



# An introduction to Responsible Artificial Intelligence

Professor Ciprian Daniel Neagu

Professor of Computing

AI Research (AIRe) Group Leader

Ciprian Daniel Neagu | LinkedIn

Ciprian Daniel Neagu, PhD - Google Scholar

Ciprian Daniel Neagu - ORCID

*Delivery based on resources produced by a team of AI  
Research (AIRe) students and experts including:*

Dr Paul Trundle  
Dr Amr Abdullatif  
John Marko, PhD student

## In this Webinar we will:

### - Introduce:

- the Big Data rationale
  - the Artificial Intelligence history
- and
- their main concepts

that are the foundations of Artificial Intelligence (AI)

### - Review:

- State of the Art and Reality
- Expectations
- Challenges
- Dilemmas

for Responsible AI

### - Conclude by looking forward to the Future of Technology and Society

# Concepts: General Definitions

- **Data** is defined as facts regarding things (such as people, objects, events) which can be **digitally stored, transmitted or processed**.
- **Information** is generally defined as **data** that have been processed and presented in a form suitable for **human interpretation** with the purpose of revealing **meanings** (such as patterns or rules).
- **Knowledge**: the theoretical and practical comprehension of a certain domain (objects, concepts and relationships) that supports making decisions.
- **Intelligence**: the capability of learning, understanding and finding solutions for problems in a specific domain.
- **Models** are compact representations of patterns.



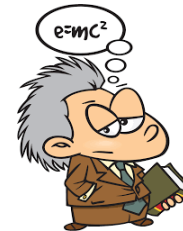
The bank balance of £1234567.89 is **data**.



"balance has jumped 8087% to £1234567.89" is **information**.



"Nobody owes me that much money" is **knowledge**.



"I'd better talk to the bank before I spend it, because of what has happened to other people" is **intelligence**.



"why did this happen, let's define the relationship between cause and outcome" is **modelling**

<https://foldoc.org/> formerly <http://foldoc.doc.ic.ac.uk>

**Intelligence** is a complex and multifaceted ability to:

- acquire
- understand,
- process,
- apply,
- and retain

**Knowledge.**

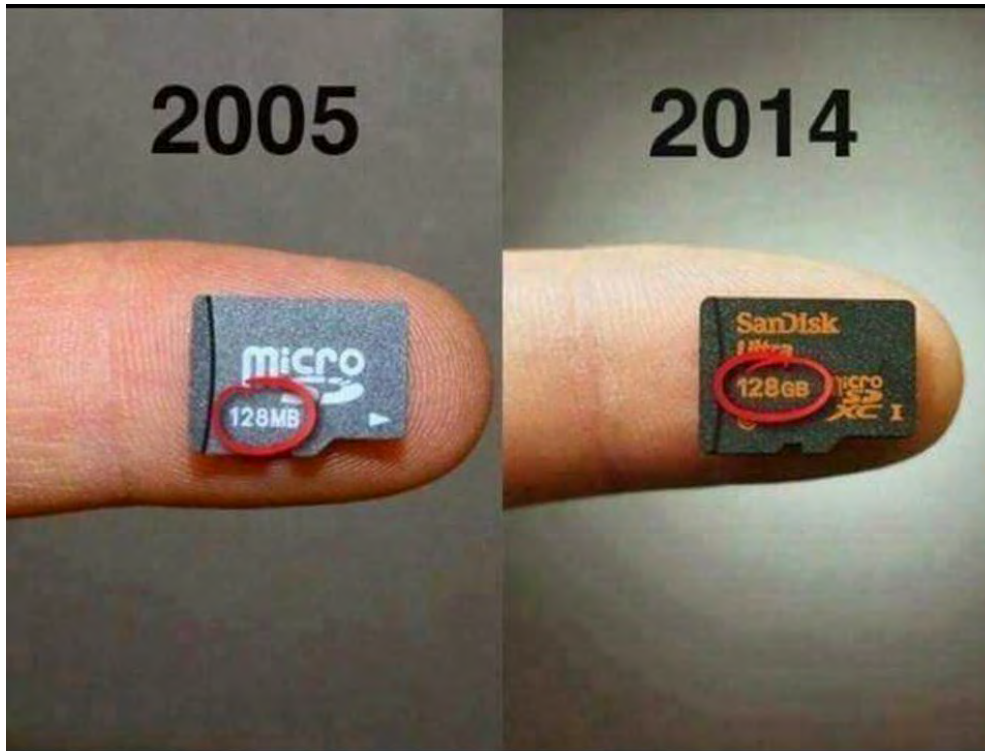
**Intelligence** encompasses a wide range of mental abilities and skills, allowing individuals to:

- adapt to their environment,
- solve problems,
- learn from experience,
- reason,
- and engage in abstract thinking.

## Technology Catalysts

By 2000 there will be "computers with a storage capacity of about  $10^9$ " bits (128MB): Alan Turing, [Computing Machinery and Intelligence](#), 1950)

The (hard drive) storage capacity ([Kryder's Law](#)) follows a similar progress with the increase rate in transistor density ([Moore's Law](#)).



