HUMANS IN CHARGE –

Steering the AI Age Responsibly

02/10/2023

NMHH
National Media and Infocommunications Authority • Hungary

NMHHSYMPOSIUM.COM

Adrián González Sánchez Microsoft

Public-Private Collaboration Approaches for Responsible AI Knowledge Transfer

Microsoft RAI, Modern AI, and the role of AI Sandboxes

02/10/2023

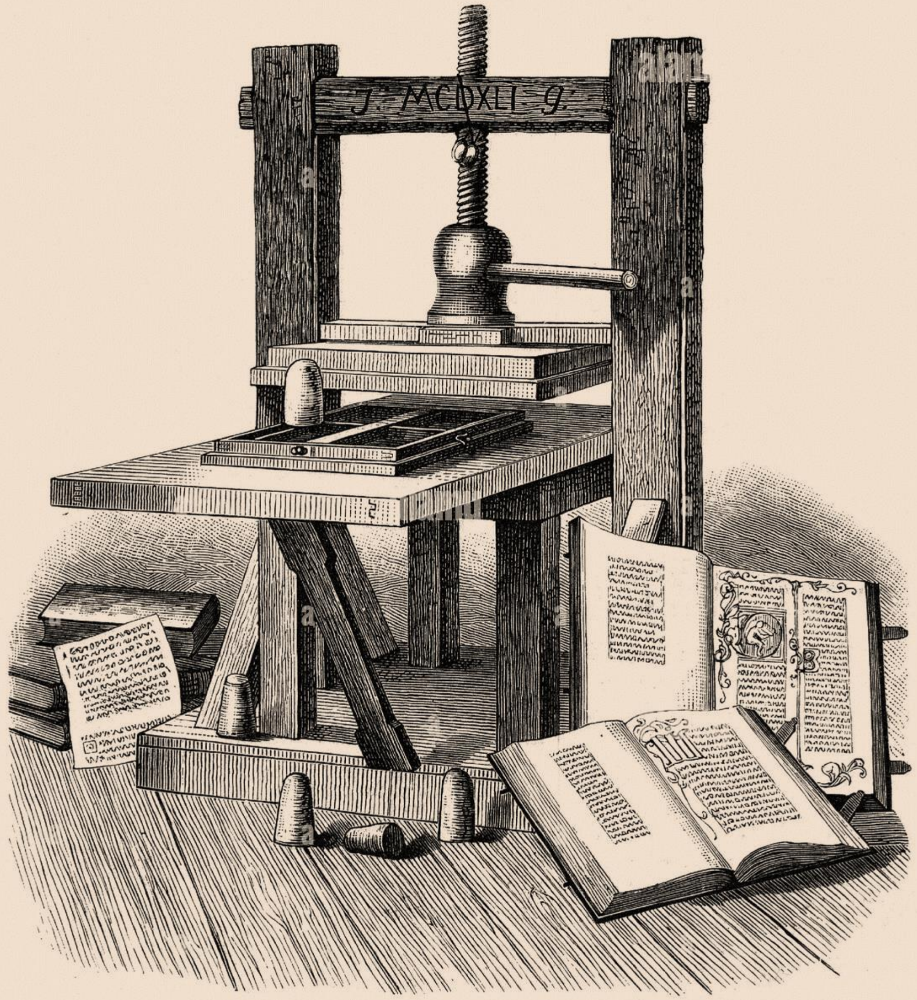
National Media and Infocommunications Authority • Hungary

