



Emerging trends and considerations on how to govern AI

Owen Larter
Director of Responsible AI Public Policy
Microsoft

A brief history of AI

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI

1950s

Artificial Intelligence

the field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

1959

Machine Learning

subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions.

2017

Deep Learning

a machine learning technique in which layers of neural networks are used to process data and make decisions.

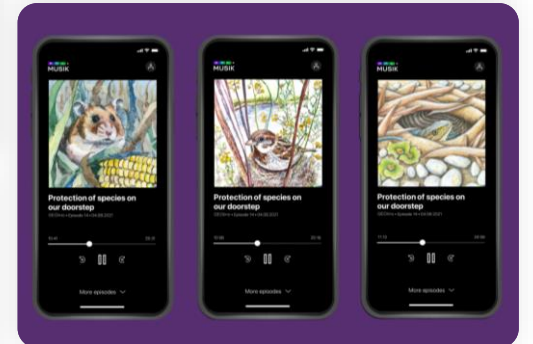
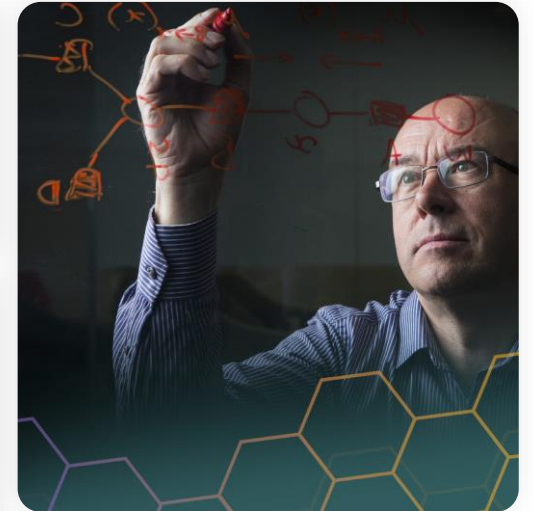
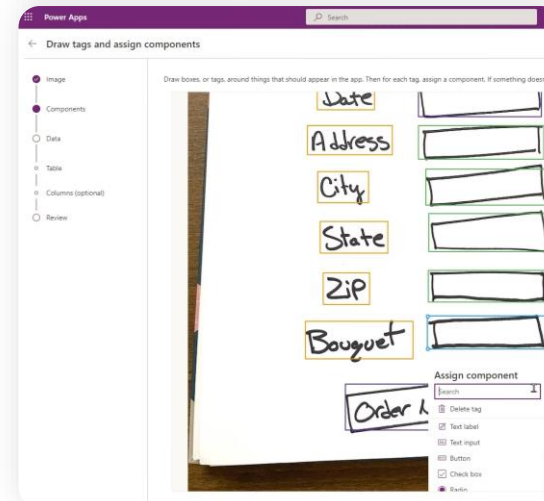
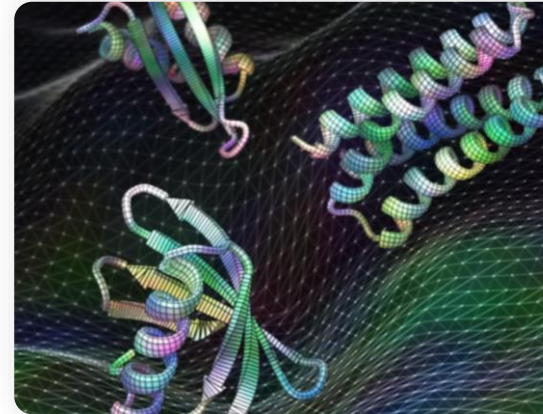
2021

Generative AI

create new written, visual, and auditory content given prompts or existing data.

What can these new **AI services** do?

- ✓ **Natural language chatbots:** everything from customer service to personalized tutoring to training for suicide prevention counselors.
- ✓ **Advanced search capabilities:** searching not just for links, but for answers.
- ✓ **Generate text:** emails, draft legislation, contracts, screenplays.
- ✓ **Real time** translation, transcription, analysis, and summaries of content (audio, video, text).
- ✓ **Image** recognition and generation.