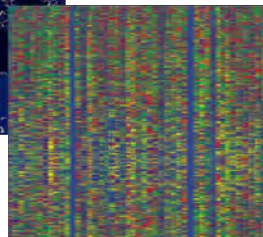




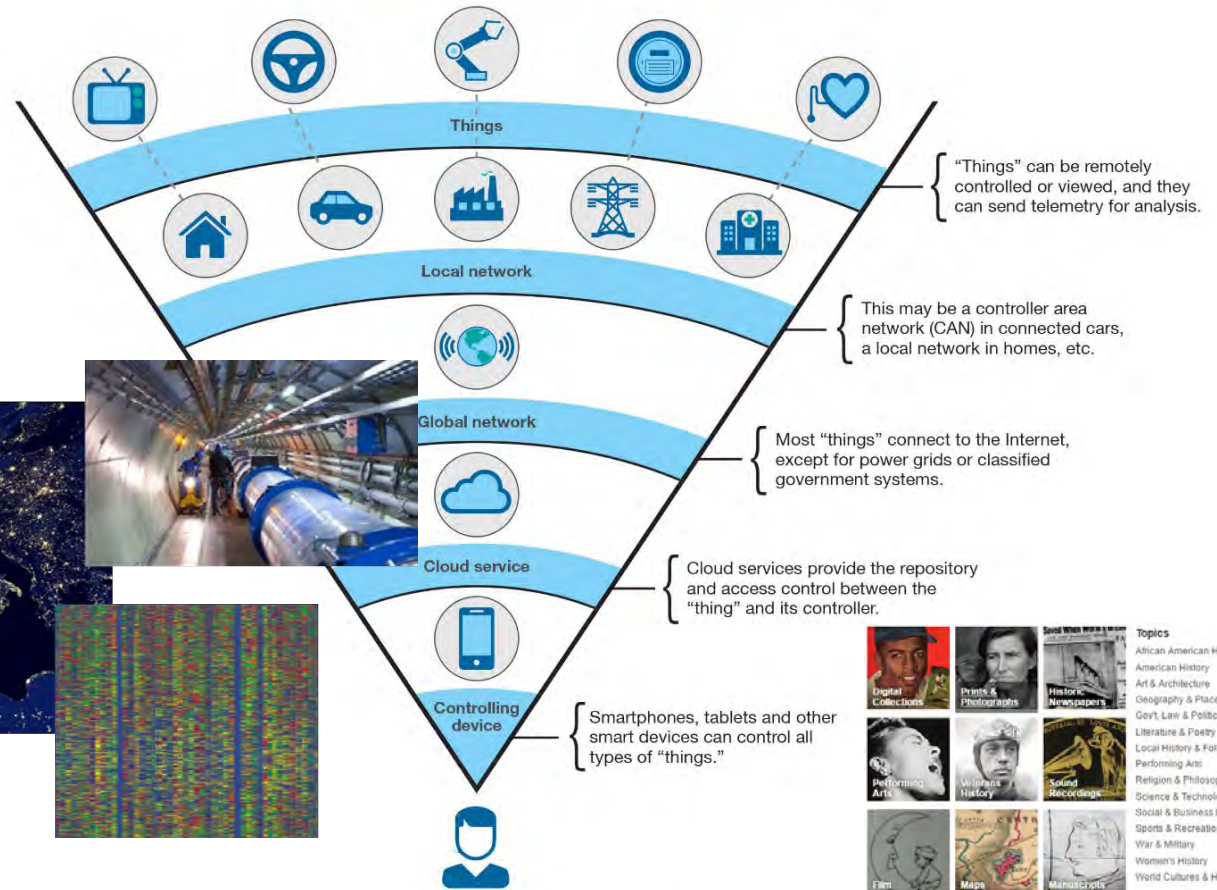
# The Rise of Big Data

## Big Data Catalysts:

- IoT
- IoX (Internet of Everything)
- Smart Cities
- Online Social Networks
- Public data resources (Wikipedia, Human Genome, Weather, Government, CERN, UN, libraries, museums, PubMed, arxiv, ...)
- Private data resources (personal, institutional, sensitive)



