Universidade do Minho
Escola de Engenharia
Departamento de Informática

A Responsible AI for Social Good

LICENCIATURA EM ENGENHARIA INFORMÁTICA
MESTRADO integrado EM ENGENHARIA INFORMÁTICA
Inteligência Artificial
2022/23

ISLab
Synthetic Intelligence Lab

Main concerns

"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

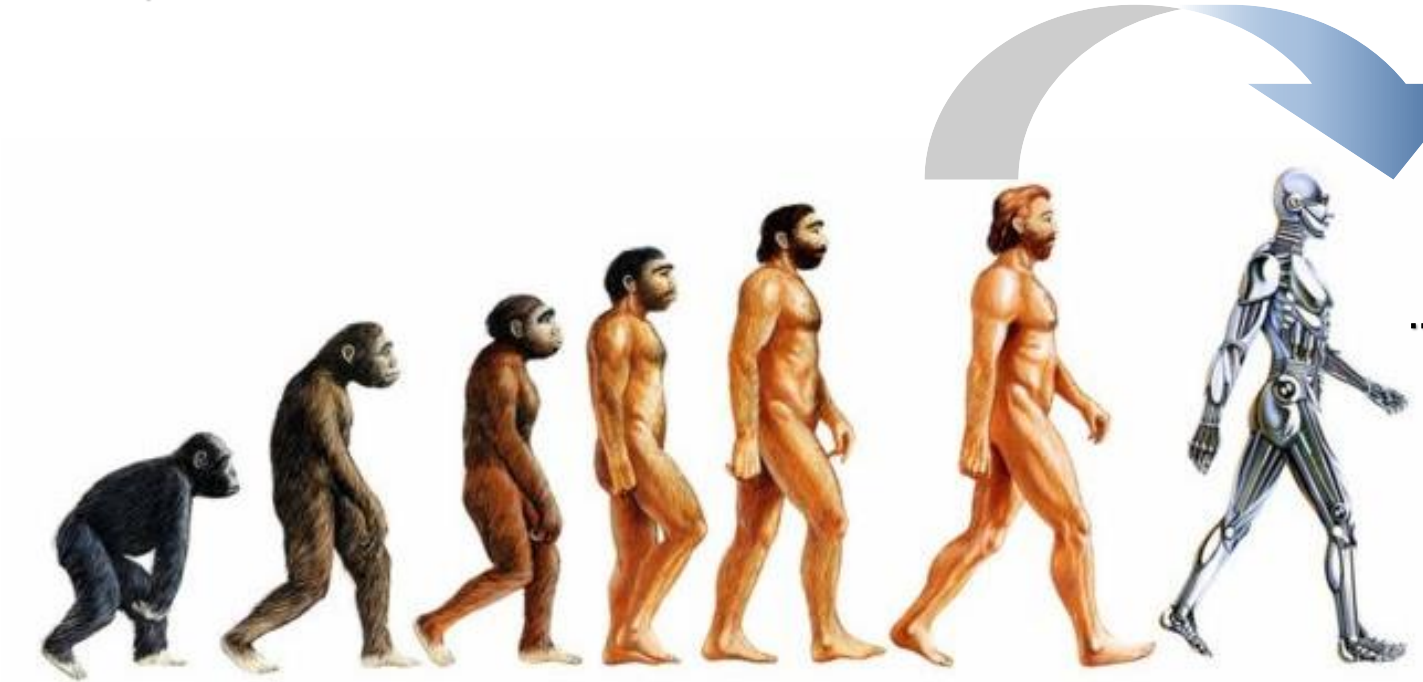
This New World

Source: The New Yorker – September 30, 2019

ISLab
Synthetic Intelligence Lab

AI: Hollywood or Reality?

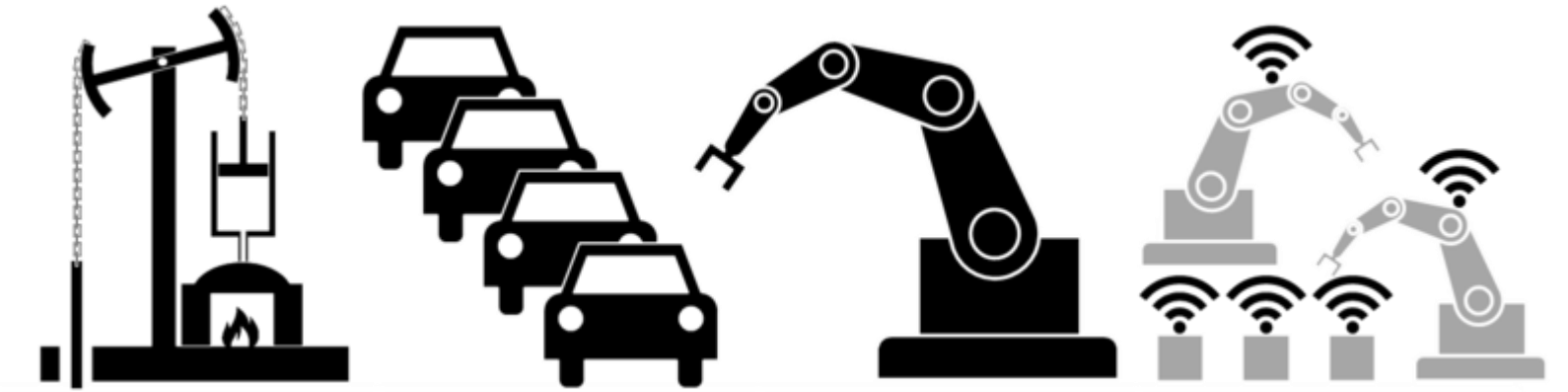
How far is the present



... of the future?