NMHHSYMPOSIUM.COM

Microsoft

HUMANS
IN CHARGE –

The New Era of AI

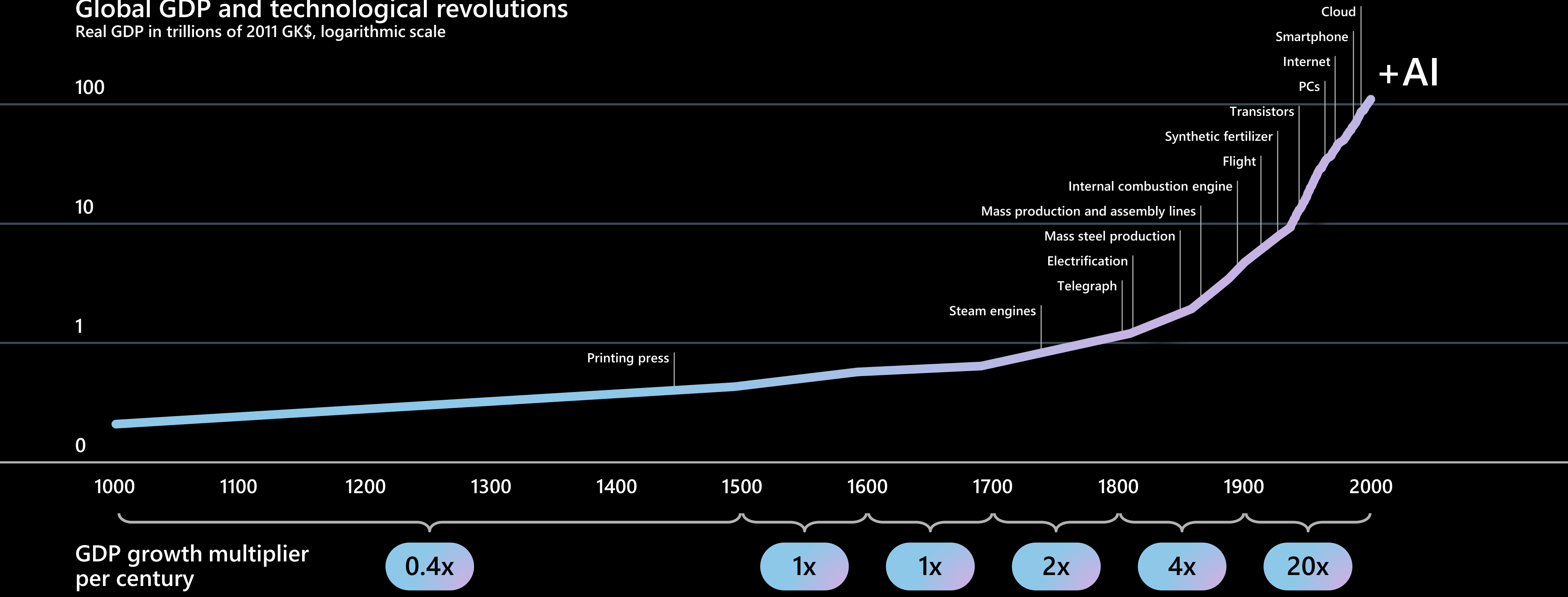


Die Gutenberg-Preſſe im Deutſchen Bibliographiſchen Muſeum zu Dresden.

Technology drives GDP growth, and the pace of change is accelerating

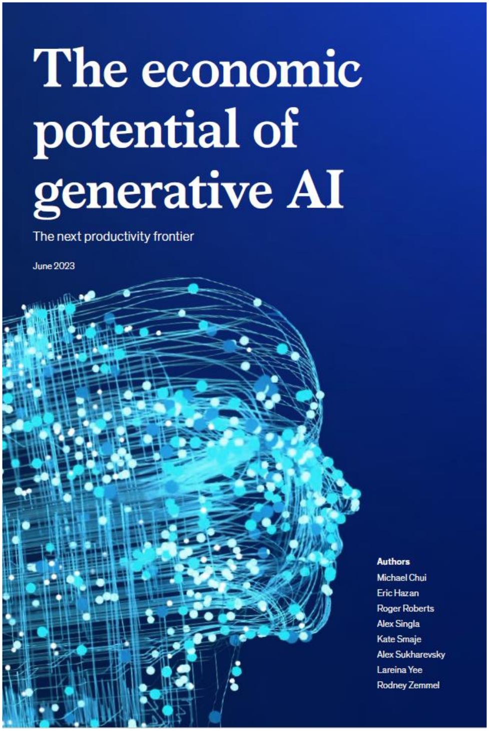
Global GDP and technological revolutions

Real GDP in trillions of 2011 GK$, logarithmic scale



Source: Maddison Project, Ourworldindata

McKinsey
& Company



The economic potential of generative AI

The next productivity frontier

June 2023

Authors
Michael Chui
Eric Hazan
Roger Roberts
Alex Singla
Kate Smaje
Alex Sukharevsky
Lareina Yee
Rodney Zemmel

- Generative AI's impact on productivity could add up to **$4.4 trillion annually** in value to the global economy.

