QuantumBlack
AI by McKinsey

AI Assurance

McKinsey perspectives

August 2023

Topics

AI defined

Digital trust

Algorithm fairness and bias

Responsibilities

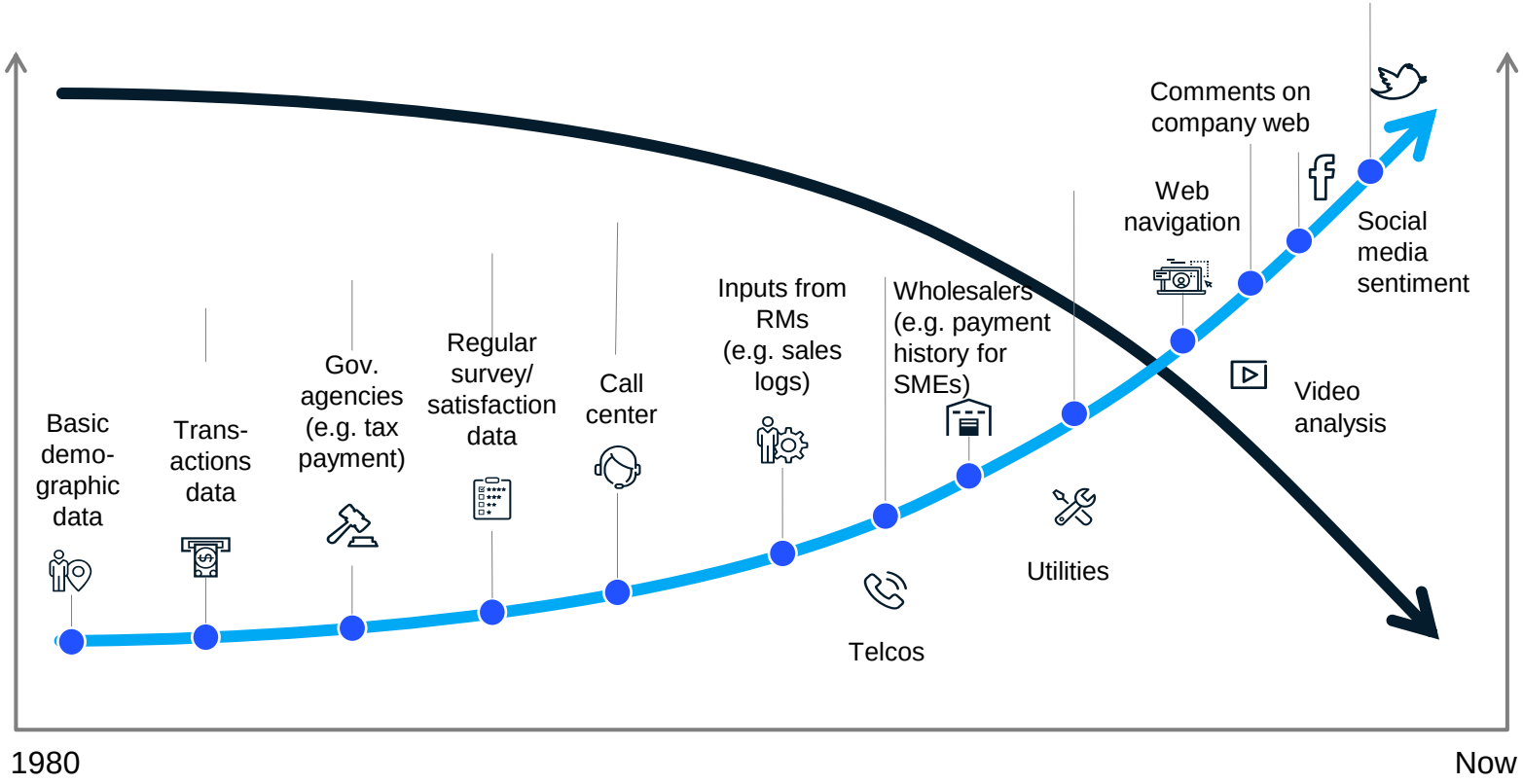
AI defined

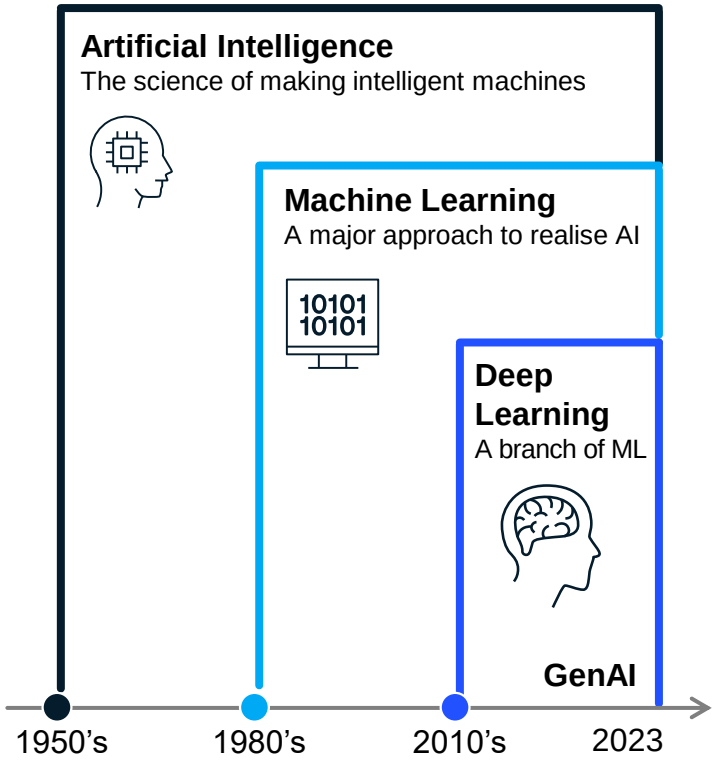
Three trends have put advanced analytics/ AI within everyone's reach…

