ARCHIVE360

Navigating the AI Governance Landscape

George T. Tziahanas
VP of Compliance

Agenda

- Emerging AI Landscape
- Existing Statutory and Regulatory Authority
- AI Governance Frameworks and EU Act
- Proposed U.S. Legislation (State Level)
- NIST AI Risk Management Framework
- What it All Means

Emerging AI Landscape

Explosive Rate of AI Adoption

- Adoption rate is unprecedented for new technology

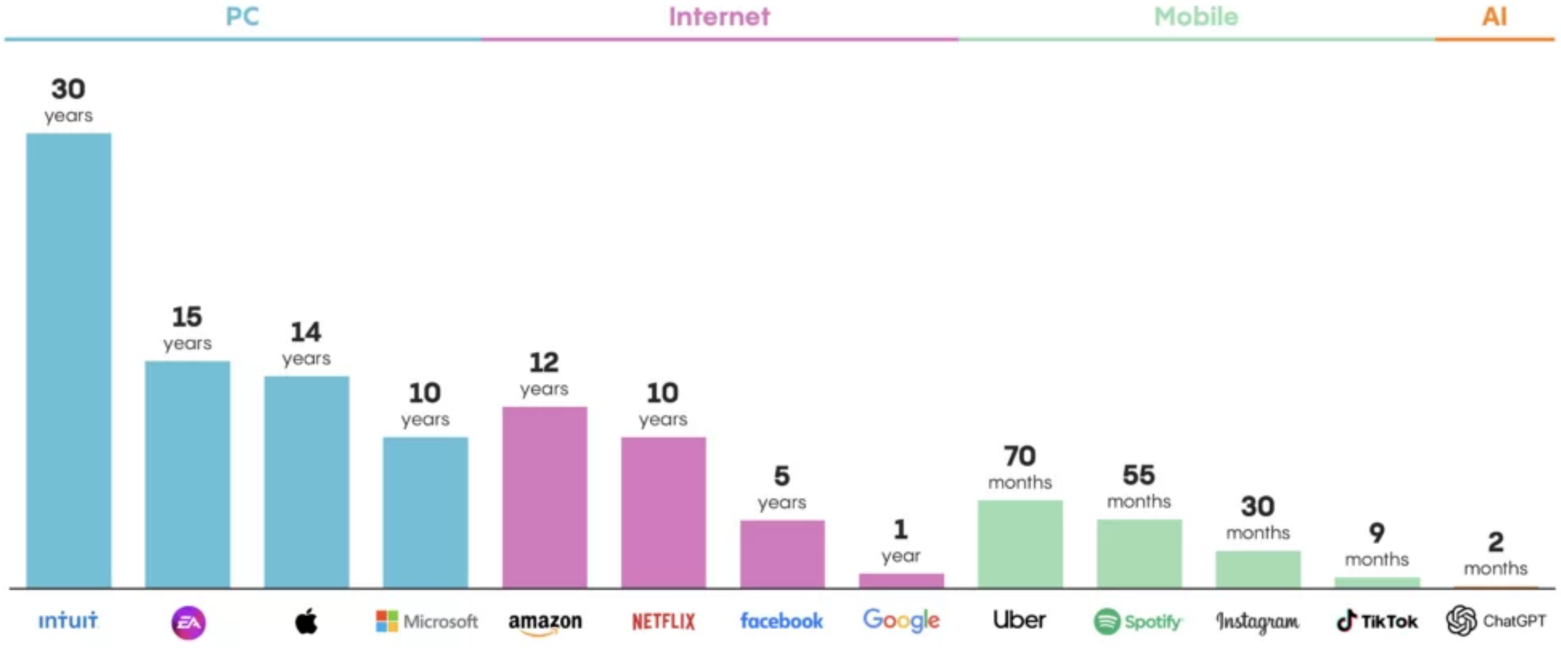
- Organizations and individuals in early stages of use

- Enterprises “invest first,” sort out ”governance” later

- **Historically, rapid adoption rates have outpaced regulatory and statutory frameworks; but they catch-up eventually**

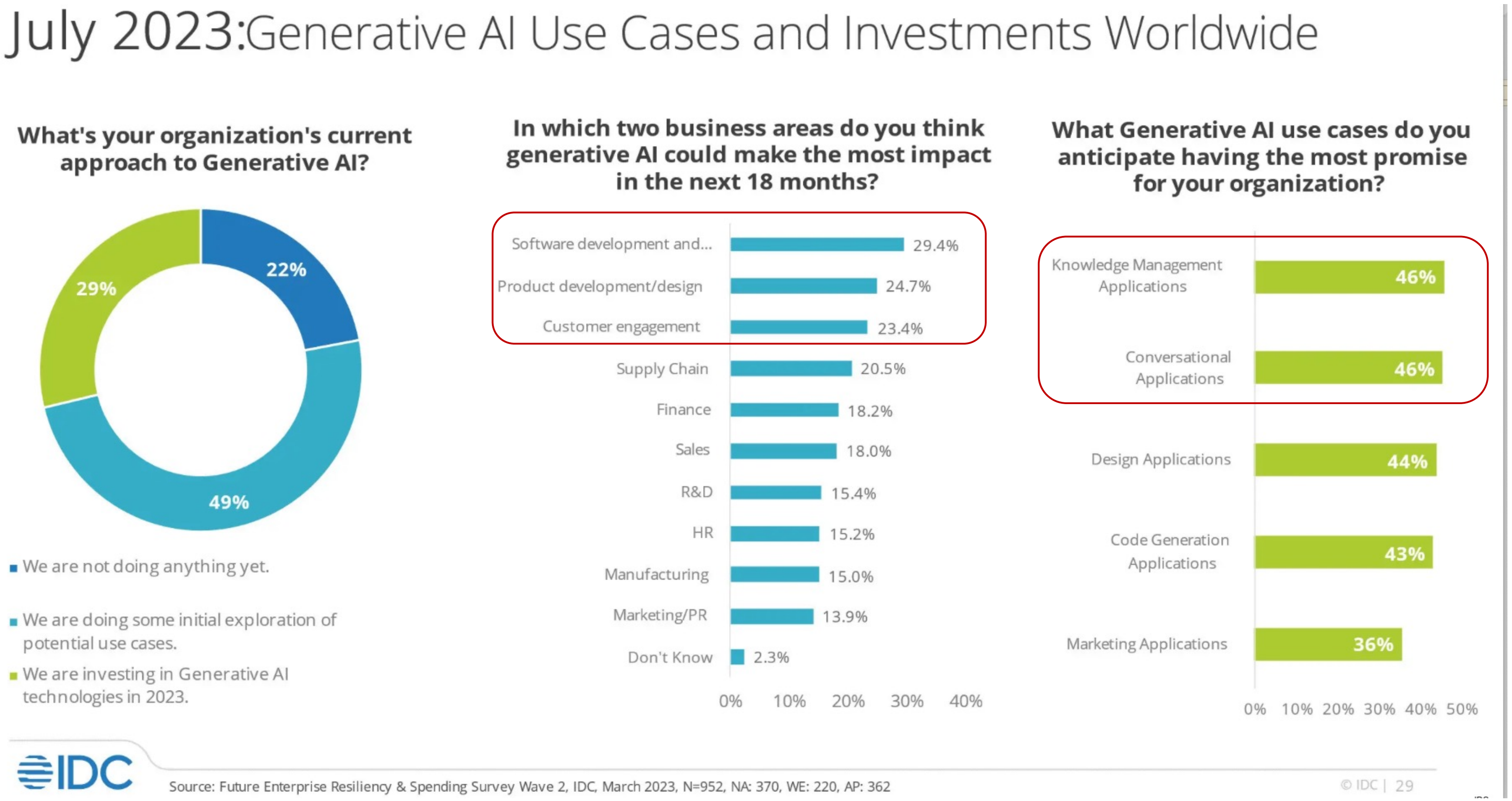
Generative AI Growth Is in a League of Its Own

ChatGPT is the fastest ever to 100M users—things are happening on a very compressed timeline



© 2023 Menlo Ventures

Gen AI Top Use Cases Areas Requiring Retention and Governance



July 2023:Generative AI Use Cases and Investments Worldwide

What's your organization's current approach to Generative AI?

- We are not doing anything yet.
- We are doing some initial exploration of potential use cases.
- We are investing in Generative AI technologies in 2023.

In which two business areas do you think generative AI could make the most impact in the next 18 months?

What Generative AI use cases do you anticipate having the most promise for your organization?

