Building More Responsible AI Systems

A Webinar featuring Gartner®, IBM and Casper Labs

Thursday, November 2, 12:00 - 01:00 PM ET

casper Labs

IBM

watsonx

Featured Speakers

Avivah Litan
Gartner
Distinguished VP Analyst

Mrinal Manohar
Casper Labs
Co-founder and CEO

Heather Gentile
watsonx
Executive Director of
Product Management,
Data and AI

Shyam Nagarajan
IBM
Executive Partner, Blockchain,
Web3.0, Metaverse and
Sustainability

Scan to participate in the Q&A session
-- ask your questions at any time!

What We'll Focus On

- Dispelling one big myth about AI
- Overcoming AI adoption hurdles
- Demonstrating good AI governance

Myth: AI and blockchain are not compatible technologies.

3 Ways Blockchain Augments Generative AI

Completely **tamper-proofs** and **decentralizes** governance process

Eradicates need to manually **serialize** sequence of events

Automates key processes that save time and overhead

Gartner's Perspective

Manage GenAI Risks Before They Manage You

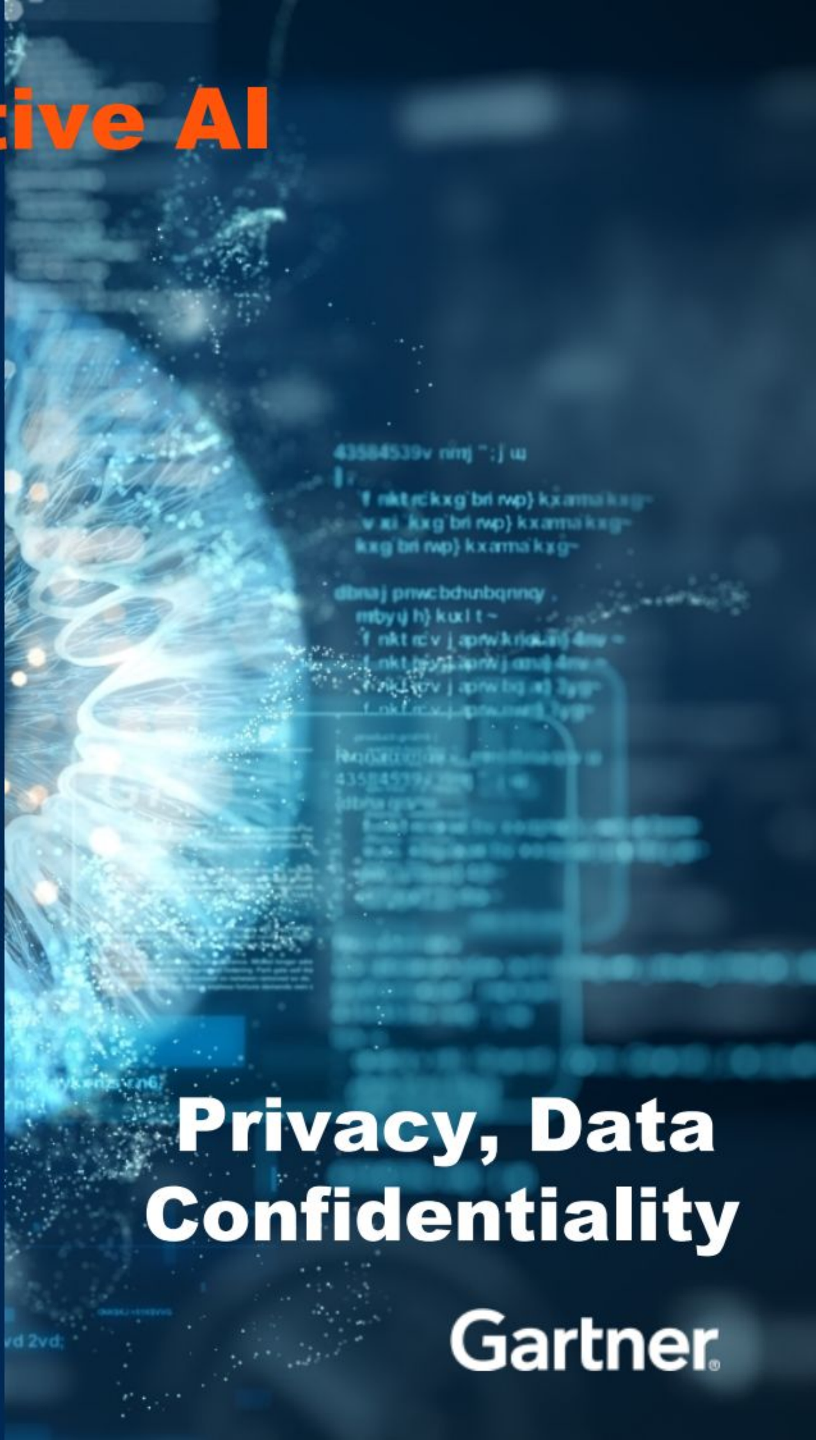
Avivah Litan

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."

Plenty to Worry About With Generative AI

Cybersecurity Threats

Unwanted, Inaccurate Outputs



Privacy, Data Confidentiality

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Key Issues

1

GenAI Risks and Attack Surface

2

Best Practices for Mitigating GenAI Risks

3

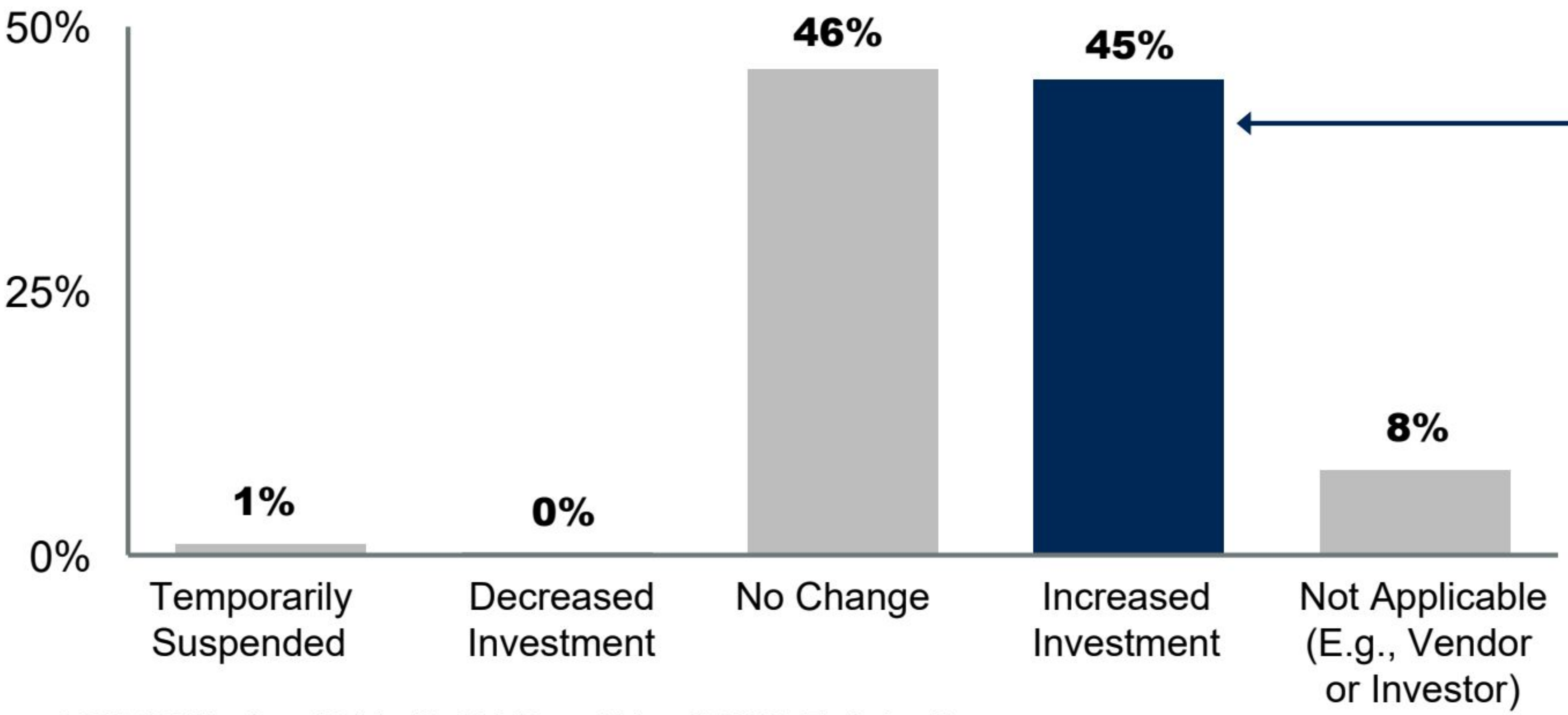
Blockchain for Managing Risks

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

OpenAI ChatGPT Spurred a Major Increase in AI Investment

Change in AI Investment Since ChatGPT

Percentage of Respondents



Nearly Half of Executives Report ChatGPT Spurred Increase in AI Investment

73%

of Organizations Have Hundreds or Thousands of Models Deployed

Source: 2021 Gartner AI in Organizations Survey

n = 2,554 (30 March and 21 April); 324, Base: Using AI (S08), Excludes Unsure

Q: How have your AI investment strategies changed since the recent publicity of ChatGPT?