- About 75% of the value that generative AI use cases could deliver falls across **customer operations, marketing, sales, software engineering**, and **R&D** .

- Generative AI can revolutionize work by automating a significant portion of employees' activities, **up to 60-70% of their current workload**. This augmentation of individual capabilities has the potential to transform the way we work.

- Generative AI is expected to have **$ 250-410 billions economic impact on banking and insurance**, accounting for 1-1.6% of total industry revenue.

Cloud Computing

❯

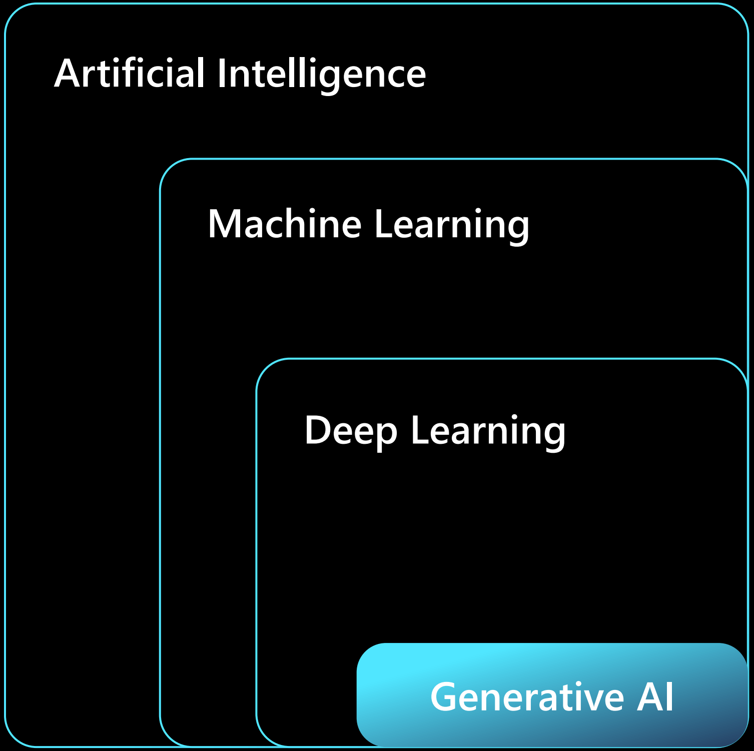
Digital Transformation

❯

AI Transformation

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.

Microsoft

2016

Jan
Object recognition
human parity

Oct
Speech recognition
human parity

2017

Dec
Reading Comprehension
human parity

2018

Mar
Machine translation
human parity

Sept
Speech Synthesis
human parity

2019

Jun
General Language Understanding
human parity

2020

Oct
Image caption
human parity

2021

Nov
Azure OpenAI
announcement

2022

Nov
ChatGPT
announcement

2023

Mar
GPT-4
announcement

- Uniform Bar Exam
- LSAT
- SAT Reading & Writing
- SAT Math
- GRE Quantitative
- GRE Writing
- GRE Verbal
- AP Art History
- AP Environmental Science
- AP Psychology
- AP US History
- AP World History
- AP Biology
- AP Microeconomics
- AP Macroeconomics
- AP Statistics
- AP Chemistry
- AP US Government
- AP Physics 2
- Leetcode hiring test