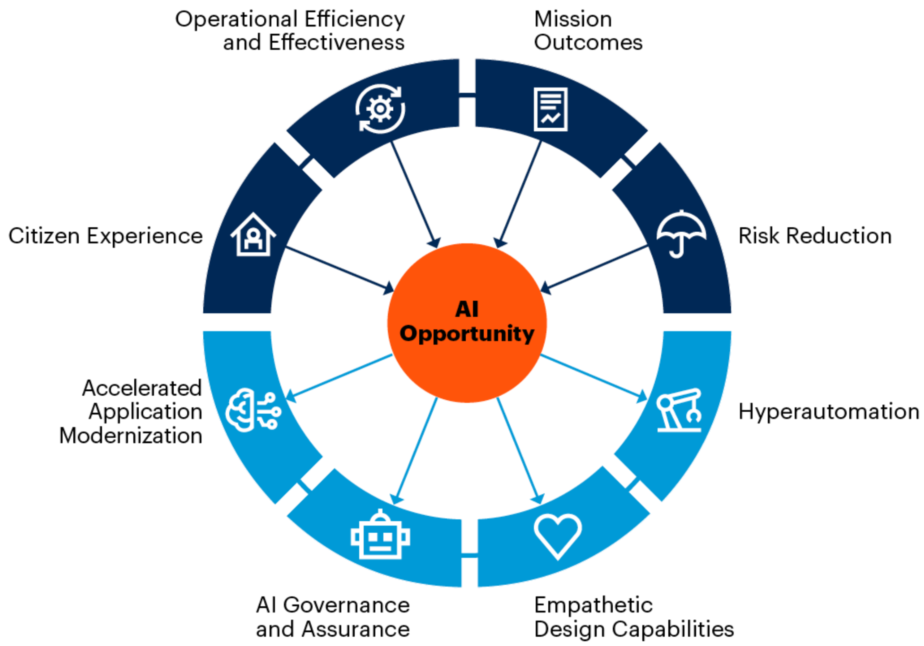
Knowledge Management Applications 46%
Conversational Applications 46%
Design Applications 44%
Code Generation Applications 43%
Marketing Applications 36%
0%
10%
20%
30%
40%
50%

IDC

Source: Future Enterprise Resiliency & Spending Survey Wave 2, IDC, March 2023, N=952, NA: 370, WE: 220, AP: 362

Government AI Opportunities Aligns with Private Sector

Government Outcome Driving AI Opportunities and Indirect Impacts



Operational Efficiency and Effectiveness

Mission Outcomes

Citizen Experience

Risk Reduction

AI Opportunity

Accelerated Application Modernization

Hyperautomation

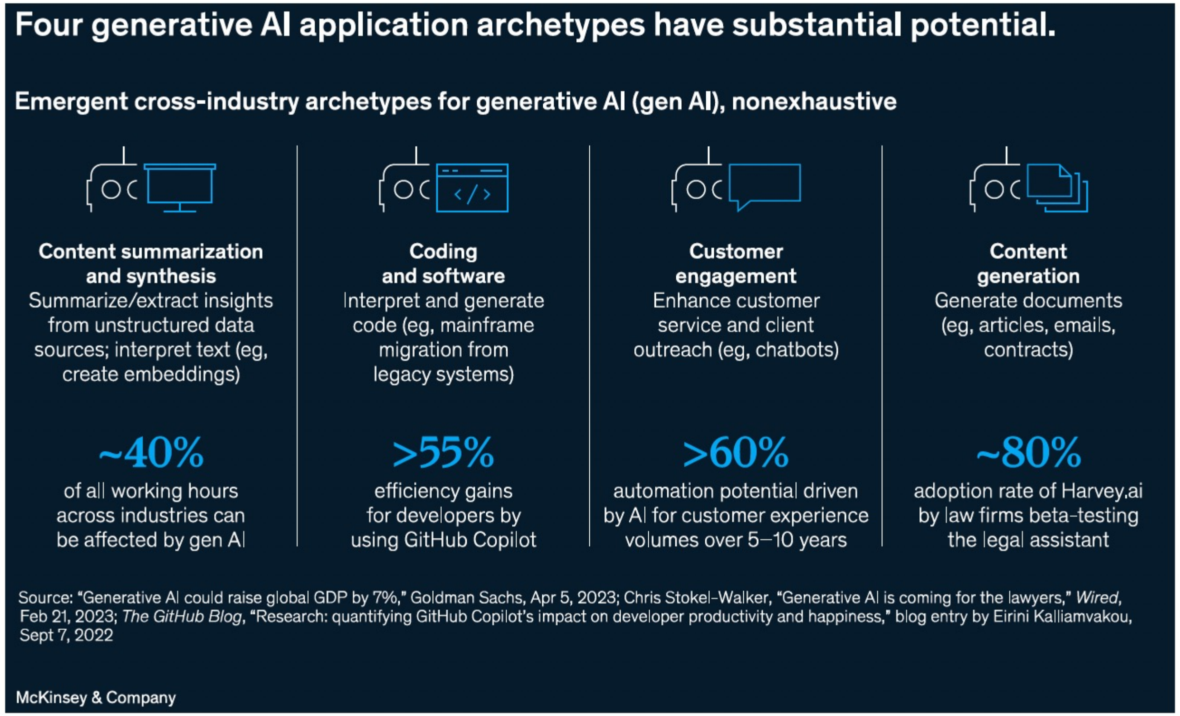
AI Governance and Assurance

Empathetic Design Capabilities

Source: Gartner

805005_C

Exhibit 2



Four generative AI application archetypes have substantial potential.

Emergent cross-industry archetypes for generative AI (gen AI), nonexhaustive

Content summarization and synthesis
Summarize/extract insights from unstructured data sources; interpret text (eg, create embeddings)

~40%
of all working hours across industries can be affected by gen AI

Coding and software
Interpret and generate code (eg, mainframe migration from legacy systems)

>55%
efficiency gains for developers by using GitHub Copilot

Customer engagement
Enhance customer service and client outreach (eg, chatbots)

>60%
automation potential driven by AI for customer experience volumes over 5–10 years

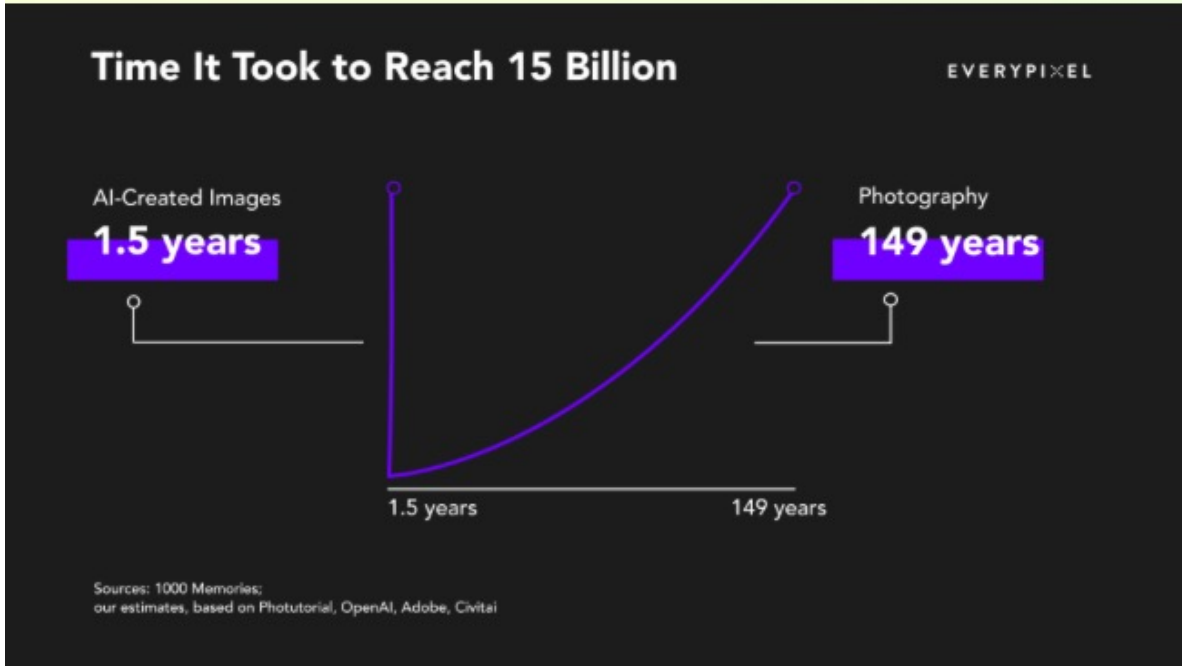
Content generation
Generate documents (eg, articles, emails, contracts)

~80%
adoption rate of Harvey.ai by law firms beta-testing the legal assistant

Source: "Generative AI could raise global GDP by 7%," Goldman Sachs, Apr 5, 2023; Chris Stokel-Walker, "Generative AI is coming for the lawyers," *Wired*, Feb 21, 2023; *The GitHub Blog*, "Research: quantifying GitHub Copilot's impact on developer productivity and happiness," blog entry by Eirini Kalliamvakou, Sept 7, 2022