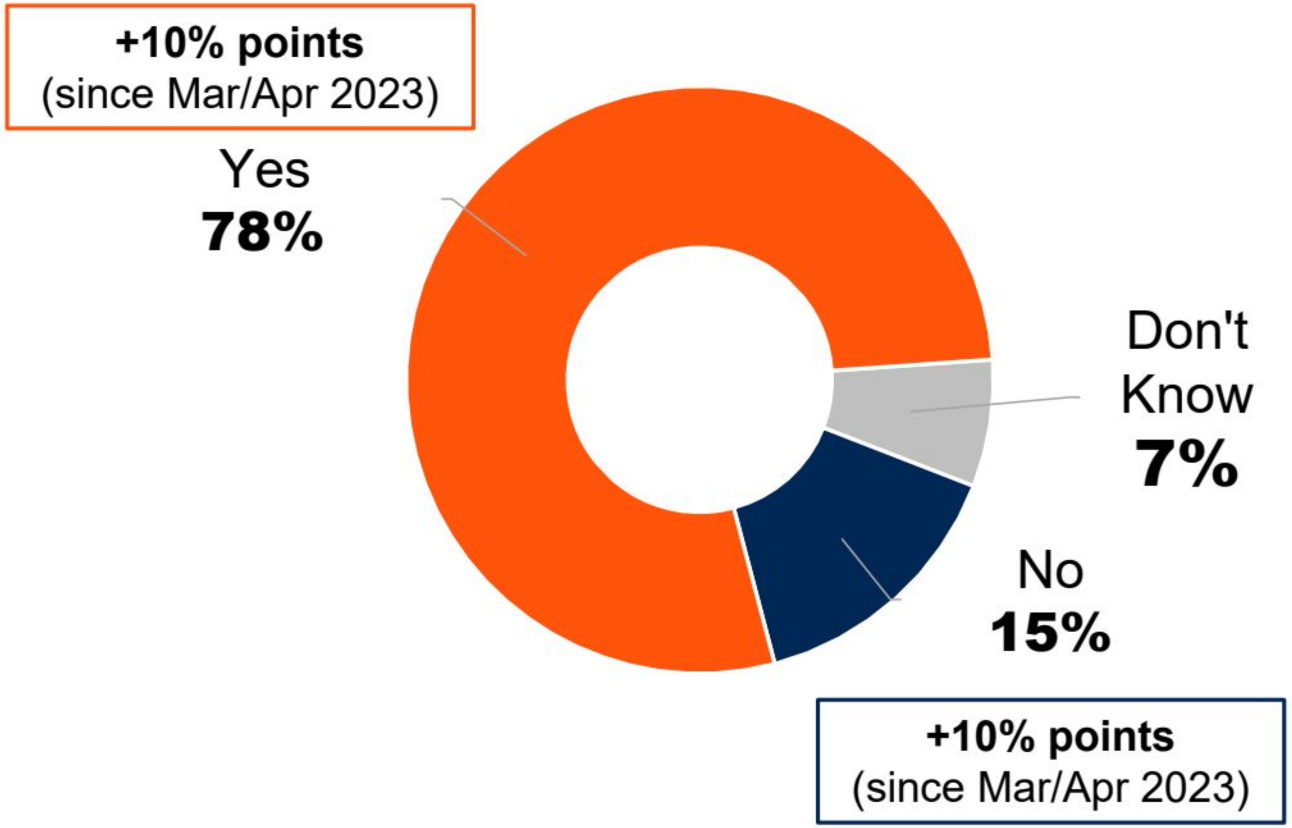
Q13A: How many AI models has your organization deployed to date?

Source: Beyond the Hype: Enterprise Impact of ChatGPT and Generative AI

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Most Say Benefits of GenAI Outweigh the Risks

Question: Based on everything you learned in the last 10 months regarding Generative AI, do you believe the benefits of Generative AI outweigh its risks?
Percentage of respondents



89%

of business technologists would bypass cybersecurity guidance to meet a business objective.

Source: Infographic: Build Business Technologists' Cyber Judgment to Improve Risk Decision Making (G00780945)

n = 1,419 (September); 2,544 (March and April)
Source: Generative AI Realities: Proactive Approaches for Quantifiable Business Results Webinar Polling September 2023;
Source: Beyond the Hype: Enterprise Impact of ChatGPT and Generative AI Polling March and April 2023

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

LLM Enterprise Risks Are Plentiful and Documented

1

Unreliable Outputs

Lawyer Used ChatGPT in Court — And Cited Fake Cases. A Judge Is Considering Sanctions (Forbes)

2

Data Privacy, Confidentiality

Samsung Bans ChatGPT, AI Chatbots After Data Leak Blunder (Mashable India)

Apple Joins a Growing List of Companies Cracking Down on Use of ChatGPT by Staffers — Here's Why (Forbes)

3

Intellectual Property

Getty Images Sues AI art Generator Stable Diffusion in the U.S. for Copyright Infringement (The Verge)

4

Liability

OpenAI Sued for Defamation After ChatGPT Fabricates Legal Accusations Against Radio Host (The Verge)

Sarah Silverman Sues OpenAI and Meta Over Copyright Infringement (Engadget)

5

Cybersecurity and Fraud

Cyberattacks Are Accelerating With AI's Help (BenefitsPRO.com)

Over 100,000 ChatGPT Accounts Stolen Via Info-Stealing Malware (Bleeping Computer)

6

Consumer Protection

Eating Disorder Helpline Disables Chatbot for 'Harmful' Responses After Firing Human Staff (VICE)

7

Regulatory Compliance

An Insider's View on AI Regulation (The Hill)

FTC Investigates OpenAI Over Data Leak and ChatGPT's Inaccuracy (The Washington Post)

8

Misinformation

How Verified Accounts Helped Make Fake Images of a Pentagon Explosion Go Viral (ABC News)

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Key Issues

1

GenAI Risks and Attack Surface

2

Best Practices for Mitigating GenAI Risks

3

Blockchain for Managing Risks

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Gartner's MOST Framework for Managing AI Trust, Risk and Security

	New to AI			
Compromise Vector	Query Attacks	Prompt Injections, Malicious, Mistaken Inputs	System Faults	Human Error or Compromise
Types of Compromise	Model Manipulation, Theft or Poor Performance	Data Poisoning or Data Drift	Asset Damage or Manipulation	Theft of Data or Money Loss
MOST Risk Management Measures	**Mo**del Mgmt. AI Model Integrity	**Mo**del Mgmt. AI Data Protection/Privacy	**S**ecurity Endpoint, Network, IAM, Data Protection	**T**rustworthiness Ethics, Bias Mitigation, ERM

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

3 Main Gaps in Hosting GenAI Provider Controls

Anomaly Detection

- Unacceptable use that compromises enterprise decision making and confidentiality
- Data compromise of confidential inputs
- Inaccurate, hallucinations, copyright, illegal and otherwise unwanted outputs

Data Protection

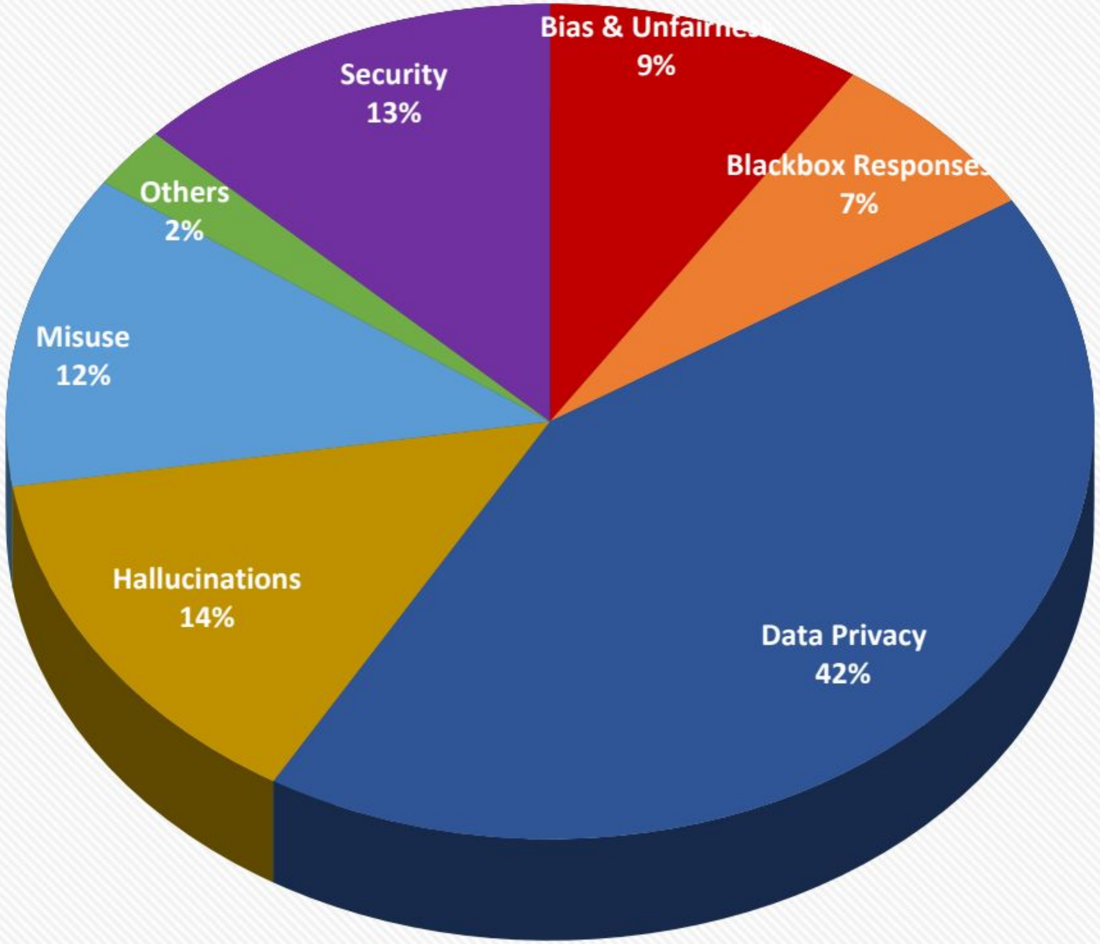
- Data leakage and potential compromise in hosted LLM vendor environment
- Completeness of data privacy and protection policies
- Hard to meet regulatory compliance e.g., for GDPR

