



**Universidade do Minho**  
Escola de Engenharia  
Departamento de Informática

# **Inteligência e a Sociedade**

## **A Responsible AI for Social Good**

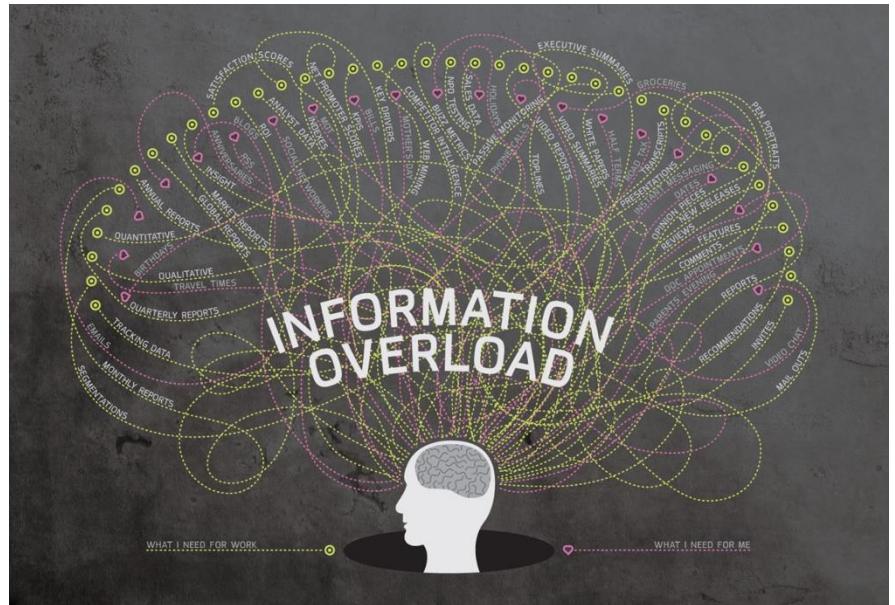
**LICENCIATURA EM ENGENHARIA INFORMÁTICA**  
**MESTRADO integrado EM ENGENHARIA INFORMÁTICA**

**In inteligência Artificial**  
**2025/26**

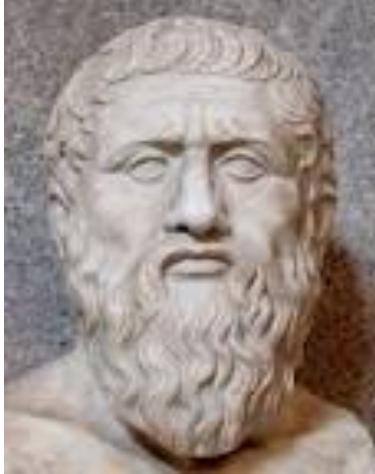


Source: Woman reading  
Chronicles of the King of France, Robert Gaguin, 1514  
<https://www.historytoday.com/reviews/history-knowledge>

## Then and Now

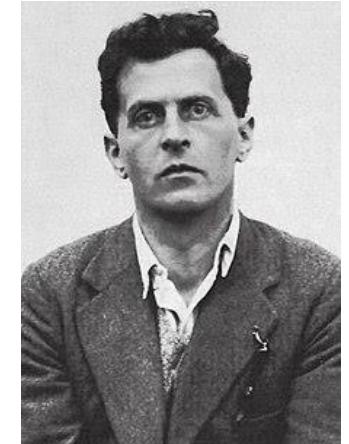


## Source of inspirations



The search for truth

Since **Plato and Aristotle**, truth would be the “exact agreement” of a statement with the reality of the thing expressed by it.



20<sup>th</sup> century **Ludwig Wittgenstein** left aside the notion of a single essence of things identified by reason. Knowledge is to be searched in generalizations about similarities between phenomena.

Image Source:

<https://pt.wikipedia.org/wiki/Platão>

[https://pt.wikipedia.org/wiki/Ludwig\\_Wittgenstein](https://pt.wikipedia.org/wiki/Ludwig_Wittgenstein)

- The AI approach to knowledge representation is exposed (normally) through two paradigms:

- Symbolic

- It is based on logic to represent knowledge;
    - It bases the reasoning in the construction of inference systems;

- Non-symbolic, or connectionist

- It bases the functioning of the system on the ability to learn, generalizing;
    - Solves problems based on past knowledge or data on solving other problems;



John McCarthy



Marvin Minsky

**p se q.**



## Some new sources of inspiration

- **Jean-Paul Sartre – Existentialism**



- **Claude Lévi-Strauss - Structuralism**



- **Zygmunt Bauman - Liquid modernity**



- **Michele Serres - The value of the "new" work**



### Individually

AI will augment us individually as people (deepening our memory, speeding our recognition), in our activity:

- day-to-day life;
- leisure activities;
- job.

### Collectively

As a way to expand our skills as a species.



## Environments with Intelligence Ambient Intelligence



Source: Novais et al., (2010) Inter-organization cooperation for ambient assisted living. *J. Ambient Intell. Smart Environ.* 2(2): 179-195.

Person centric computing

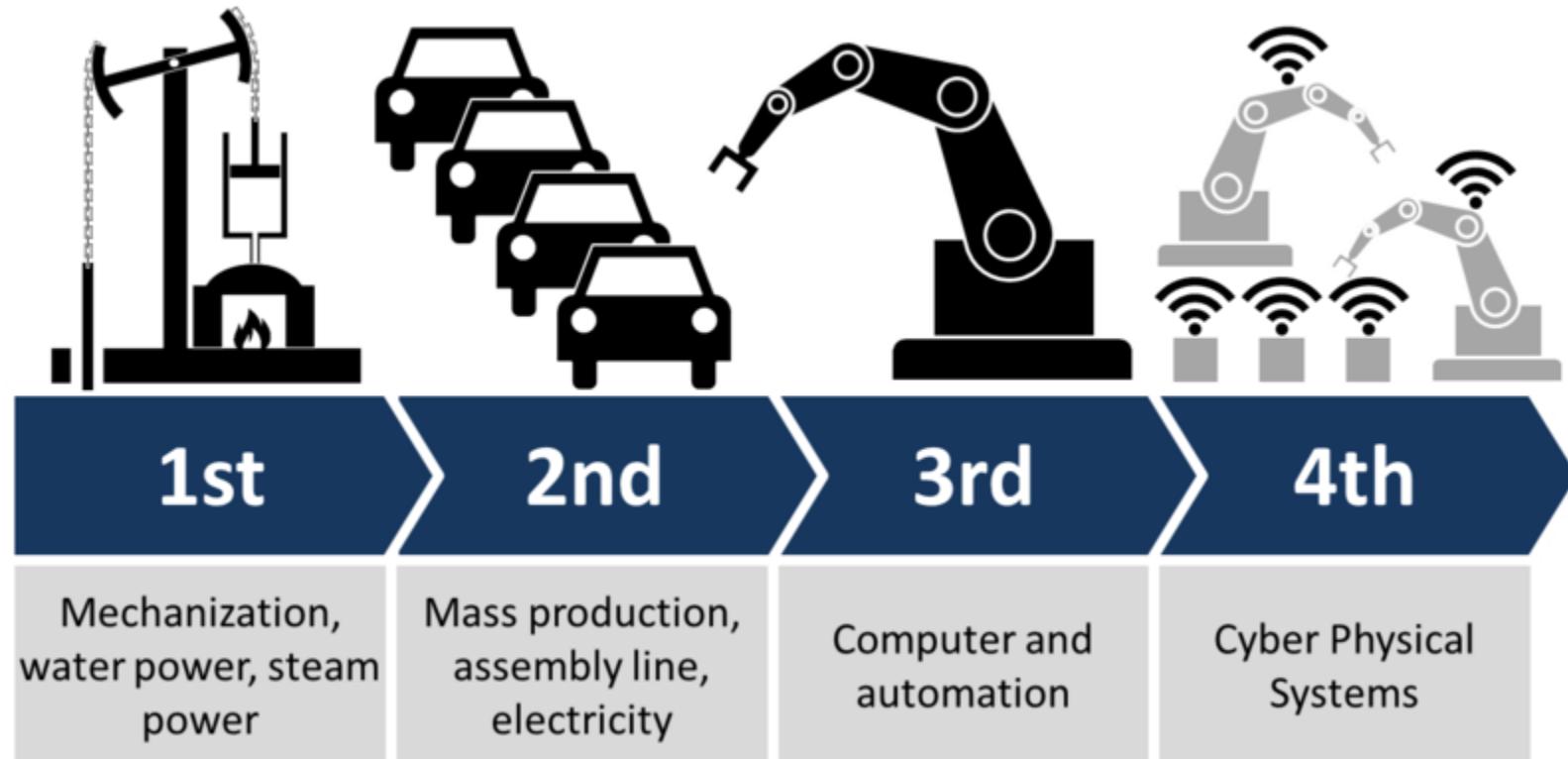


**“Our Intelligence is what makes us human, and AI is an extension of that quality”.**

Yann LeCun (A.M. Turing Award 2018)



## The Revolution



*Source: The 4 Industrial Revolutions (by Christoph Roser at AllAboutLean.com)*

Historically the acceptance and diffusion of technology depends on two factors:

- Labour price;
- Scalability.