1 Tumbling costs of data storage and processing

2 Rapidly expanding data availability

3 Advanced, new algorithms

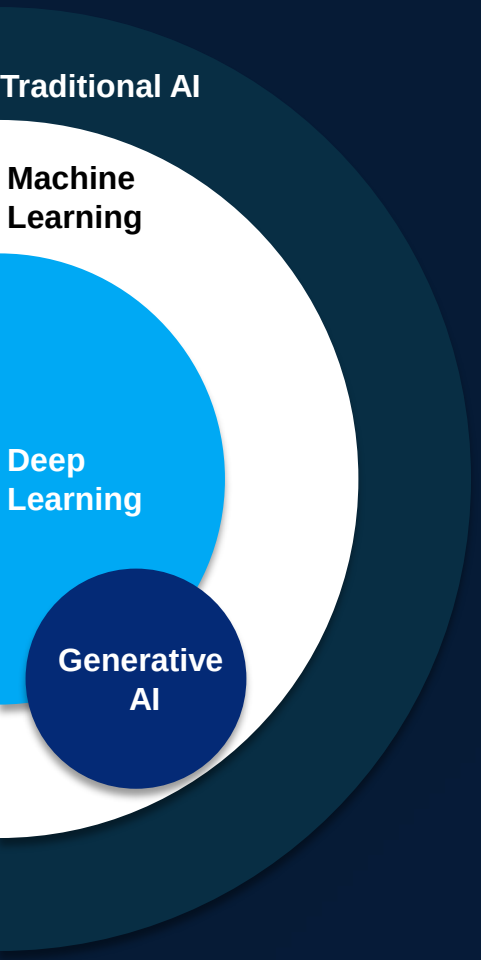




Source: Dave Evans "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything"

AI encompasses a spectrum approaches

Illustrative



		Definition	Advantages	Example use cases
1	**Traditional AI**	Rules-based systems that **mimic human cognitive functions**	**Interpretable**, can make decisions based on explicit rules	Engineering design optimization
2	**Machine Learning**	Algorithms that can **analyze** and **learn** from **data**	Can **learn** and **improve performance** over time, can handle **large data sets**	Manufacturing defect detection
3	**Deep Learning**	A subset of ML algorithms that use **neural networks** with multiple layers (3+)	Can **learn complex patterns** in data, can **handle unstructured data**	Computer vision in factories
4	**Generative AI**	Group of ML/DL algorithms that **generate outputs** based on **data** they have been **trained on**	Can **generate new and original content**	Software development augmentation / co-pilot Regulatory filing draft creation

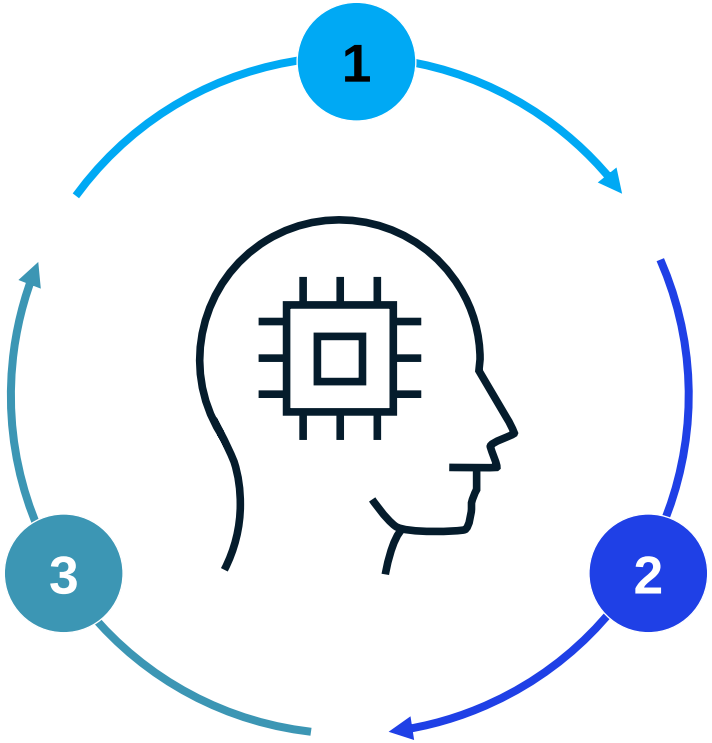
Source: IBM - AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?; WEF - What is generative AI?; McKinsey Analysis

Similar to how the human brain works, AI receives information, learns from it, improves its model and makes better decisions over time

Decide

Machines do tasks or make decisions based on the output of algorithms and their predictions

Machines identify patterns and make decisions similar to how humans when faced with something new, compare it to a known pattern to make sense of it before acting



Receive information

Machines do tasks or make decisions based on the output of algorithms and their predictions