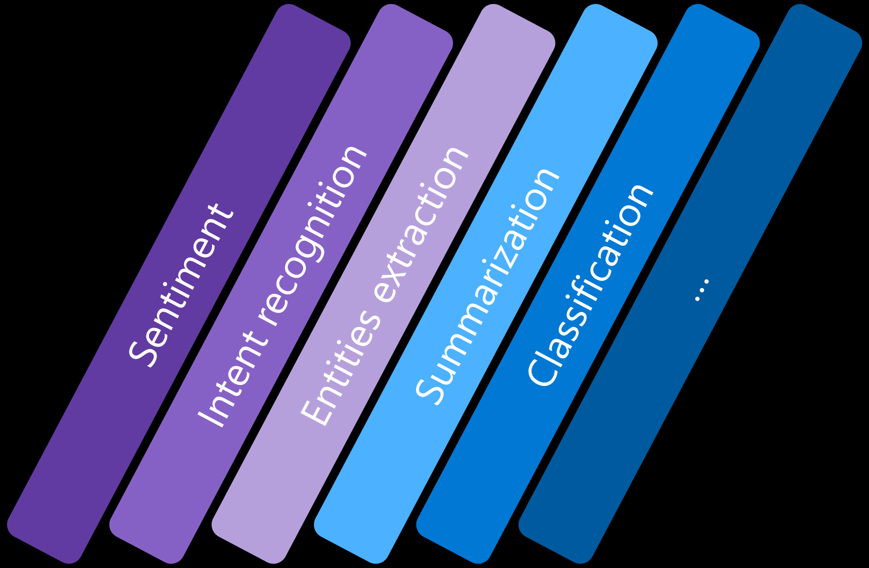
Source: “GPT-4 technical Report” (OpenAI), “Sparks of Artificial General Intelligence” (Microsoft Research)

Why is Generative AI different?

Microsoft

Traditional Models

(one model per capability)

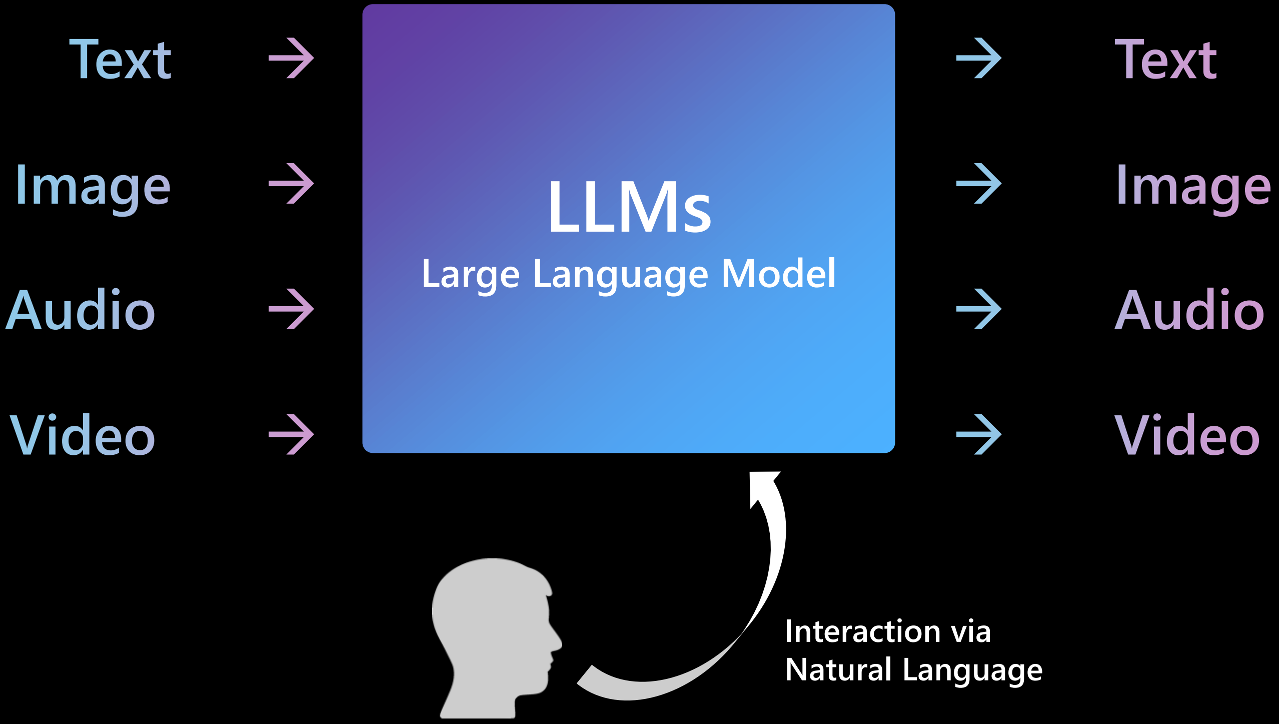

Sentiment
Intent recognition
Entities extraction
Summarization
Classification
…

New Foundation Models

(one single model for everything)

Sentiment
Intent recognition
Entities extraction
Summarization
Questions & Answers
Style transfer
Rewriting
Code generation
…

Why is Generative AI different?