Application Security

- Prompt Injection attacks
- Vector database attacks
- Hacker access to model states and parameters

© 2022 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

GenAI Data Privacy Risks #1 Concern



Poll: Which Risks of GenAI are You Most Worried About?

N = 713; Source: Gartner IT Executives Webinar Poll, August 2023

Range of Methods can Reduce Unwanted Outputs, e.g., Copyright Materials, Hallucinations, etc.

Aligned Models

Prompt Engineering

Tools and Frameworks

Content Authentication

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Key Issues

1

GenAI Risks and Attack Surface

2

Best Practices for Mitigating GenAI Risks

3

Blockchain for Managing Risk

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Blockchain Can Uniquely Transform Digital Business; Byzantine Fault Tolerance

Current State

Lack of Trust Across Ecosystems
No trust in others' data and processes

Data Reconciliation Issues
Different data in different siloed systems for same events or entities

No Process Automation Across Entities
Automation limited to single entities or organizations

Blockchain Features

Shared System of Record
Anytime, anywhere

Immutable Data
Append only

Distributed Ledger
Independently verifiable

Tokens
Digital or physical assets

Smart Contracts
Process automation

Transformed State

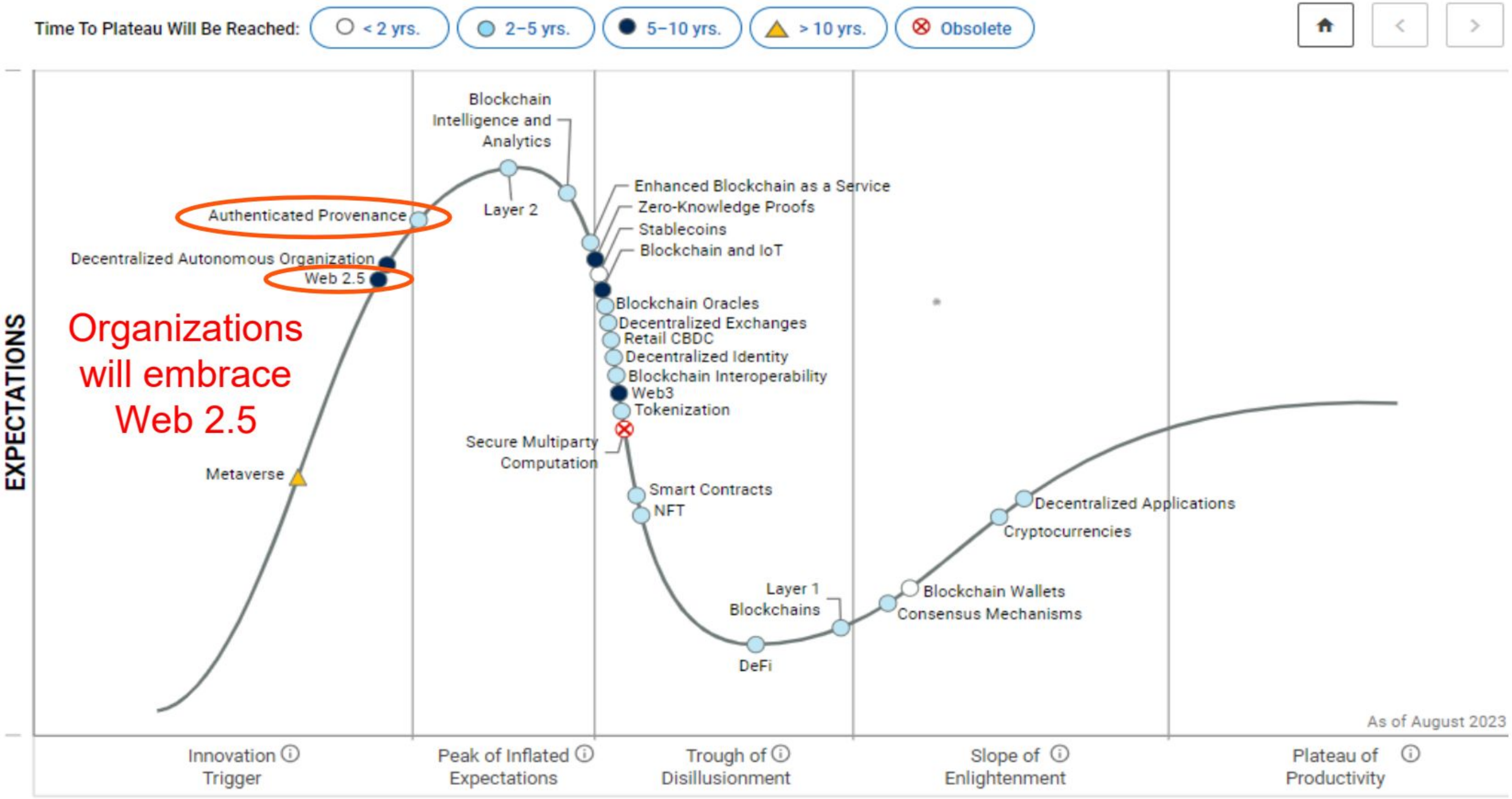
Trust & Transparency

Programmable Economy

More Efficiencies and Opportunities

© 2019 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Gartner Hype Cycle for Blockchain

