



PEAs in Pods

Community | Engagement | Research

PEAs in Pods
Community | Engagement | Research



EPSRC PEAs in Pods: Co-production Of Community Based Public Engagement For Data And AI Research

Funded by EPSRC Grant Ref: EP/W033488/1 2022



Engineering and
Physical Sciences
Research Council



Manchester
Metropolitan
University

Activity 2.2: AI Research and Ethics



Image Source: FreePik – Image generated by AI

Activity aims

- To explore different definitions of AI from different perspectives
- To define ethics in the context of AI and briefly review core ethical AI principles
- To provide an overview of emerging and current AI Legislation*
- To explore why an ethical approach is essential to building trustworthy AI



*As at May 2025



unesco

Motivation

“Inclusion is not only participation but also engagement, representation, and the empowerment of disabled and unrepresented groups and populations.”

Jake Okechukwu Effoduh, Nigerian lawyer and activist

“We need to identify community champions to effectively communicate with the under-represented, understand their perspectives and empower them.”

Natalie Pang, Centre for Trusted Internet and Community Singapore

“To bridge the digital divide, there should be a “creation of mutual language and understanding of AI””

Session Activity 1

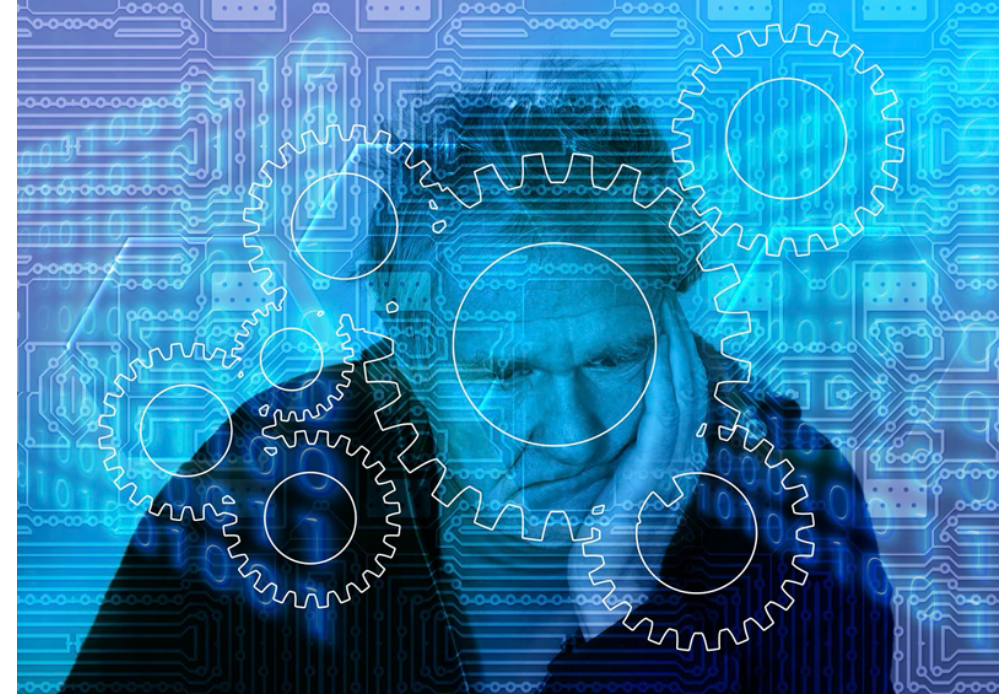
What is your definition of Artificial Intelligence ?



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

Session Activity 1

Explain in a few words, what is AI to your elderly relative.



[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Public Perceptions of AI

The branch of computer science dealing with the reproduction or mimicking of human-level intelligence, self-awareness, knowledge, conscience, thought in computer programs.

do not understand a lot but i think it's to do with your brain and what you know

A self-aware computer program

Computers which learn how to be more like humans

What is used by Tony Stark in the Avengers to control the robots and they seem to have a mind of their own

Stuff that make computers do things better than humans

the discipline concerned with what is morally good and bad and morally right and wrong

AI is the ability to learn human behaviour. Often achieved through machine learning from data which has been labelled.

It is the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

Robots - Machines that can the jobs of humans

Intelligence that was created by other intelligence rather than naturally developed

UK Definition of AI

The UK uses the definition of AI adopted by OECD countries:

“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment”



[This Photo](#) by Unknown Author is licensed under [CC BY](#)

Source: UK Government AI Playbook (2025), [online], Available: [Artificial Intelligence Playbook for the UK Government \(HTML\)](#) - GOV.UK

What is Artificial Intelligence ethics?

“AI ethics is a set of values, principles, and techniques that employ widely accepted standards to guide moral conduct in the development and use of AI systems.”

Source Gov.UK

New technology = new ethical problems

Should a device, a technique or technology be restricted because people can use it for illegal or harmful actions as well as beneficial ones?