Multimodal models

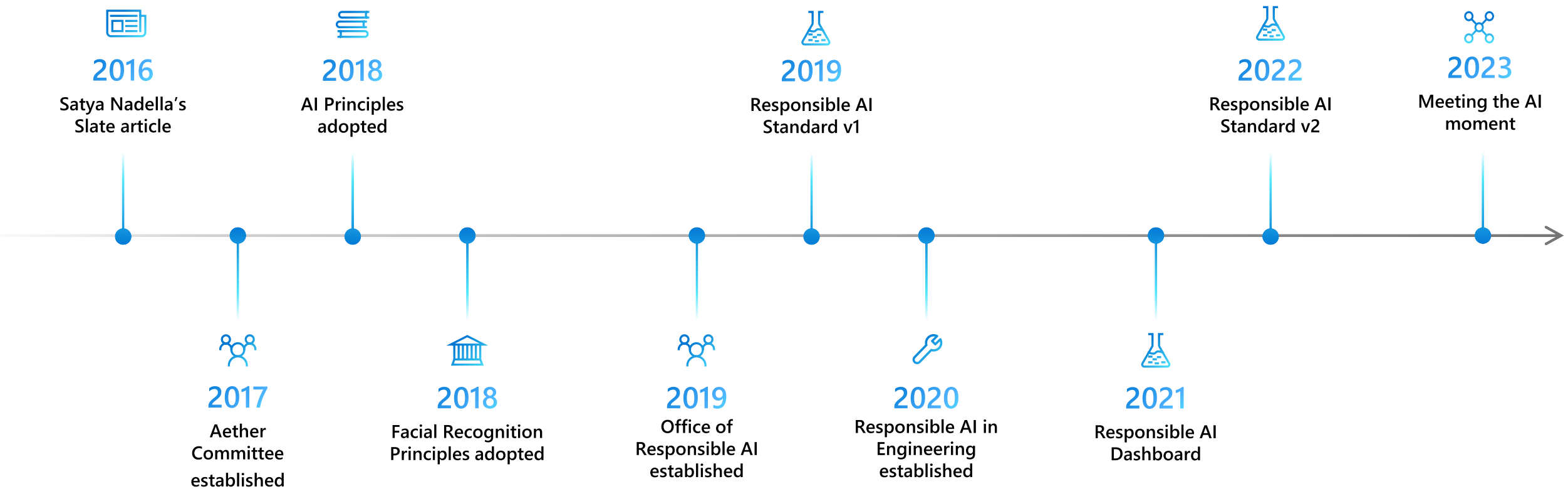


AI systems that learn about concepts in several modalities, such as text, images, and videos, in order to better understand the world

GPT-4 is a large multimodal model: it accepts image and text inputs and emits text outputs

Lessons from our experience on AI governance

Our responsible AI journey



Microsoft's AI Principles



Fairness



Reliability
& Safety



Privacy &
Security



Inclusiveness



Transparency



Accountability

Rules:

Responsible AI Standard



Records

our practice of **Responsible AI by Design** – the proactive ways in which we guide the design, build, and testing of AI systems.