Text → → Text

Image → → Image

Audio → → Audio

Video → → Video

LLMs

Large Language Model

Interaction via
Natural Language

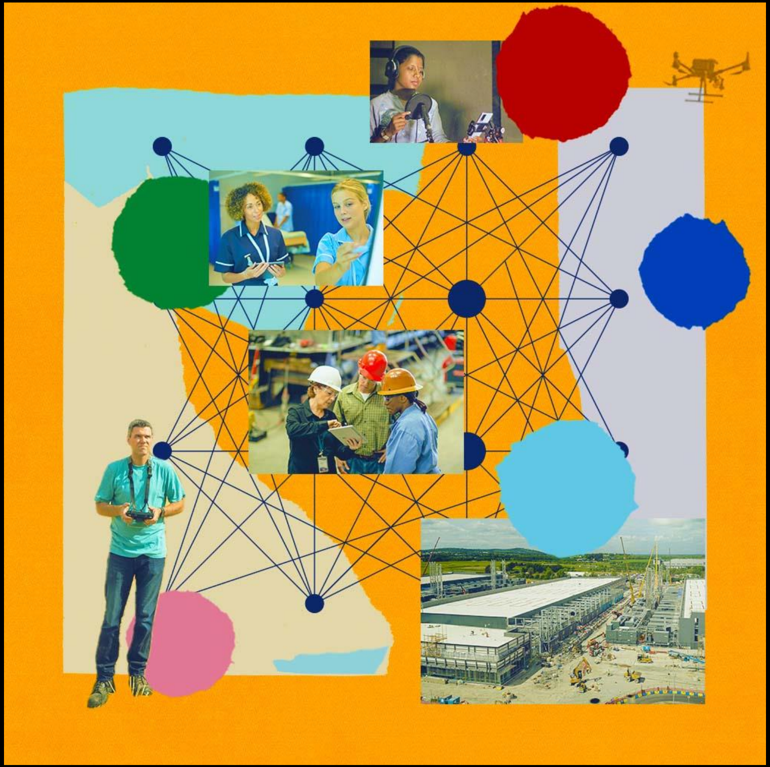
Microsoft

HUMANS
IN CHARGE –

The Microsoft Approach

Microsoft’s Approach to AI

Microsoft



> We have made huge investments in AI because we’re **optimistic about what it can do to help people, industry and society**, and because we’re committed to **bringing technology and people together to realize the promises of AI responsibly**.