Machines continue to receive new data, structured and unstructured, similar to how humans build experience as they grow

Learn

Machines process the data through algorithms changing them as they learn more about the information they are processing

Algorithms performance improves as they are exposed to more data over time

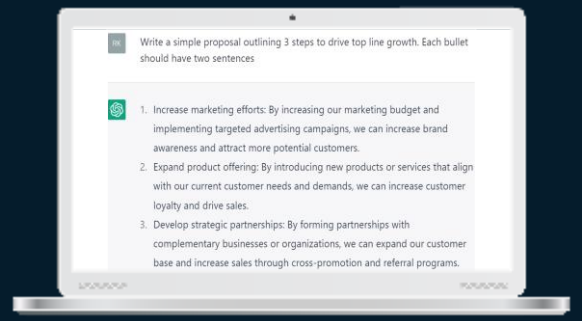
And then … Generative AI

Non-exhaustive

Generative AI (GenAI) enables the **creation of new unstructured content**, such as text, images, etc.

GenAI is powered by Foundation Models (artificial intelligence models) trained on a **broad set of data** that can be adapted to a wide range of tasks

These models are typically also **better at interpreting / labelling unstructured data than traditional AI**



Generate marketing or social media copy in ”house style” using ChatGPT, Copy.A, etc.

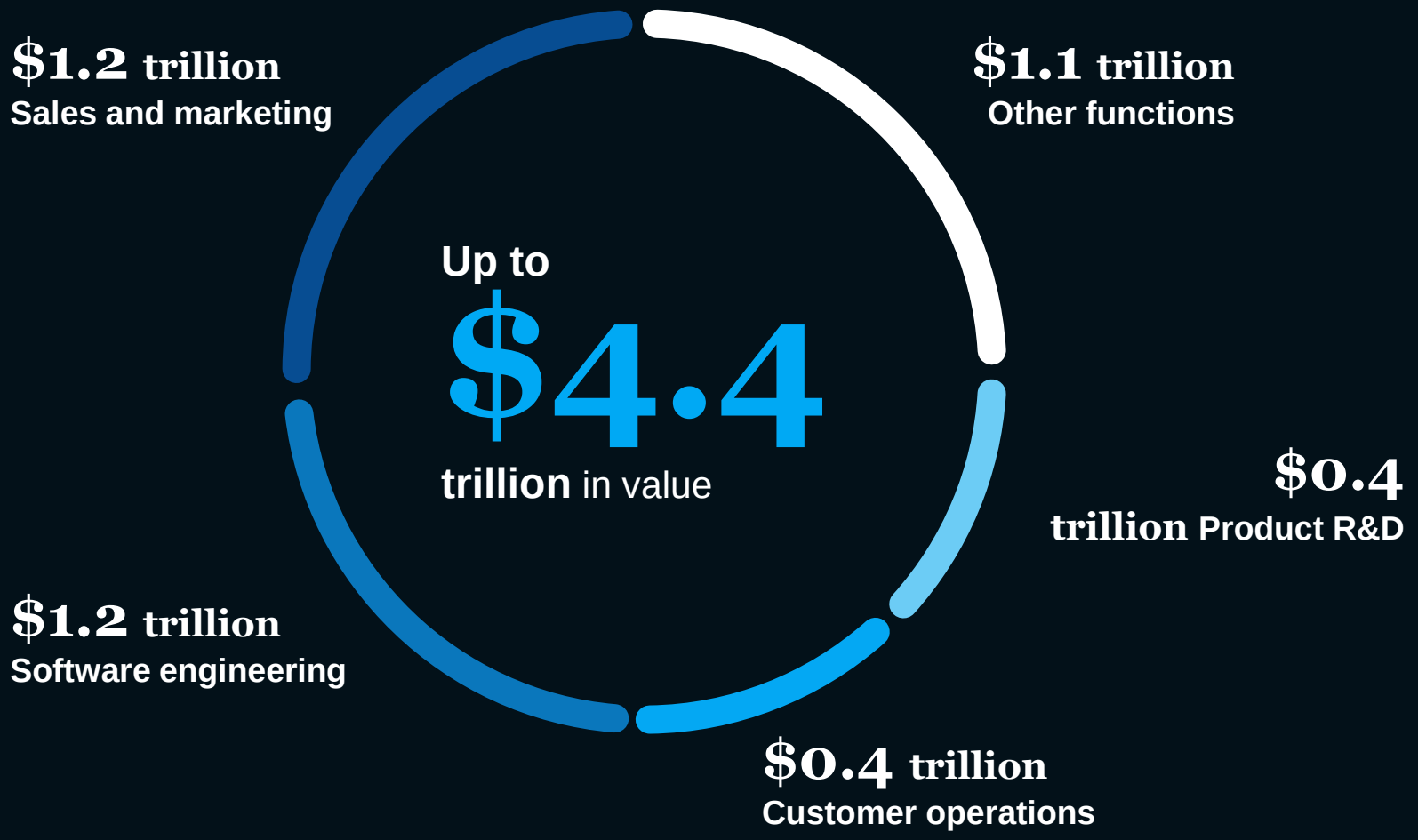
Create new product design concepts using DALL-E2, Stable Diffusion, etc.

Accelerate the drug discovery process, reducing time in laboratories with ABSCI, etc.

Automate code generation in programming languages like Python with Codex/Github Copilot, etc.