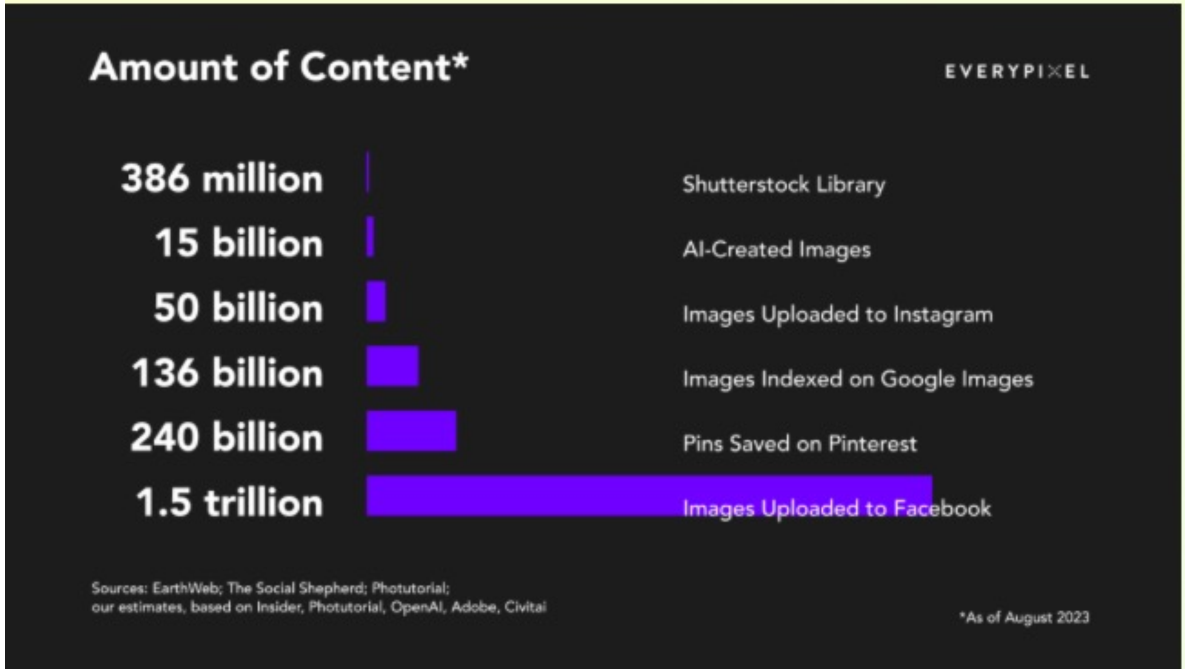
McKinsey & Company

AI Becoming Multi-Modal



Time It Took to Reach 15 Billion

Sources: 1000 Memories;
our estimates, based on Photutorial, OpenAI, Adobe, Civitai



Amount of Content*

Sources: EarthWeb; The Social Shepherd; Photutorial;
our estimates, based on Insider, Photutorial, OpenAI, Adobe, Civitai

*As of August 2023

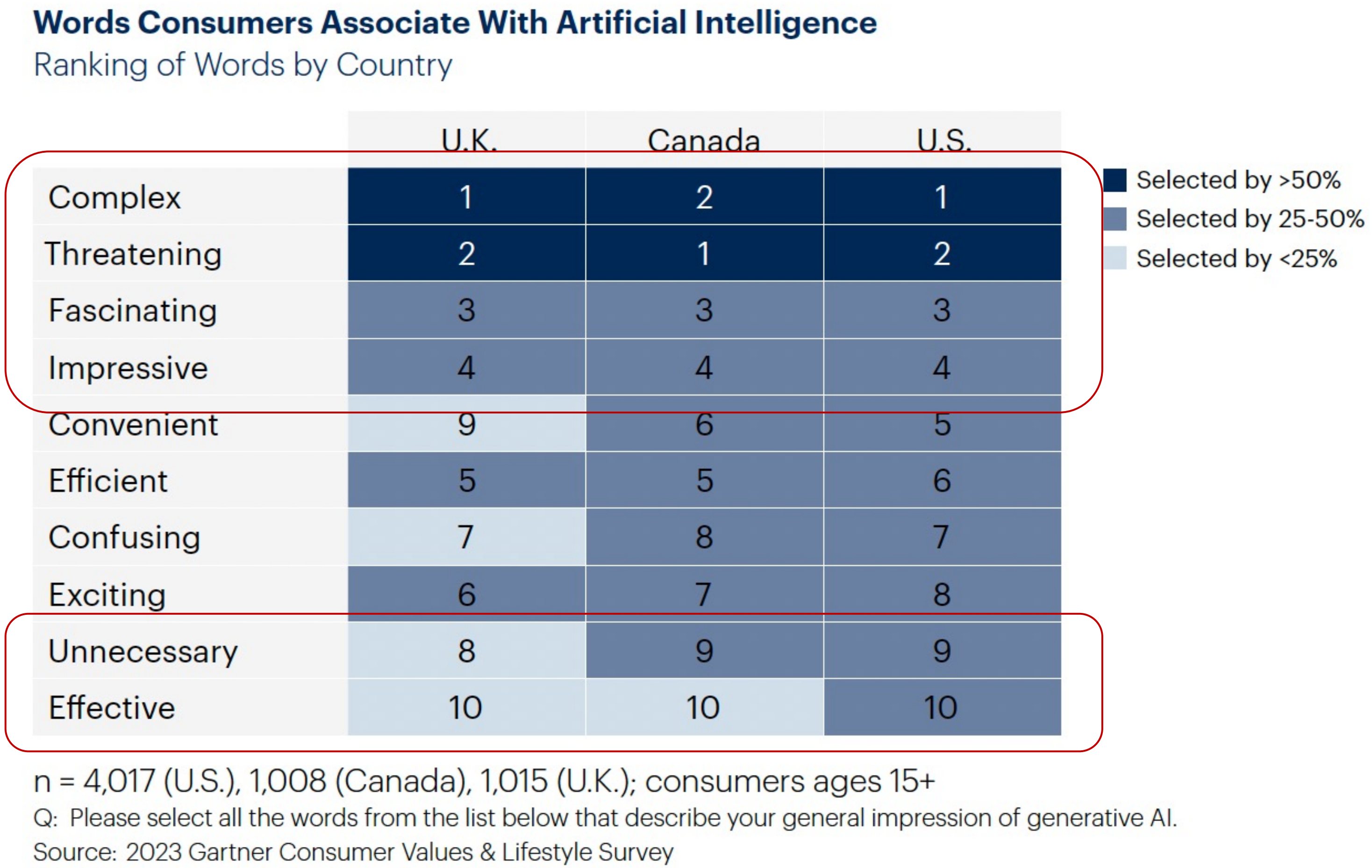
By 2026, single-modality AI models will lose out to multimodal AI models (text, image, audio and video) in over 60% of GenAI solutions, up from less than 1% in 2023.

Source: Gartner

https://journal.everypixel.com/ai-image-statistics

View of AI from Mortal Humans

- Top 2 ranking terms are negative
- Next 2 ranking terms are positive
- Last 2 ranking terms diametrically opposed
- Indication that sentiment is dynamic and uncertain
- **Concern over AI will drive legislatures and regulators**



Words Consumers Associate With Artificial Intelligence

Ranking of Words by Country

	U.K.	Canada	U.S.
Complex	1	2	1
Threatening	2	1	2
Fascinating	3	3	3
Impressive	4	4	4
Convenient	9	6	5
Efficient	5	5	6
Confusing	7	8	7
Exciting	6	7	8
Unnecessary	8	9	9
Effective	10	10	10

■ Selected by >50%
▨ Selected by 25-50%
□ Selected by <25%

n = 4,017 (U.S.), 1,008 (Canada), 1,015 (U.K.); consumers ages 15+

Q: Please select all the words from the list below that describe your general impression of generative AI.

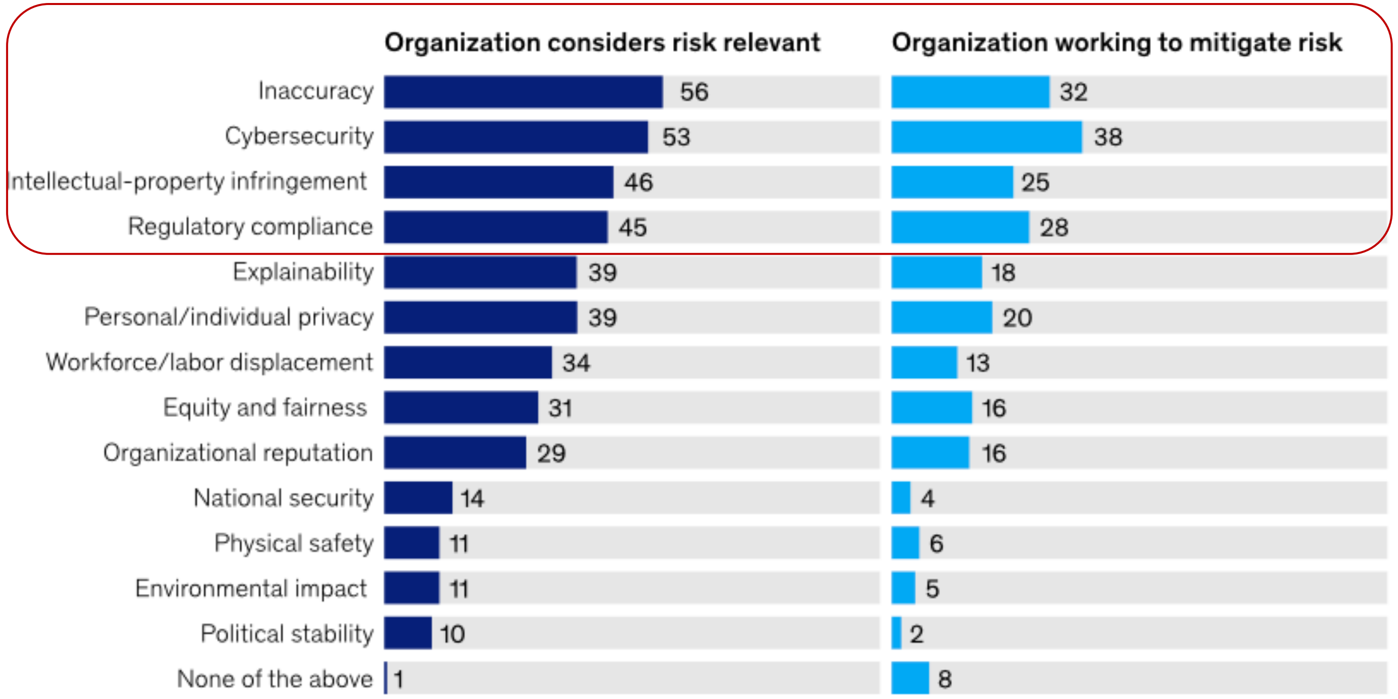
Source: 2023 Gartner Consumer Values & Lifestyle Survey