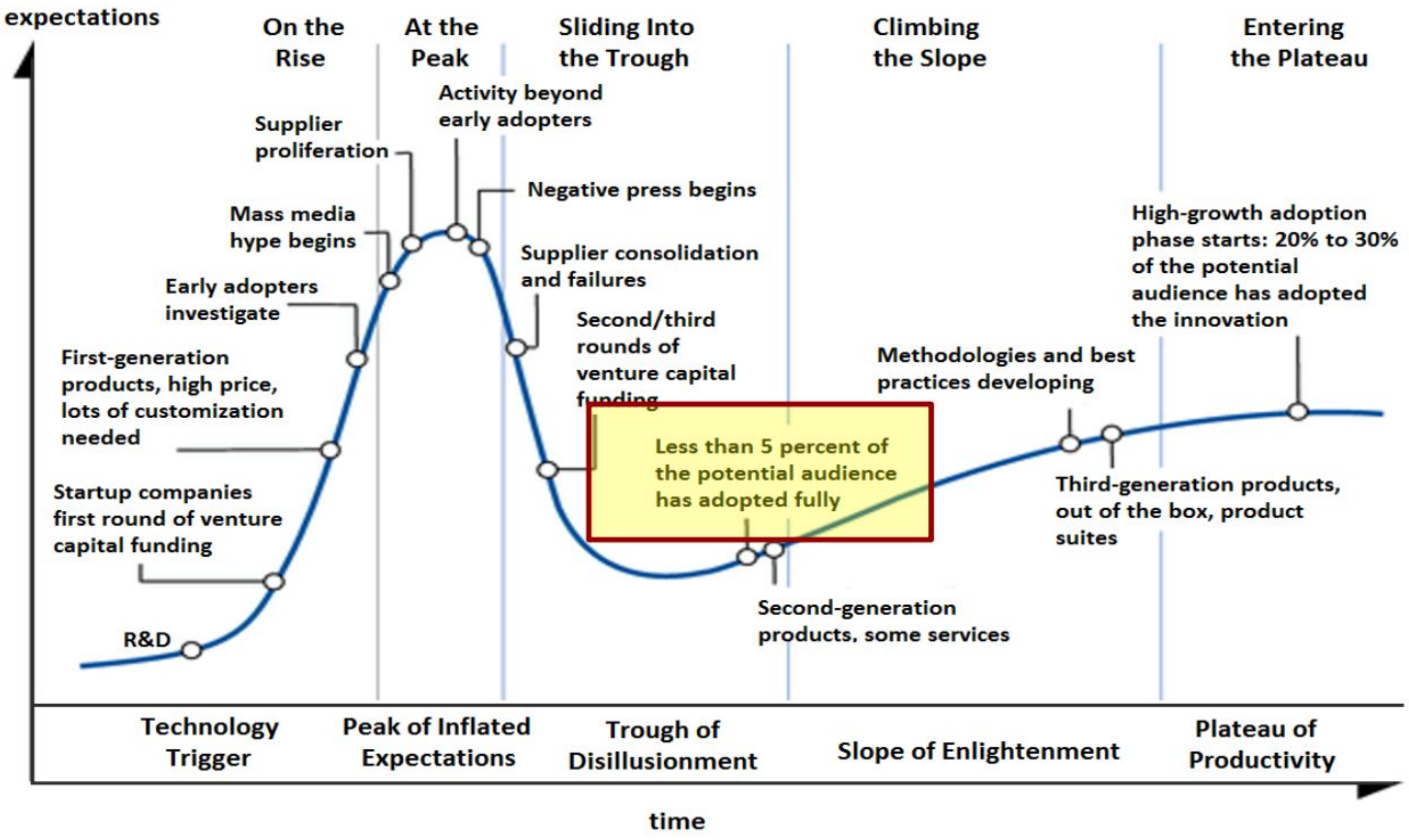


Organizations will embrace Web 2.5

N = 774; Source: Gartner IT Executives Webinar Poll, August 2023

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

Milestones in the Journey to the Plateau of Productivity


expectations
On the Rise
At the Peak
Sliding Into the Trough
Climbing the Slope
Entering the Plateau
Activity beyond early adopters
Supplier proliferation
Negative press begins
Mass media hype begins
Supplier consolidation and failures
Early adopters investigate
Second/third rounds of venture capital funding
First-generation products, high price, lots of customization needed
Startup companies first round of venture capital funding
R&D
High-growth adoption phase starts: 20% to 30% of the potential audience has adopted the innovation
Methodologies and best practices developing
Third-generation products, out of the box, product suites
Second-generation products, some services
Technology Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time

Less than 5 percent of the potential audience has adopted fully

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

watsonx on AI Governance

The speed, scope, and scale of generative AI impact is unprecedented

Massive early adoption

80%

of enterprises are working with or planning to leverage foundation models and adopt generative AI

Broad-reaching and deep impact

Generative AI could raise global GDP by

7%

within 10 years

Critical focus of AI activity and investment

Generative AI expected to represent

30%

of overall market by 2025

Sources: Statista; Reuters; Goldman Sachs; IBM Institute for Business Value; Gartner. Scale Zeitgeist: AI Readiness Report, a survey of more than 1,600 executives and ML practitioners

Our principles and pillars in practice

Governance, Risk and Compliance

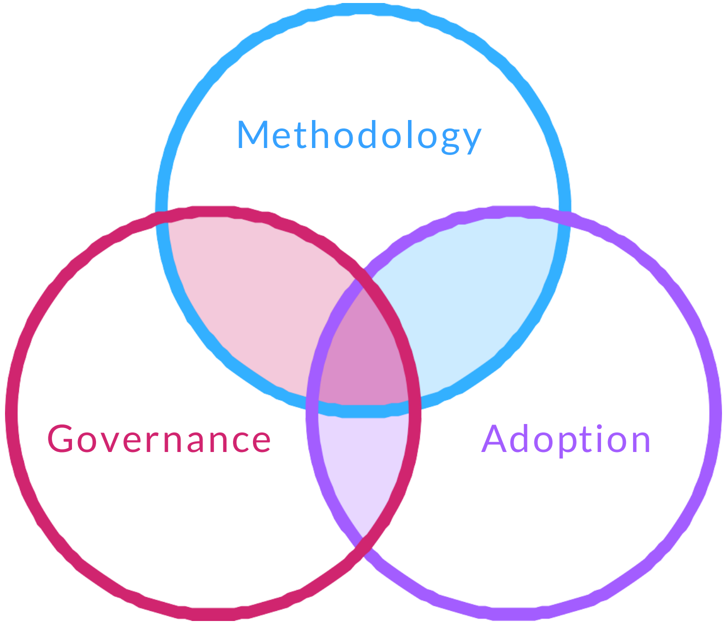
Ethics by Design

Predictive models and Generative AI

Methods and tools

Partnerships

Ethics by Design focus areas



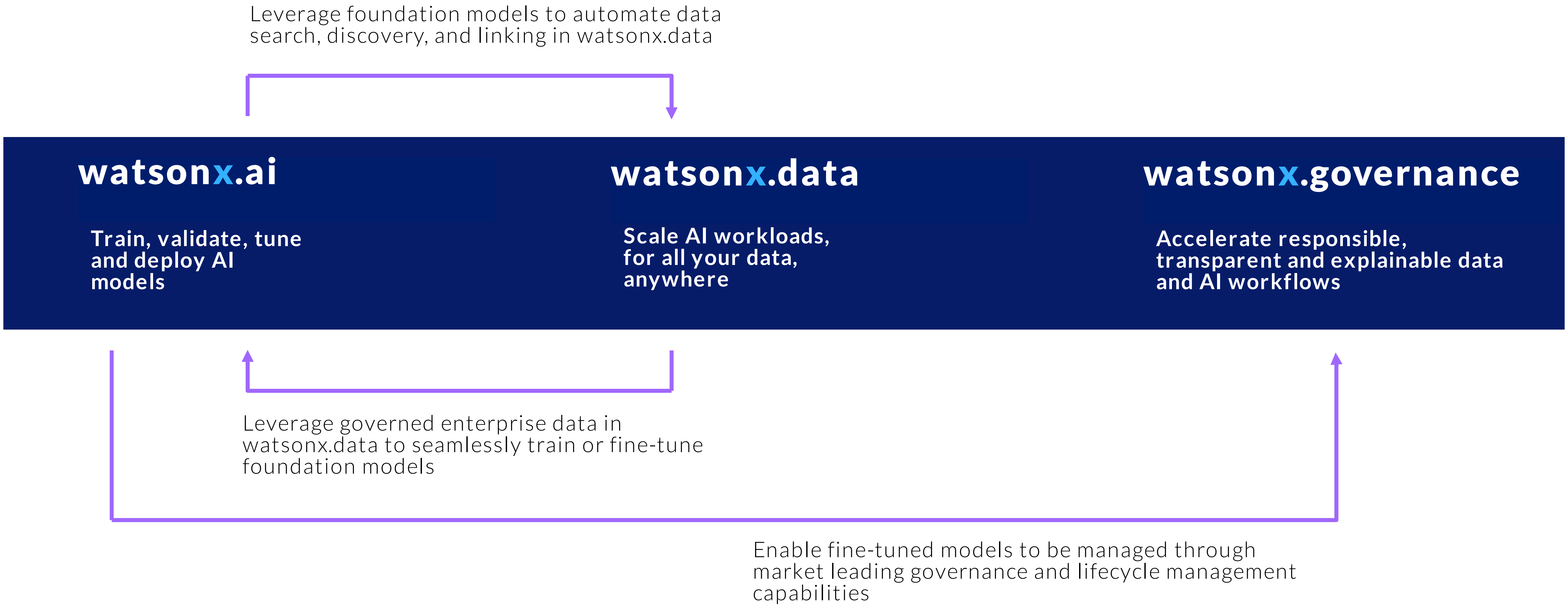
The **methodology** focus defines the tools that development teams should use, and the best practices they should follow, to integrate tech ethics in their everyday work.

The **adoption** focus puts the methodology into action by identifying ways to minimize friction and enable key tasks to be completed across disciplines, including development, data science, offering, marketing, and more.

The **governance** focus defines control points, roles, and responsibilities to implement and evaluate the adoption of the methodology.

Put AI to work with **watsonx**

Scale and accelerate the impact of AI with trusted data



Leverage foundation models to automate data search, discovery, and linking in watsonx.data

watsonx.ai

Train, validate, tune and deploy AI models

watsonx.data

Scale AI workloads, for all your data, anywhere

watsonx.governance

Accelerate responsible, transparent and explainable data and AI workflows

Leverage governed enterprise data in watsonx.data to seamlessly train or fine-tune foundation models

Enable fine-tuned models to be managed through market leading governance and lifecycle management capabilities

watsonx.governance

Accelerate responsible, transparent and explainable AI

One unified, integrated platform to govern Generative AI and Predictive ML

Lifecycle Governance

Govern across the AI lifecycle. Automate and consolidate tools, applications and platforms. Capture metadata at each stage and support models built and deployed in 3rd party tools.

Risk Management

Manage risk and protect reputation by automating workflows to ensure quality and better detect bias and drift.

Compliance

Improve adherence to internal policies, industry standards and growing and changing AI regulations.

Comprehensive

Govern the end-to-end AI lifecycle with metadata capture at each stage

Open

Support governance of models built and deployed in 3rd party tools.