Summary of Key Ethical Principles

- AI should **not be used to harm or kill** any human and respect human rights.
- AI systems must always be **fair, unbiased and transparent** in the decision-making process
- AI systems and solutions should always operate **within the law** and have **human accountability**
- **Data Governance and Data Privacy** should be incorporated into the AI life cycle
- Humans should always **know when** they have **interactions** with an AI system
- AI systems should be **inclusive to all**
- **Human-centered** AI design
- Appropriate levels of **Explainability** should be provided on AI decision making.
- Humans must always be **in the loop** when an AI is making a decision that effects other humans.
- Humans responsible for designing, developing and operating AI systems should be **competent in the skills and knowledge** required
- AI systems should be **sustainable** and work to benefit humans, the society and the environment

unesco

Recommendation on
the Ethics of Artificial Intelligence

Adopted on 23 November 2021

HM Government

National AI Strategy

Explaining decisions made with AI
Draft guidance for consultation

Part 1:
The basics of explaining AI

ico. The Alan Turing Institute

Equality and Human Rights Commission

Strategic plan 2022-25

equalityhumanrights.com

BLUEPRINT FOR AN AI BILL OF RIGHTS

MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE

Government Digital Service | **Department for Science, Innovation & Technology**

Putting the OECD AI Principles into practice: progress and future perspectives

Session summary of this high-level panel discussion from the OECD MCM 2021
With the generous support of the Government of the Republic of Korea

OECD | 6C

Article 22
GDPR

Artificial Intelligence Playbook for the UK Government

Created by the Government Digital Service
February 2020

Stepoe

THE EU ARTIFICIAL INTELLIGENCE ACT

Key aspects

On April 21, 2021 the EU Commission published its proposal for an Artificial Intelligence Act.
Here is what you need to know

- 1 Limited prohibited uses
- 2 Regulation of high-risk AI systems
- 3 Ex-ante/ex-post requirements
- 4 GDPR-trained?
- 5 What comes next?

Version 2 - For Public Discussion

IEEE
Advancing Technology for Humanity

ETHICALLY ALIGNED DESIGN
A Guide for Embedding Human Well-being with Autonomous and Intelligent Systems

DRCF

Digital Regulation Cooperation Forum

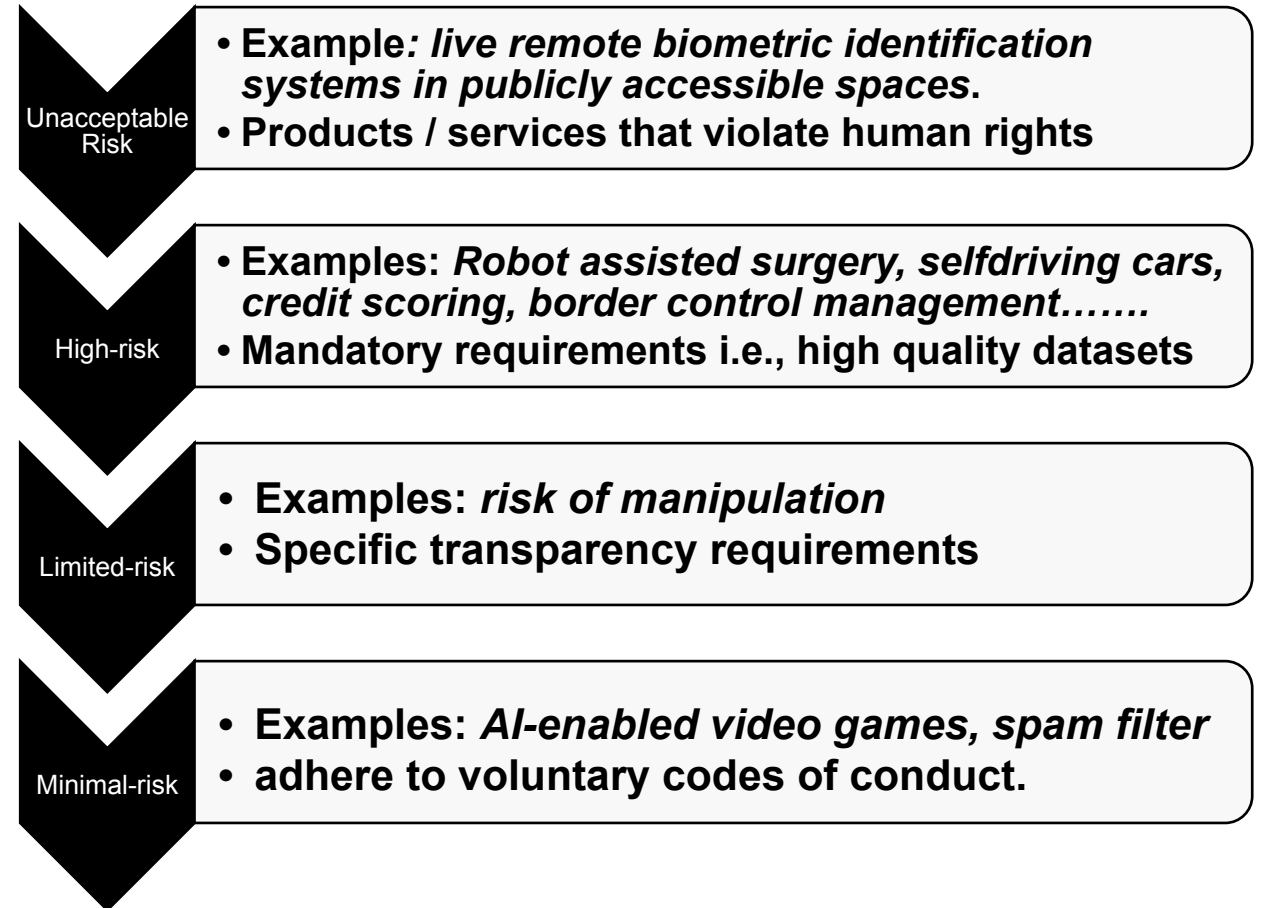
CMA | **Ofcom** | **ico.** | **FCA** FINANCIAL CONDUCT AUTHORITY

AI STANDARDS HUB

The EU AI Act 2024

Who is it for?