Source: Press search, Expert interviews; DALLE

Generative AI is poised to boost performance and unlock trillions of dollars across functions



Source: McKinsey report: The Economic Potential of Generative AI: The Next Productivity Frontier

Four main GenAI use case archetypes ("4 C's") are demonstrating significant value in industry

● Explored next

Content Synthesis (insight generation AI)

- ChatGPT — ChatGPT able to act as a virtual assistant
- AutoGPT — Self-prompting ChatGPT for executing complex tasks
- UiPath — RPA for business processes

~14%
Productivity increase for customer support agents using AI assistants[1]

Coding & Modelling (content generation AI)

- GitHub Copilot — GitHub's co-pilot for coding
- Cody — Google's co-pilot for coding
- MOSTLY·AI — Synthetic data for machine learning models

>55%
Productivity gains for developers utilizing coding co-pilots such as Github Copilot[2]

Creative Generation (content generation AI)

- Jasper — Copywriter and content generator
- Midjourney — Artificial image generator
- IIElevenLabs — Artificial voice generator

>10x
Expected increase in AI generated outbound marketing messages by 2025[3]

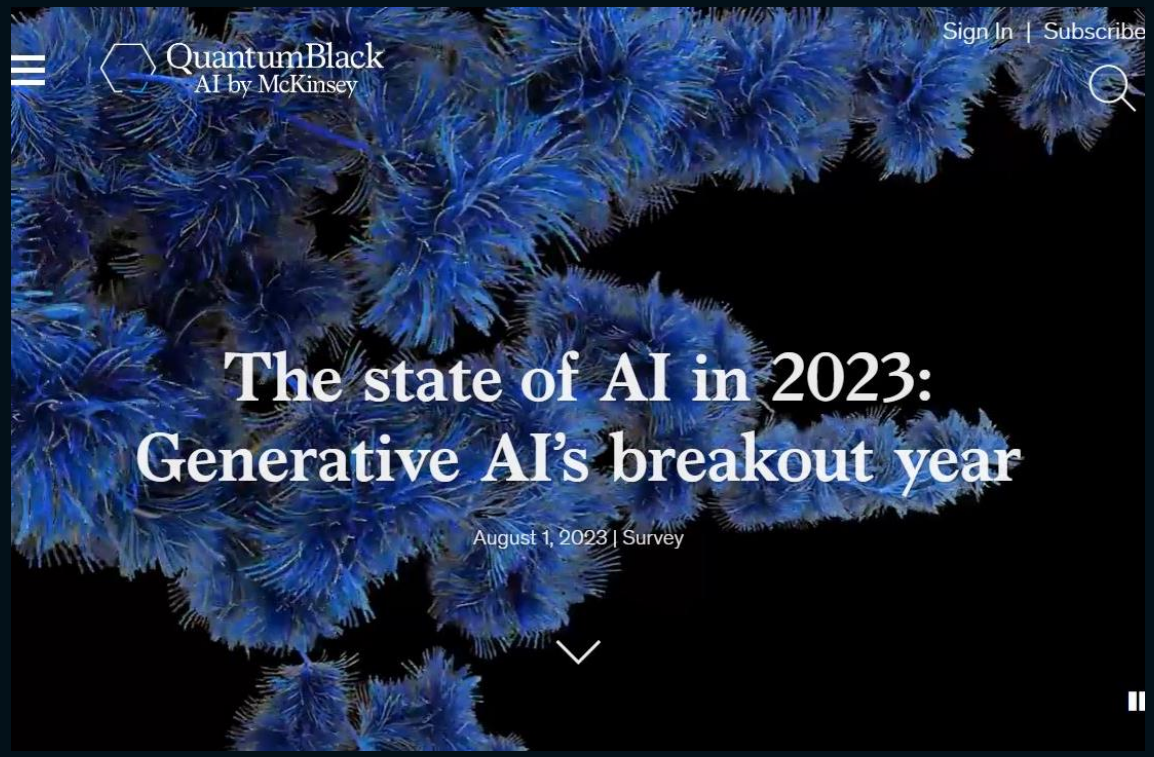
Customer Engagement (interaction AI)

- Dialogflow — Google's chatbot service
- ChatSpot — Hubspot's CRM chat service
- yellow.ai — Conversational CX service

>60%
AI-driven automation potential of customer interaction volumes over next 5-10 years[4]

1. Stanford and MIT study, Generative AI at Work, NBER, Apr 2023
2. GitHub, Research: quantifying GitHub Copilot's impact on developer productivity and happiness, Sep 2022
3. Gartner, Beyond ChatGPT: The Future of Generative AI for Enterprises, Jan 2023 – increase from 2% in 2022 to 30% in 2025
4. McKinsey analysis on automation potential in CX BPO, 2023 – today 20-30% of volume is automated