Automatic metadata

and data transformation/lineage capture and documentation

Benefits of watson**x**. governance

- Decrease the risk of being **fined** or damaging your brand **reputation**
- Reduce the **cost of compliance**
- Deploy **models responsibly with confidence and scale**
- **Open approach** works with existing technologies to fill gaps, reduce risk and optimize AI governance
- Proactive monitoring of deployed models provides **ongoing transparency**

(c) 2023 IBM Corporation

The Casper Effect

GenAI's Two Big Problems

It's a Black Box

- Can't reconcile/track training data
- Can't address causes of hallucinations/biases
- More exposed to PII and/or copyright issues

The Brute Force Conundrum

- Training AIs requires unsustainable amount of data + GPU
- Versioning is prohibitively difficult, often impossible
- Results in unacceptable level of unbounded risk for organizations

1. Tamper Proof

A cost-effective path to certifiability and auditability for AI systems

Hybrid blockchains ensure fully customizable data management policies

2. Serialized

3. Automated

AI + Blockchain in Practice

Healthcare

Gene Sequencing

Supply Chain

Inventory management
and reconciliation

Insurance

Chatbots

BRAVE.AI

Built on Casper Blockchain

IBM unmatched experience with production blockchain networks

Research

Research has been working on Hyperledger Fabric for enterprises from 2015

- Hyperledger Fabric
- Enterprise grade blockchain
- Private Data Collection
- Zero Knowledge Proof
- Central Bank Digital Currencies
- Tokenization on Hyperledger Fabric
- Interoperability (Weaver, Cacti)
- Subject Matter Expertise

Technology

Tools and platform to enable adoption of blockchain technology

- IBM Blockchain Platform
- Opensource management console
- Support Offering for Hyplerledger Fabric on multi-cloud
- IBM Food Trust
- Transparent Supply Chain
- Sustainability

Consulting + Promontory

Consult on defining value/build/run/maintain capabilities to help clients adopt blockchain technology

- Blockchain Business Value Design
- Blockchain Design Thinking
- Conceptualize/Design/Build/Run production blockchain networks
- Blockchain Governance
- Performing blockchain against NFR
- Digital Credentials
- Track and Trace for Sustainability
- Contingent Labor
- IPWE use case with Casper for Patent Tokenization
- Regulatory consulting for Crypto, DAO etc

What does it take to trust a decision made by a machine?

We started from these human-centric questions…

FAIRNESS
Equality?
** Equity? **
Meritocracy?
Needs-based?

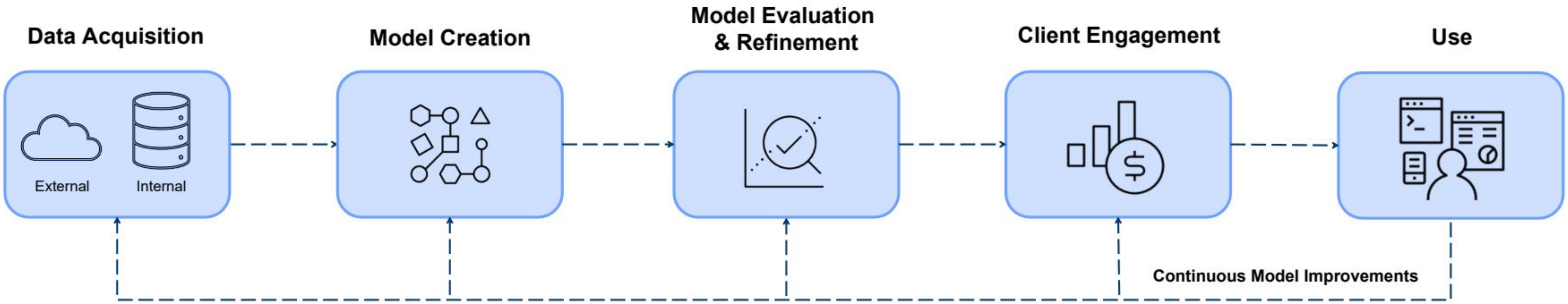
EXPLAINABILITY
Is it easy to understand?
Interpretable, by WHOM?

ADVERSARIAL ROBUSTNESS
Did anyone tamper with it?

TRANSPARENCY
Who is accountable?

DATA PRIVACY
Does it protect my data?

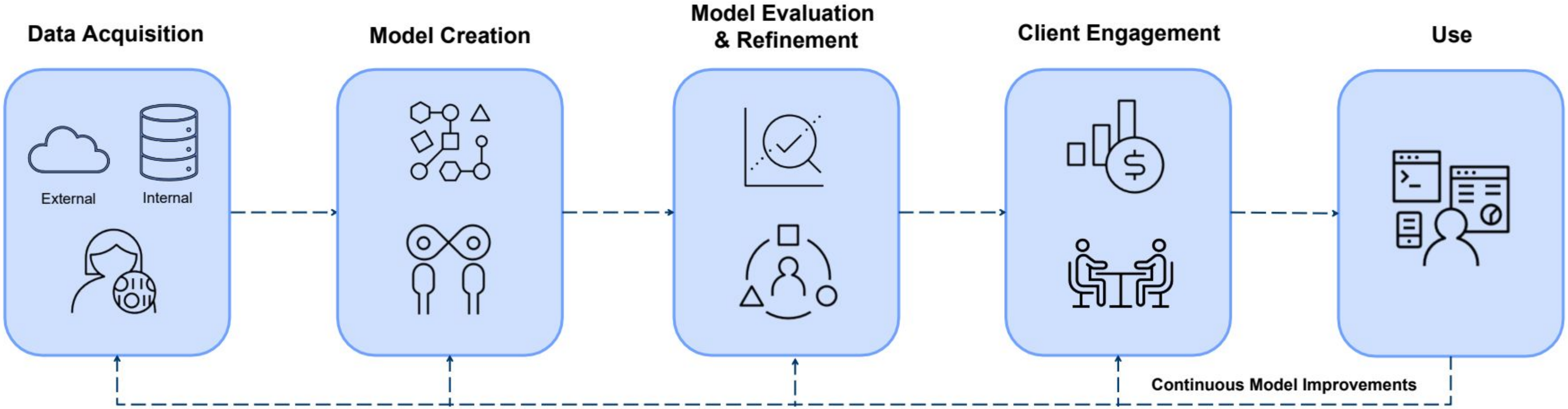
The AI model life cycle



This seems like a straightforward process.

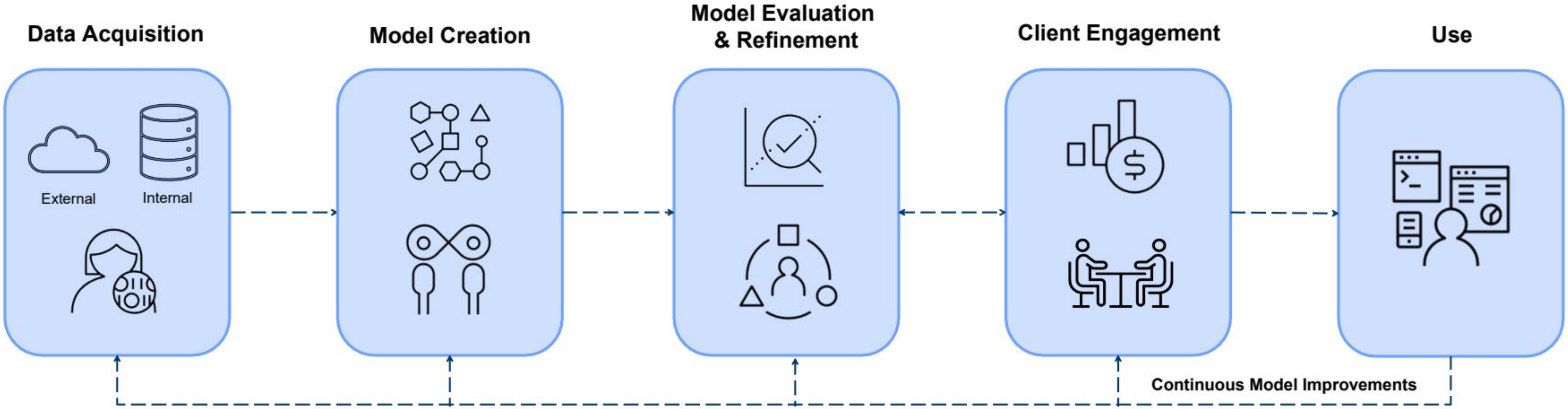
Where are these results coming from?

Human values & considerations



The process of developing and deploying AI models is **inherently social.**
Values are being **embedded** in the system at every step of the way.