Microsoft

Our Responsible AI journey

2016
Satya Nadella's Slate article

2017
Aether Committee established

2018
AI Principles adopted

2018
Facial Recognition Principles adopted

2019
Office of Responsible AI established

2019
Responsible AI Standard v1

2020
Responsible AI Strategy in Engineering

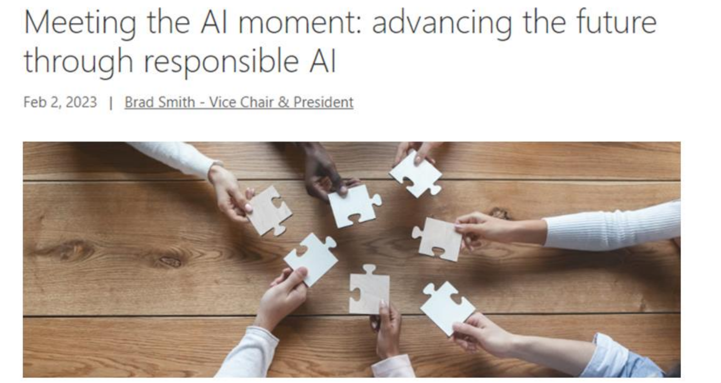
2021
Responsible AI Dashboard

2022
Responsible AI Standard v2

2023
Meeting the AI moment

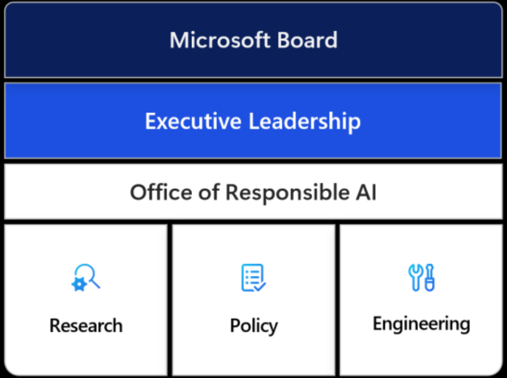
Microsoft's Approach to AI

- AI is built and used responsibly and ethically
- AI advances international competitiveness and national security
- AI serves society broadly, not narrowly