- “legal framework will apply to both public and private actors inside and outside the EU as long as the AI system is placed on the Union market, or its use affects people located in the EU”.
- A new European Artificial Intelligence Board (EAIB), will also be set up to support a consistent application of the regulation
- EU database for stand-alone high-risk AI systems



Source: [The AI Act Explorer](#) | [EU Artificial Intelligence Act](#)

High Risk

AI systems that have significant implications for health, safety, or fundamental rights are classified as high risk. These systems must comply with stringent requirements, including [risk management](#), [data governance](#), transparency, and human oversight. High-risk AI applications include biometric identification, critical infrastructure, and employment-related AI systems.

Minimal Risk

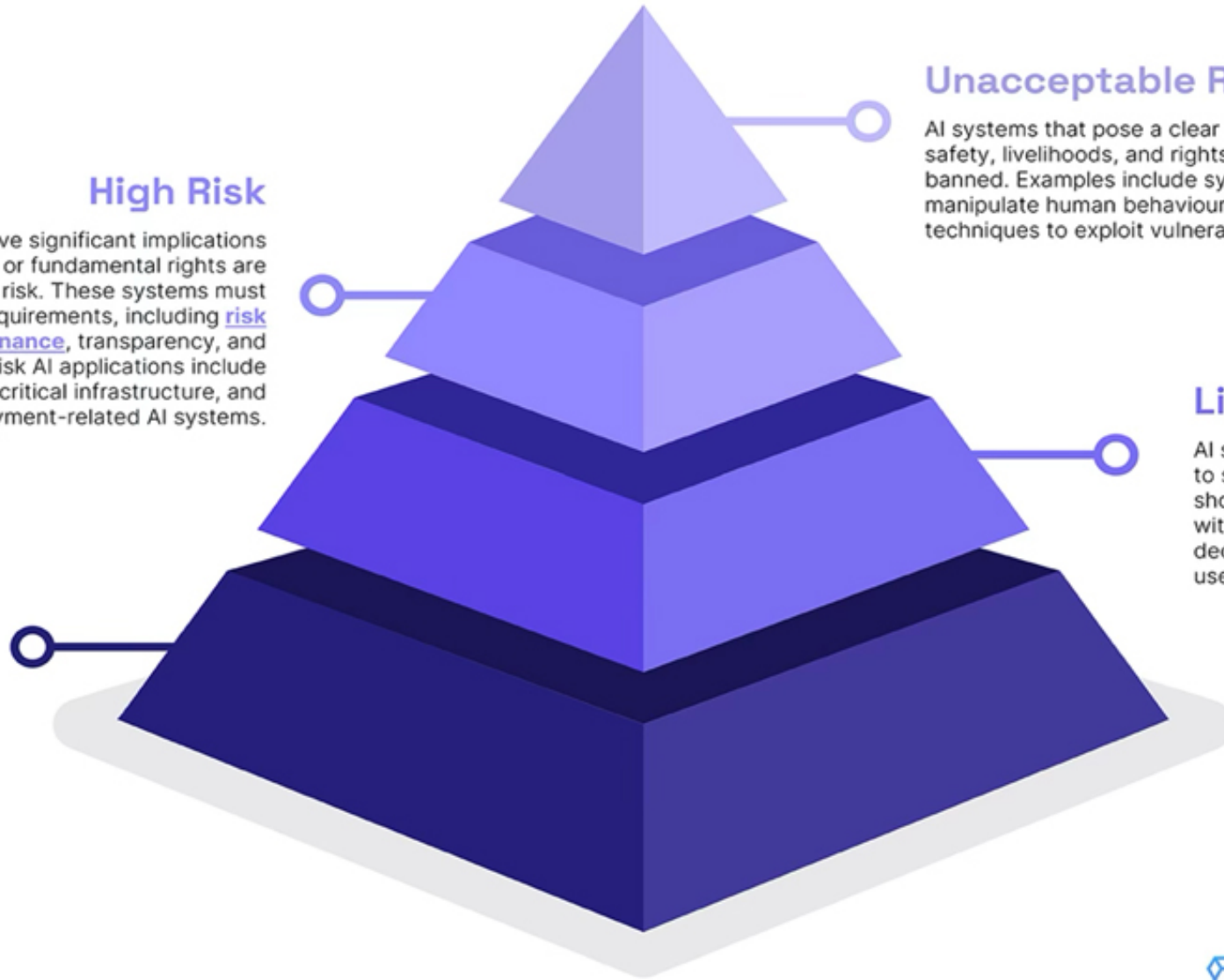
AI systems with minimal risk are subject to minimal regulatory intervention, focusing on voluntary codes of conduct and self-regulation. These systems include AI applications like spam filters and video games.

Unacceptable Risk

AI systems that pose a clear threat to the safety, livelihoods, and rights of people are banned. Examples include systems that manipulate human behaviour or use subliminal techniques to exploit vulnerable individuals.