Example:

Henry Bessemer (1856) - the steelmaking process

But however the invention remote to 2000 a.C. Anatolia.

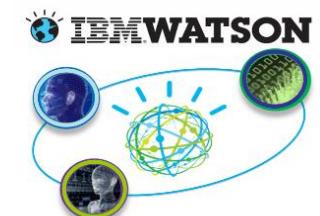
**This is the AI moment.**



Source: IBM Deep Blue versus Kasparov (1996/97)

### ▪ Watson (IBM)

- In February 2011, Watson beat the two best players in the USA program TV Jeopardy (Brad Rutter and Ken Jennings);
- Watson represents an important step in the development of cognitive systems.
- It use Natural language processing, generation and even hypotheses and learning.
- Deep QA



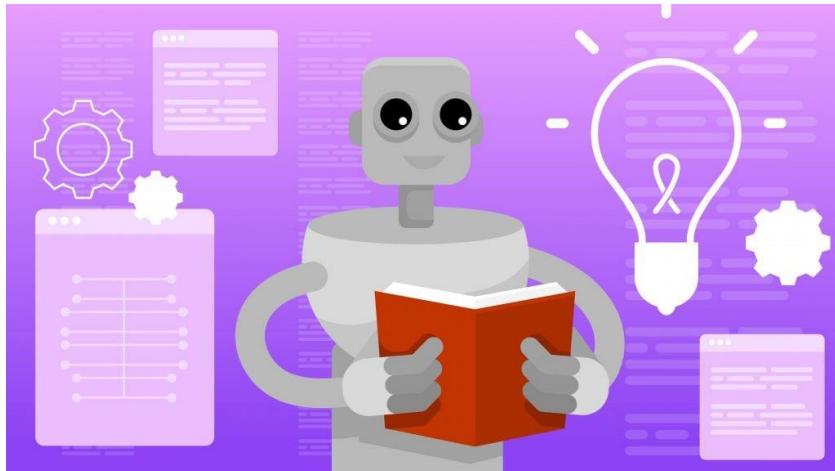
- **AlphaGo (Google DeepMind) - 2016**

- The match between man and the machine did not go well for Lee Se-dol the Go game world champion;
- Go is a board game for 2 players that is similar to chess but more complex in relation chess.
- AlphaGo combines deep neural networks (evaluation) and Monte Carlo tree search (choice). With a combination of supervised learning and reinforcement.



## Generative Pre-trained Transformer 3

An autoregressive language model that uses deep learning to produce human-like text



Source: <https://www.rev.com/blog/what-is-gpt-3-the-new-openai-language-model>

# Suddenly, the world has changed Generative AI ....ChatGPT, etc

- **Generative AI** is a type of AI technology that can produce various types of content, including text, images, audio and synthetic data;
- It does this by **learning patterns** from existing data, and then using this knowledge to generate new and unique outputs in response to prompts;
- Output: highly realistic and complex content that **mimics human creativity**;
- Using **Large language models (LLM)** are language models consisting of a very large neural network with billions of weights, trained on very large quantities of unlabelled text using self-supervised learning.



<a href="#">DALL-E 2</a> <a href="#">stable Diffusion</a> <a href="#">craiyon</a> <a href="#">Lexica</a> <a href="#">MidJourney</a> <a href="#">Imagen</a> <a href="#">Wombo</a> <a href="#">NightCafe</a> <a href="#">GauGAN2</a> <a href="#">DeepAI</a> <a href="#">Jasper</a> <a href="#">artbreeder</a> <a href="#">Wonder</a> <a href="#">pixray-text2image</a> <a href="#">neurallove</a> <a href="#">Omneky</a> <a href="#">alpaca</a> <a href="#">image.space</a> <a href="#">KREA</a> <a href="#">Nyx + gallery</a> <a href="#">&gt;ROSEBUDAI</a> <a href="#">PhotoRoom</a>											
Text-to-Image (T2I)											
<a href="#">runway</a> <a href="#">Flikr</a> <a href="#">O-synthesis</a> <a href="#">MetaAI</a> <a href="#">Google AI</a> <a href="#">Phenaki</a> <a href="#">CONTENDA</a> <a href="#">Play.jht</a> <a href="#">MURFA</a> <a href="#">RESEMBLE.AI</a> <a href="#">WELLSAID</a> <a href="#">descript</a> <a href="#">Algorithmic</a>											
Text-to-Video (T2V)											
<a href="#">Simplified</a> <a href="#">Jasper</a> <a href="#">frase</a> <a href="#">eleutherAI</a> <a href="#">Regustry</a> <a href="#">letterdrop</a> <a href="#">grammarly</a> <a href="#">copy.ai</a> <a href="#">MarketMuse</a> <a href="#">AI21Labs</a> <a href="#">HubSpot</a> <a href="#">NovelAI</a> <a href="#">InferKit</a> <a href="#">GooseAI</a> <a href="#">ResearchAI</a> <a href="#">Writsonic</a> <a href="#">cohere</a> <a href="#">CHIBI</a> <a href="#">Ideas AI</a> <a href="#">copysmith</a> <a href="#">Flowrite</a> <a href="#">WICHNESS</a> <a href="#">sudo</a> <a href="#">write</a> <a href="#">Rytr</a> <a href="#">ideasbyai</a> <a href="#">text.cortex</a> <a href="#">OpenAI GPT-3</a> <a href="#">Blog Idea Generator</a> <a href="#">HyperWise</a> <a href="#">Subtxt</a> <a href="#">WRITER</a> <a href="#">wordline</a> <a href="#">LATKA</a> <a href="#">COMPOSE AI</a> <a href="#">Moonbeam</a> <a href="#">BerthaAI</a> <a href="#">anyword</a> <a href="#">Hypotenuse AI</a> <a href="#">Peppertype.ai</a>											
Text-to-Text (T2T)											
<a href="#">TREE</a> <a href="#">MDM: Human Motion Diffusion Model</a> <a href="#">replit</a> <a href="#">Ghostwriter</a> <a href="#">GitHub Copilot</a> <a href="#">MUTABLE AI</a> <a href="#">tobin91</a> <a href="#">Amazon CodeWhisperer</a>											
Text-to-Motion (T2M)											
<a href="#">LensAI</a>											
Text-to-3D (T2D)											
<a href="#">DreamFusion</a> <a href="#">CLIP-Mesh</a> <a href="#">GET3D</a>											
Audio-to-Text (A2T)											
<a href="#">descript</a> <a href="#">AssemblyAI</a> <a href="#">Whisper</a>											
Audio-to-Audio (A2A)											
<a href="#">AudioLM</a> <a href="#">NN-VOICEMOD</a>											
Brain-to-Text (B2T)											
<a href="#">speech from brain</a> <a href="#">non-invasive brain recordings</a>											
Image-to-Text (A2I)											
<a href="#">neurallove</a> <a href="#">GPT-3 x Image Captions</a>											



Source: <https://www.futuristgerd.com/2019/06/at-work-expertise-is-falling-out-of-favor/>  
Gerd Leonhard