IBM model for the Internet of Things

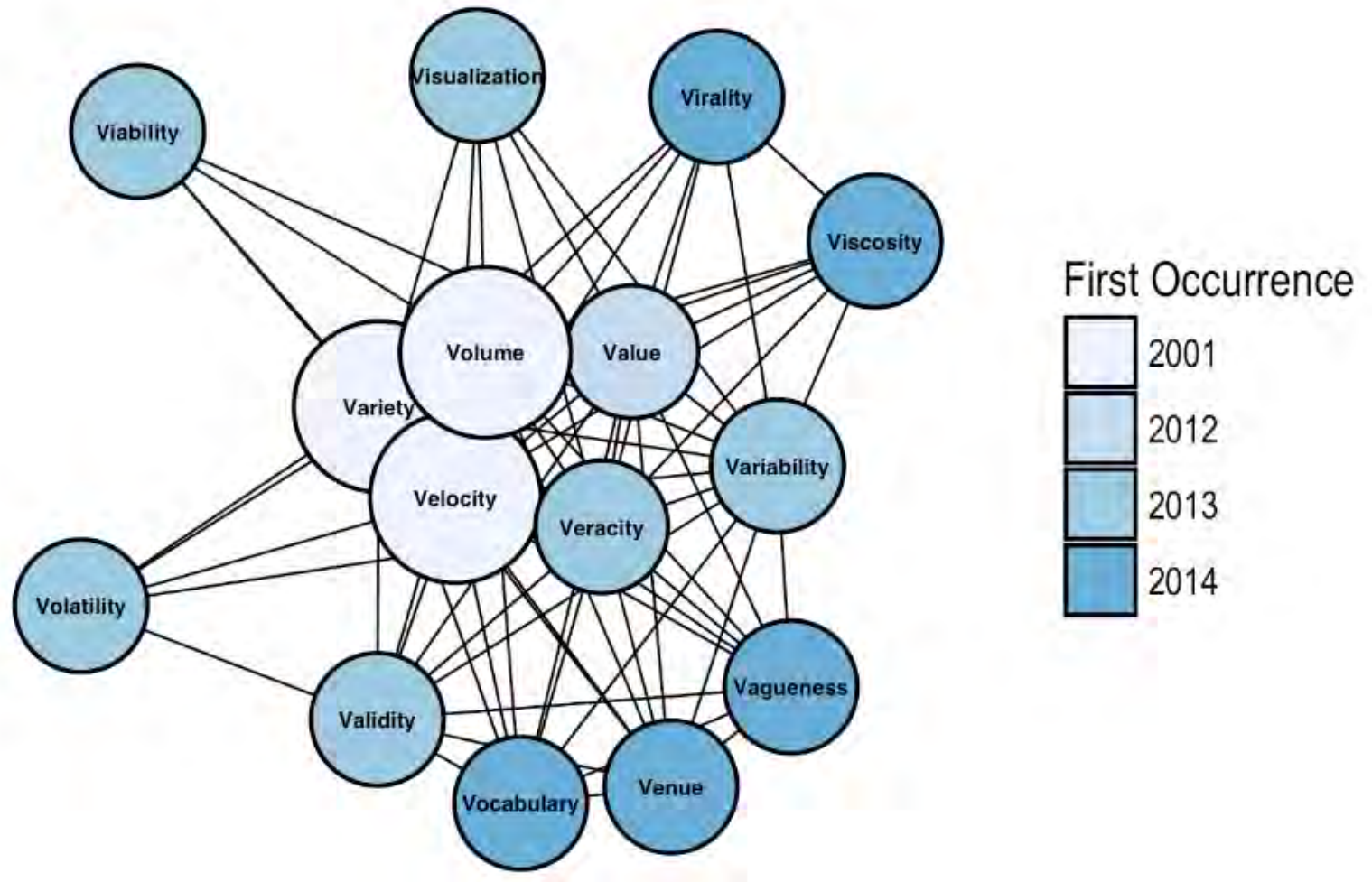


Graphic 1. IBM model for the Internet of Things

Source: IBM X-Force® Research and Development



# The 42 V's of Big Data



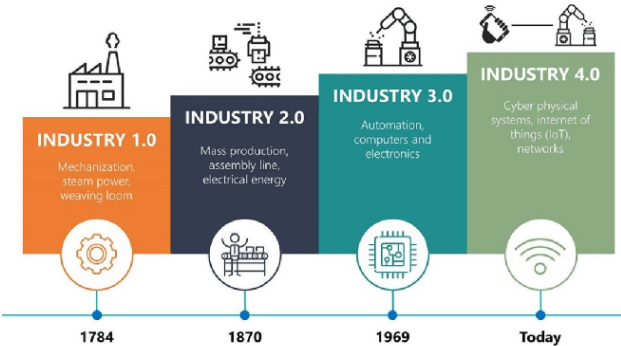
Tom Shafer, 2017

<https://www.elderresearch.com/blog/42-v-of-big-data>



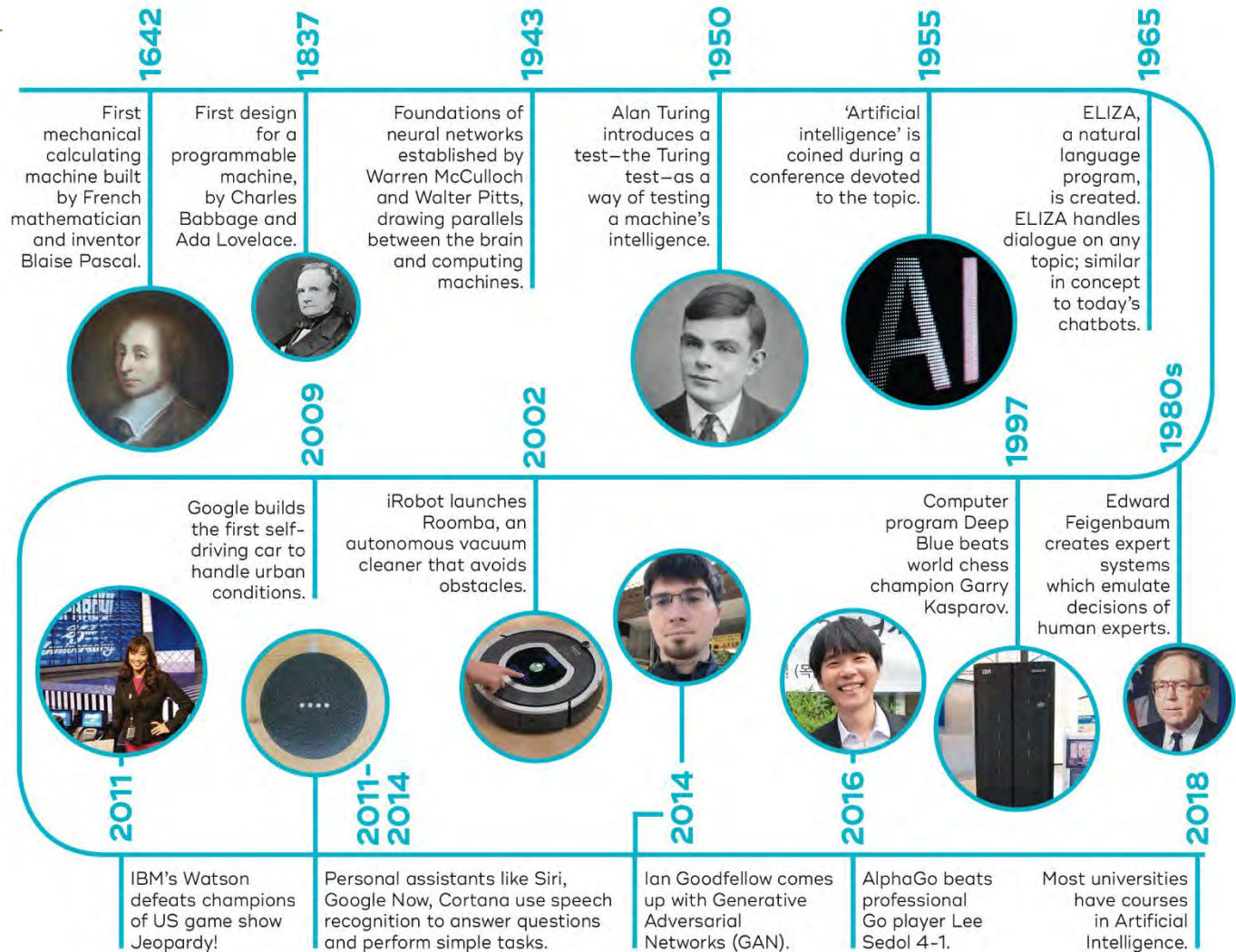


# History of Artificial Intelligence?



Optimization of the Job Shop Scheduling Problem in Industry 4.0 (researchgate.net)