Limited Risk

AI systems with limited risk must adhere to specific transparency obligations. Users should be informed that they are interacting with an AI system to make informed decisions. Examples include AI systems used in chatbots or virtual assistants.



www.holisticai.com



Holistic AI



Engineering and
Physical Sciences
Research Council



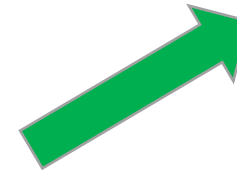
Manchester
Metropolitan
University

Post-Market Monitoring for High-Risk AI

Business / Providers
monitor AI system in
market



Breach of Human Rights
or Safety Laws



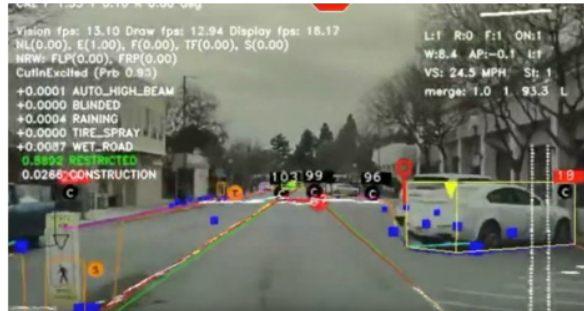
Reported Breach –
National Supervisory
Body / EU



Violation of the EU
AI Act

DRCF

Digital Regulation Cooperation Forum



AI system forcibly withdrawn
from the market.

Regulators can access source code



This Photo by Unknown Author is licensed under CC BY-SA

UK Algorithmic Transparency Standard

- The Algorithmic Transparency Standard helps **public sector** organisations provide clear information about the algorithmic tools they use, and why they're using them.
- Algorithmic transparency means being open about how algorithmic tools support decisions. This includes providing information on algorithmic tools and algorithm-assisted decisions in a complete, open, understandable, easily-accessible, and free format.
- The standard is made up of an:
 - algorithmic transparency data standard
 - algorithmic transparency template and guidance that helps **public sector organisations** provide information to the data standard



Source: [Algorithmic Transparency Standard - GOV.UK \(www.gov.uk\)](https://www.gov.uk)

The UKs Pro-Innovation Framework – the 5 principles

- Safety, security and robustness
- Appropriate transparency and explainability
- Fairness
- Accountability and governance
- Contestability and redress



Source: [A pro-innovation approach to AI regulation - GOV.UK \(www.gov.uk\)](https://www.gov.uk)

The UKs Pro-Innovation Framework – 3 objectives

1. Drive growth and prosperity
- 2. Increase public trust in AI**
3. Strengthen the UK's position as a global leader in AI

"Providing education and awareness to give clarity to businesses and empower citizens to make their voices heard as part of the ongoing iteration of the framework."

Source: [A pro-innovation approach to AI regulation - GOV.UK \(www.gov.uk\)](https://www.gov.uk)

AI Regulations Globally: Same Goal, Different Paths: December 2024

North America

- United States
 - AI Regulations, Federal Trade Commission
 - AI Principles, The US DoD
 - AI RMF, NIST
- Canada
 - Artificial Intelligence and Data Act



South America

- Brazil
 - Proposed AI Regulations

Europe

- European Union
 - EU AI Act
- United Kingdom
 - A pro-innovation approach to AI Regulation



Middle East

- Saudi Arabia
 - AI Ethics Principles

Africa

- NA

Asia

- China
 - Interim Measures for the Management of Generative Artificial Intelligence Services
- Japan
 - AI Guidelines for Business
- Singapore
 - The Model AI Governance Framework

Australia & Oceania

- Australia
 - AI Ethics Principles

Source: <https://www.linkedin.com/pulse/ai-regulations-globally-same-goal-different-paths-dr-irina-steenbeek-1cipe/>

OECD.AI National AI Polices and Strategies - LIVE

OECD.org | Going Digital Toolkit | EN

Blog | Experts | AI Principles | Policy areas | Trends & data | Countries | About

Home > National strategies & policies

National AI policies & strategies

This section provides a live repository of over 800 AI policy initiatives from 69 countries, territories and the EU. Click on a country/territory, a policy instrument or a group targeted by the policy.

Countries & territories | Policy instruments | Target Groups

Search for a specific dashboard ...

Argentina	Croatia	India	Mauritius	Saudi Arabia	Uganda
Armenia	Cyprus	Ireland	Mexico	Serbia	Ukraine
Australia	Czech Republic	Israel	Morocco	Singapore	United Arab Emirates
Austria	Denmark	Italy	Netherlands	Slovakia	United Kingdom
Belgium	Egypt	Japan	New Zealand	Slovenia	United States
Brazil	Estonia	Kazakhstan	Nigeria	South Africa	Uruguay
Bulgaria	Finland	Kenya	Norway	Spain	Uzbekistan
Canada	France	Korea	Peru	Sweden	Viet Nam
Chile	Germany	Latvia	Poland	Switzerland	European Union
China	Greece	Lithuania	Portugal	Thailand	
Colombia	Hungary	Luxembourg	Romania	Tunisia	
Costa Rica	Iceland	Malta	Rwanda	Türkiye	



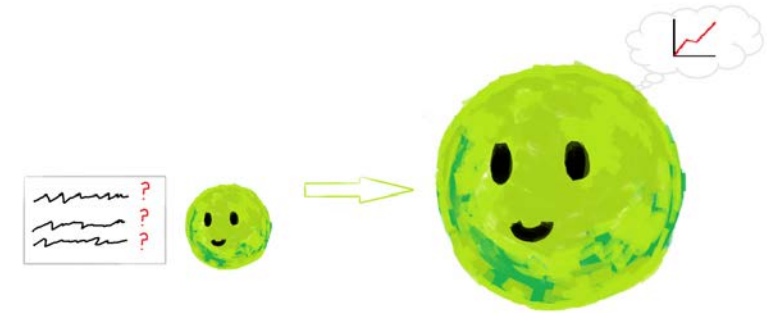
Are you confused yet ?