Meeting the AI moment: advancing the future through responsible AI

Feb 2, 2023 | Brad Smith - Vice Chair & President



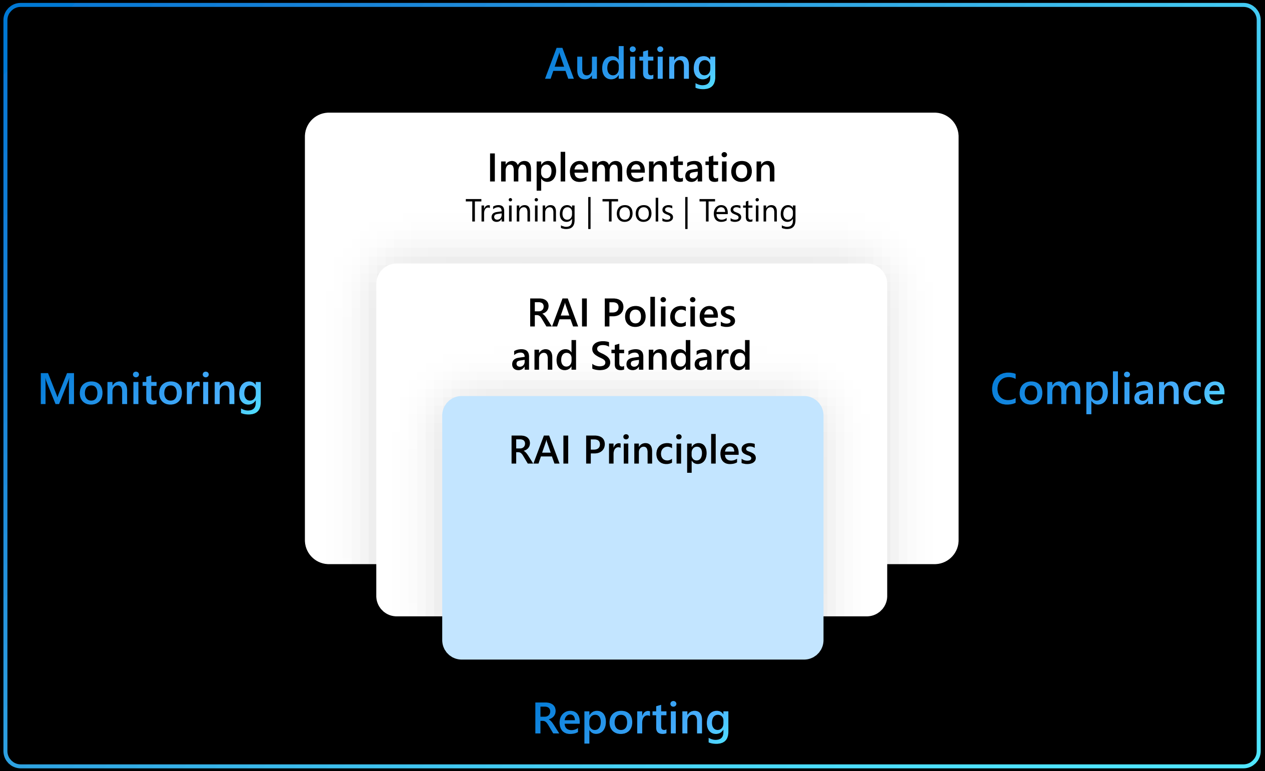
Microsoft Board

Executive Leadership

Office of Responsible AI

Research | Policy | Engineering

Responsible AI Governance Framework



Auditing

Implementation

Training | Tools | Testing

RAI Policies and Standard

RAI Principles

Monitoring

Compliance

Reporting

Microsoft’s AI Principles

Fairness

Reliability & Safety

Privacy & Security

Inclusiveness

Transparency

Accountability

Tools: Responsible AI Standard & Others

Records

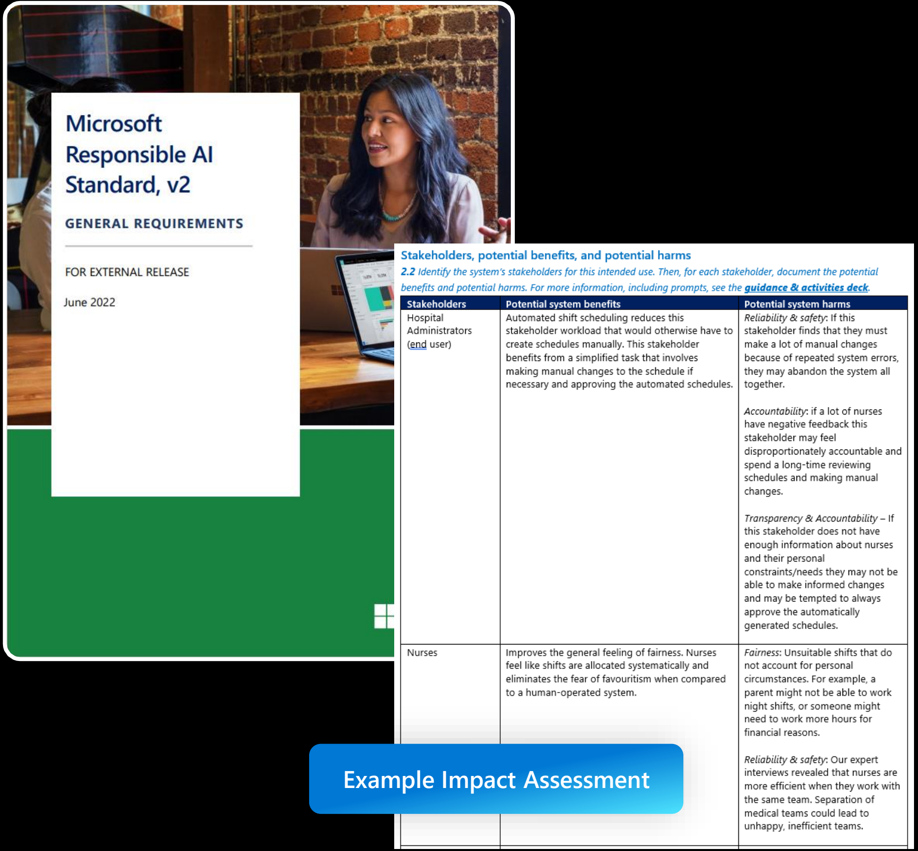
our practice of **Responsible AI by Design** – the proactive ways in which we guide the design, build, and testing of AI systems.

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.



Microsoft
Responsible AI
Standard, v2

GENERAL REQUIREMENTS

FOR EXTERNAL RELEASE

June 2022

Stakeholders, potential benefits, and potential harms

2.2 *Identify the system's stakeholders for this intended use. Then, for each stakeholder, document the potential benefits and potential harms. For more information, including prompts, see the* ***guidance & activities deck****.*

Stakeholders	Potential system benefits	Potential system harms
Hospital Administrators (end user)	Automated shift scheduling reduces this stakeholder workload that would otherwise have to create schedules manually. This stakeholder benefits from a simplified task that involves making manual changes to the schedule if necessary and approving the automated schedules.	*Reliability & safety*: If this stakeholder finds that they must make a lot of manual changes because of repeated system errors, they may abandon the system all together. *Accountability*: if a lot of nurses have negative feedback this stakeholder may feel disproportionately accountable and spend a long-time reviewing schedules and making manual changes. *Transparency & Accountability* – If this stakeholder does not have enough information about nurses and their personal constraints/needs they may not be able to make informed changes and may be tempted to always approve the automatically generated schedules.
Nurses	Improves the general feeling of fairness. Nurses feel like shifts are allocated systematically and eliminates the fear of favouritism when compared to a human-operated system.	*Fairness*: Unsuitable shifts that do not account for personal circumstances. For example, a parent might not be able to work night shifts, or someone might need to work more hours for financial reasons. *Reliability & safety*: Our expert interviews revealed that nurses are more efficient when they work with the same team. Separation of medical teams could lead to unhappy, inefficient teams.