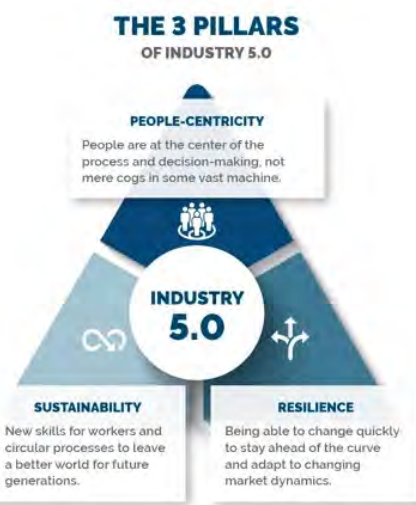
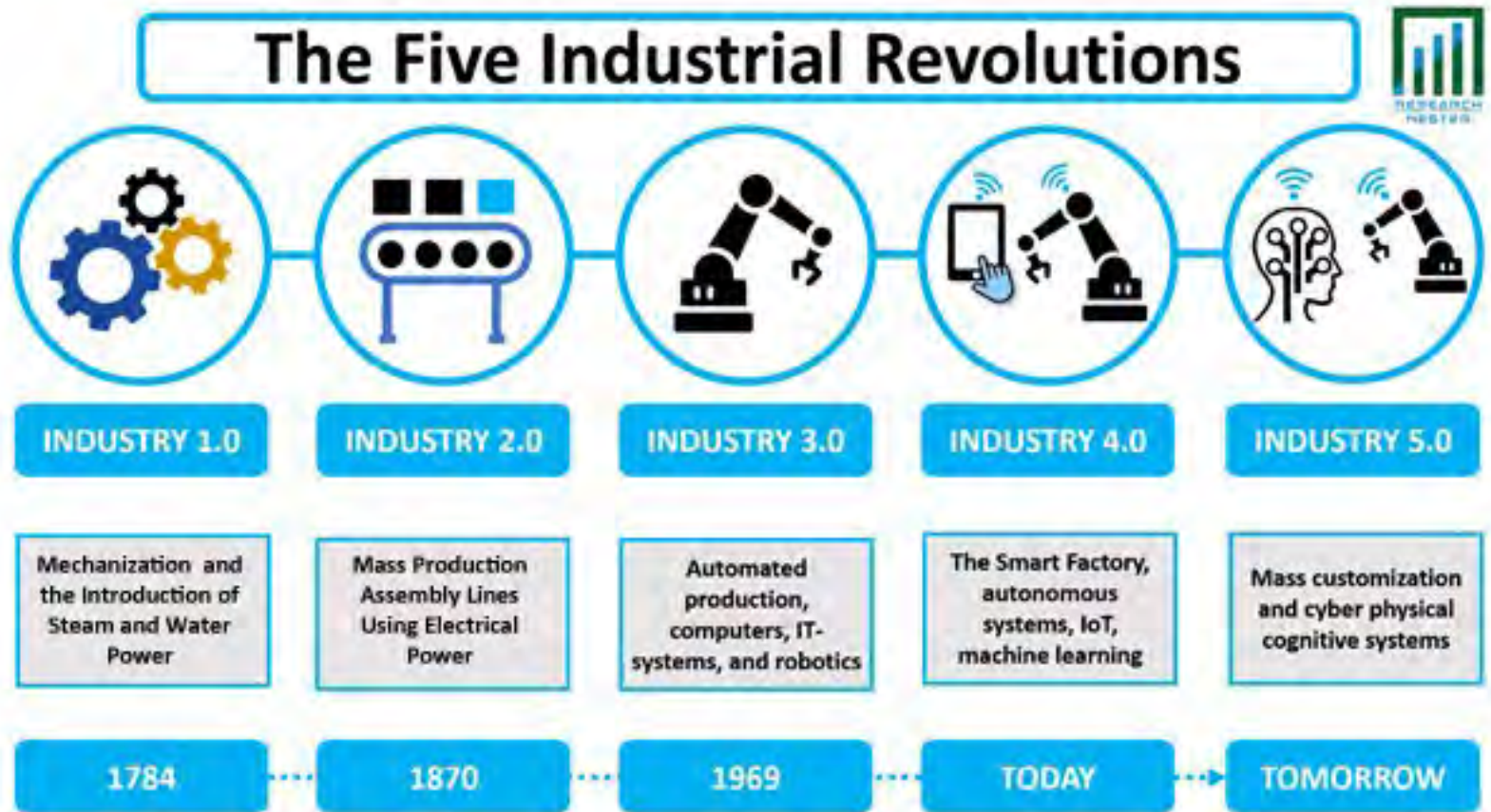
History of Artificial Intelligence - Queensland Brain Institute - University of Queensland (uq.edu.au)







# The 5 Industry Revolutions



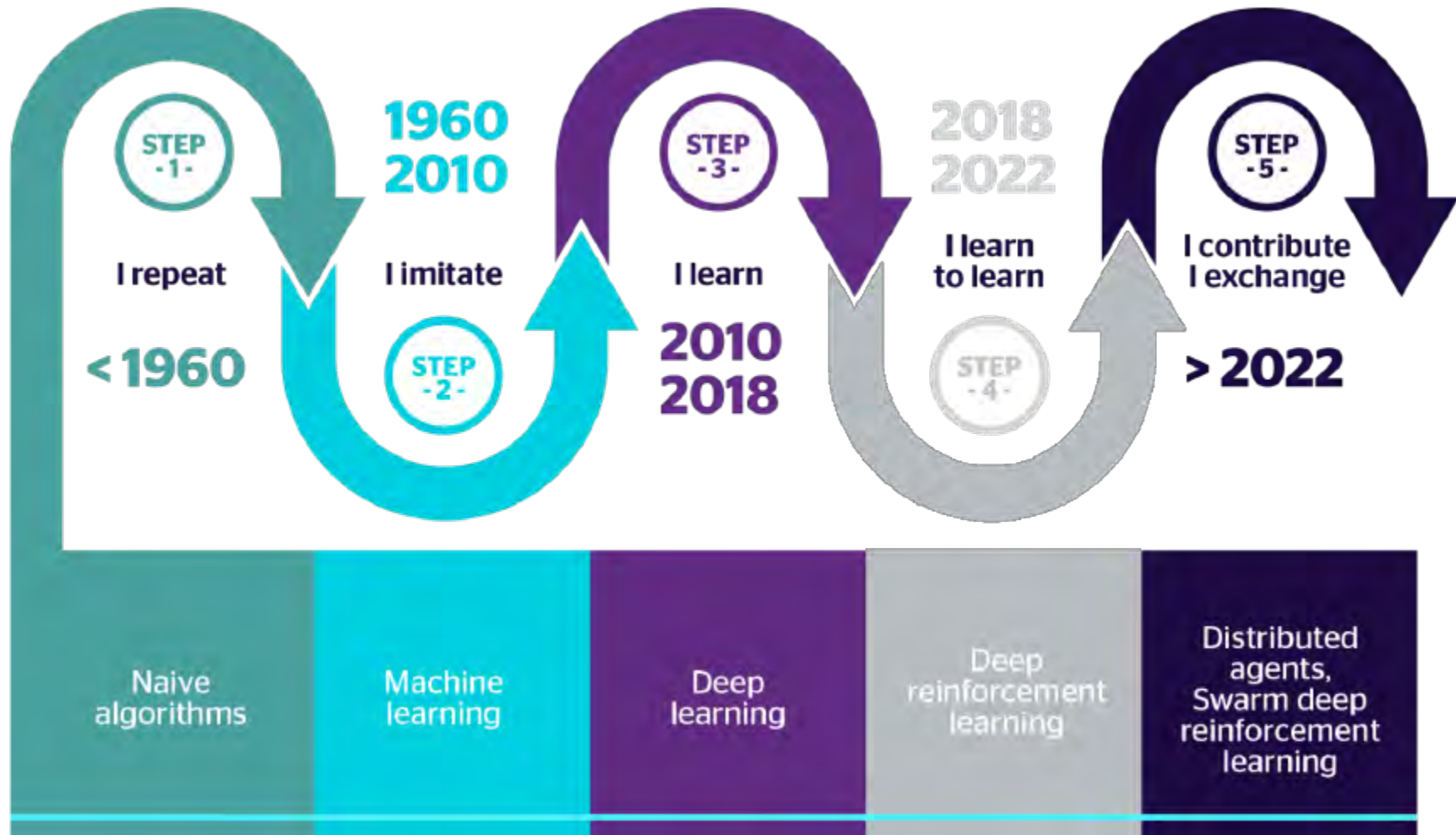
[Industry 5.0 \(europa.eu\)](https://www.europa.eu)

Source: Research Nester

[Industry 5.0 Market Size, Share, Growth And Global Trends Analysis. 2030 \(researchnester.com\)](https://www.researchnester.com)