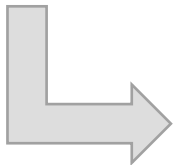
***“The only true wisdom is in knowing you know nothing.”
– Socrates***

This Photo by Unknown Author is licensed under [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)

A Risk Based Approach to Artificial Intelligence



How do you assess risk?



Over counter pain medication

Package leaflet: Information for the user

Ibuprofen 200 mg film-coated tablets Ibuprofen 400 mg film-coated tablets

Read all of this leaflet carefully before you start taking this medicine because it contains important information for you.

- Keep this leaflet. You may need to read it again.
- If you have any further questions, ask your doctor or pharmacist.
- This medicine has been prescribed for you only. Do not pass it on to others. It may harm them, even if their signs of illness are the same as yours.
- If you get any side effects, talk to your doctor or pharmacist. This includes any possible side effects not listed in this leaflet. See section 4.

Legal Requirement
Patient Information
Leaflet



4. Possible side effects

Like all medicines, this medicine can cause side effects, although not everybody gets them.

With regard to the following side effects, it must be considered that they are largely dependent on the dose and that they vary from patient to patient.

The most commonly observed side effects are gastrointestinal in nature. Peptic ulcers, perforation or gastrointestinal bleeding, sometimes fatal, particularly in the elderly, may occur. Nausea, vomiting, diarrhoea, flatulence, constipation, indigestion, abdominal pain, melaena, haematemesis, ulcerative stomatitis, exacerbation of colitis and Crohn's disease have been reported following administration. Less frequently, inflammation of the stomach lining (gastritis) has been observed.

Medicines such as Ibuprofen may be associated with a small increased risk of heart attack (myocardial infarction) or stroke. Water retention (oedema), high blood pressure and heart failure have been reported in association with NSAIDs.

The side effects are stated according to the frequency they occur. The following convention has been used:

Very common: may affect more than 1 in 10 people

Common: may affect up to 1 in 10 people

Uncommon: may affect up to 1 in 100 people

Rare: may affect up to 1 in 1,000 people

Very rare: may affect up to 1 in 10,000 people
Not known: frequency cannot be estimated from the available data

The following side effects are important and will require immediate action if you experience them. You should stop taking Ibuprofen and see your doctor immediately if the following symptoms occur:

Common:

- black tarry stools or blood-stained vomit (digestive tract ulcer with bleeding)

Side effects

Common:

- Digestive tract ulcer with or without perforation
- Bowel inflammation and worsening of inflammation of the colon (colitis) and digestive tract (Crohn's disease) and complications of diverticula of the large bowel (perforation or fistula)
- Microscopic bleeding from the intestine which may result in anemia
- Mouth ulcers and inflammation (ulcerative stomatitis)
- Headache, sleepiness, vertigo, dizziness, fatigue, agitation, insomnia and irritability

Uncommon:

- Inflammation of the stomach lining
- Kidney problems including development of oedema, inflammation of the kidneys and kidney failure
- Runny nose
- Difficulty breathing (bronchospasm)

Rare:

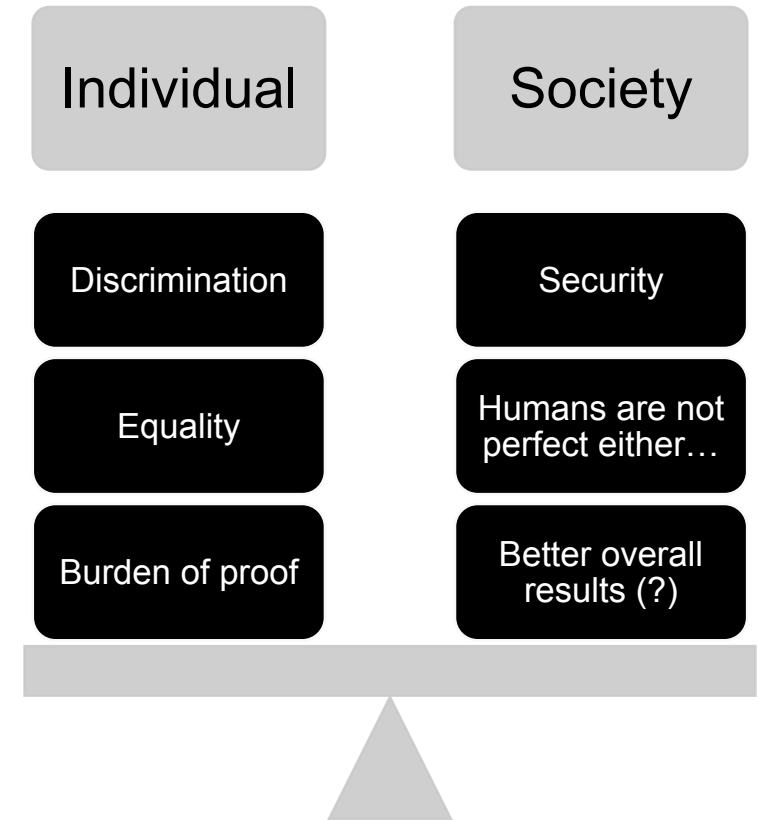
- Depression, confusion, hallucinations
- Lupus erythematosus syndrome
- Increase of blood urea nitrogen, serum transaminases and alkaline phosphatase decrease in haemoglobin and haematocrit values, inhibition of platelet aggregation and prolonged bleeding time, decrease of serum calcium and increase in serum uric acid values
- Damage of the kidney tissue