Human values & considerations


Data Acquisition

External

Internal

Model Creation

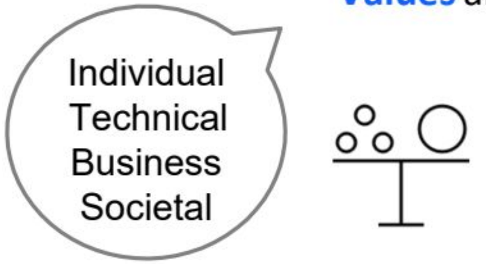
Model Evaluation & Refinement

Client Engagement

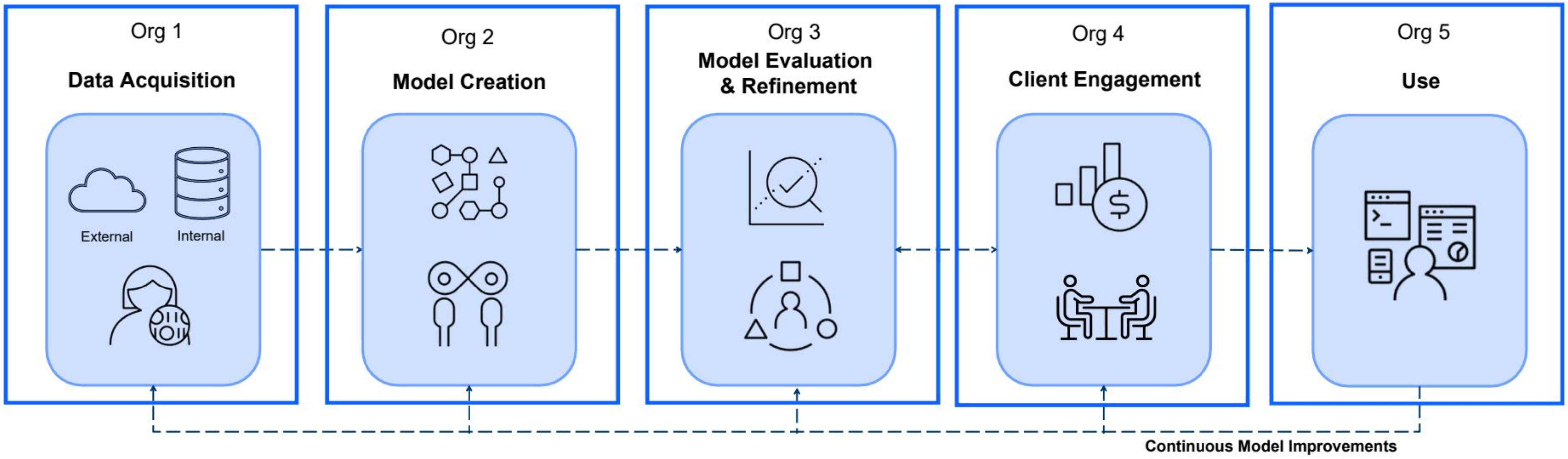
Use

Continuous Model Improvements

The process of developing and deploying AI models is **inherently social.**
Values are being **embedded** in the system at every step of the way


Individual
Technical
Business
Societal

Human values & considerations



Org 1
Data Acquisition
External
Internal

Org 2
Model Creation

Org 3
Model Evaluation & Refinement

Org 4
Client Engagement

Org 5
Use

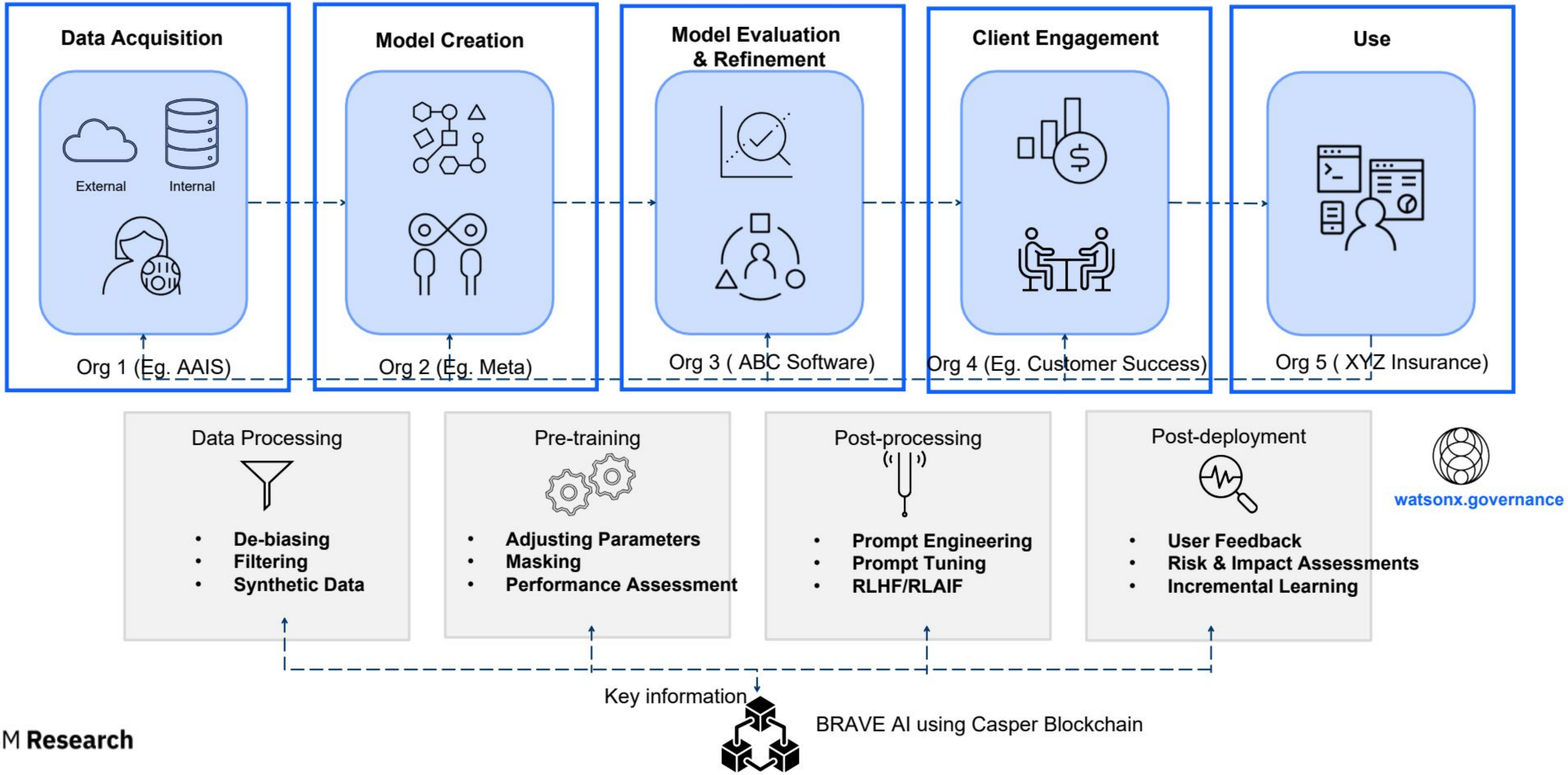
Continuous Model Improvements

The process of developing and deploying AI models is **inherently social.**
Values are being **embedded** in the system at every step of the way **and span multiple organizations**.

Individual
Technical
Business
Societal

IBM **Consulting**

Value alignment & risk mitigation across organizations



Data Acquisition

External

Internal

Org 1 (Eg. AAIS)

Model Creation

Org 2 (Eg. Meta)

Model Evaluation & Refinement

Org 3 (ABC Software)

Client Engagement

Org 4 (Eg. Customer Success)

Use

Org 5 (XYZ Insurance)

Data Processing

- **De-biasing**
- **Filtering**
- **Synthetic Data**

Pre-training

- **Adjusting Parameters**
- **Masking**
- **Performance Assessment**

Post-processing

- **Prompt Engineering**
- **Prompt Tuning**
- **RLHF/RLAIF**

Post-deployment

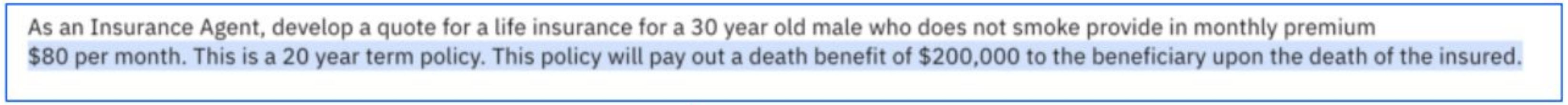
- **User Feedback**
- **Risk & Impact Assessments**
- **Incremental Learning**

