Government Leaders Forum

Responsible AI and Gen AI

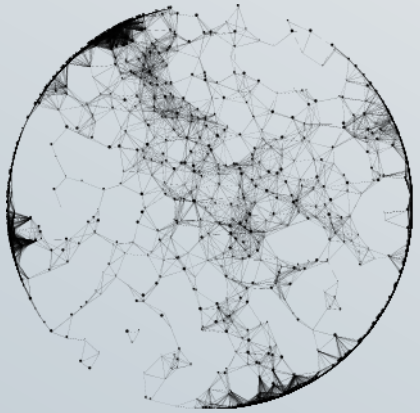
Friday, September 6

The Forum Hotel Grove Ballroom II

FIFTY MILLER CENTER UNIVERSITY of VIRGINIA YEARS

1975–2025

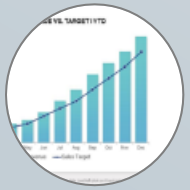
As AI expands beyond analytical AI towards generative AI, a new set of opportunities is emerging



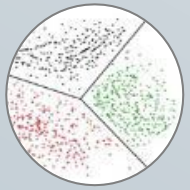
Analytical AI

Analytical AI algorithms are used to solve analytical tasks faster and more efficiently than humans — e.g., classify, predict, cluster, or evaluate data

Examples of use



Forecasting economic growth



Public health data analysis

Sentiment analysis

Generative AI

Generative AI algorithms are used to create new content on par or beyond human capabilities — e.g., generate audio, code, images, text, and videos

Examples of use

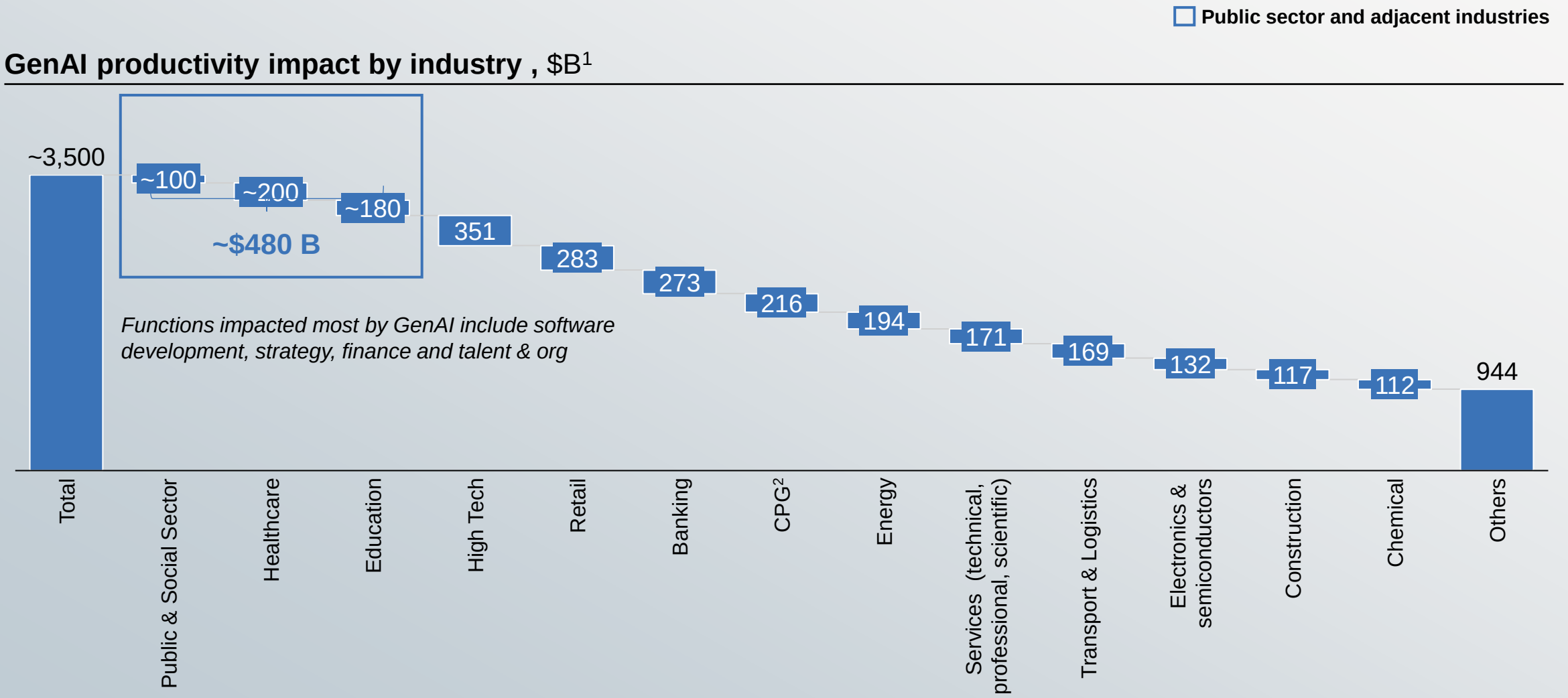
Citizen services: Digital citizen assistant

Policy analysis: Analysis of large unstructured data (e.g., parliament records)

Internal operations: Automated contract writing

Modernization of legacy IT systems: GitHub coding co-pilot

GenAI could unlock up to ~$480B in productivity gains for the global public sector and adjacent industries



1. Excluding implementation costs (e.g., training, licenses)
2. Consumer Packaged Goods

Source: McKinsey.com, "The economic potential of generative AI", June 14, 2023 (link) | Databases: CBF, CIS / IHS, Oxford Economics, Sales Navigator, MSC360