About the Research

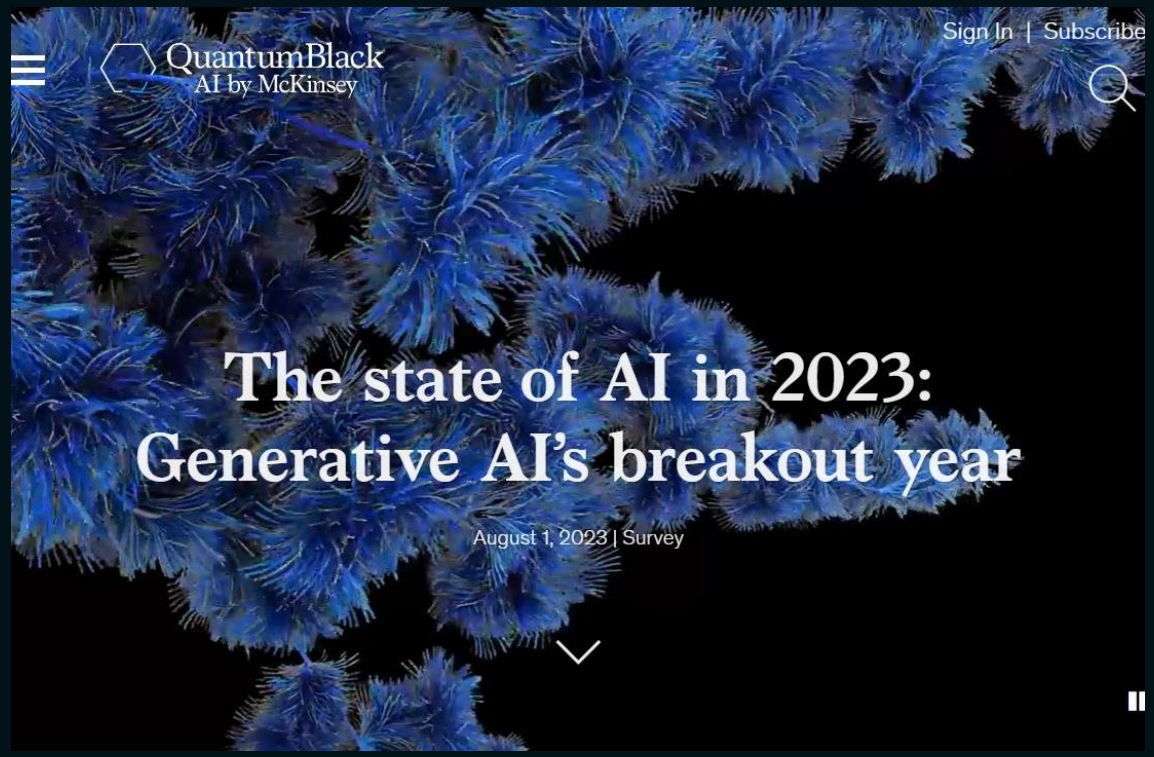


The State of AI in 2023

The online survey was in the field April 11 to 21, 2023, and garnered responses from 1,684 participants representing the full range of regions, industries, company sizes, functional specialties, and tenures.

Source: McKinsey Global Survey, The State of AI in 2023 – Generative AI's Breakout Year