# What is Artificial Intelligence?

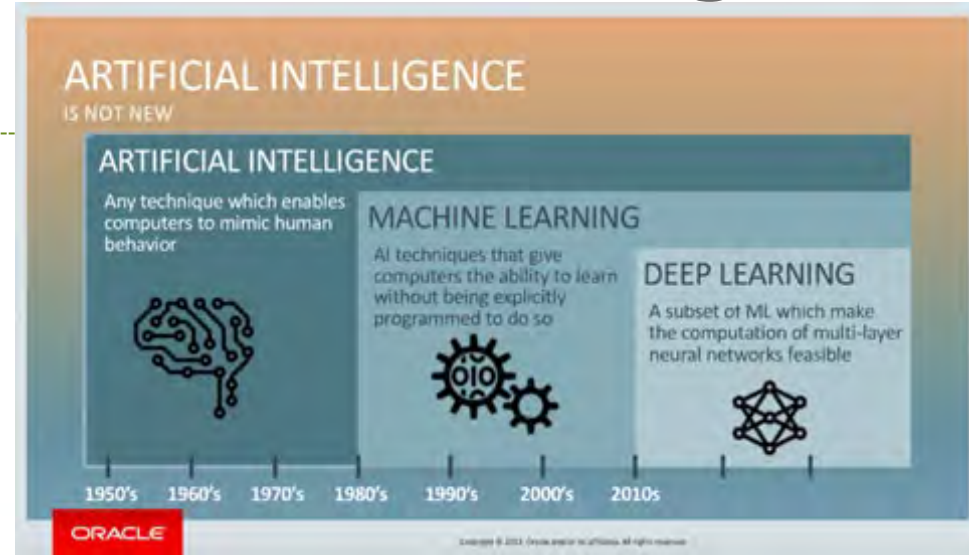


History of Artificial Intelligence with Briefed Examples - Defining AI

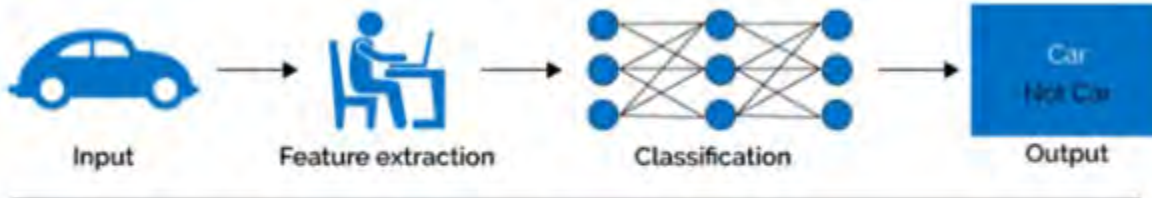


# What is Artificial Intelligence?

AI refers to the **(weak/strong)** ability of computer systems to perform tasks that would normally require human intelligence to accomplish: learning, problem-solving, decision-making, perception, NLP. AI algorithms can process large amounts of data, identify patterns, and make predictions or decisions based on that data. (Ref: ChatGPT 3.5)



## Machine Learning



## Deep Learning



- Labeled data
- Direct feedback
- Predict outcome/future

- No labels
- No feedback
- "Find hidden structure"

- Decision process
- Reward system
- Learn series of actions

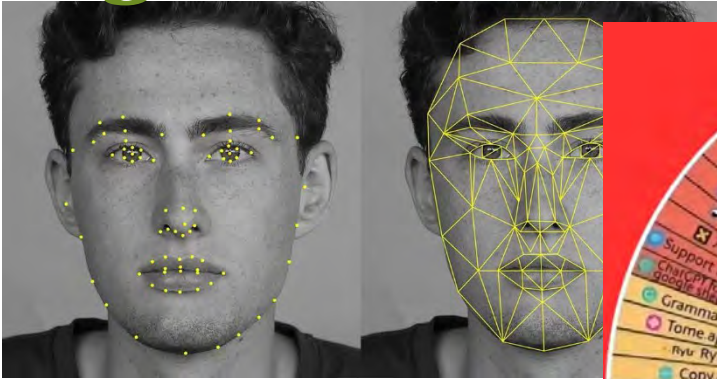
<https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning>

[https://www.slideshare.net/SebastianRaschka/nextgen-talk-022015/8-Learning\\_Labeled\\_data\\_Direct\\_feedback](https://www.slideshare.net/SebastianRaschka/nextgen-talk-022015/8-Learning_Labeled_data_Direct_feedback)

<https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c>



## A great addition to humanity ever known?

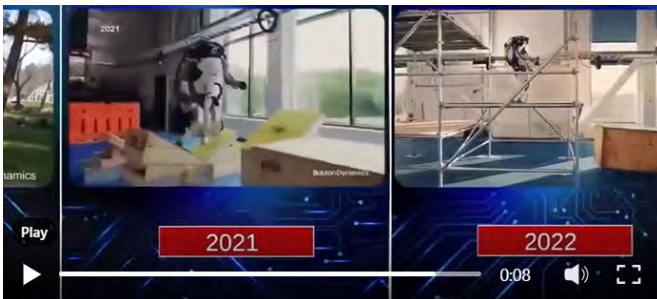
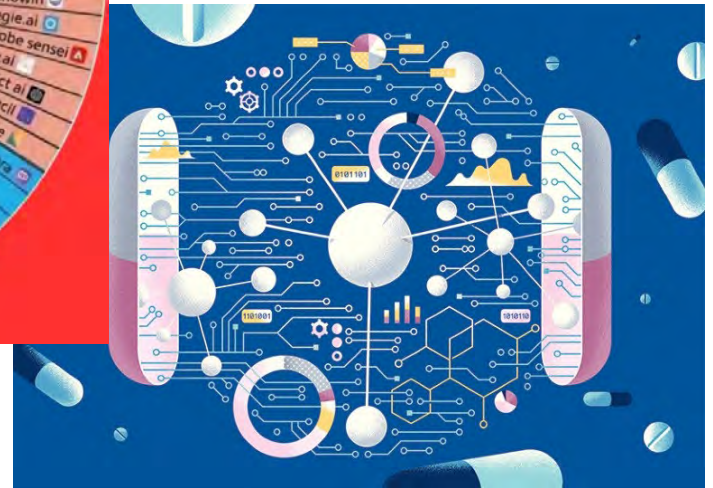


Mind-Blowing AI Tools for Video, Productivity, Marketing, Chatbot, Design & Writing | LinkedIn



Proving AI in the Clinic: An Algorithm That Accurately Evaluates Heart Failure (stanford.edu)