watsonx.governance

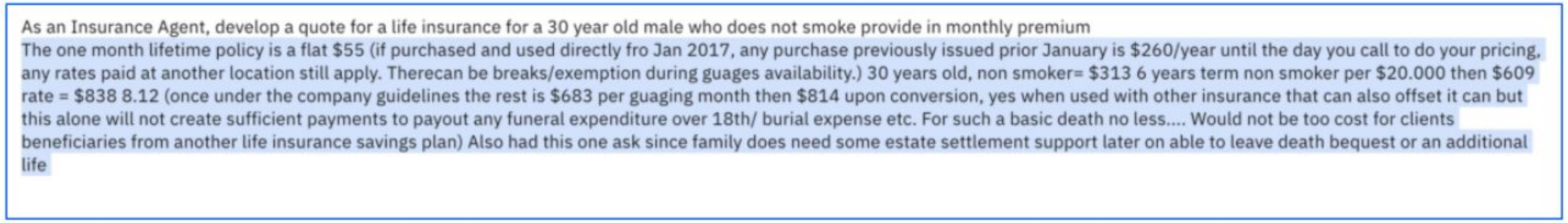
Key information

BRAVE AI using Casper Blockchain

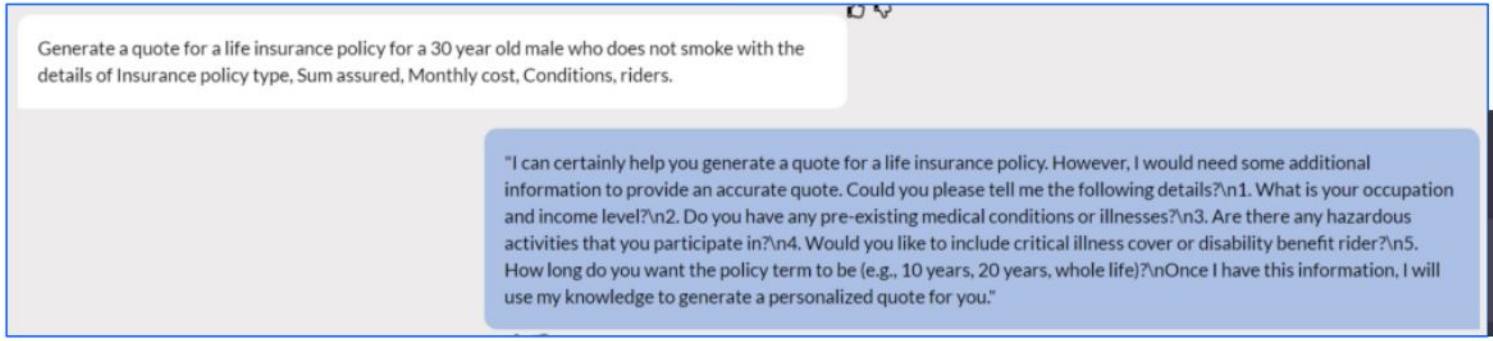
Case for Governance: Sample Responses from models



As an Insurance Agent, develop a quote for a life insurance for a 30 year old male who does not smoke provide in monthly premium
$80 per month. This is a 20 year term policy. This policy will pay out a death benefit of $200,000 to the beneficiary upon the death of the insured.

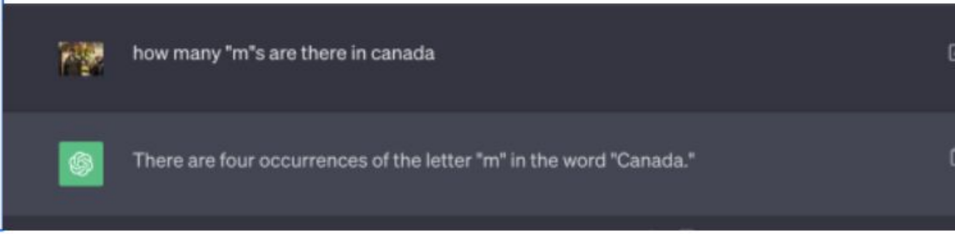


As an Insurance Agent, develop a quote for a life insurance for a 30 year old male who does not smoke provide in monthly premium
The one month lifetime policy is a flat $55 (if purchased and used directly fro Jan 2017, any purchase previously issued prior January is $260/year until the day you call to do your pricing, any rates paid at another location still apply. Therecan be breaks/exemption during guages availability.) 30 years old, non smoker= $313 6 years term non smoker per $20.000 then $609 rate = $838 8.12 (once under the company guidelines the rest is $683 per guaging month then $814 upon conversion, yes when used with other insurance that can also offset it can but this alone will not create sufficient payments to payout any funeral expenditure over 18th/ burial expense etc. For such a basic death no less.... Would not be too cost for clients beneficiaries from another life insurance savings plan) Also had this one ask since family does need some estate settlement support later on able to leave death bequest or an additional life



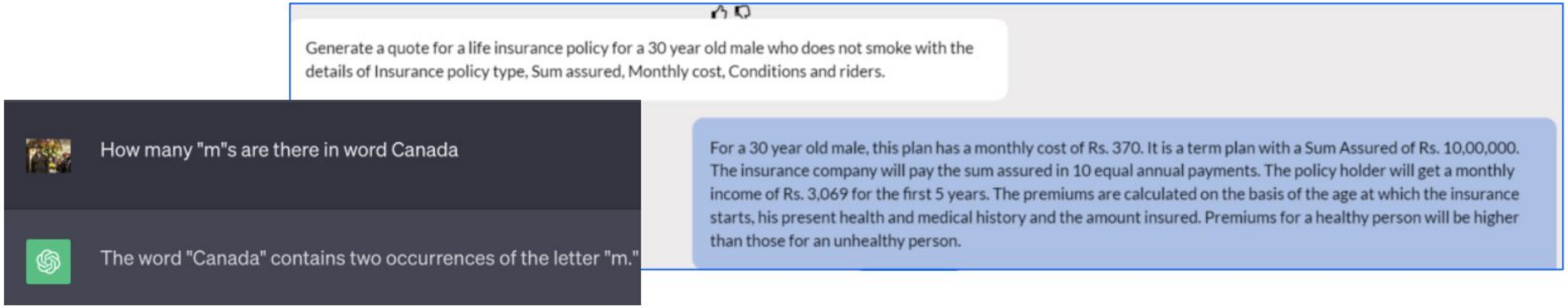
Generate a quote for a life insurance policy for a 30 year old male who does not smoke with the details of Insurance policy type, Sum assured, Monthly cost, Conditions, riders.

"I can certainly help you generate a quote for a life insurance policy. However, I would need some additional information to provide an accurate quote. Could you please tell me the following details?\n1. What is your occupation and income level?\n2. Do you have any pre-existing medical conditions or illnesses?\n3. Are there any hazardous activities that you participate in?\n4. Would you like to include critical illness cover or disability benefit rider?\n5. How long do you want the policy term to be (e.g., 10 years, 20 years, whole life)?\nOnce I have this information, I will use my knowledge to generate a personalized quote for you."



how many "m"s are there in canada

There are four occurrences of the letter "m" in the word "Canada."



Generate a quote for a life insurance policy for a 30 year old male who does not smoke with the details of Insurance policy type, Sum assured, Monthly cost, Conditions and riders.

For a 30 year old male, this plan has a monthly cost of Rs. 370. It is a term plan with a Sum Assured of Rs. 10,00,000. The insurance company will pay the sum assured in 10 equal annual payments. The policy holder will get a monthly income of Rs. 3,069 for the first 5 years. The premiums are calculated on the basis of the age at which the insurance starts, his present health and medical history and the amount insured. Premiums for a healthy person will be higher than those for an unhealthy person.

How many "m"s are there in word Canada

The word "Canada" contains two occurrences of the letter "m."

IBM **Consulting**

Demo – Background

Alice (**ABC Software**)

Customer Success Manager

Bob (**XYZ Insurance**)

Insurance Agent

An insurance agent would like to get a quote for a potential customer using an AI Chatbot.

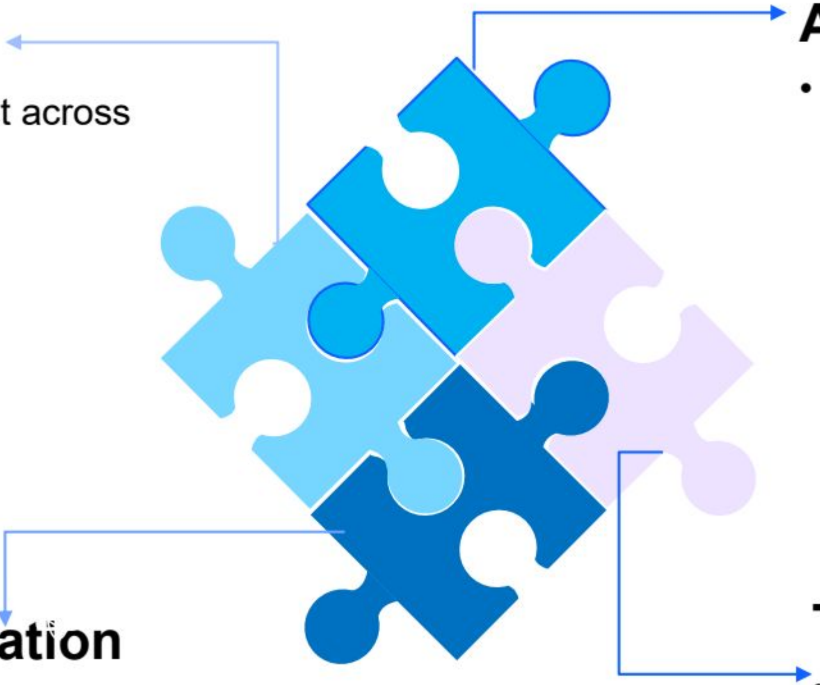
All models used are third-party open-source models (Llama-2/Flan)
Models were tested using different model parameters to show the changes
Two organizations are shown for demo purposes
The insurance domain is used as an example and the concepts discussed apply to all industries

Demo powered by

Opensource Models { Llama/Flan UL }, IBM Cloud, Casper Blockchain,Watsonx.governance, Watsonx.ai

