ongoing monitoring, feedback, and evaluation

What harms might this stakeholder experience related to system changes and operation after release, especially related to identification of issues, maintenance, and improvement over time?

Source: [Microsoft Responsible AI Impact Assessment Guide](#), June 2022

Dependent on quality inputs and broad assessor perspectives

Case Study: Hospital Employee and Resource Optimization System (HEROS)

“Potential benefits and harms for the stakeholder ‘scheduled surgery patient’, ...the evaluation or decision subject...

...asking how this stakeholder could benefit directly or indirectly from using the system... summarized benefits as:

...better understanding of the length of hospital stay and better able to plan for things like childcare or house sitting”

Source: [Microsoft Responsible AI Impact Assessment Guide](#), June 2022

Accountability

POTENTIAL HARM

Fit for purpose

What harms might this stakeholder experience if the system does not effectively solve the intended problem?

E.g., If the system is unable to accurately predict the length of hospital stays for scheduled surgery patients (the intended problem), then decision makers will either make poor decisions based on the system outputs or stop using the system.

Accountability

POTENTIAL HARM

Data governance and management

What harms might this stakeholder experience if the data used to train the system have not been sufficiently managed or evaluated in relation to the system's intended use(s)?

E.g., If the system is trained using data from all types of hospital stays it may not accurately represent hospital stays specifically for scheduled surgery patients.

Reliability & Safety

POTENTIAL HARM

Ongoing monitoring, feedback, and evaluation

What harms might this stakeholder experience related to system changes and operation after release, especially related to identification of issues, maintenance, and improvement over time?

E.g., It's possible that practices within the hospital shift over time, and a model trained on the original data set could become less accurate over time. Predictions would be less reliable, potentially compromising patient (decision subjects) care.

Health Insurance AI - What is the risk and downstream Impact?



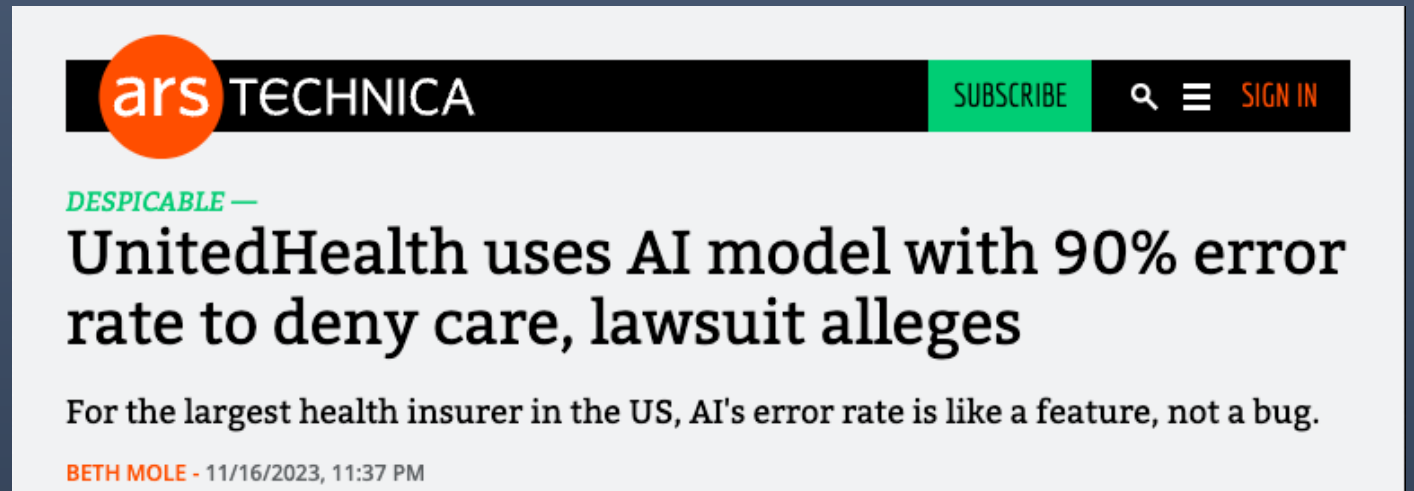
The screenshot shows the top portion of a news article from the American Medical Association (AMA). The header is purple with the AMA logo on the left, a 'Join / Renew' button, a search icon, and a user profile icon. The article title is 'Oversight needed on payers' use of AI in prior authorization' in large, bold, black text. Below the title, it says 'JUN 14, 2023 • 3 MIN READ' and 'By Tanya Albert Henry, Contributing News Writer'.

Source: [American Medical Association, June 2023](#)

“NaviHealth employees have been told to hew closer and closer to the algorithm's predictions. In 2022, case managers were told to keep patients' stays in nursing homes to within 3 percent of the days projected by the algorithm, according to documents obtained by Stat. In 2023, the target was narrowed to 1 percent”

“*ProPublica* revealed that over a period of two months in 2022, Cigna doctors denied more than 300,000 claims as part of a review process that used artificial intelligence, with Cigna doctors spending an average of 1.2 seconds on each case,”

UnitedHealthcare has said it uses technology enabling it to make “fast, efficient and streamlined coverage decisions.”



The screenshot shows the top portion of a news article from Ars Technica. The header is black with the 'ars TECHNICA' logo on the left, a green 'SUBSCRIBE' button, a search icon, and a 'SIGN IN' button. The article title is 'UnitedHealth uses AI model with 90% error rate to deny care, lawsuit alleges' in large, bold, black text. Below the title, it says 'For the largest health insurer in the US, AI's error rate is like a feature, not a bug.' and 'BETH MOLE - 11/16/2023, 11:37 PM'.

Source: [Ars Technica, November 2023](#)

Are health triage chatbots adequately governed?

Sep 18, 2023 - Technology

New AI tools are helping doctors screen for mental health conditions

Source: [Axios, September 2023](#)

“The lack of transparency and methodological flaws are concerning, as they delay AI’s safe, practical implementation.

Also, data engineering for AI models seems to be overlooked or misunderstood, and data is often not adequately managed. These significant shortcomings may indicate overly accelerated promotion of new AI models without pausing to assess their real-world viability,” Dr Novillo-Ortiz.”

Source: [World Health Organisation, February 2023](#)

“[Kintsugi](#), an American startup that has raised more than \$28 million from investors and the National Science Foundation, uses its AI-powered voice analysis tool looks for signs of clinical depression and anxiety in short clips of speech.”

“[Grace Chang](#), Kintsugi's founder and CEO, told Axios, "It's not what somebody says, it's how they're saying it that really matters." Kintsugi's system uses data from 250,000 people who made voice journals to identify "[voice biomarkers](#).”



Progress governing Software as a Medical Device (SaMD)

Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan

The U.S. Food and Drug Administration (FDA) issued the "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" from the Center for Devices and Radiological Health's Digital Health Center of Excellence.

The Action Plan is a direct response to stakeholder feedback to the April 2019 discussion paper, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device" and outlines five actions the FDA intends to take.



Source: [FDA, January 2021](#)

GUIDANCE DOCUMENT

Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions

Draft Guidance for Industry and Food and Drug Administration Staff

APRIL 2023

Source: [FDA, April 2023](#)

DECEMBER 14, 2023

Delivering on the Promise of AI to Improve Health Outcomes



► [BRIEFING ROOM](#) ► [BLOG](#)

Source: [The White House, December 2023](#)

“The commitments received today will serve to align industry action on AI around the “FAVES” principles—that AI should lead to healthcare outcomes that are Fair, Appropriate, Valid, Effective, and Safe”

Source: [The White House, December 2023](#)