- **Customers**

- Customer Experience and Customization:
- Algorithms track customers journeys and help them to find the right product/service.
  - **Increase the level of Satisfaction**

- **Human resource**

- an organization should keep employees willing to achieve organizational goals, this is crucial for her survival!
  - **Empower the employee; talent retention and attraction.**

- **Product**

- Embedded to existing products or services to make them more effective, reliable, safer, and to enhance their longevity.
  - **Value creation.**

- **Process**

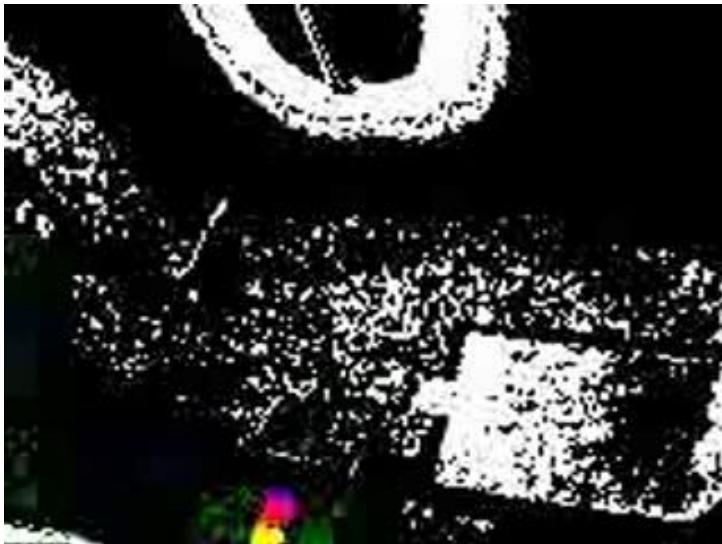
- Automation (e.g., Digital monitoring and control, task automation, human-robot collaboration);
  - **Productivity improvement.**

- **Knowledge (discovery)**

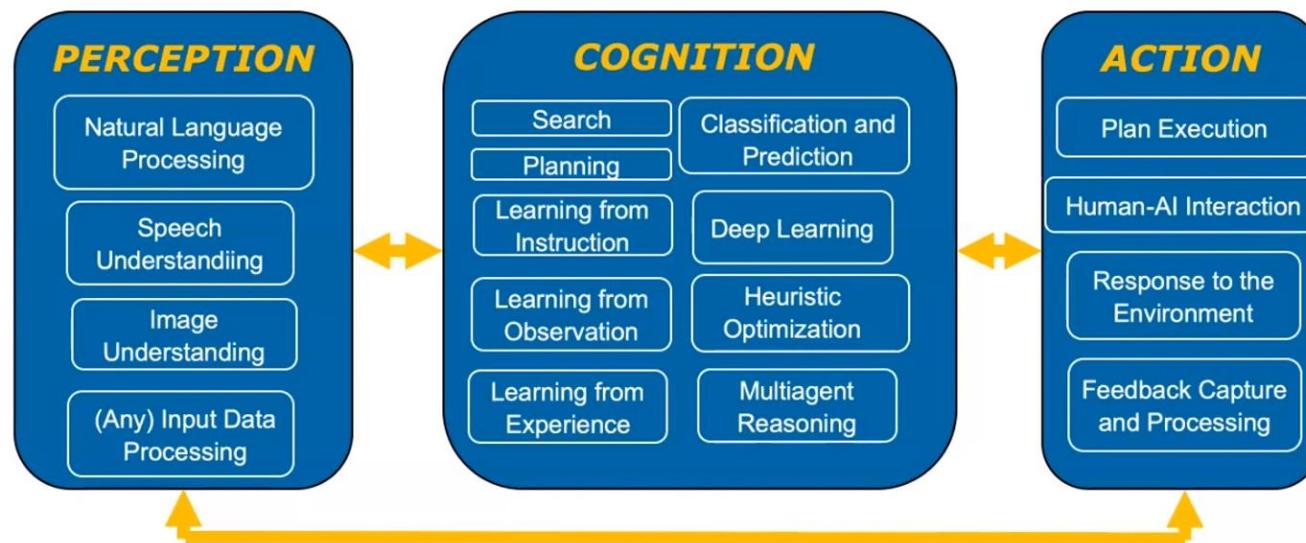
- Identifying insights in engineering systems (e.g., emerging production faults, use and performance of their products);
- Predictive and preventive maintenance.
  - **Efficiency and quality improvement.**

- **"Computers" are almost at our (human) level in certain basic functions**
  - Today computer vision is (probably) better than human vision, language translation is already very close to human ability.
  - Vision, writing (text), images and speech are practically at human levels.
- **In the short term there will be productivity gains in repetitive operations that can be automated.**
- **Routine and repetitive operations**
  - Computers have incredibly "good" memories and are "fantastic" in pattern-recognition tasks, which makes them suitable for automating "any" routine and repetitive operation.
- **We are automating tasks (we are not automating jobs !!!)**
  - There is no evidence that there will be mass unemployment.
- **What we are talking about and seeing is thinking in an absolutely different way.**
- **Most of our middle class and working class jobs are disappearing**
  - The labour market is changing as it is polarizing between highly skilled and low skilled jobs

**Put into context!**



## AI as a Science and an Engineering Generic architecture of an “intelligent entity”



Source: Manuela Veloso, 2023



DALLE-2 generated image

**“As the use and impact of autonomous and intelligent systems become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles.”**

[IEEE Standards Association](#) - Webinar: Ethical Considerations for System Design

## Learning with AI

- Using AI-driven tools to support teaching and learning;

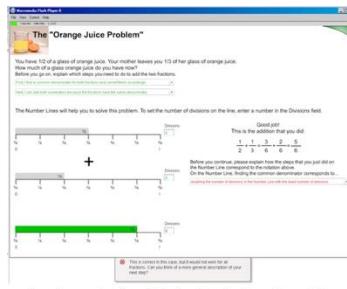
## Learning about AI

- Teaching how AI works and how to create (with) AI;

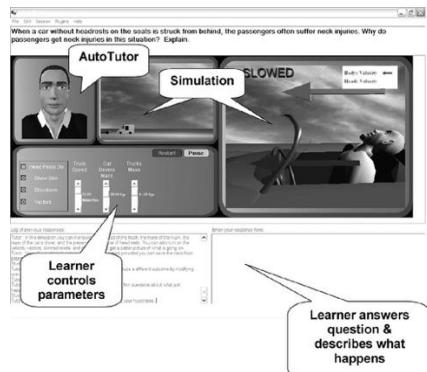
## Learning for AI

- Enable everyone to understand the possible impact of AI on their lives i.e. the ethical and

Source: Wayne Holme, Maya Balik and Charles Racll (2019), Artificial Intelligence in Education. Promise and Implications for Teaching and Learning, ISBN: 978-1784293700



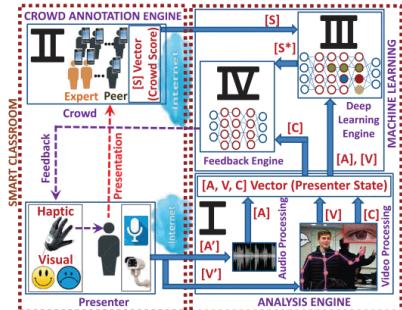
**Intelligent Tutoring Systems**  
 CTAT - The Cognitive Tutor  
 Authoring Tools  
<http://ctat.pact.cs.cmu.edu/>



Source: An example-tracing tutor for 6th-grade fractions learning  
 I. J. Artificial Intelligence in Education 2009  
 A New Paradigm for Intelligent Tutoring Systems: Example -Tracing Tutors  
 Vincent Aleven, Bruce M. McLaren, Jonathan Sewall, Kenneth R. Koedinger

**Pedagogical agents**  
 AutoTutor  
<http://ace.autotutor.org/>

Source: IEEE Transactions on Education 2005  
 AutoTutor: an intelligent tutoring system with mixed-initiative dialogue,  
 Arthur C. Graesser, Patrick Chipman, Brian C. Haynes, Andrew Olney