The State of AI in 2023 – Generative AI's Breakout Year

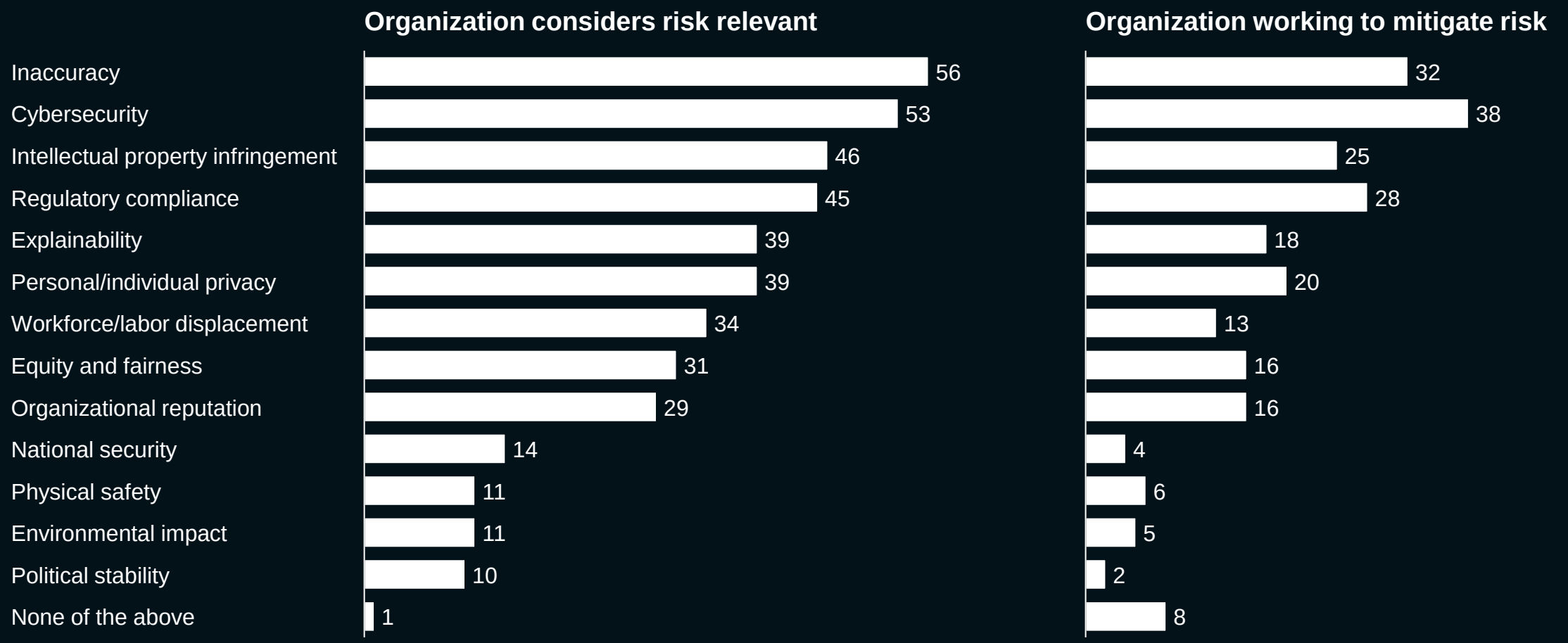


The State of AI in 2023

1. **It's early days still, but use of gen AI is already widespread**
2. **Leading companies are already ahead with gen AI**
3. **AI-related talent needs shift, and AI's workforce effects are expected to be substantial**
4. **With all eyes on gen AI, AI adoption and impact remain steady**

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. Only asked of respondents whose organizations have adopted AI in at least one function. For both risks considered relevant and risks mitigated, n = 913.

Source: McKinsey Global Survey, The State of AI in 2023 – Generative AI's Breakout Year

Digital trust

What is Trustworthy Artificial Intelligence?

Source: EU Commission High-Level Expert Group on Artificial Intelligence

“”

Trustworthy AI has three components: (1) it should be **lawful, ensuring compliance** with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.

Source: This is informed by the Ethics Guidelines for Trustworthy AI, accessible here

What is Digital Trust?

Impact of Digital Trust

Digital Trust helps identify, assess, and mitigate risk

- **Elevated digital and analytical risks**
- **Higher public sensitivity**

Upward opportunity

- Higher performance
- Lower operational costs
- Greater acceptance

Downward protection

- Decreased likelihood of disruption
- Reduced operational risk
- Reduced regulatory and reputational risk

In our increasingly digital world, 'trust' is a differentiator for high performers…

Preliminary **Not exhaustive**

87%

of **customers believe it is important** for them to be able to **review and control their personal data online**[1]

71%

of customers factor a company's ability to keep information safe into buying decisions[1]

282%

Cumulative stock market returns of **trusted brands** in 2022 **compared to 36% for untrusted brands**[2]

…What is Digital Trust?

Digital Trust is an individual's confidence in an organization to protect their data, enact effective cybersecurity measures, offer trustworthy AI-powered products and services and provide transparency around AI and data usage.

Digital Trust addresses digital and analytic risk across AI/ML & analytics, data, technology & cloud, and risk capabilities & culture

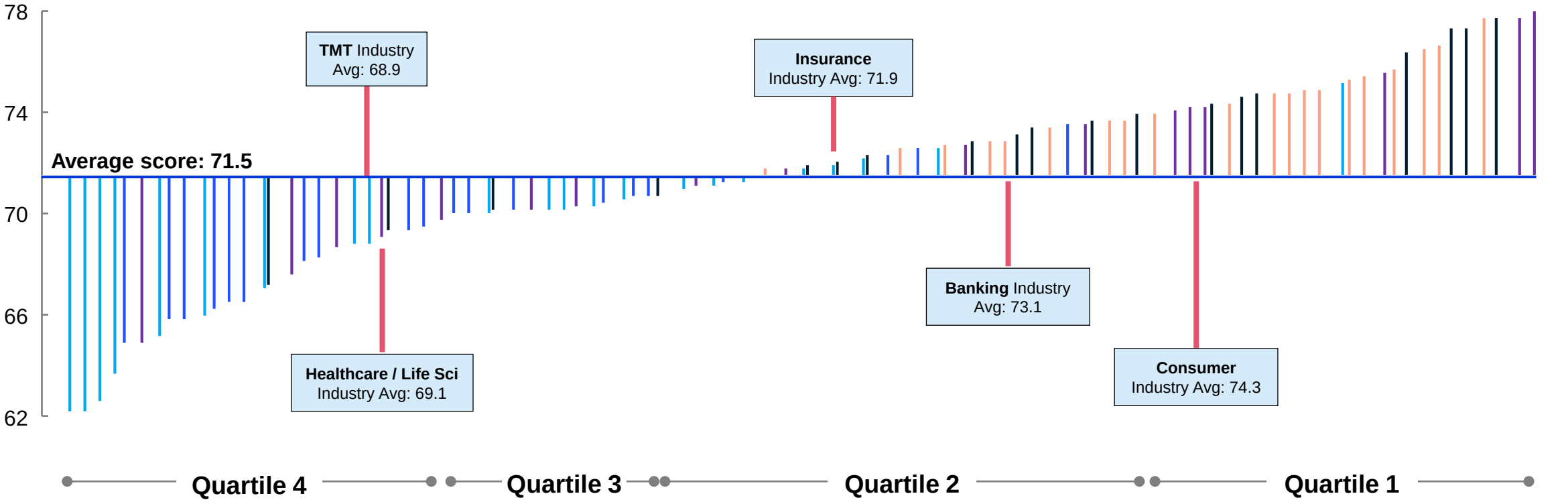
Source: 1. PEW 2020 customer survey; 'Trust and Predictive Technologies 2016' by Edelman and The University of Cambridge Psychometrics Centre, McKinsey 2. Aggregate performance (e.g., each company was indexed to 2017 and weighted by market capitalization)

Digital Trust performance of 100 selected companies in Digital Trust Index compared to industry peers and Digital Trust leaders