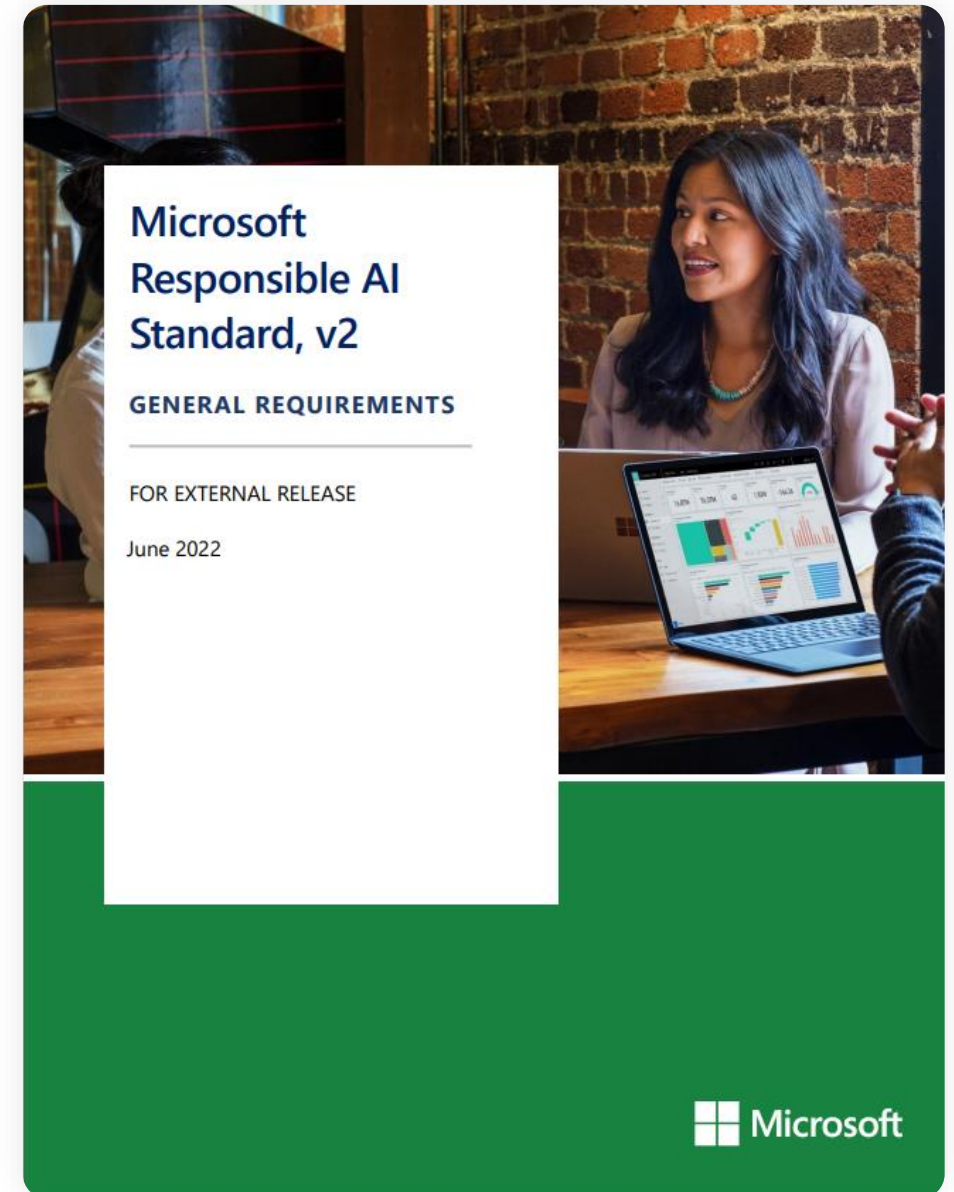
Establishes

a **durable framework** for the maturing practice of responsible AI and evolving regulatory requirements.



Reflects

our **deeper exploration** of what our six AI principles mean and the steps we must taken to uphold them.



Sensitive Uses:

A rule-making and oversight process



**Consequential
Impact**

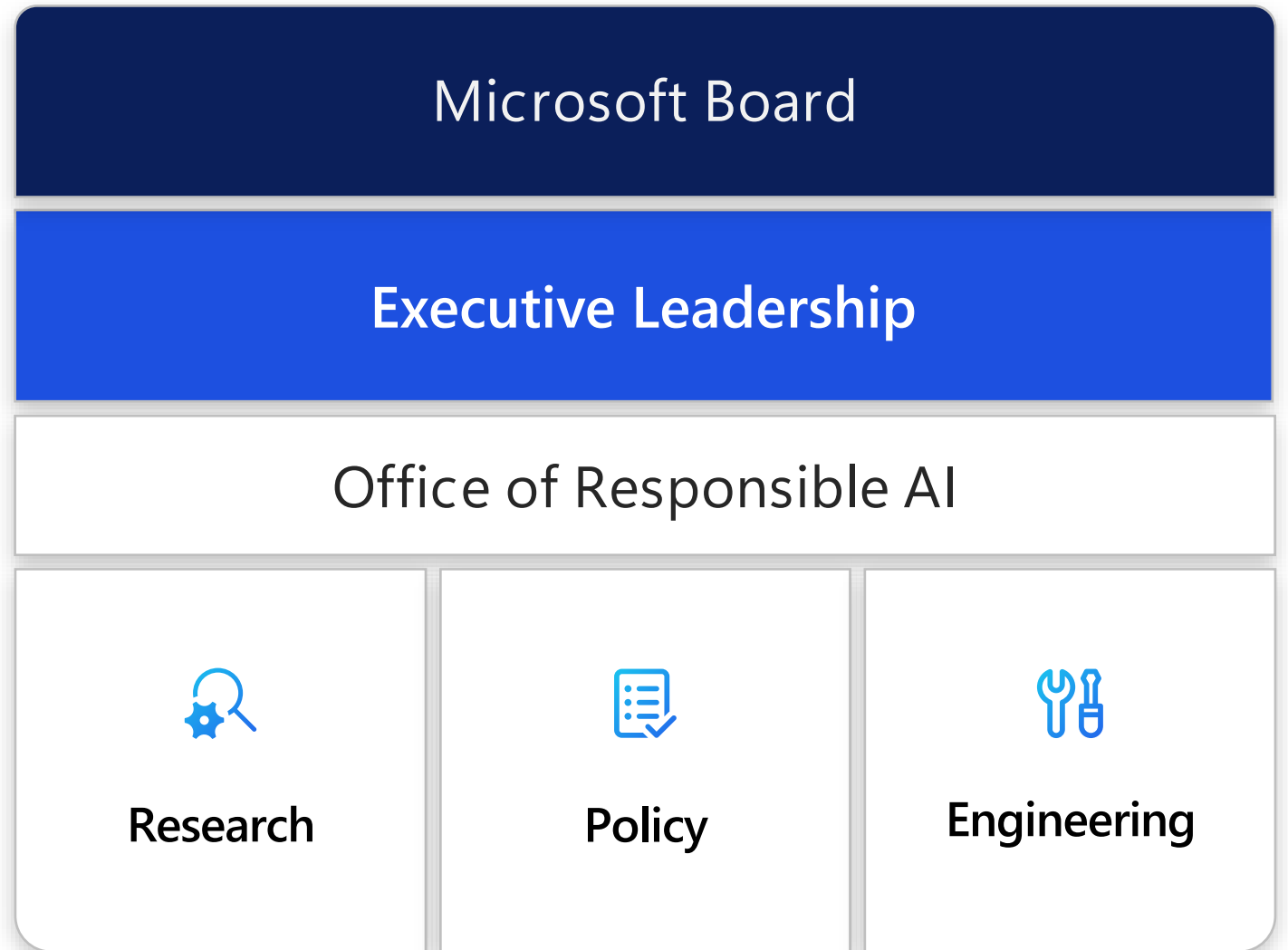


**Physical or
Psychological Injury**



**Threat to
Human Rights**

Our ecosystem



Governing AI: A Blueprint for the future

Governing AI: A Blueprint for the Future

May 25, 2023

Foreword: How Do We Best Govern AI?



Brad Smith, Vice Chair
and President, Microsoft

“Don’t ask what computers can do, ask what they should do.”

That is the title of the chapter on AI and ethics in a book I coauthored in 2019. At the time, we wrote that “this may be one of the defining questions of our generation.” Four years later, the question has seized center stage not just in the world’s capitals, but around many dinner tables.

As people have used or heard about the power of OpenAI’s GPT-4 foundation model, they have often been surprised or even astounded. Many have been enthused or even excited. Some have been concerned or even frightened. What has become clear to almost everyone is something we noted four years ago—we are the first generation in the history of humanity to create machines that can make decisions that previously could only be made by people.

Countries around the world are asking common questions. How can we use this new technology to solve our problems? How do we avoid or manage new problems it might create? How do we control technology that is so powerful?

These questions call not only for broad and thoughtful conversation, but decisive and effective action. This paper offers some of our ideas and suggestions as a company.

These suggestions build on the lessons we’ve been learning based on the work we’ve been doing for several years. Microsoft CEO Satya Nadella set us on a clear course when he [wrote in 2016](#) that “perhaps the most productive debate we can have isn’t one of good versus evil: The debate should be about the values instilled in the people and institutions creating this technology.”

Since that time, we’ve defined, published, and implemented ethical principles to guide our work. And we’ve built out constantly improving engineering and governance systems

to put these principles into practice. Today we have nearly 350 people working on responsible AI at Microsoft, helping us implement best practices for building safe, secure, and transparent AI systems designed to benefit society.

New opportunities to improve the human condition

The resulting advances in our approach have given us the capability and confidence to see ever-expanding ways for AI to improve people’s lives. We’ve seen AI help save individuals’ eyesight, make progress on new cures for cancer, generate new insights about proteins, and provide predictions to protect people from hazardous weather. Other innovations are fending off cyberattacks and helping to protect fundamental human rights, even in nations afflicted by foreign invasion or civil war.