How artificial intelligence is changing drug discovery



Evolution of Boston Dynamics: 40 Years of Development ( 1983 - 2023 ) - YouTube





## A great threat to humanity ever known? Creative or Lying?

*What book written by Professor Daniel Neagu focuses on machine learning?*

Neagu's book on machine learning. ([openai.com](https://openai.com))

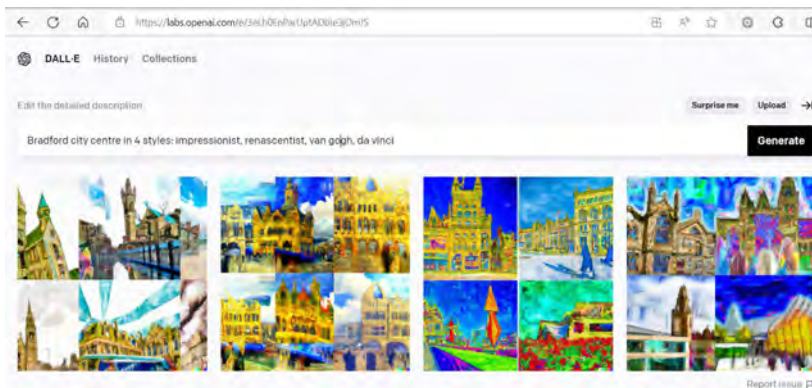


# What is Artificial Intelligence?

## A great threat to humanity ever known?

[DALL·E 2 \(openai.com\)](https://openai.com)

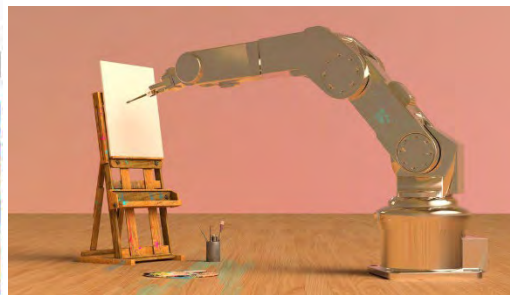
[Renascentist image of Bradford?](#)



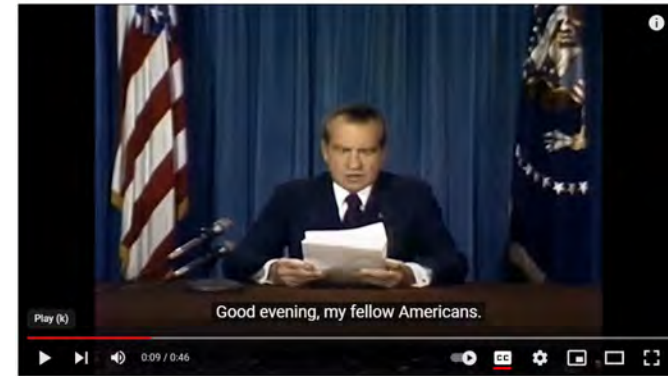
[AI: Digital artist's work copied more times than Picasso - BBC News](#)

[Artists fight AI programs that copy their styles](#)

["Art is dead Dude" - the rise of the AI artists stirs debate - BBC News](#)



[Deep Fake Videos and Audios](#)



[In Event of Moon Disaster - Nixon Deepfake Clips](#)

<https://www.bbc.com/news/uk-66220781>

[AI quiz: Can you tell which person is real?](#)

17 July



How much do you know about Artificial Intelligence? As the technology rapidly advances, test your knowledge of how AI affects life now and its possible impacts in the near future.



# What is AI? Turing Test time!

https://www.bbc.com/news/uk-66220781

## AI quiz: Can you tell which person is real?

17 July

<https://www.bbc.com/news/uk-66220781>



How much do you know about Artificial Intelligence? As the technology rapidly advances, test your knowledge of how AI affects life now and its possible impacts in the near future.



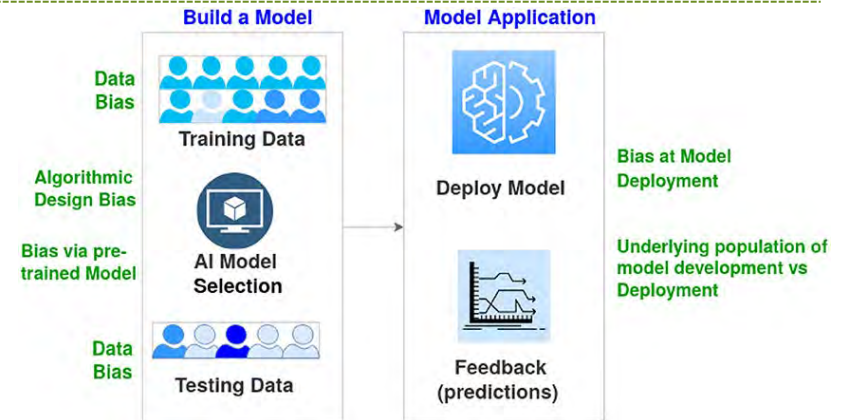


# What challenges are brought by AI?

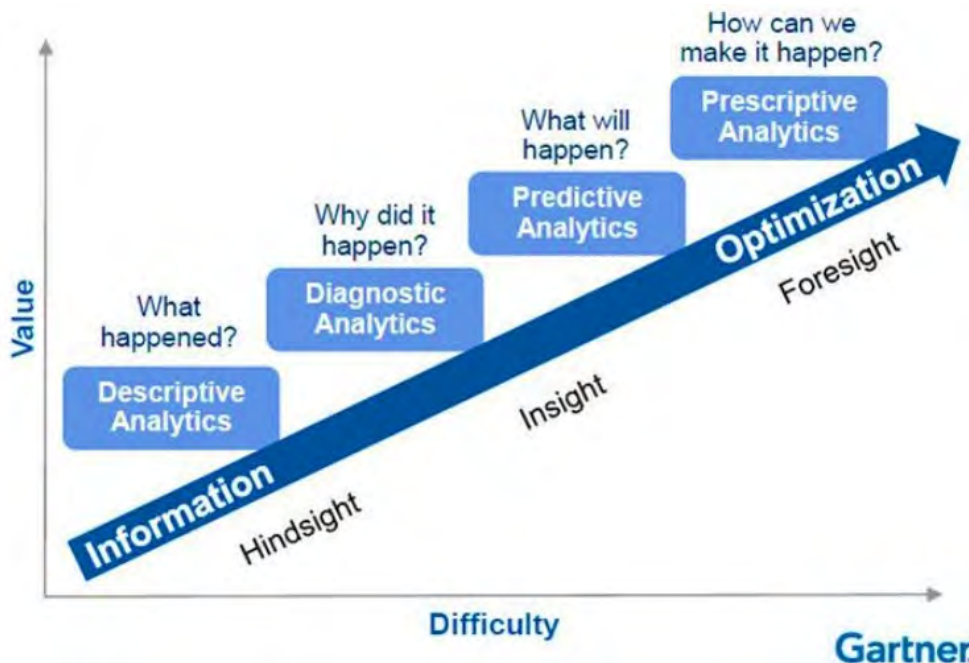
AI is a risk for human extinction!