Source: <https://www.mheducation.com/ideas/three-levels-learning-analytics-adaptive-learning.html>  
 Source: Kim, Y., Soyata, T., & Behnagh, R. F. (2018).  
 Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education.  
 IEEE Access, 6, 5308-5331.  
 doi: 10.1109/ACCESS.2018.2791861

# Learning with AI

**Smart classrooms and Learning environments**

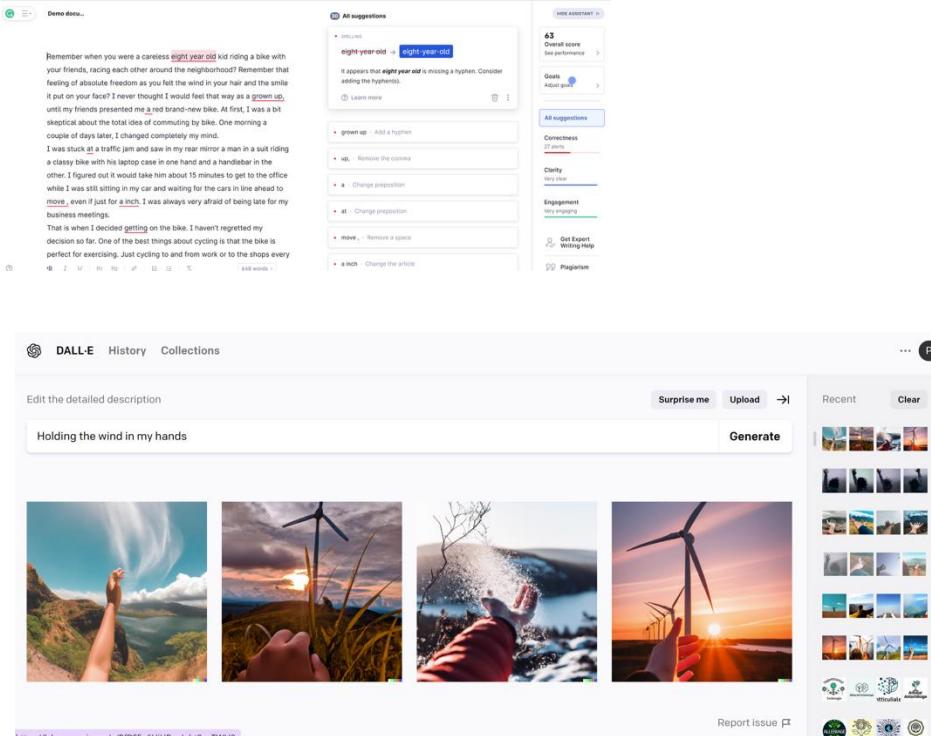
**Adaptive learning and learning analytics**

## Learning about AI

### Real-World Use Cases for Large Language Models

- **Search** (Search engines like Google and Bing already use LLMs to offer better user results);
- **Generate Content** (including conversational artificial intelligence and chatbots, the creation of marketing copy, code assistants, and high-quality content such as articles, summaries, captions, and even music);
  - Content creation;
  - Dialogue generation;
  - Storytelling;
  - Content augmentation; ...
- **Extract and Expand**
  - Extraction from data sets;
  - Expand the content.
- **Answering Questions**
  - Customer support systems;
  - Legal and financial analysis;
  - Language Translation;
  - Market Research and Competitor Analysis; ...

**Human resource empowerment**  
**Boosts productivity**



Demo doc... ...

Overall score: 63

Goals: Adjust goals

Overall score: 63

See performance

All suggestions

eight-year-old → eight-year-old

It appears that eight-year-old is missing a hyphen. Consider adding the hyphen.

Learn more

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

grown up → Add a hyphen

It was short at a traffic jam and saw in my rear mirror a man in a suit riding a classy bike with his laptop case in one hand and a handbag in the other. I figured out it would take him about 15 minutes to get to the office while I was still sitting in my car and waiting for the cars in line ahead to move, even if just for a inch. I was always very afraid of being late for my business meetings.

That is when I decided getting on the bike. I haven't regretted my decision so far. One of the best things about cycling is that the bike is perfect for exercising. Just cycling to and from work or to the shops every couple of days later, I changed completely my mind.

I was short at a traffic jam and saw in my rear mirror a man in a suit riding a classy bike with his laptop case in one hand and a handbag in the other. I figured out it would take him about 15 minutes to get to the office while I was still sitting in my car and waiting for the cars in line ahead to move, even if just for a inch. I was always very afraid of being late for my business meetings.

That is when I decided getting on the bike. I haven't regretted my decision so far. One of the best things about cycling is that the bike is perfect for exercising. Just cycling to and from work or to the shops every couple of days later, I changed completely my mind.

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

up → Remove the comma

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

at → Change preposition

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

move, → Remove a space

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

Get Expert Writing Help

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

in → Change the article

Overall score: 63

See performance

Goals: Adjust goals

All suggestions

Paraphraser

Grammar Checker

Plagiarism Checker

Co-Writer

Summarizer

Citation Generator

Translator

QuillBot Premium

Chrome Extension

Word Extension

Paraphrase

Try Sample Text

Paste Text

Help Center

DALL-E History Collections

Edit the detailed description

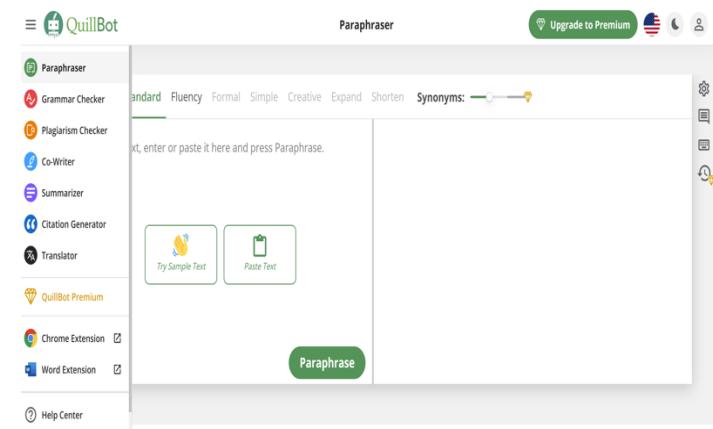
Surprise me Upload → Recent Clear

Generate

Holding the wind in my hands

Report issue ...

## Grammarly, Quillbot, Dall-E 2, Copilot



QuillBot

Paraphraser

Grammar Checker

Plagiarism Checker

Co-Writer

Summarizer

Citation Generator

Translator

Standard Fluency Formal Simple Creative Expand Shorten

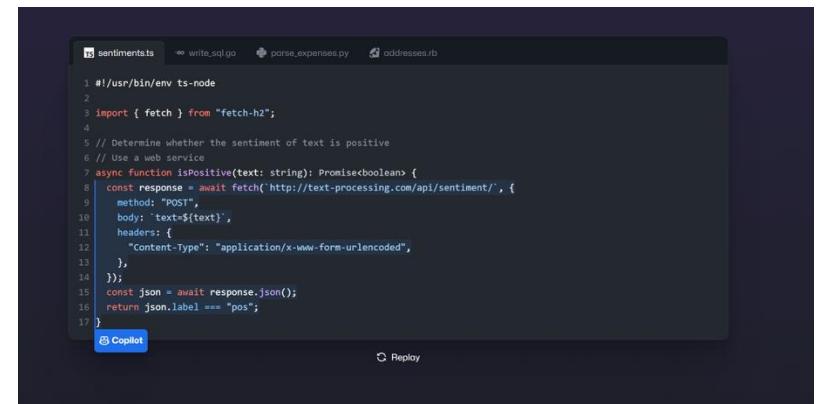
Synonyms: ...

Text, enter or paste it here and press Paraphrase.

Try Sample Text

Paste Text

Paraphrase



```

sentiments.ts  ↗ write_sql.go  ↗ parse_expenses.py  ↗ addresses.rb
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17}

```

Copilot

Replay

# Learning for AI Concerns



AP photo  
Elementary school teachers picket against use of calculators in grade school  
The teachers feel if students use calculators too early, they won't learn math concepts

## Math teachers protest against calculator use

By JILL LAWRENCE

"My older kids don't pay any attention to an answer being absurd," he said. "Teachers are shy."