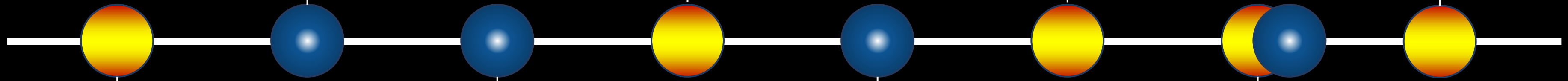
Example Impact Assessment

Microsoft HUMANS IN CHARGE –

The Opportunity

3

The EU-Spain Case



Research, Development and Innovation (RDI) Strategy in AI

Spanish Government
(March 2019)

Ethics guidelines for trustworthy AI

European Commission
(April 2019)
* Updated Feb 2020

Whitepaper on AI - A European approach to excellence and trust

European Commission
(February 2020)

Spanish National AI Strategy (ENIA)

Spanish Government
(November 2020)

EU AI Act Proposal (new AI Regulation)

European Commission
(April 2021)

A Coruña as first National AI Oversight Agency (AESIA)

Spanish & Galicia Government
(December 2022)

Spanish Presidency EU

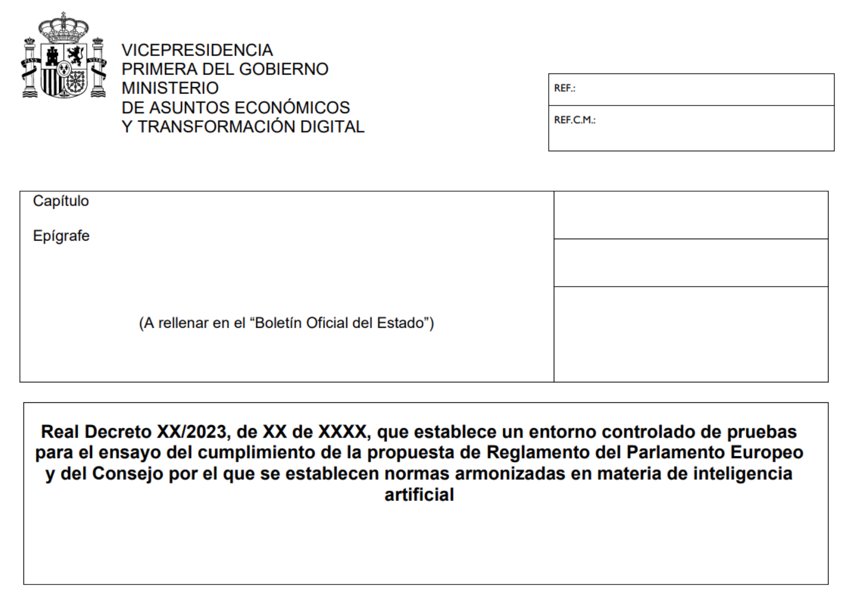
Council of the European Union
(July 2023)

First AI Regulatory Sandbox for the EU AI Act

Spanish Government
(2023+)

AI Sandboxes

The opportunity for public-private collaboration



VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

REF.:

REF.C.M.:

Capítulo

Epígrafe

(A rellenar en el “Boletín Oficial del Estado”)

Real Decreto XX/2023, de XX de XXXX, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial

- An AI sandbox to enable adopter and provider companies to implement the EU AI Act measures
- The Spanish AI Sandbox as a replicable model for implementation in Hungary
- The upcoming Hungarian presidency of the Council of the EU as an opportunity to leverage this approach
- A unique moment for governments and companies to collaborate and work together

Microsoft

Let’s work together!

- Governance systems
- New research and policies
- Innovative technologies
- Tools for adopters

aka.ms/ResponsibleAIResources

HUMANS
IN CHARGE –