Balancing AI Use against Risk

Outlined risks do not appear to be slowing AI hype, but can already see influence with regulators and increasingly enterprises

Inaccuracy, cybersecurity, and intellectual-property infringement are the most-cited risks of generative AI adoption.

Generative AI–related risks that organizations consider relevant and are working to mitigate, % of respondents[1]

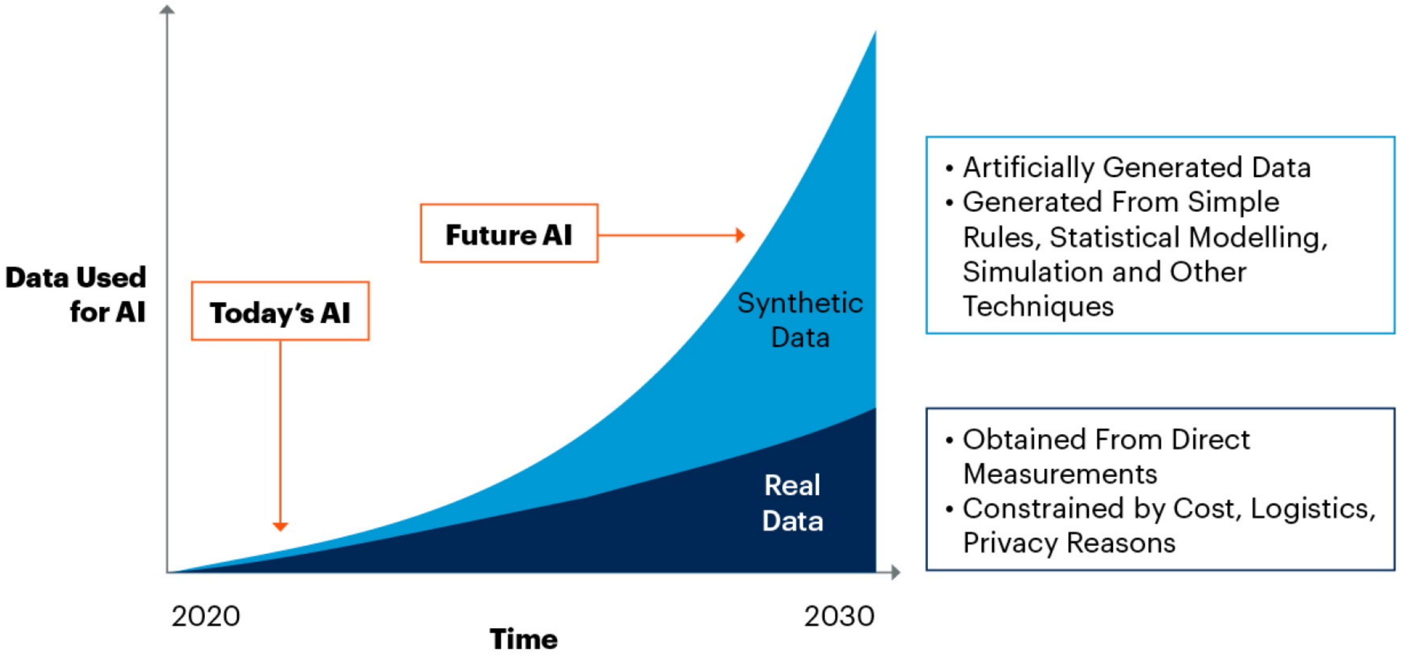


[1] Asked only of respondents whose organizations have adopted AI in at least 1 function. For both risks considered relevant and risks mitigated, n = 913.
Source: McKinsey Global Survey on AI, 1,684 participants at all levels of the organization, April 11–21, 2023

Synthetics feeding Synthetics

Gartner predicts that in the next three and a half years, generative AI will account for 10% of all data produce compared to less than 1% at present (end of 2022)

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



- Artificially Generated Data
- Generated From Simple Rules, Statistical Modelling, Simulation and Other Techniques

- Obtained From Direct Measurements
- Constrained by Cost, Logistics, Privacy Reasons

Source: Gartner

750175_C

Data used to train will increasingly be created by the robots, to train other robots

Represents explosive growth in new sets of data, much subject to governance requirements

Humans Say The Robots are Already Regulated

AI Joint Statement: Enforcement Efforts Against Discrimination and Bias in Automated Systems

> “Although many of these tools offer the promise of advance, their use also has the potential to perpetuate unlawful bias, unlawful discrimination, and produce other ***harmful outcomes***”
>
> –CFPB, DOJ, FTC, EEOC

https://files.consumerfinance.gov/f/documents/cfpb_joint-statement-enforcement-against-discrimination-bias-automated-systems_2023-04.pdf

AI Joint Statement

Letter specifically identifies sources of potential problems, which include:

- Data and Datasets
- Model Opacity and Access
- System Design and Use

AKA Potential Records

AI, analytics, and automation solutions are a construct of all three; and issues with any of these can have *a harmful outcome*

Harmful Outcomes: Wrong at the Speed of AI

Plaintiff (Mata) filed suit against airline Avianca, for alleged injuries from metal serving cart on an international flight

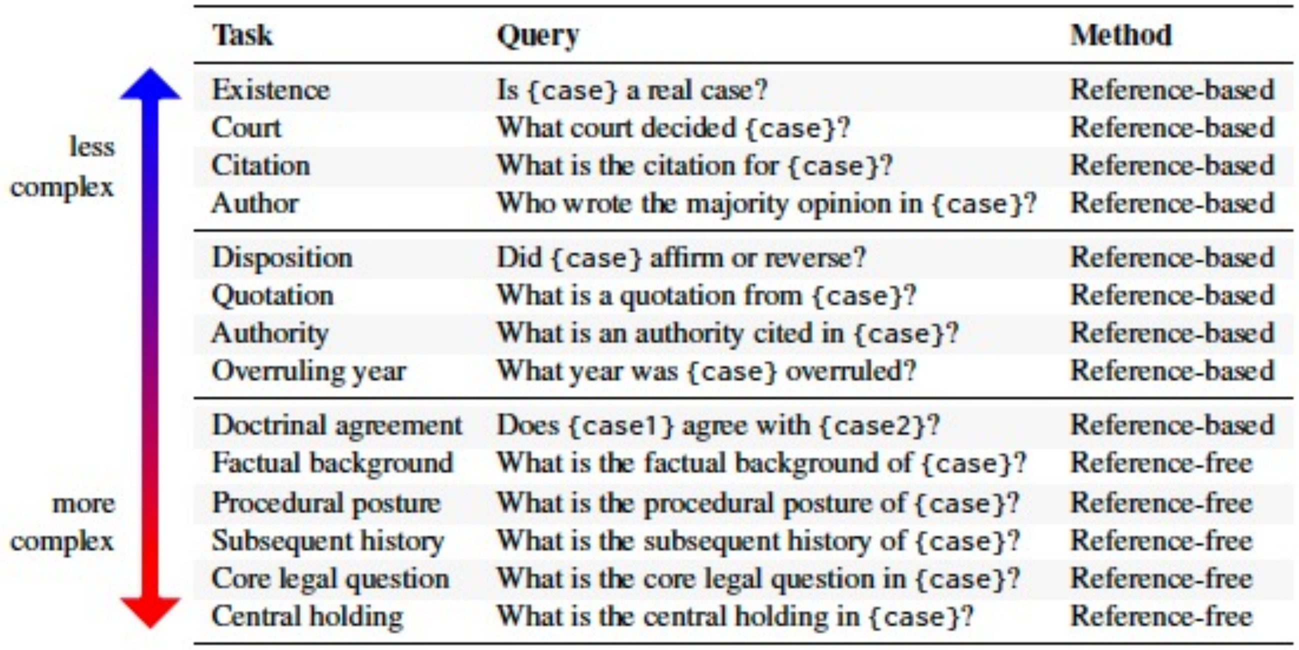
Plaintiff's attorney submitted brief related to "tolling effect of bankruptcy under the Montreal Convention," and asked ChatGPT to draft the filing