## ChatGPT banned in Italy over privacy concerns

By Shiona McCallum  
Technology reporter

31 March 2023  
Updated 1 April 2023

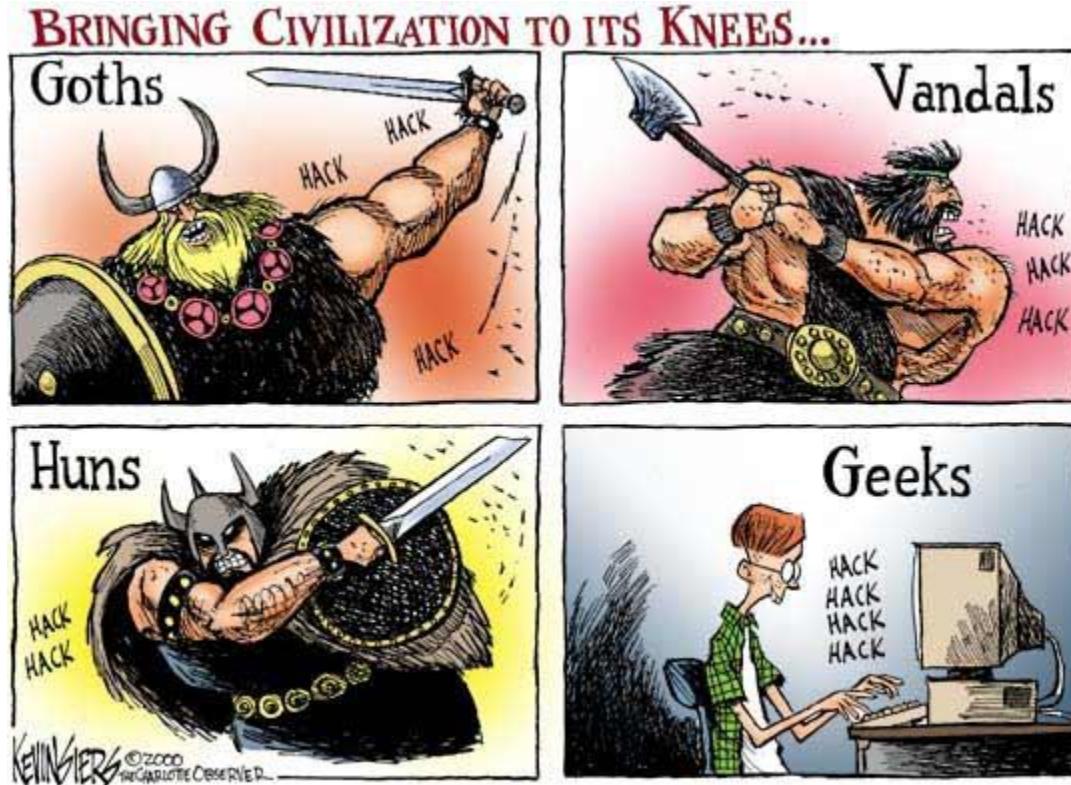


OpenAI launched ChatGPT last November

**Italy has become the first Western country to block advanced chatbot ChatGPT.**

The Italian data-protection authority said there were privacy concerns relating to the model, which was created by US start-up OpenAI and is backed by Microsoft.

## New threats





RECODE THE GOODS FUTURE PERFECT THE HIGHLIGHT FIRST PERSON PODCASTS VIDEO MORE ▾ [Twitter](#) [Facebook](#) [RSS](#) [User](#) [Search](#)

### Death by algorithm: the age of killer robots is closer than you think

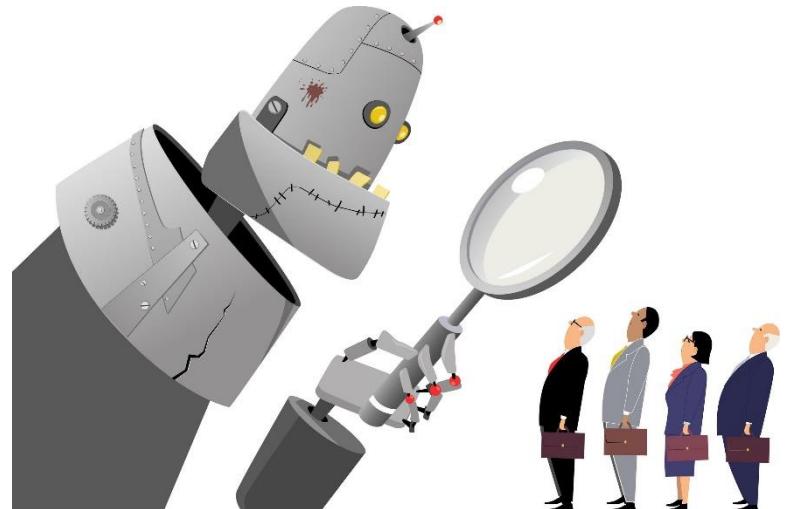
We have the technology to make robots that kill without oversight. But should we?

By Kelsey Piper | Jun 21, 2019, 8:20am EDT

[f](#) [Twitter](#) [SHARE](#)



A US Marine-operated Raven surveillance drone prepares to land outside a Marine base on March 21, 2009, near the remote village of Baqwa, Afghanistan, after flying a mission. | John Moore/Getty Images



Source: <https://medium.com/@turalt/ai-isnt-biased-we-are-b74ec94d1698>

**Computers are not immune to human imbecility**

Source: <http://expresso.sapo.pt/sociedade/2016-04-03-Os-computadores-nao-sao-imunes-a-imbecilidade-humana>

BUSINESS NEWS OCTOBER 10, 2018 / 4:12 AM / 6 MONTHS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

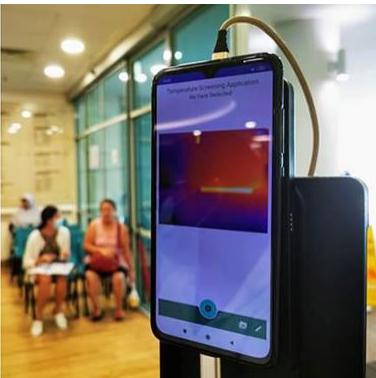
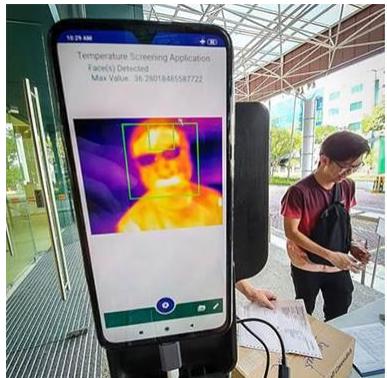
8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's ([AMZN.O](#)) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Source: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

- **FACIAL RECOGNITION AND FEVER DETECTOR**



Source: REUTERS

Source:  
<https://www.bbc.com/portuguese/geral-43011505>

- **INTELLIGENT DRONES & ROBOTS**



Source: <https://www.fierceelectronics.com/electronics/dragonfly-drones-could-hover-over-crowds-to-detect-coronavirus>



Source: The Wall Street Journal  
Fever-Detecting Goggles and Disinfectant Drones: Countries Turn to Tech to Fight Coronavirus

### Singapore - a data-controlled society

Started as a program to protect its citizens from terrorism has ended up influencing economic and immigration policy, the property market and school curricula.