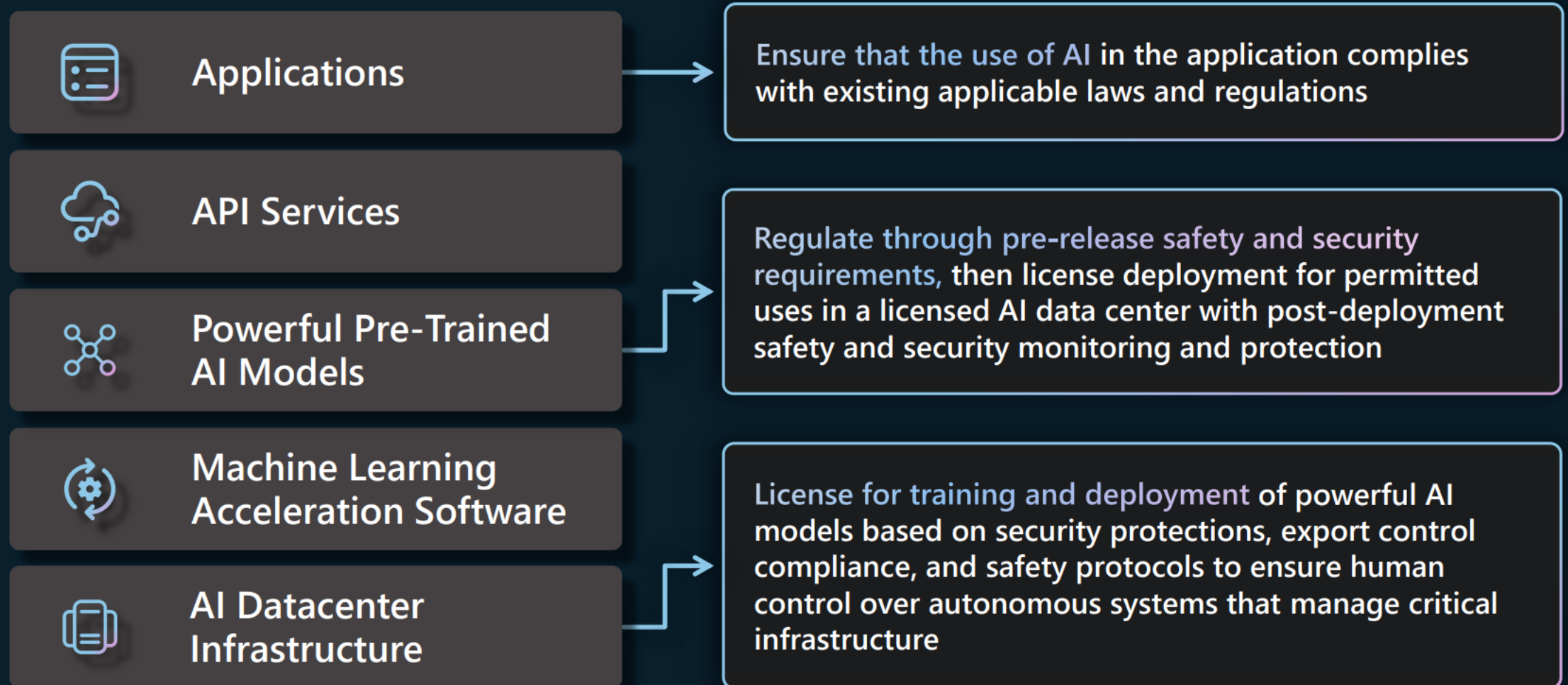
Everyday activities will benefit as well. By acting as a copilot in people’s lives, the power of foundation models like GPT-4 is turning search into a more powerful tool for research and improving productivity for people at work. And for any parent who has struggled to remember how to help their 13-year-old child through an algebra homework assignment, AI-based assistance is a helpful tutor.

In so many ways, AI offers perhaps even more potential for the good of humanity than any invention that has preceded it. Since the invention of the printing press with movable type in the 1400s, human prosperity has been growing at an accelerating rate. Inventions like the steam engine, electricity, the automobile, the airplane, computing, and the internet have provided many of the building blocks for modern civilization. And like the printing press itself, AI offers a new tool to genuinely help advance human learning and thought.

The technology stack for AI foundation models

	Applications	Software programs where the output of an AI model is put to work
	API Services	APIs (Application Program Interfaces), or endpoints, through which applications access pre-trained models
	Powerful Pre-Trained AI Models	Pre-trained models like GPT-4 that can be used to solve similar problems without starting from scratch
	Machine Learning Acceleration Software	Software that speeds up the process of developing and deploying large AI models
	AI Datacenter Infrastructure	Advanced supercomputing infrastructure, including clusters of advanced GPUs (Graphics Processing Units) with high bandwidth network connections

A proposed AI regulatory architecture



A five-point blueprint for governing AI

1

Implement and build upon new government-led AI safety frameworks

2

Require effective safety brakes for AI systems that control critical infrastructure

3

Develop a broader legal and regulatory framework based on the technology architecture for AI

4

Promote transparency and ensure academic and public access to AI

5

Pursue new public-private partnerships to use AI as an effective tool to address the inevitable societal challenges that come with new technology