Geoffrey Hinton, Bill Gates etc:  
Statement on AI Risk | CAIS (safe.ai)

Andrew Ng said the technology poses 'no  
extinction risk' for humans



Frontiers | Data and model bias in artificial intelligence for  
healthcare applications in New Zealand (frontiersin.org)



Artificial intelligence: Experts propose  
guidelines for safe systems - BBC News

What is AI, is it dangerous and what jobs  
are at risk? - BBC News

AI models fall short of draft EU rules,  
researchers say | Financial Times

Elon Musk claims more trust can be put in  
his xAI than OpenAI and Google | Financial  
Times (ft.com)



**Responsibility** refers to the duty or obligation of an individual or group to fulfil certain roles, tasks, or duties in a reliable and accountable manner.

Responsibility is an important aspect of **ethical and moral behaviour**, as it requires individuals to take ownership of their actions and acknowledge the impact they have on themselves, others, and the wider community.

Key aspects of Responsibility include:

**Accountability:** recognising and accepting the consequences of one's actions.

**Reliability:** trustworthy in fulfilling commitments, obligations and meeting expectations.

**Ethical Decision-making:** avoiding actions that harm others or violate societal norms.

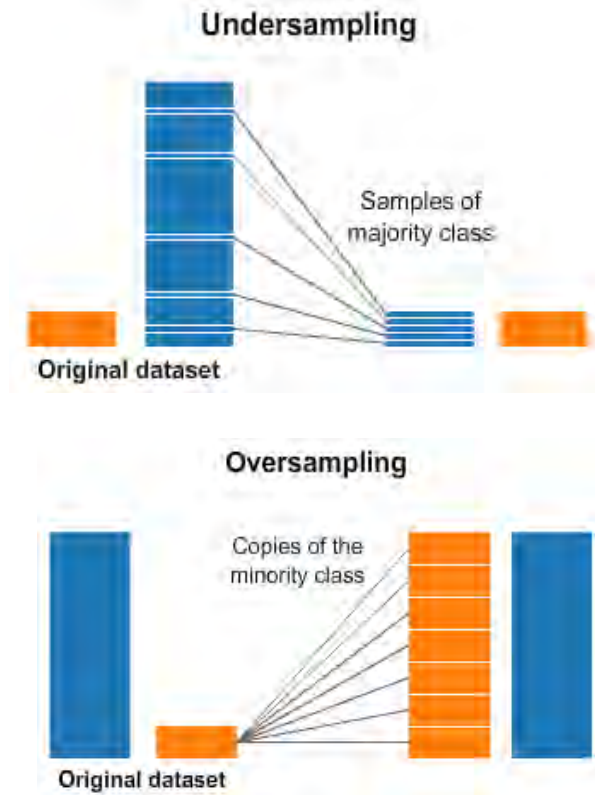
**Learning from Mistakes:** using them as opportunities for growth and improvement.

Responsibility plays a **significant role** in building **trust and respect** in relationships and contributes to a sense of **integrity and dignity** in individuals and communities.

## Where are the AI risks and challenges generated from?

AI is shifting from Human Expert Knowledge to Machine Learning Models:

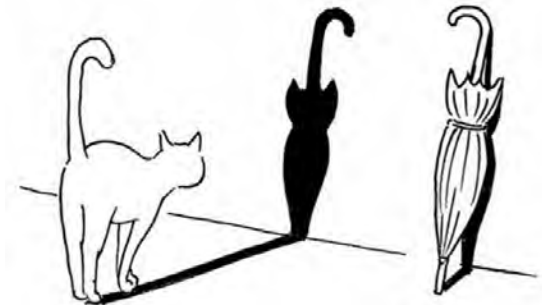
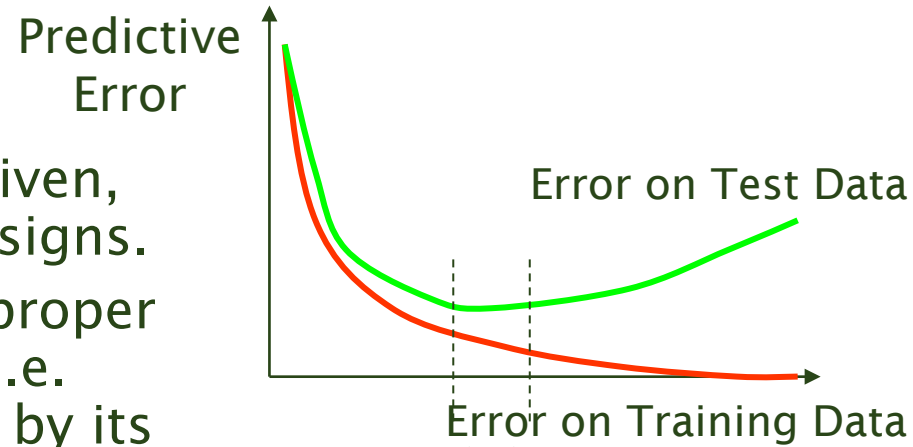
- **Quality and Relevant Data** to Any (publicly) available Big Data due to digital resource availability and business expectations, **using any imbalanced, biased, low quality, irrelevant training data**
- Statistical Learning grounds to Machine Learning automated solutions replacing Result Confidence with Model Output Performance and Accuracy
- Replacing Validation with Testing
- Expert Systems Industry Revolution to (Big) Data-driven/ -centric/ -enabled/ - enhanced AI models
- Lack of (Big) Data and AI Models Governance sustainable standards
- Decision support with robust models is replaced with Governing topics through numbers



## So, how are AI risks and challenges generated?

AI is shifting from Human Expert Knowledge to Machine Learning Models:

- Initial **Weak AI** models were problem driven, using strong statistical experimental designs.
- Many current AI models claim (without proper validation or evidence) to be **Strong AI** i.e. identical with Human Intelligence either by its organisation or its results.
- Misrepresentation and misunderstanding of Artificial Narrow Intelligence and Artificial General Intelligence:
- Why general artificial intelligence will not be realized | Humanities and Social Sciences Communications (nature.com)



TANGO

@tango2030weibo.tumblr.com

## First Law

*A robot may not injure a human being or, through inaction, allow a human being to come to harm.*