Very rare:

- Unpleasant awareness of heart beat, heart failure or heart attack
- Disorders of blood cell formation (anaemia, leucopenia, thrombocytopenia, pancytopenia, agranulocytosis, neutropenia)
- The first symptoms or signs may include fever, sore throat, surface mouth ulcers, flu-like symptoms, severe fatigue, nasal and skin bleeding)
- Ringing or buzzing in the ears

Risk Factors in Artificial Intelligence

- ✓ Bias
- ✓ Fairness
- ✓ Explainable AI (XAI)
- ✓ Understandable AI (UAI)
- ✓ Responsibility and Accountability
- ✓ Societal impacts (potential benefits and harms)
- ✓ Usability and reliability
- ✓ Data governance and data privacy
- ✓ AI Governance
- ✓ AI technical safety (autonomous systems)
- ✓ Trust



How can we design, implement, and deploy AI systems to be trustworthy?



This Photo by Unknown Author is licensed under [CC BY-SA-NC](#)

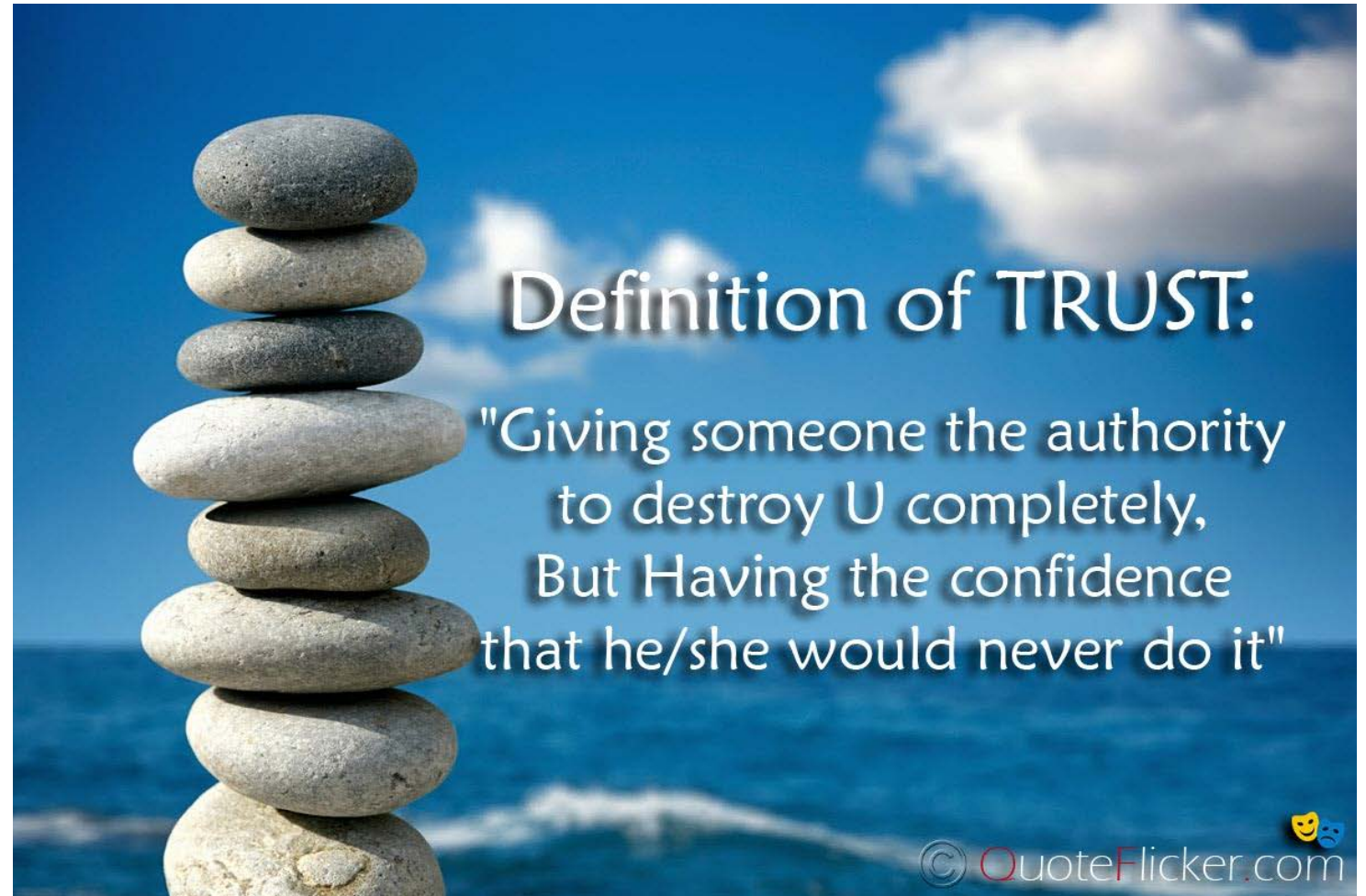
What is Trust?