- Document included references to a number of cases supporting plaintiff's position; yet defense counsel was unable to locate many of the cases cited by ChatGPT
- Plaintiff's counsel even asked ChatGPT if the citations were "real," and received assurance they were

- Unfortunately, the cases were completely fabricated by ChatGPT. Lawyers were ultimately sanctioned; it was their duty to understand (and supervise) AI
- The lesson is useful to compliance, legal, and records professionals more broadly; cannot "outsource" obligations to AI without knowledge and oversight

https://www.acc.com/resource-library/practical-lessons-attorney-ai-missteps-mata-v-avianca

Not Hallucinating Negative Outcomes


less
complex
more
complex

Task	Query	Method
Existence	Is {case} a real case?	Reference-based
Court	What court decided {case}?	Reference-based
Citation	What is the citation for {case}?	Reference-based
Author	Who wrote the majority opinion in {case}?	Reference-based
Disposition	Did {case} affirm or reverse?	Reference-based
Quotation	What is a quotation from {case}?	Reference-based
Authority	What is an authority cited in {case}?	Reference-based
Overruling year	What year was {case} overruled?	Reference-based
Doctrinal agreement	Does {case1} agree with {case2}?	Reference-based
Factual background	What is the factual background of {case}?	Reference-free
Procedural posture	What is the procedural posture of {case}?	Reference-free
Subsequent history	What is the subsequent history of {case}?	Reference-free
Core legal question	What is the core legal question in {case}?	Reference-free
Central holding	What is the central holding in {case}?	Reference-free

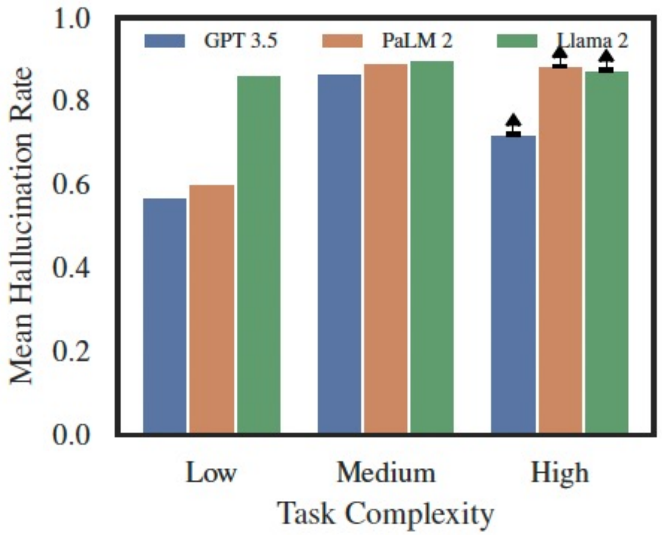


Figure 3: Relationship between task complexity and mean hallucination rate. Higher values indicate a greater likelihood of factually incorrect LLM responses. High complexity tasks include several reference-free tasks, so those reported hallucination rates are lower bounds on the true rates. Contra-factual tasks are excluded from this comparison.

AI can fabricate information while making it seem authentic; and **can do so on a frequent basis**

Concern about accuracy, bias, and harmful outcomes central to regulatory and legislative activity

https://arxiv.org/abs/2401.01301

Governance of AI: Early Actions in EU

First Actions on AI-Privacy Focused

- Unsurprisingly, EU taking aggressive stance over AI and potential privacy issues
- EU Data Protection Board launched dedicated task force to coordinate potential enforcement actions against ChatGPT
- Italy briefly banned ChatGPT until they made changes to address privacy and youth interaction issues

Introduced proposed AI Legislation, which was recently adopted

https://www.complianceweek.com/regulatory-enforcement/edpb-task-force-latest-scrutinizing-chatgpt-ai-accountability/32954.article#toggle
https://www.complianceweek.com/data-privacy/chatgpt-back-in-italy-after-user-privacy-updates/33019.article

Robots Cannot Vote (Yet): AI Legislation Begins

AI and Government

Governments Initially Focused on Responsible AI Frameworks (not legislation)

- Relative (in)maturity of AI market and rate of growth make any prescriptive legal or compliance language difficult
- Australia, UK, and US (via Executive Order) proposals include:
  - https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework
  - https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
  - https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Responsible AI Frameworks: High-Level Objectives

Australia	United States
• Achieve safe, more reliable and fairer outcomes for all Australians • Reduce the risk of negative impact on those affected by AI applications • Help businesses and governments to practice the highest ethical standards when designing, developing, and implementing AI	• AI must be safe and secure • Requires addressing AI systems pressing security risks with respect to biotechnology, cybersecurity, critical infrastructure, and other national security dangers- • While navigating AI's opacity and complexity • Will not tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice

https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework

https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

EU First to Legislate the Robots

Establishes class of AI and use cases that are prohibited (e.g. social scoring, real-time biometric surveillance, health and safety systems)

Creates another class of AI considered “High-Risk” when used with respect to:

- Critical infrastructure
- Employment/worker decisions
- Essential private services (healthcare and financial services)
- Law enforcement and immigration

Act applies to developers/deployers located in the EU, and third-party countries where the AI system’s output is used in the EU

https://artificialintelligenceact.eu/annex/3/

EU AI Act-Highlights for Organizations

Scope/Requirement	Description
Impact Assessments	• Organizations **must conduct an impact assessment** for systems exempt from Annex III (defining a high-risk system) • Impact assessments **subject to retention obligations and disclosure** to authorities
Record Keeping	• Design their high-risk AI system for **record-keeping** to enable it **to automatically record events relevant** for identifying national level risks and substantial modifications throughout the system's lifecycl
Data Governance	• Conduct **data governance**, ensuring that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose
Accuracy, Robustness and Security	• High-risk **AI systems shall be designed and developed** in such a way that they achieve an appropriat**e level of accuracy, robustness, and cybersecurity**, and perform consistently in those respects throughout their lifecycle

Robots Watching Legislation in Process by U.S. States

Example State Level Legislation: GDPR 2.0

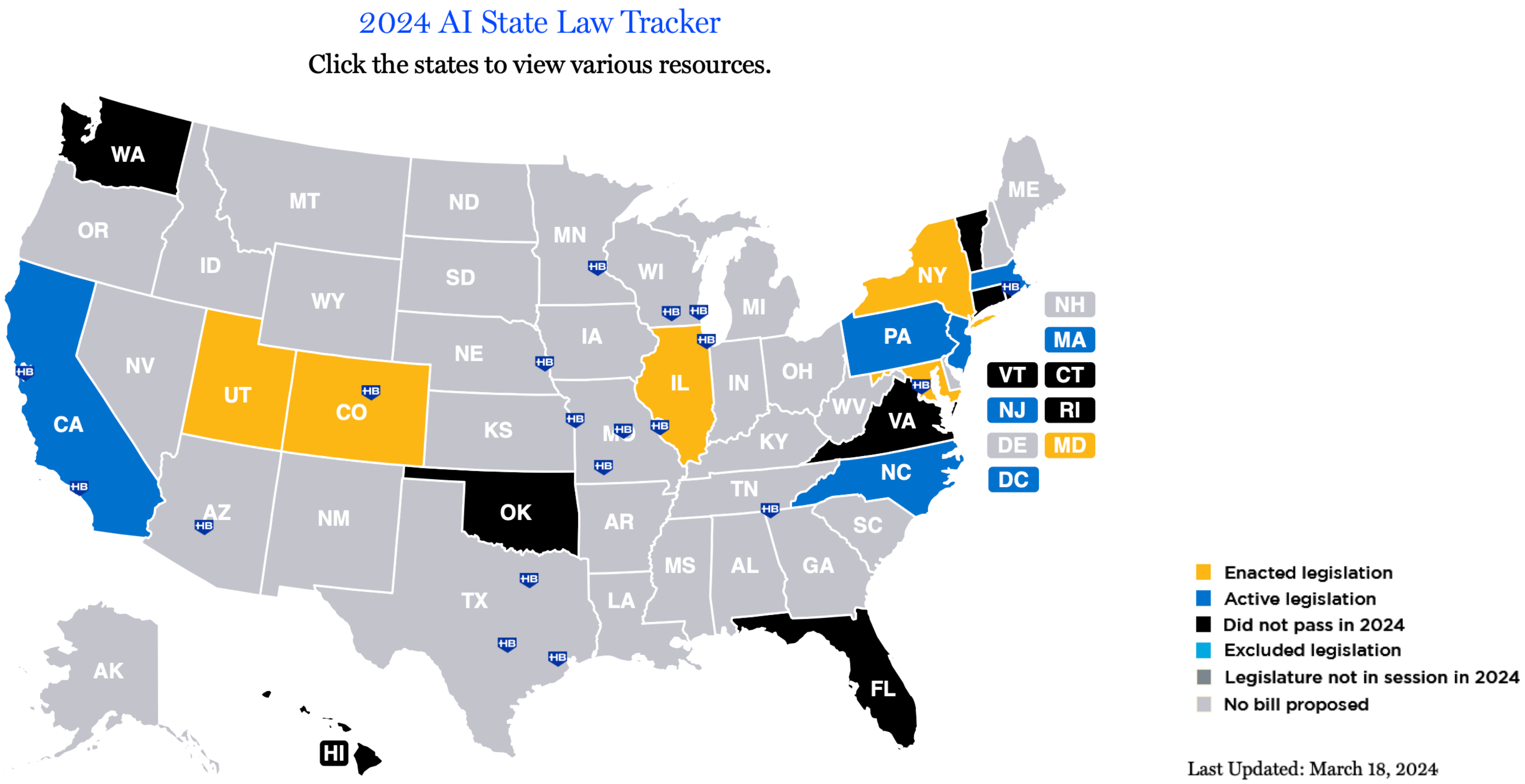
Automated Employment Decision Tool Legislation

- Illinois
- California
- New York
- Et al.