### China:

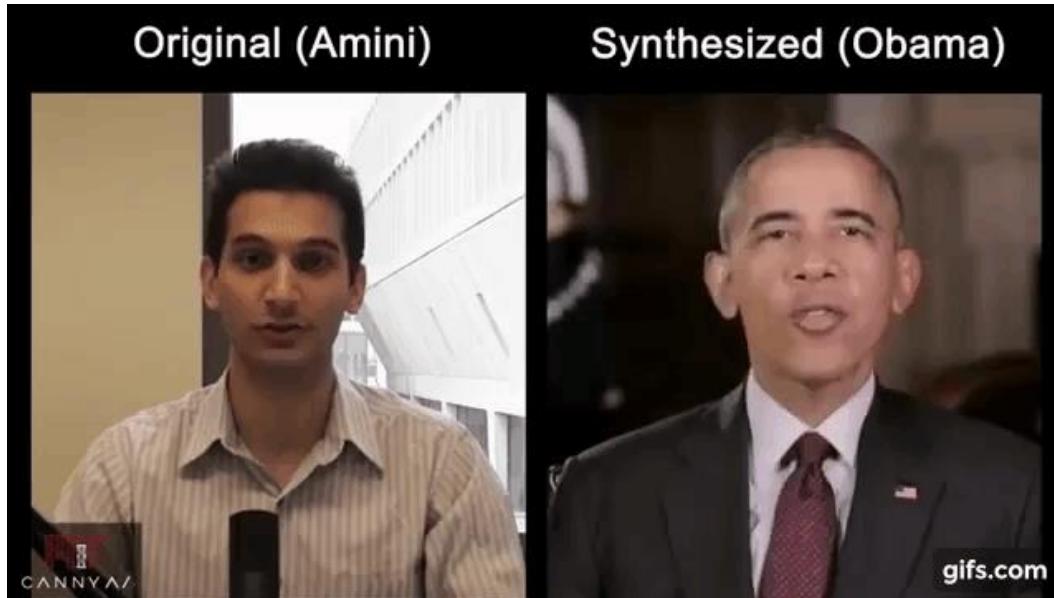
Every citizen will receive a so-called "Citizen Score", which will determine under what conditions they may get loans, jobs, or travel visa to other countries. This kind of individual monitoring would include people's Internet surfing and the behavior of their social contacts

### UK:

In 2015 when details of the British secret service's "Karma Police" program became public, showing the comprehensive screening of everyone's Internet use.

**Is Big Brother a reality? Programmed society, programmed citizens**

A deepfake ("deep learning" and "fake") are synthetic media in which a person in an existing image or video is replaced with someone else's likeness.





## **Bad Data**

- Bad data causing bad outcomes.

## **Malicious data**

- Malicious data could cause malicious outcomes.

## **Transparency**

- The ‘black box’ problem. Explanation and interpretation.

## **Misuse**

- AI system is not transparent, it is (somewhat) unpredictable;

## **Data curation for training**

e.g. has bias been addressed, good quality, relevant?

## **Ethical procurement**

was cheap labour involved, was copyright infringed?

## **Carbon implication**

how much computing power was used to build it and how much will be needed to run it?

## **Key performance indicators**

how good/useful is the AI?

## **Economic exclusion**

“Responsible AI is really all about the how: how do we design, develop and deploy these systems that are **fair, reliable, safe** and **trustworthy**. And to do this, we need to think of Responsible AI as a set of socio-technical problems. We need to go beyond just improving the data and models. We also have to think about the people who are ultimately going to be interacting with these systems.”

*Saleema Amershi, Principal Researcher at Microsoft Research and Co-chair of the Aether Human-AI Interaction & Collaboration Working Group*

It should be taken into account:

- AI applications are systems designed and created by humans, they are in practice artifacts.
- We have to guarantee and make sure that its purpose, the objective for which it was built was in the 1<sup>st</sup> place guaranteed and in 2<sup>nd</sup> is in fact what we want.
- Don't forget it is an engineering creation ... an **artefact**

Key issues:

- Data;
- Autonomy;
- Learning;
- among others....

**In many applications areas, as we have seen, AI systems are better than humans (or can be or will be).**

## Responsibility in AI system

- **in** Design

Ensuring that development processes take into account ethical and societal implications of AI

- **by** Design

Integration of ethical reasoning abilities as part of the behaviour of artificial autonomous systems

- **for** Design(ers)

Research integrity of researchers and manufacturers, and certification mechanisms

Source: Virginia Dignum (2019) Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer Artificial Intelligence: Foundations, Theory, and Algorithms, <https://doi.org/10.1007/978-3-030-30371-6>.

- **Accountability**

- ability to explain and justify its decisions to users and other stakeholders

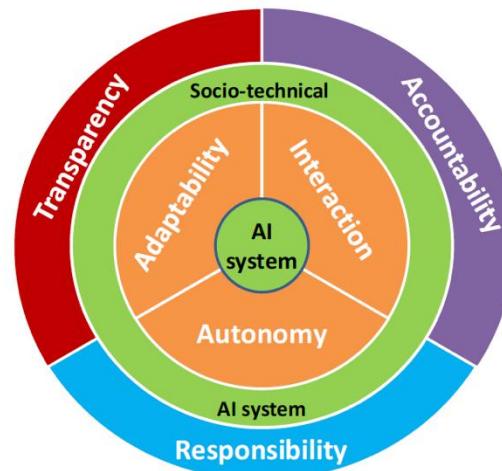
- **Responsibility**

- the role of people themselves in their relation to AI systems
  - is not just about making rules to govern intelligent machines
  - whole socio-technical system englobes people, machines and institutions

- **Transparency**

- ability to describe, inspect and reproduce the mechanisms through which AI systems make decisions and learn
  - will increase trustiness in system
  - explicit and open about choices and decisions concerning data sources and development processes and stakeholders

- Should be developed with responsibility and incorporating social and ethical values
  - **Autonomy** - Responsibility
  - **Adaptability** - Transparency
  - **Interaction** - Accountability



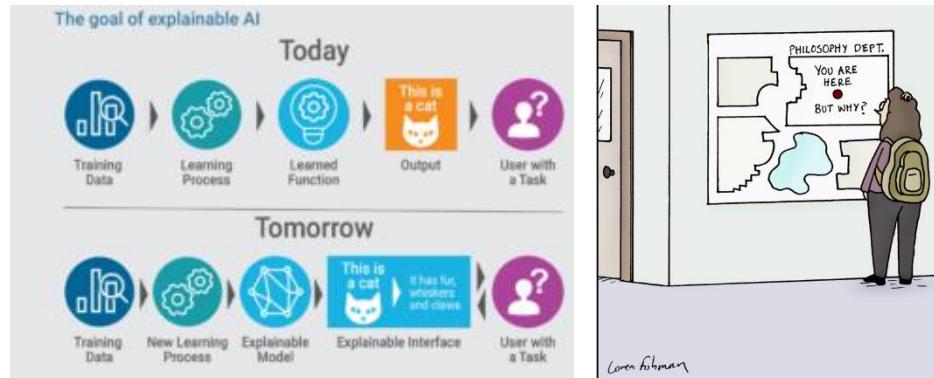
Source: Virginia Dignum (2019) Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer Artificial Intelligence: Foundations, Theory, and Algorithms, <https://doi.org/10.1007/978-3-030-30371-6>.

## We need! Explainable AI

### (Explainable/Interpretable/Transparent)\* AI

We need to apply techniques which can be trusted and easily understood by humans (a transparent "black box").

- **Deep Explanation** — understand what a system is doing through introspection or justification;
- **Model Induction** — observing the behavior of a system and using that to infer the model that can be used to explain that behavior.



Source: [AI and Machine Learning: Key FICO Innovations](https://medium.com/@BonsaiAI/what-do-we-want-from-explainable-ai-5ed12cb36c07)

<https://medium.com/@BonsaiAI/what-do-we-want-from-explainable-ai-5ed12cb36c07>



Transparent, auditable, explainable systems ...