“

Quote by Trust is the glue of life. It's the most essential ingredient in effective communication. It's the foundational principle that holds all relationships.

—
STEPHEN COVEY

GRACIOUSNOTES.COM

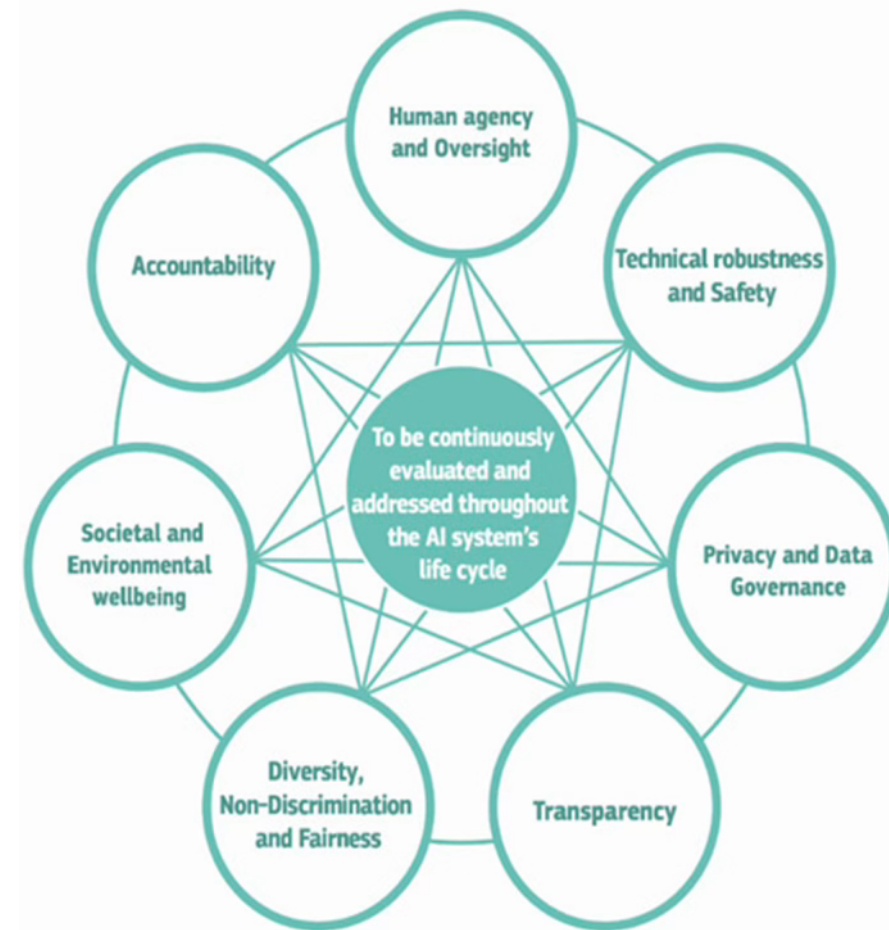


EU INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

Trustworthy AI has three components, which should be met throughout the system's entire life cycle:

1. It should be lawful, complying with all applicable laws and regulations;
2. It should be ethical, ensuring 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: [ai-ethics-guidelines.pdf \(aepd.es\)](#)