Algorithmic Bias Legislation:

- Florida
- California
- New York
- Massachusetts
- Illinois
- Virginia
- New Jersey

https://www.huschblackwell.com/2024-ai-state-law-tracker



2024 AI State Law Tracker

Click the states to view various resources.

🟨 Enacted legislation
🟦 Active legislation
⬛ Did not pass in 2024
🟦 Excluded legislation
◼ Legislature not in session in 2024
◻ No bill proposed

Last Updated: March 18, 2024

Proposed Legislative Frameworks: Commonality with EU AI Act

Example: New York Algorithmic Bias Legislation

Scope: Among these rights and protections are (i) the right to safe and effective systems; (ii) protections against algorithmic discrimination; protections against abusive data practices; (iv) the right to have agency over one's data; (v) the right to know when an automated system is being used…

- "Equal opportunity" means equal access to education, housing, credit, employment, and other programs
- "Access to critical resources or services" including but not limited to:
  - Healthcare
  - Financial Services
  - Safety
  - Social Services
  - Government benefits

New York Proposed AI Act Continued

Additional Requirements

- Automated systems shall undergo **pre-deployment and ongoing disparity testing and mitigation**, under clear organizational oversight.

- **Independent evaluations** and plain language reporting in the form of an **algorithmic impact assessment**, including disparity testing results and mitigation information, **shall be conducted** for all automated systems
- California and some states include an affirmative disclosure requirement of algorithmic impact assessments to designated state agencies; other states require upon request

California, Illinois, and other U.S. states proposed legislation incorporates similar requirements; and often much of the same language

Key Takeaways on Regulatory and Statutory Landscape

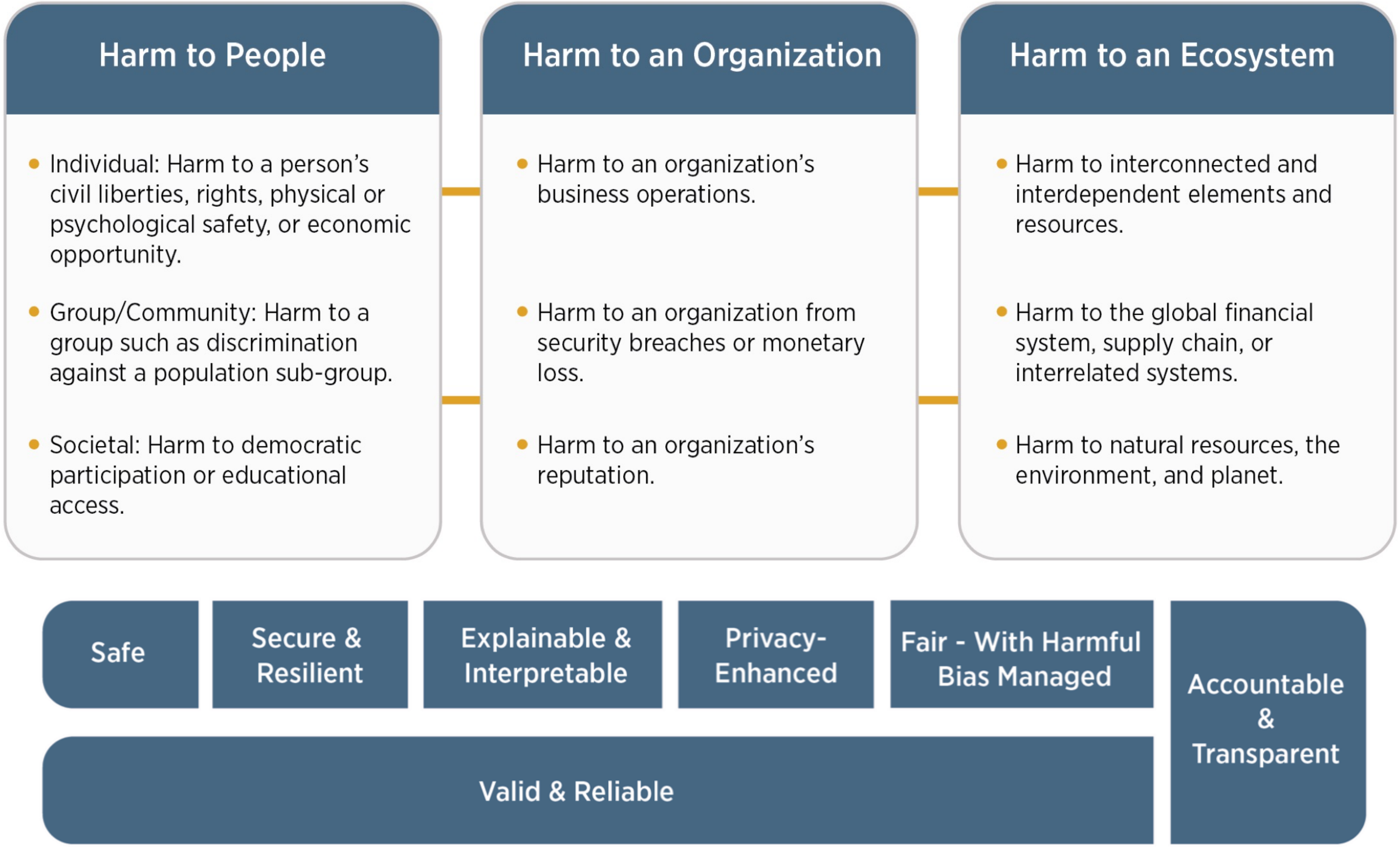
U.S Regulators are taking the position they have sufficient authority under existing laws and regulations to broadly govern AI

The EU AI Act is expansive and will impact European and multi-national organizations

U.S. State statutes share characteristics of EU AI Act, and likely to make it into law before any federal U.S. law. See also GDPR/CPRA 2.0

AI Risk Management Framework for Government Agencies and Private Sector

NIST AI Risk Management Model



Harm to People

- Individual: Harm to a person's civil liberties, rights, physical or psychological safety, or economic opportunity.
- Group/Community: Harm to a group such as discrimination against a population sub-group.
- Societal: Harm to democratic participation or educational access.

Harm to an Organization

- Harm to an organization's business operations.
- Harm to an organization from security breaches or monetary loss.
- Harm to an organization's reputation.