### Principles for AI development

- Fairness - AI systems should treat all people fairly
- Reliability & Safety - AI systems should perform reliably and safely
- Privacy & Security - AI systems should be secure and respect privacy
- Inclusiveness - AI systems should empower everyone and engage people
- Accountability - People should be accountable for AI systems
- Transparency - AI systems should be understandable

Source: Microsoft AI principles  
<https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primary>

### Asilomar principles: Ethics and Values

- **Safety:** AI systems should be safe and secure
- **Failure Transparency:** If an AI system causes harm, it should be possible to ascertain why
- **Judicial Transparency:** Provide a satisfactory explanation auditable by a competent human authority
- **Responsibility:** Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications
- **Value Alignment:** Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation
- **Human Values:** AI systems should be compatible with ideals of human dignity, rights, freedoms, and cultural diversity

Source: ASILOMAR AI PRINCIPLES

<https://futureoflife.org/ai-principles/>

- **Personal Privacy:** People should have the right to access, manage and control the data they generate
  - **Liberty and Privacy:** The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
  - **Shared Benefit:** AI technologies should benefit and empower as many people as possible
  - **Shared Prosperity:** The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
  - **Human Control:** Humans should choose how and whether to delegate decisions to AI systems
  - **Non-subversion:** The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends
- **AI Arms Race:** An arms race in lethal autonomous weapons should be avoided

Source: ASILOMAR AI PRINCIPLES  
<https://futureoflife.org/ai-principles/>

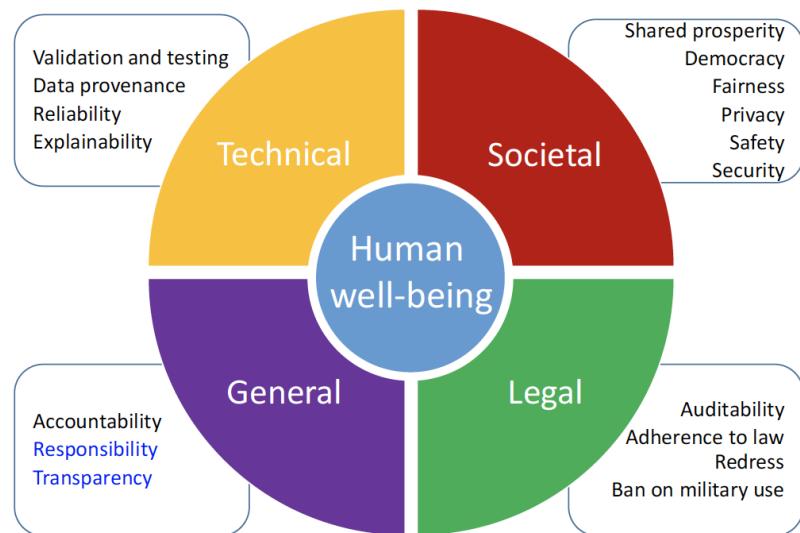
### **The Barcelona declaration for the proper development and usage of artificial intelligence in Europe**

- **Prudence** - The leap forward in AI has been caused by a maturation of AI technologies, but we must be aware of the still existent limitations
- **Reliability** - Determine AI systems reliability and security., particularly in domains like medicine or autonomous robots
- **Accountability**
- **Responsibility**
- **Constrained Autonomy**
- **Human Role** - All AI systems critically depend on human intelligence

Source: The Barcelona declaration for the proper development and usage of artificial intelligence in Europe  
<https://content.iospress.com/articles/ai-communications/aic180607>

“is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human flourishing and well-being in a sustainable world.”

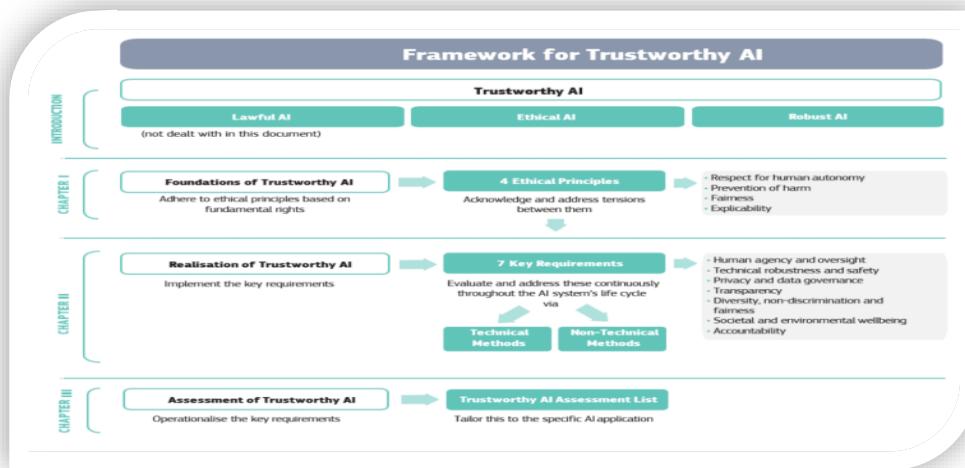
Dignum, 2019



Source: Virginia Dignum (2019) Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer Artificial Intelligence: Foundations, Theory, and Algorithms, <https://doi.org/10.1007/978-3-030-30371-6>.

- **AI confluence with other emergent technologies:**
  - Blockchain, IoT, Quantum Computing, etc.
- **Emerging Technologies in Schools:**
  - Virtual Reality, Augmented Reality
- **Wise AI:**
  - (Value learning, Anomaly detection, Fairness, Governance and policy, ...).
- **AI with Ethics (Ethical AI):**
  - IEEE global initiative on ethics - Global initiative for ethically aligned design of autonomous and intelligent systems: Legal accountability; Transparency; Policies; Embedding values into AI applications; Governance frameworks.
  - EU HIGH LEVEL EXPERT GROUP ON AI

# The Guidelines as a framework for Trustworthy AI



Source: EU - The Guidelines as a framework for Trustworthy AI  
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

## Data protection and privacy

- Data is a valuable asset for the organisations that collect it, as it can be used for advertising, marketing, data analysis and product development...
- The European Union implemented the **General Data Protection Regulation (GDPR)** in 2018, which establishes clear rules for the collection, use and protection of personal data in the EU. This has led companies around the world to adopt stricter privacy and data protection policies to comply with the GDPR.
- Organisations must develop and adapt technologies that protect data (particularly personal data) through data obfuscation and confidential computing, which allow data to be used

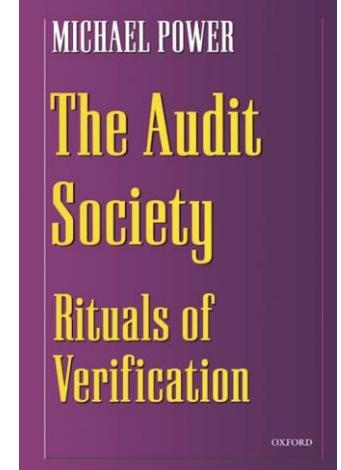


- Regulation of the European Parliament and of the Council establishing harmonised rules for Artificial Intelligence (2021):
- The regulation aims to establish:
  - Harmonised rules for the placing on the market, putting into service and use of artificial intelligence systems in the EU;
  - Bans on certain AI practices;
  - Specific requirements for high-risk AI systems and obligations for operators of such systems;
  - Harmonised transparency rules for certain AI systems;
  - Market surveillance and oversight rules.

# THE AI ACT

---

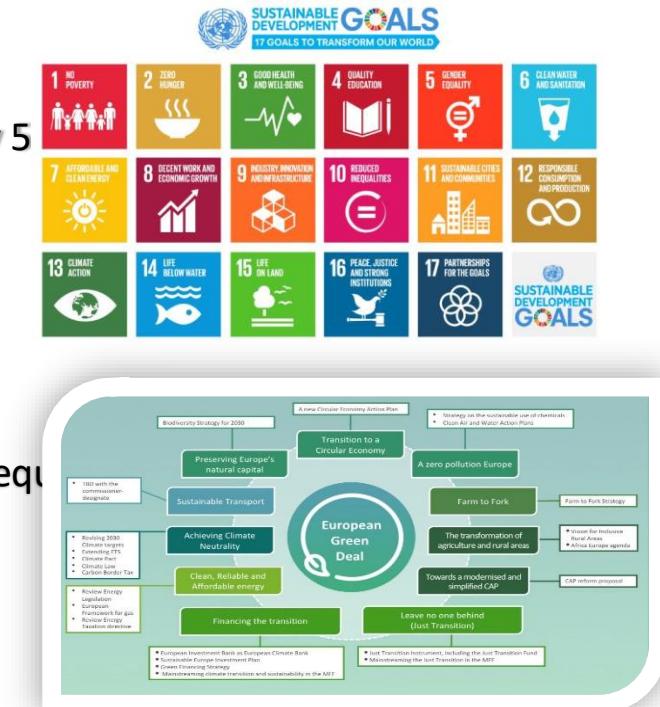
- To avoid **stifling innovation**, regulation must be proportionate and adapted to the nature of the sector and the stage of development of technologies;
- The **approach** must be **flexible** and allow for rapid adaptation to changes and technological developments;
- Regulators can also seek the **involvement of technology experts** to help develop more accurate and appropriate regulations;
- **Companies** operating in these emerging sectors should also be encouraged to **participate in the regulatory process** and provide constructive feedback on proposed regulations;
- An environment of collaboration between public administration, companies, regulators and other stakeholders to promote responsible innovation and ensure that new technologies are developed safely and ethically.