Public attitudes to data and AI: Tracker survey (Wave 4)
Published: 16/12/24

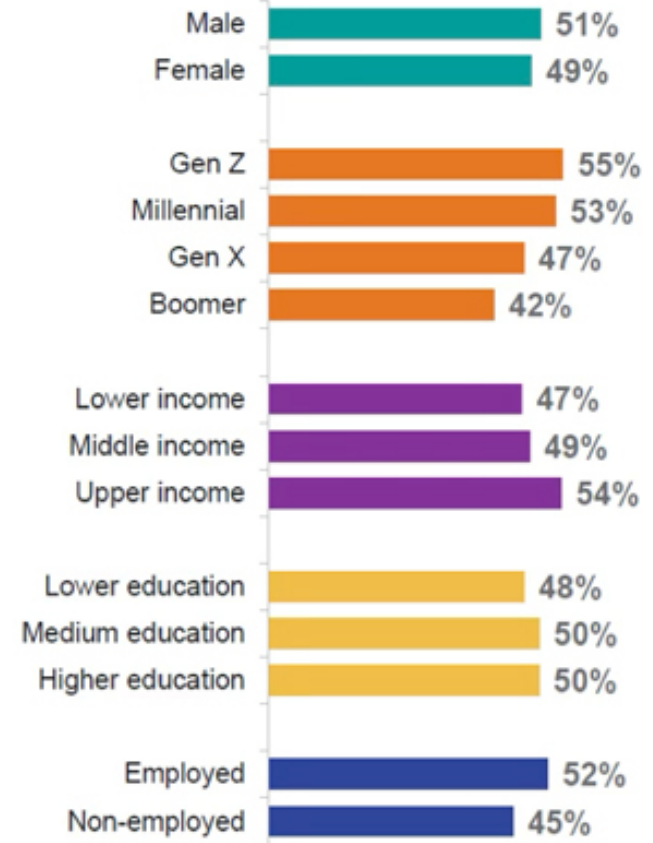
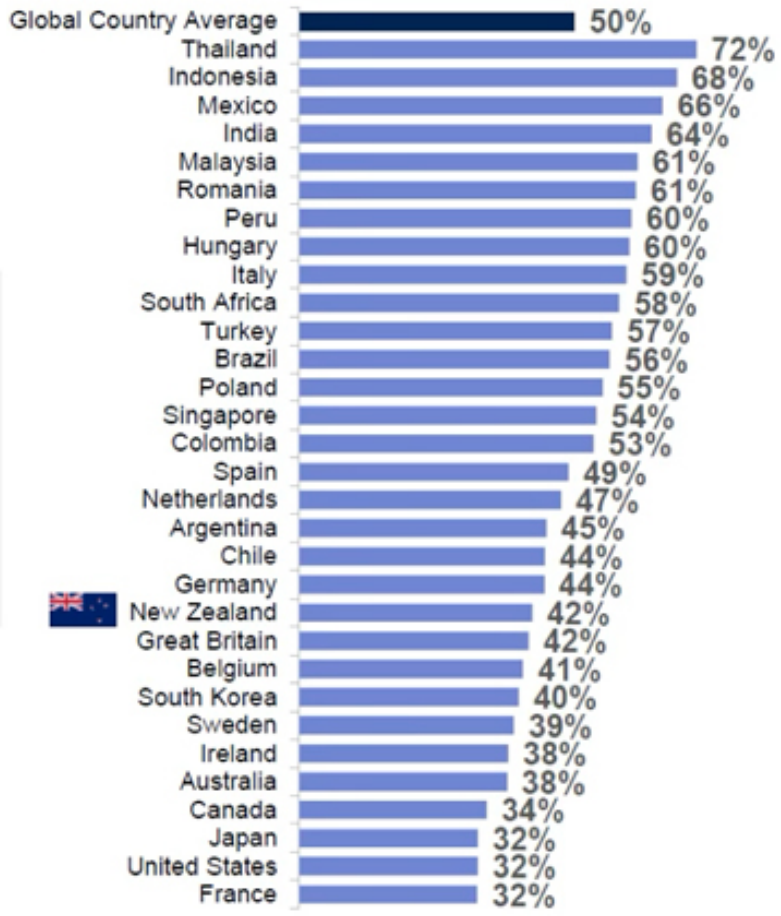
Table 2: Trust in organisations, and in their actions with data (Showing % Sum: Trust)

	Data actors									
	The NHS	Academic researchers at universities	Regulators	Banks and other financial institutions	Researchers at pharmaceutical companies	Utilities providers	Big technology companies	HR and recruitment services	The Government	Social media companies
Effectively use the data to improve the products or services you receive	73%	64%	62%	64%	64%	54%	55%	48%	46%	39%
Keep data about you safe	72%	62%	64%	67%	59%	55%	49%	51%	48%	33%
Use your data to benefit society	72%	64%	58%	50%	62%	44%	42%	40%	46%	30%
Be open and transparent about what they do with data about you	69%	59%	58%	58%	53%	45%	42%	45%	41%	30%
Let you make decisions about how your data is used	66%	55%	54%	56%	49%	46%	43%	44%	41%	34%

Source: [Public attitudes to data and AI: Tracker](#)

Trust in AI

Q. How much do you agree or disagree with the following:
I trust that companies that use artificial intelligence will protect my personal data
 % Agree (very/somewhat)



Base: 22,816 adults under the age of 75 across 31 countries, interviewed May 26 – June 9, 2023 – online only in all countries except India.
 The "Global Country Average" reflects the average result for all the countries where the survey was conducted. It has not been adjusted to the population size of each country or market and is not intended to suggest a total result.
 The samples in Brazil, Chile, Colombia, India, Indonesia, Ireland, Malaysia, Mexico, Peru, Romania, Singapore, South Africa, Thailand, and Turkey are more urban, more educated, and/or more affluent than the general population



[Présentation PowerPoint \(ipsos.com\)](https://www.ipsos.com)

People perspective - What are the barriers to trustworthiness?

- Understanding of AI – what it is? – how does it learn?
- Knowledge of what is and what isn't AI
- Concerns about the use of their data
 - Data privacy
 - How data is used to build models
- Ethical issues such as bias, fairness.....
- Not understanding how and why an automated decision has been made.
- Risk factors (Human V Machine)
- Digital literacy
- Digital poverty
- Media portrayal
- Lived experiences
- Removal of services



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

Your Perception of Trust

6 questions to explore your opinion of trust in AI



[This Photo](#) by Unknown Author is licensed under [CC BY](#)

Summary



What is AI and Why Ethics Matter – Looked at different ways to define AI and why ethics are important in AI.



AI Rules and Laws – Reviewed new and existing laws that guide how AI should be used responsibly.



Making AI Trustworthy – Discussed why AI needs to be fair, transparent, and ethical so people can trust it.