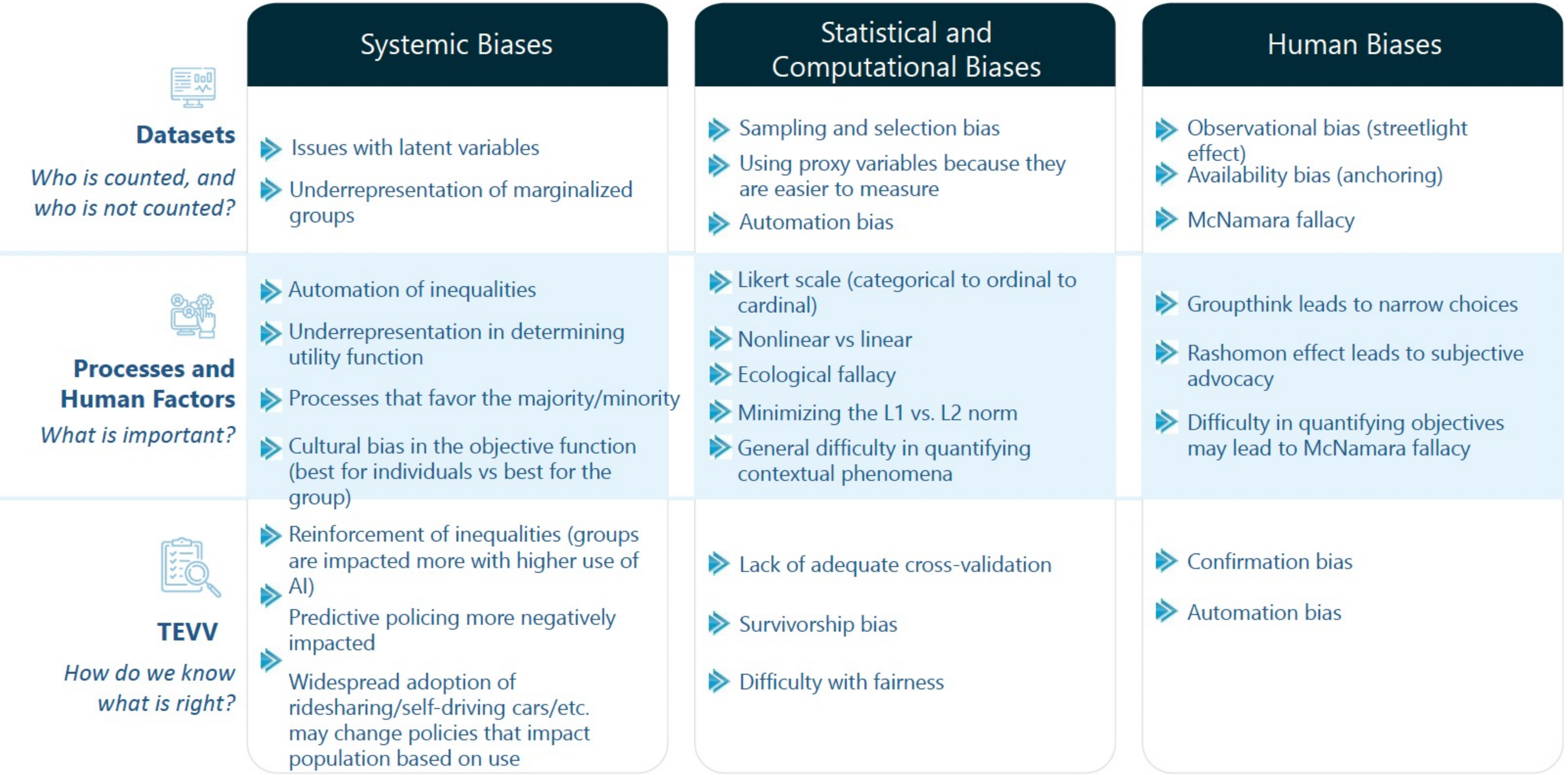
Harm to an Ecosystem

- Harm to interconnected and interdependent elements and resources.
- Harm to the global financial system, supply chain, or interrelated systems.
- Harm to natural resources, the environment, and planet.

Safe
Secure & Resilient
Explainable & Interpretable
Privacy-Enhanced
Fair - With Harmful Bias Managed
Accountable & Transparent
Valid & Reliable

https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Managing AI Bias



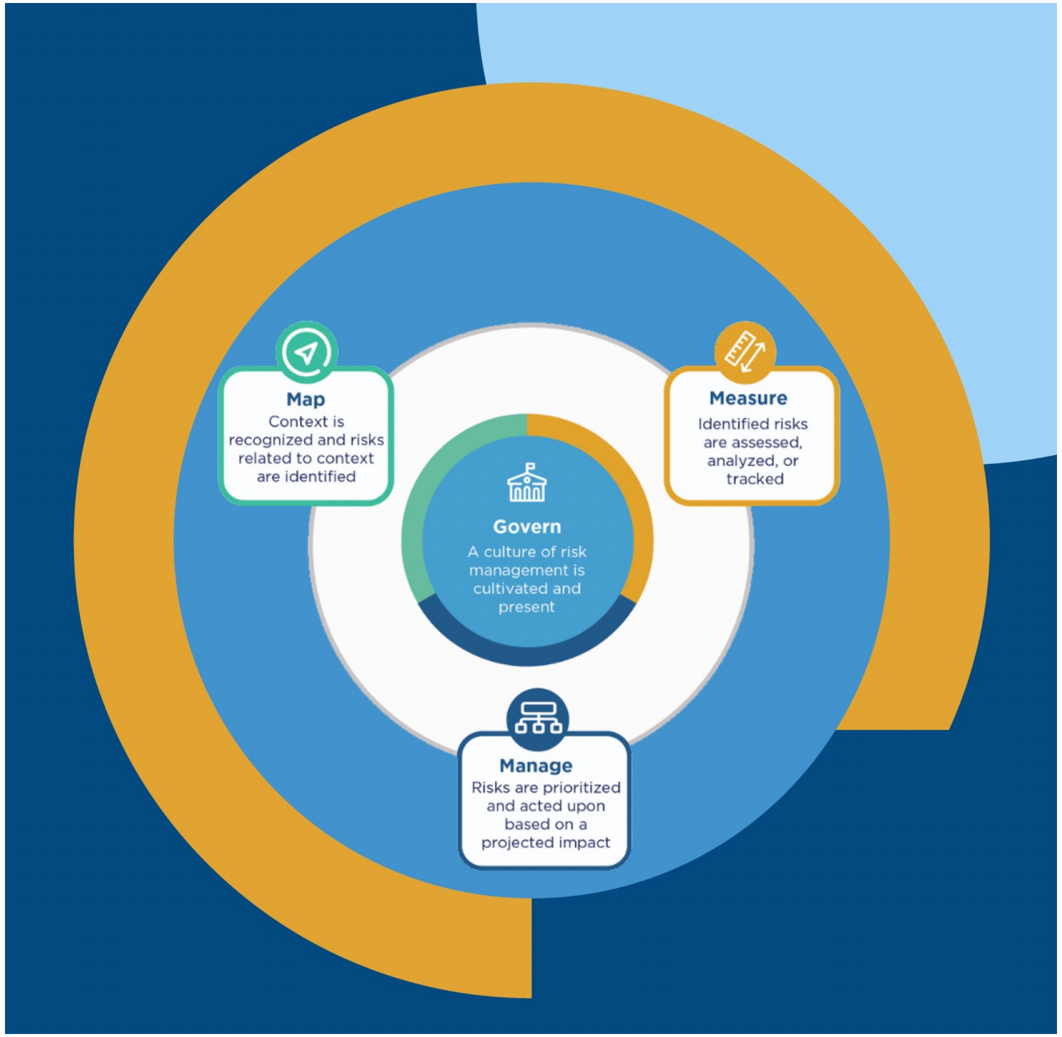
	Systemic Biases	Statistical and Computational Biases	Human Biases
Datasets *Who is counted, and who is not counted?*	Issues with latent variables Underrepresentation of marginalized groups	Sampling and selection bias Using proxy variables because they are easier to measure Automation bias	Observational bias (streetlight effect) Availability bias (anchoring) McNamara fallacy
Processes and Human Factors *What is important?*	Automation of inequalities Underrepresentation in determining utility function Processes that favor the majority/minority Cultural bias in the objective function (best for individuals vs best for the group)	Likert scale (categorical to ordinal to cardinal) Nonlinear vs linear Ecological fallacy Minimizing the L1 vs. L2 norm General difficulty in quantifying contextual phenomena	Groupthink leads to narrow choices Rashomon effect leads to subjective advocacy Difficulty in quantifying objectives may lead to McNamara fallacy
TEVV *How do we know what is right?*	Reinforcement of inequalities (groups are impacted more with higher use of AI) Predictive policing more negatively impacted Widespread adoption of ridesharing/self-driving cars/etc. may change policies that impact population based on use	Lack of adequate cross-validation Survivorship bias Difficulty with fairness	Confirmation bias Automation bias

Fig. 5. How biases contribute to harms

https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

NIST AI Risk Management Framework

Category	Description
Govern	Policies, processes, procedures and practices across the organization related to the mapping, measuring and managing of AI risks are in place, transparent, and implemented effectively.
Manage	AI risks based on assessments and other analytical output from the Map and Measure functions are prioritized, responded to, and managed.
Map	Context is established and understood. Intended purpose, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented
Measure	Appropriate methods and metrics are identified and applied Approaches and metrics for measurement of AI risks enumerated during the Map function are selected for implementation starting with the most significant AI risks.


Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

ARCHIVE360

https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

Key Takeaways on AI Governance

Existing and proposed regulations and statutes will **require governance/retention of enormous sets** of new information

- **Datasets used to train models**, or at least sufficient information that describes all sources
- Data/information **created via generative AI subject to existing retention** requirements (including chat-based interactions)
- System **designs using AI/automation** for covered applications
- **AI and ML models themselves**, and logs associated with their use
- **Testing** done to verify the accuracy, safety, and potential disparity
- **Impact and security assessments** (internal or third-party)

Backup

SEC Provides Glimpse into Scope of Data Subject to Retention and Disclosure for AI Systems

Reportedly initiated a “Street Sweep” in 2023 for Registered Investment Advisors and Use of AI

- **Used existing authority** to request books and records of RIA’s
- **Exceptionally broad requests**, which generally aligned with the categories laid-out in the Joint Statement
- However, **provided a much more granular** view into the training, design, use, and supervision of AI

https://www.natlawreview.com/article/us-securities-exchange-commission-targets-ai-multiple-fronts-ai-sweep-examination

SEC Record Request for AI from RIA’s

Requested Information from Securities and Exchange Commission: Investment Advisor “Street Sweep”	
A description of the **AI models and techniques** used by the advisers	**Contingency plans** in case of **AI system failures** or inaccuracies
A list of **algorithmic trading signals** and associated models	Client **profile documents used by the AI** system to understand a client's risk tolerance and investment objectives
The **sources and providers of their data**	AI-related security measures
Internal reports of **any incidents where AI** use raised regulatory, ethical, or legal issues	A list and description **of all data acquisition errors** and/or adjustments to algorithmic modifications due to data acquisition errors
Copies of **any AI compliance written supervisory policies** and procedures	**Samples of any reports detailing the validation process** and performance of robo-advisory algorithms
A list of those who **develop, implement, operate, manage, or supervise AI** software systems	A list of all board, management, or staff committees with specific AI-related responsibilities, the frequency of any meetings, a list of the members of each committee, and whether minutes are kept
All **disclosure and marketing documents** to clients where the **use of AI by the adviser** is stated or referred to specifically in the disclosure, including **audio and video marketing in which the adviser's use of AI** is mentioned	A list of all media used to advertise, market or promote products and services, including social media, chat forums, websites, due diligence questionnaire responses, PPMs, pitch books, presentations, newsletters, annual reports, and podcasts and/or other video or audio marketing, and two recent examples of each kind of ad

EU AI Act: Record Keeping Detail

Record Keeping Requirements for High-Risk Systems

High-risk AI systems shall technically allow for the automatic recording of events (‘logs’) over the duration of the lifetime of the system.
2. In order to ensure a level of traceability of the AI system’s functioning that is appropriate to the intended purpose of the system, logging capabilities shall enable the recording of events relevant for:
2a. (i) identification of situations that may result in the AI system presenting a risk within the meaning of **Article 65**(1) or in a substantial modification;
(ii) facilitation of the post-market monitoring referred to in **Article 61**;
and (iii) monitoring of the operation of high-risk AI systems referred to in **Article 29**(4).
3. [deleted].
4. For high-risk AI systems referred to in paragraph 1, point (a) of **Annex III**, the logging capabilities shall provide, at a minimum:
(a) recording of the period of each use of the system (start date and time and end date and time of each use);
(b) the reference database against which input data has been checked by the system;
(c) the input data for which the search has led to a match;
(d) the identification of the natural persons involved in the verification of the results, as referred to in **Article 14** (5).

EU AI Act: Data Governance Detail

High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in paragraphs 2 to 5 whenever such datasets are used.

2. Training, validation and testing data sets shall be subject to appropriate data governance and management practices appropriate for the intended purpose of the AI system. Those practices shall concern in particular:
(a) the relevant design choices;
(aa) data collection processes and origin of data, and in the case of personal data, the original purpose of data collection;
(b) [deleted];
(c) relevant data preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation;
(d) the formulation of assumptions, notably with respect to the information that the data are supposed to measure and represent;(e) an assessment oof the availability, quantity and suitability of the data sets that are needed;

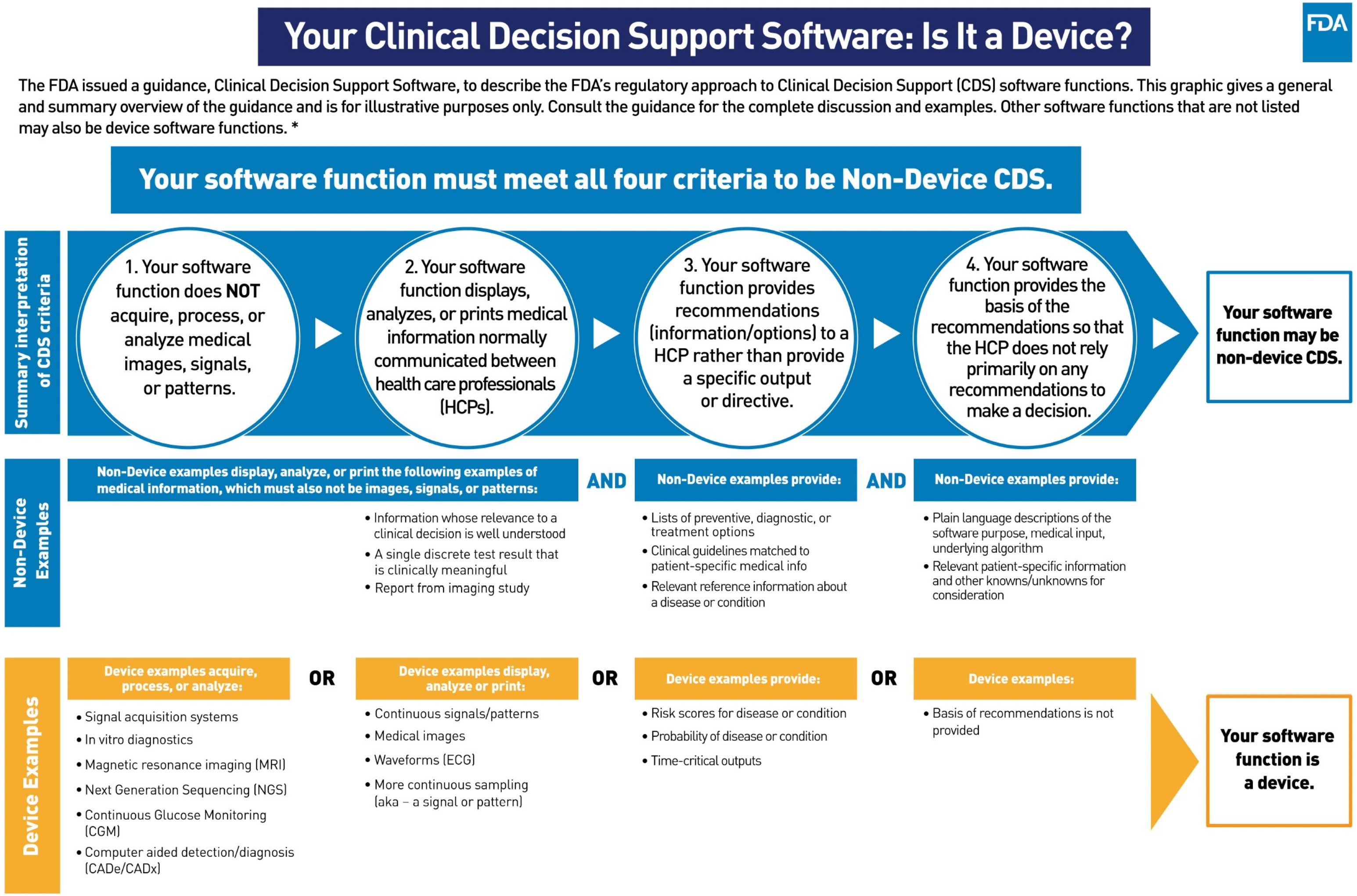
AI Governance: Deep Dive Into Your Business

General AI governance objectives and programs will only go so far

Understanding the business and regulatory framework is critical

- Healthcare
- Software
- Construction
- Energy
- Mining

Will need to map use of AI to compliance requirements



Your Clinical Decision Support Software: Is It a Device?

The FDA issued a guidance, Clinical Decision Support Software, to describe the FDA's regulatory approach to Clinical Decision Support (CDS) software functions. This graphic gives a general and summary overview of the guidance and is for illustrative purposes only. Consult the guidance for the complete discussion and examples. Other software functions that are not listed may also be device software functions. *

Your software function must meet all four criteria to be Non-Device CDS.

Summary interpretation of CDS criteria

1. Your software function does **NOT** acquire, process, or analyze medical images, signals, or patterns.
2. Your software function displays, analyzes, or prints medical information normally communicated between health care professionals (HCPs).
3. Your software function provides recommendations (information/options) to a HCP rather than provide a specific output or directive.
4. Your software function provides the basis of the recommendations so that the HCP does not rely primarily on any recommendations to make a decision.

Your software function may be non-device CDS.

Non-Device Examples

Non-Device examples display, analyze, or print the following examples of medical information, which must also not be images, signals, or patterns:

- Information whose relevance to a clinical decision is well understood
- A single discrete test result that is clinically meaningful
- Report from imaging study

AND

Non-Device examples provide:

- Lists of preventive, diagnostic, or treatment options
- Clinical guidelines matched to patient-specific medical info
- Relevant reference information about a disease or condition

AND

Non-Device examples provide:

- Plain language descriptions of the software purpose, medical input, underlying algorithm
- Relevant patient-specific information and other knowns/unknowns for consideration

Device Examples

Device examples acquire, process, or analyze:

- Signal acquisition systems
- In vitro diagnostics
- Magnetic resonance imaging (MRI)
- Next Generation Sequencing (NGS)
- Continuous Glucose Monitoring (CGM)
- Computer aided detection/diagnosis (CADe/CADx)

OR

Device examples display, analyze or print:

- Continuous signals/patterns
- Medical images
- Waveforms (ECG)
- More continuous sampling (aka – a signal or pattern)

OR

Device examples provide:

- Risk scores for disease or condition
- Probability of disease or condition
- Time-critical outputs

OR

Device examples:

- Basis of recommendations is not provided

Your software function is a device.