## Second Law

*A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.*

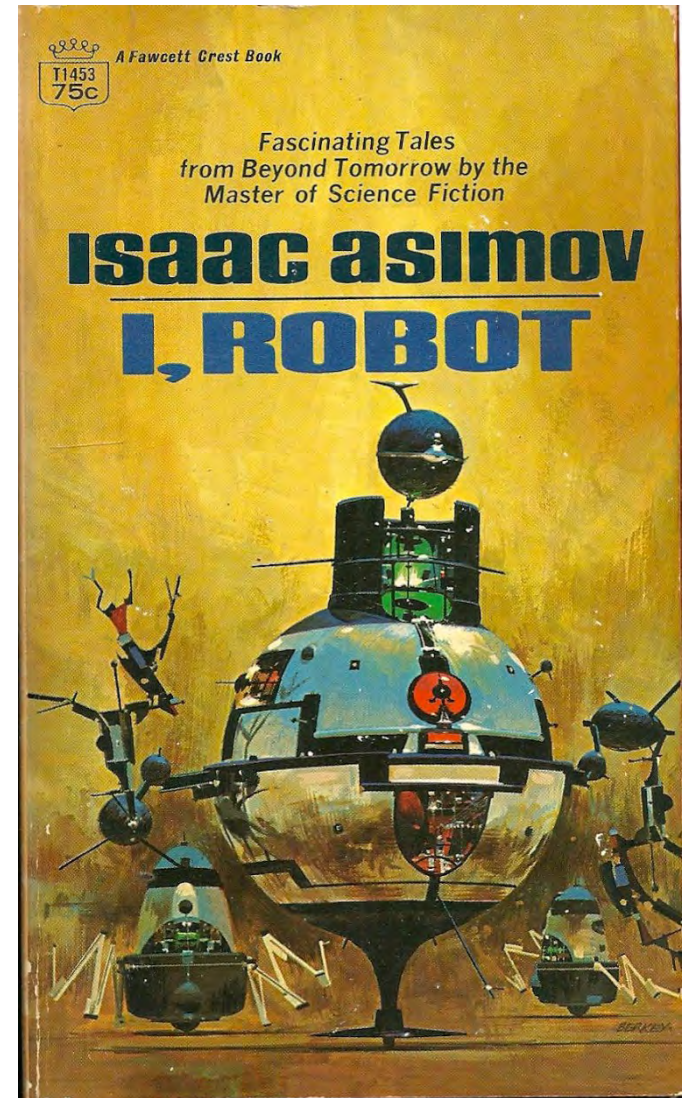
## Third Law

*A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

## Zeroth Law (added):

*A robot may not harm humanity, or, by inaction, allow humanity to come to harm.*

*Can we identify any Responsibility features?  
Accountability, Reliability, Ethics, Learning from Mistakes,  
Trust, Respect.*





Join at **menti.com** and use code **3295 3022**

Choose top 5 most relevant concepts for Responsible Artificial Intelligence from:

ACCOUNTABILITY

AWARENESS

BALANCE

CONSCIENCE

CONSIDERATION

EXPLAINABILITY

EFFECTIVENESS

EFFICIENCY

EMPATHY

ETHICS

FAIRNESS

RELIABILITY

RESPONSIBILITY

ROBUSTNESS

SAFETY

SUSTAINABILITY

TRANSPARENCY

TRUSTWORTHINESS

VISION

UNBIAS

<https://www.mentimeter.com/app/presentation/alxjbeux5fykiemoamtbeb8cw12ht4wd>

# What is Responsible AI?

RAI is the duty, obligation or expectation from technology (designers, owners or users) to provide or demonstrate features of:



# What is Responsible AI?



## **Responsible AI:**

[Edinburgh Declaration on Responsibility for Responsible AI | by Shannon Vallor | Jul, 2023 | Medium](#)

[SHIFTing artificial intelligence to be responsible in healthcare: A systematic review - ScienceDirect](#)

## **XAI: Explainable, Transparent**

[IBM AI explainability 360 toolkit](#)

[Explainable AI in Industry: Practical Challenges and Lessons Learned | Companion Proceedings of the Web Conference 2020 \(acm.org\)](#)

## **TAI: Trustworthy, Reliable, Privacy-preserving, Human-centric**

[Trustworthy AI \(itu.int\)](#)

[Generative Search Engines: Beware the Facade of Trustworthiness \(stanford.edu\)](#)

[Humane AI | Human-Centered Artificial Intelligence \(humane-ai.eu\)](#)

[How do you design AI that reflects human values? \(linkedin.com\)](#)



# What is Responsible AI?

## Ethical AI: Fair, Inclusive, Empathic, Unbiased, Accountable

<https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-techniques>

<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>

<https://artificialintelligenceact.eu/>

[Artificial intelligence: Partnership between UNESCO and the EU to speed up the implementation of ethical rules | UNESCO](#)

[Preliminary study on the Ethics of Artificial Intelligence - UNESCO Digital Library](#)

[Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library](#)

[Decoding the EU Artificial Intelligence Act \(stanford.edu\)](#)

[Institute for Technology, Ethics, and Culture - Markkula Center for Applied Ethics \(scu.edu\)](#)

[A Survey on Bias and Fairness in Machine Learning | ACM Computing Surveys](#)

[Continuous Auditing of Artificial Intelligence: a Conceptualization and Assessment of Tools and Frameworks | SpringerLink](#)

[AI models fall short of draft EU rules, researchers say | Financial Times](#)

[capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act by Luciano Floridi, Matthias Holweg, Mariarosaria Taddeo, Javier Amaya Silva, Jakob Mökander, Yuni Wen :: SSRN](#)

[View of Ethical Review in The Age of Artificial Intelligence \(aiej.org\)](#)



**Efficient:**

**Safe, Sustainable**

**Robust (to noise, errors, and adversarial attacks),**

[Training Search - AI Standards Hub](#)

[Artificial intelligence: Experts propose guidelines for safe systems - BBC News](#)

**Effective: Reliable, Clear, Accurate, Relevant, Balanced**

[A pro-innovation approach to AI regulation - GOV.UK](#)

[www.gov.uk](http://www.gov.uk)

[A Survey on Bias and Fairness in Machine Learning | ACM](#)

[Computing Surveys](#)

[AI for Good - All Year Always Online \(itu.int\)](#)



# What is Responsible AI?

mentimeter Results:

<https://www.mentimeter.com/app/presentation/alxjbeux5fykiemoamtbeb8cw12ht4wd>



Join at [menti.com](https://www.menti.com) use code 3295 3022



List top 5 most relevant concepts for Responsible Artificial Intelligence:  
205 Responses







## **AIRe Responsible AI in Healthcare Reading Club:**

Microsoft Teams meeting [Click here to join the meeting](#)

**Meeting ID:** 356 431 233 098 **Passcode:** EohtuX

[Download Teams](#) | [Join on the web](#)