Illustrative **Not exhaustive**

Banking | Healthcare / Life Sciences | Consumer | Insurance | TMT

Distribution of Digital Trust (DT) score (out of 100)



Full list of measurement indicators in back-up

1.Digital Trust Emergents are defined as the bottom quintile companies in the Digital Trust Index database
2.Digital Trust Leaders are defined as the top quintile companies in the Digital Trust Index database

Source: McKinsey expert analysis

Digital trust levers and priorities

Theme	Immediate priorities	Questions for tomorrow
Vision and strategy	Does your organization **have a digital trust vision** (e.g., what constitutes ethical use of data, tech, and responsible AI)**?** How is **data and technology ethics being built** into strategy? (e.g., repeatable playbooks, governance mechanisms, etc.)	Is your **vision & mission future-proofed** (e.g., trust infrastructure in place to support imminent innovations)? How are you **planning talent strategy for the future (e.g., with Gen AI's ability to automate and augment** tasks)?
Framework and Taxonomy	Does your organization have a **data and privacy risk management function** (e.g., common language across the organization, risk taxonomy)?	Does your organization **have a mechanism to ensure that the taxonomy is continuously updated**?
Policies and Standards	How is your organization's **data managed / structured** to enable **quick and consistent access** to insights (e.g., data quality standards, tooling)?	**What new and evolving data, AI and technology policies do you need** (e.g., policies around managing third-party risks in foundation models)?
Capabilities & controls	How is **data being safeguarded**, and are **the strongest controls being applied to the most critical data**? (e.g., Data Leakage Prevention program)	What **new capabilities (e.g., hardware, cloud, MLOps) and controls (e.g., privacy enhancing technologies) are** needed to enable digital trust long-term?
Adoption and Scaling	What governance & **change management practice** is in place to address new data use cases?	How do teams **keep pace with translating data and AI trust decisions into technical requirements**, especially in a rapidly evolving business landscape?

AI/ML in digital impacts existing risks and brings new ones

Increase in existing risks

Extensive data requirements

AI/ML techniques require vast amount of data and management of hundreds of features

Algorithmic transparency and complexity

There are a large number of available machine learning algorithms, many of which are complex and opaque

Exploitation or misuse potential

AI/ML systems have characteristics, such as frequent retraining, scalability and anonymity, that increase risks of misuse

Appearance of new risks

Bias and unethical behaviour

AI/ML models can lead to unintended results that cause bad public and regulatory reaction

Explainability

Understanding of human-machine interactions and use of explainable AI/ML is still evolving

Risk inducing scalability

AI/ML makes it possible to embed analytics deeply into process workflows and automate decisions

Fairness and bias management

Fairness in algorithms

Fairness concerns are becoming increasingly relevant for **models whose decisions directly affect customers** e.g., loan approval models, pre-screening models for hiring, personalized marketing models

While there do not yet exist widely-accepted definitions, metrics and thresholds for fairness risks, managing the risk of unfair model operation is an **important component of an overall digital ethics strategy**

If fairness risks are not monitored and managed appropriately, this may result in **reputational risks for the company**, which may significantly impact revenues.

Fairness risks appear in different ways in each business unit

Exemplary business units

	Typical questions	Fairness concerns
Human resources & talent	How can we choose the best fitting candidate? How can we ensure our hiring processes are fair?	When using algorithmic systems for classification and evaluation, discrimination and selection bias might occur
Marketing & Sales	How can we target different audiences? How can we use data to drive sales conversions?	Users' and customers' willingness to share data may be undermined due to perceptions about the collection and use of data
Customer support	How can we provide an easy self-registration process? How can we drive efficiencies through chatbots?	When digital interfaces and tools like chatbots are offered, the human-machine interactions might lead to unfair outcomes
Product / service mgmt. & development	How can we ensure that our services are fair? How can we collect data for future improvements?	When digital products are developed, they might be rejected by users due to ethical concerns, resulting in high innovation and development costs
Strategy & corporate finance	Which processes need to be digitalized? How can we adapt our business model in alignment to fairness factors?	When new business models are explored or developed, there can be unintended consequences that discriminate for or against certain populations
Core IT & security	How can we ensure our models are fair and secured? How can we ensure roles and responsibilities are clear?	When new systems are introduced, both employees and management may be affected and disadvantaged. There can be unclear responsibilities

Leadership support is critical to creating a **fairness culture** and **governance structure** for risk mitigation and harm avoidance

A Fairness Framework serves as a **strategic framework** for **operationalizing fairness principles** in a corporate context thriving for value-centered innovation

Responsibilities

Organizations starting their AI/ GenAI journey should focus on Responsible AI from day-one

Risk categories to address with Responsible AI

Impaired fairness
Algorithmic bias; misrepresentation of generated content as human-created

IP infringement
Infringement on copyrighted or otherwise legally protected materials

Privacy concerns
Unauthorized use/disclosure of personal or sensitive information

Malicious use
AI-generated promulgation of malicious content

Performance & explainability risk
Inability to explain model outputs appropriately and model inaccuracies

Security threats
Vulnerabilities in generative AI systems that may be breached or exploited