However, GenAI amplifies inherent risks, with broader possibility of organizational and societal impact

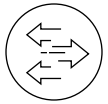
Illustrative **Non-exhaustive**

Data management | Model management | Cybersecurity

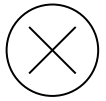
Heightened risk from GenAI

Data Privacy and quality

IP infringement



Impaired fairness



Inaccurate or misleading output

Lack of explainability & accountability

Security threats

Incorrect/ malicious use

Workforce and environmental harms

Third party risk

Macro-prudential risks

Source: McKinsey.com, “Implementing Generative AI with speed and safety”, March 13 2024 (link)

Responsible AI regulation is emerging and is slated to accelerate, driven by regulators in leading economies

Updated as of May 2024 **DOES NOT CONSTITUTE LEGAL ADVICE**

In effect

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

Passed

European Union Artificial Intelligence Act

Proposed

Canadian Artificial Intelligence and Data Act

In effect

Act on the Protection of Personal Information

Proposed

Measures for the Management of Gen AI Services

Proposed

Framework for Gen AI in Schools

Source: https://www.fairly.ai/blog/map-of-global-ai-regulations#AIRegulations

Some US states are implementing responsible AI legislation, executive actions, and policies – more will likely follow

Updated as of June 30, 2024 **Based on legislative summary prepared by the National Conference of State Legislatures** **DOES NOT CONSTITUTE LEGAL ADVICE**

Colorado - In effect

S205 – Consumer Protections for AI

Focuses on the **application of AI**

Requires developers of high-risk AI systems to **use reasonable care to avoid algorithmic discrimination**

Requires developers to:

- Implement a **risk management policy**
- Complete an **impact assessment** and annually review deployment
- **Notify** the attorney general of the discovery of algorithmic discrimination within 90 days

California - Proposed

SB 1047 – Safe and Secure Innovation

Focuses on **frontier models** underlying AI applications

Requires developers to use reasonable care to **test and report the safety of frontier models** before public release

Establishes a **Board of Frontier Models** within the government Operations Agency to **define and monitor safety tests**

Source: "AI Intelligence 2024 Legislation", National Conference of State Legislatures, Updated June 3 2024 (Link), Cate M and Cecilia K, "A California Bill to Regulate AI causes alarm in Silicon Valley", New York Times, Aug 14 2024 (Link); "Bill Summary: SB24-205 Consumer Protections for Artificial Intelligence", Colorado Legislative Assembly website, (link)

An organization's responsible AI strategy often balances its impact aspirations and risk appetite

Illustrative

RAI Use Case Prioritization Cube

Risk
Likelihood and impact of an adverse risk event

Impact
Incremental productivity, social impact, & innovation from use case

Feasibility
To implement the use case

What is the risk?
Type of risk inherent in the Responsible AI (RAI) use case

Why is it important?
Potential impact that the risk might have on the organization, its constituents, and society

How will it be measured?
Key Risk Indicators such as Bias Ratio

- Complexity / technical challenge
- Resources required to deliver & embed RAI guardrails

Source: McKinsey.com, "How generative AI can help banks manage risk and compliance", March 1 2024 (link)

Scaling AI responsibly typically involves decision making across four layers ...

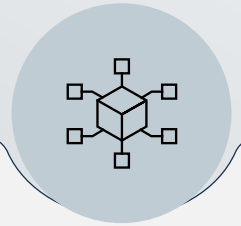
Controls[1] employed at the organizational, data, model, and application layer complement and reinforce each other

1. Organization layer

How is AI governance structured and operationalized?

Process-oriented definition of:

- *Responsible AI strategy*
- *Enablers incl. governance, tools, & training*
- *Operating model*
- *Risk measurement & reporting*

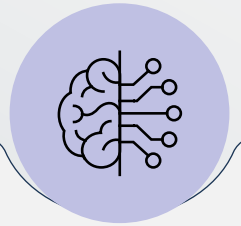


2. Data layer

How is training and input data sourced, processed and stored?

Compliance-oriented review of

- *Data sourcing, quality, and processing*
- *Data storage, management, and afterlife*



3. Model layer

How are AI models trained, tested, and distributed?

Performance-oriented review of:

- *Model design*
- *Model training and testing*
- *Documentation of model limitations and user instructions*

4. Application layer

How do AI systems impact users and society?

Impact-oriented review of:

- *Use case design & functionality*
- *Ethical impact assessments*
- *Channels for whistleblowing, complaints and redress*

1. **Technical controls:** Tools and infrastructure deployed to mitigate risk or to detect and respond to data breaches and security incidents, **Procedural controls:** Structured processes with clearly defined decision paths to manage risks and allocate accountability for potential harms, **Cultural controls**: Employee training and awareness programs to promote a proactive risk management culture

Source: McKinsey.com, "Getting to know – and manage – your biggest AI risks", May 3, 2021 (link)

…such decision-making to be driven by cross-functional SteerCo and Delivery teams

Illustrative Non-exhaustive

AI Trust Council

The AI Trust council **spearheads strategy and decision making** with a mandate to develop policies and guidelines, proactively assess and mitigate AI risk, and review/ approve AI use-cases

AI Delivery Pod

The AI Delivery Pod supports the leadership team to **implement strategy, processes, & policies at the use-case level**, ensuring responsible implementation and early risk mitigation

Capabilities and teams represented

Domain/ function owner	Data governance	IT & Cybersecurity	Risk, legal, & compliance	AI use case build team	Head of AI