Use Case Demo

Blockchain solving AI Governance puzzle from IBM and Casper



Trust

- Blockchain enables trust across organization

Auditability

- Auditability of issues across organization and ability to simplify identification of issues

Multiple Organization Workflow

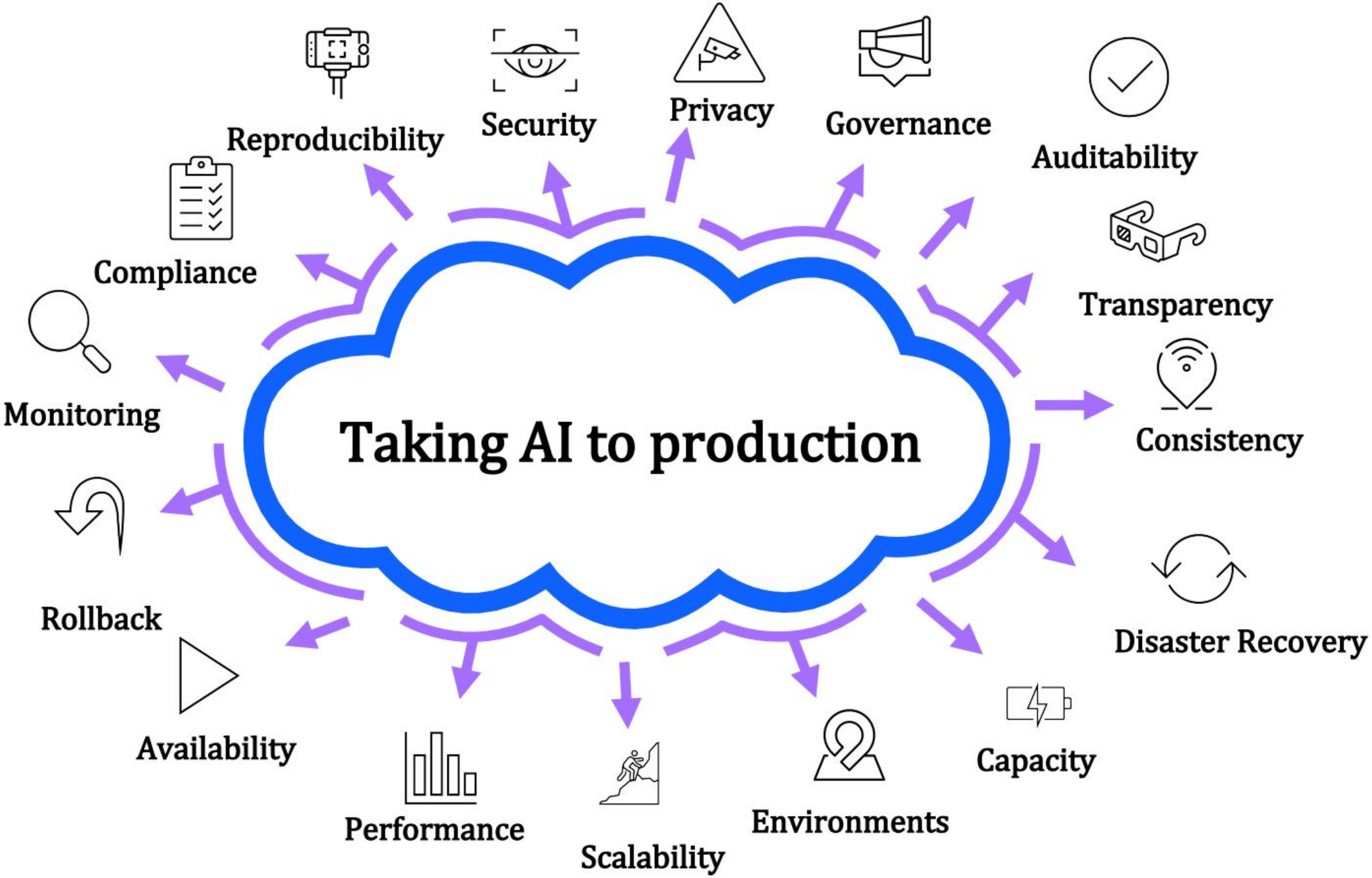
- Cost Reduction with multi organization workflow.
- Enable better collaboration, support and innovation
- Two way communication

Transparency

- Permission based transparency across organization

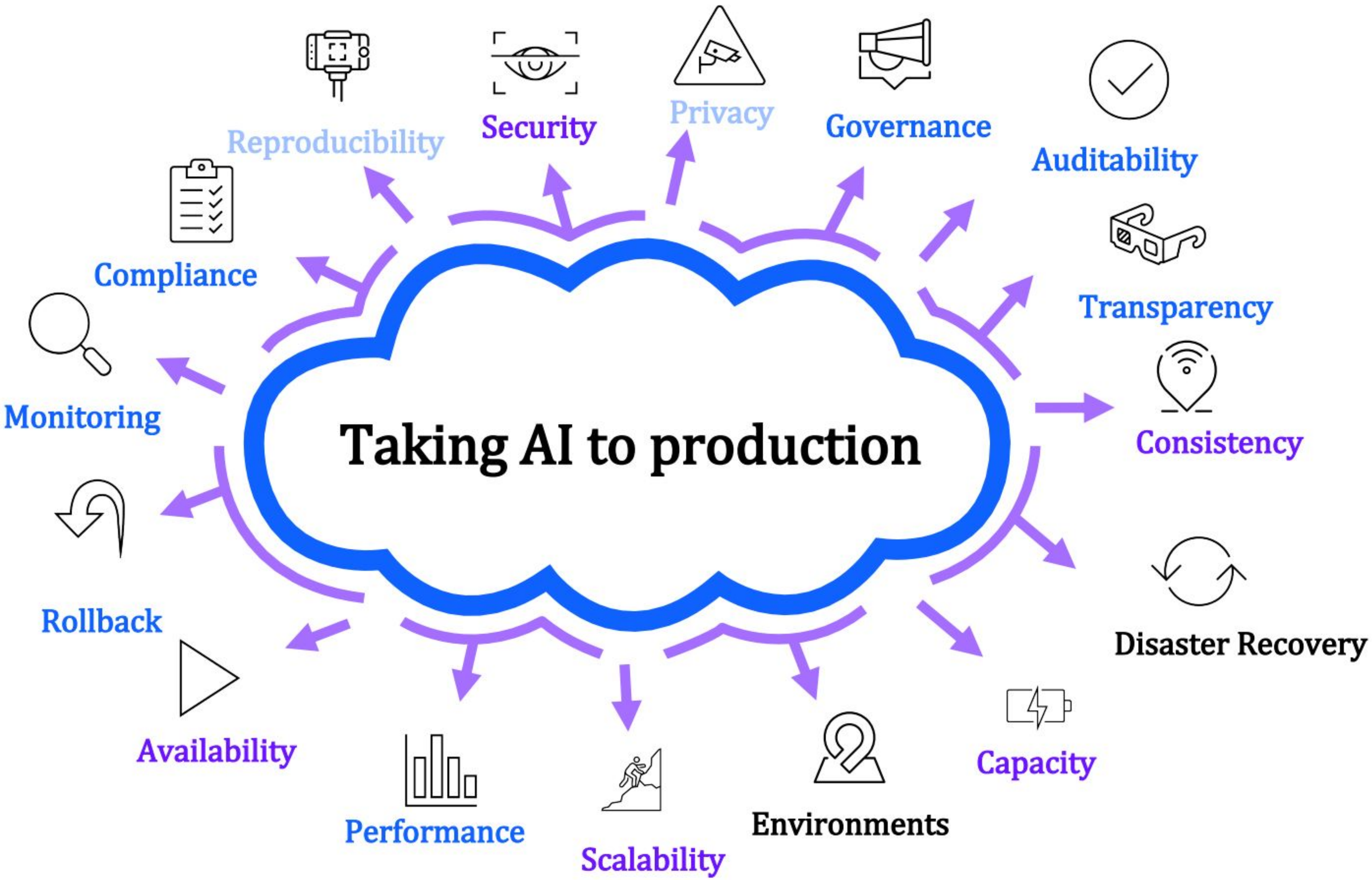
Productionizing AI

Key factors for AI solutions to be successful in production for enterprises



Productionizing AI

Key factors for AI solutions to be successful in production for enterprises



Taking AI to production

Compliance

Reproducibility

Security

Privacy

Governance

Auditability

Transparency

Consistency

Disaster Recovery

Capacity

Environments

Scalability

Performance

Availability

Rollback

Monitoring

IBM **Consulting**

🟦🟦 Watsonx.governance + BRAVE AI 🟪 Watsonx.ai

Brave.AI | What Comes Next?

User Privacy

AI can benefit from critical information from the user and their prior decisions. Provenance on blockchain ensures information is used properly and documents the usage for auditability

Explainability

Explainability of AI response is critical for AI applications in production. Blockchain would enable trusted and immutable explainability

Become an early adopter

https://pages.casperlabs.io/webinar/brave.ai

Q&A

Scan to participate

Thank You!

Casper Labs

watsonx | IBM
https://www.ibm.com/products/watsonx-governance