ESG impact
Non-compliance with ESG standards; reputational risk

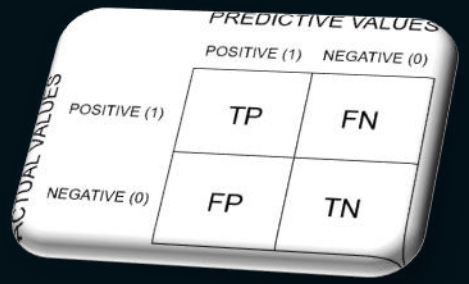
Third-party risk
Risks associated with the use of third-party AI tools

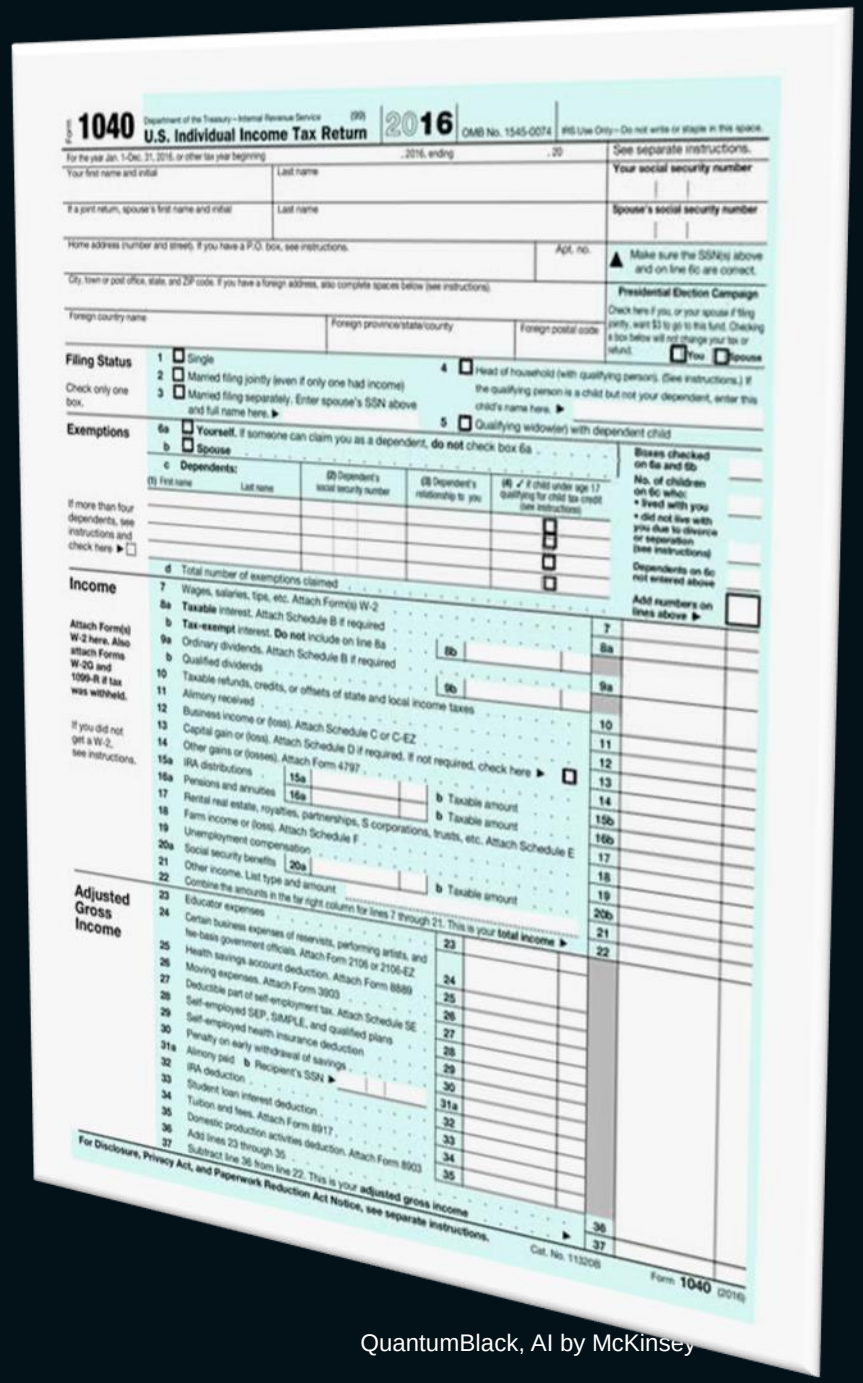
Assurance strategies

An example

IRS

Department of the Treasury
Internal Revenue Service





QuantumBlack, AI by McKinsey

In AI space, MLOps promotes assurance

ML Lifecycle

Data management
Ensure data availability, quality and control to feed the ML system

- Data discovery
- Data transformation
- Data versioning

Development
Standardize processes & tooling to improve team efficiency and solution performance

- Production ready code (incl. CI)
- Model choice & optimization
- Experiment tracking

Deployment
Standardize processes & provision tooling to reliably deploy solutions

- Model management & deployment (incl. CD)
- Model validation
- Serving infrastructure

Live Ops
Maintain reliable model performance and drive continuous improvement

- Monitoring
- Model maintenance
- Continuous improvement
- User support

Enablers

Technology Stack
Provision the environment & tooling to optimize ML workflows

Compliance, security & risk
Establishing processes, governance and tooling to control your ML system

Assetization
Assetizing and maximizing reuse of ML components across the ML portfolio

People
Establish tech talent bench & new org/op model to delivery new ways of working

QuantumBlack
AI by McKinsey