From a 4.0 Digital Society to a Society 5.0 in which Artificial Intelligence extend human capabilities and address social challenges

## Example: Social problems identified by Japan for Society 5

- Reduction of Greenhouse gas emissions
  - Increase production and reduce loss of foodstuffs
  - Migration of costs associated with ageing society
  - Promotion of sustainable industrialization
  - Redistribution of wealth and correction of regional



## The Education, Ethics and AI framework.



Source: Southgate, E., Blackmore, K., Pieschl, S., Grimes, S., McGuire, J. & Smithers, K. (2018). *Artificial intelligence and emerging technologies (virtual, augmented and mixed reality) in schools: A research report*. Newcastle: University of Newcastle, Australia.

## Main dimensions of the LASI AI strategy

### DIGITAL INFRASTRUCTURE

- We need to develop high quality digital infrastructure and to invest in new digital services and applications.

### PEOPLE

- People will be the true driving force of the AI transformation strategy.
- Increased AI proficiency in the population:
  - Collaboration with Higher Education Institutions in order to ensure the availability of qualified talent;
  - Training and reconversion actions for general people;
  - We need to attract and retain AI talent.

### ECONOMY AND BUSINESSES (Industry 4.0)

- We need urgently to increase the use of AI technologies by companies and organisations;
  - The creation of new startup companies that will explore this new business opportunity;
  - Public and private funding for AI investments;
  - The capacity of adapt the traditional jobs to this new environment;
  - Adjust human resources skills to the needs of the labour market.

### COMMUNITY (Society 5.0)

- Human-centered society;
  - Citizens and organizations will improve their living standards by increasingly using AI technologies in their social, leisure and cultural activities
  - Engagement of the society and the community to this opportunity.

- Now we need to develop other skills:
  - Curiosity.
  - Imagination.
  - Creativity.
  - Ambiguity.
- Common sense!

Recognizing situations is harder than recognizing single objects.

- **Automation of repetitive tasks:** With AI, routine and repetitive tasks can be automated, allowing professionals to focus on more complex and higher-value activities. This may include tasks such as software testing, report generation, system configuration, and more.
- **Process optimization:** AI can be used to analyze large amounts of data and identify patterns, thus optimizing software development processes. For example, machine learning algorithms can be applied to improve efficiency in code debugging or identify performance bottlenecks.
- **Assistance in software development:** AI-based tools can assist Computer Engineering professionals in various stages of software development. This includes coding suggestions, automatic error detection, documentation generation, and even user interface design.
- **Advanced data analysis:** AI can help extract valuable information and insights from large datasets. Computer Engineering professionals can use AI techniques such as data mining and machine learning to analyze complex data and make informed decisions based on these insights.
- **Enhancement in cybersecurity:** AI can be applied to detect and prevent cyber threats. AI algorithms can analyze real-time behavioral patterns, identifying suspicious or anomalous activities that may indicate a possible security breach.
- **Development of AI systems:** Computer Engineering professionals can be involved in the development of AI systems, designing and implementing machine learning algorithms, neural networks, and other AI methods. They can work on data collection and preparation, model training, and performance evaluation of these systems.

## **The jobs of the future will involve the creation of knowledge and innovation**

- **AI Knowledge**
  - It is crucial to understand the concepts and fundamentals of AI, including Machine Learning. This will enable you to grasp the capabilities and limitations of AI, as well as its applications in different fields.
- **Critical Thinking and Creativity**
  - While AI can perform many tasks, human critical thinking and creativity remain the fundamental skills.
- **Lifelong Learning**
  - AI is continuously evolving, so it's important to stay updated on the latest trends and advancements in this field.
- **Ethical Thinking**
  - Developing ethical awareness and the ability to make ethical decisions related to AI are valuable competencies.
- **Interpersonal Skills**
  - The ability to work in teams, collaborate with others, communicate complex ideas clearly, and solve problems collaboratively are skills that add value, especially in projects involving the implementation of AI systems.
- **Adaptability and Flexibility**
  - Having a **multidisciplinary approach** is crucial in an AI-dominated world, as it allows for a holistic understanding and effective problem-solving. Integrating insights from various fields enables comprehensive solutions and promotes collaboration among experts.

- We cannot replace empathy;
- Complex digital technologies require competent professionals;
- There will always be tasks algorithms and robots can never complete.

## How computers imagine humans...



Source: João Martinho Moura, 2017

- **“Collaboration between humans and technology is the ultimate response.”**

Source: <https://medicalfuturist.com/5-reasons-artificial-intelligence-wont-replace-physicians/>

**One thing is clear**

**The way in which we organize society will change deeply.**

**If we take the wrong decisions it could threaten our greatest historical achievements and (probably) our life.**

Source: Will Democracy Survive Big Data and Artificial Intelligence? By Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter on February 25, 2017.  
[https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/?WT\\_mc\\_id=SA\\_SP\\_20170227](https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/?WT_mc_id=SA_SP_20170227)



The New Yorker – October 23, 2017

- Virginia Dignum (2019) Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer Artificial Intelligence: Foundations, Theory, and Algorithms, <https://doi.org/10.1007/978-3-030-30371-6>.
- Walsh, Toby, (2018), Machines That Think: The Future of Artificial Intelligence, Amherst, MA: Prometheus Books.
- Floridi, L., (2016) Should we be afraid of AI? Aeon Essays, <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>.

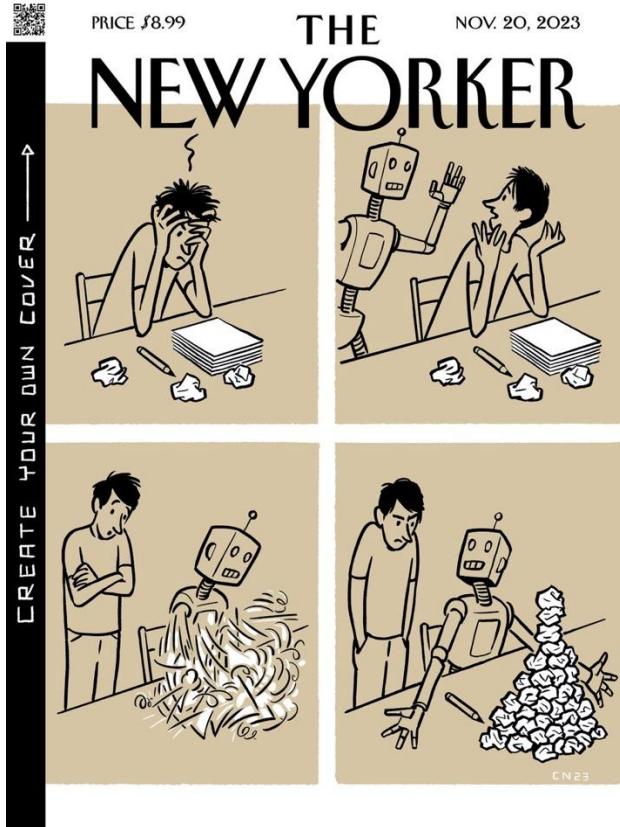
**"If you don't know where you're going, any road will do"**



Source: Lewis Carroll, Alice in Wonderland

“Cat: Where are you going?  
Alice: Which way should I go?  
Cat: That depends on where you are going.  
Alice: I don't know.  
Cat: Then it doesn't matter which way you go.”

**"If you don't know where you're going, any road will do"**





**Universidade do Minho**  
Escola de Engenharia  
Departamento de Informática

# **Inteligência e a Sociedade**

## **A Responsible AI for Social Good**

**LICENCIATURA EM ENGENHARIA INFORMÁTICA**  
**MESTRADO integrado EM ENGENHARIA INFORMÁTICA**

**Inteligência Artificial**  
**2025/26**