ISLab
Synthetic Intelligence Lab

The New Revolution



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

ISLab
Synthetic Intelligence Lab

Why AI now?

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

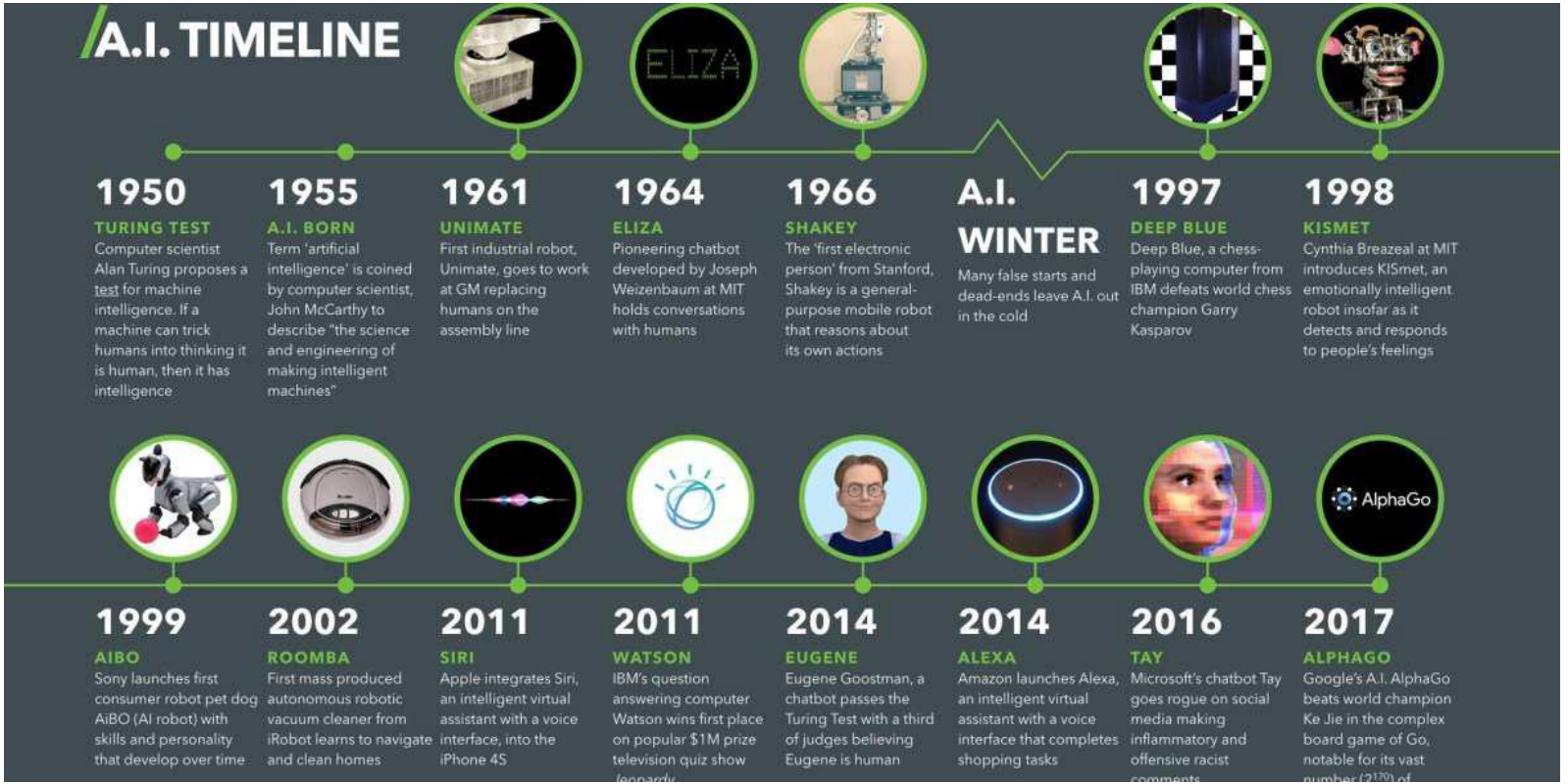
- Labour price;
- Scalability.

Example:
Henry Bessemer (1856) - the steelmaking process
But however the invention remote to 2000 a.C. Anatolia.

This is the AI moment.

ISLab
Synthetic Intelligence Lab

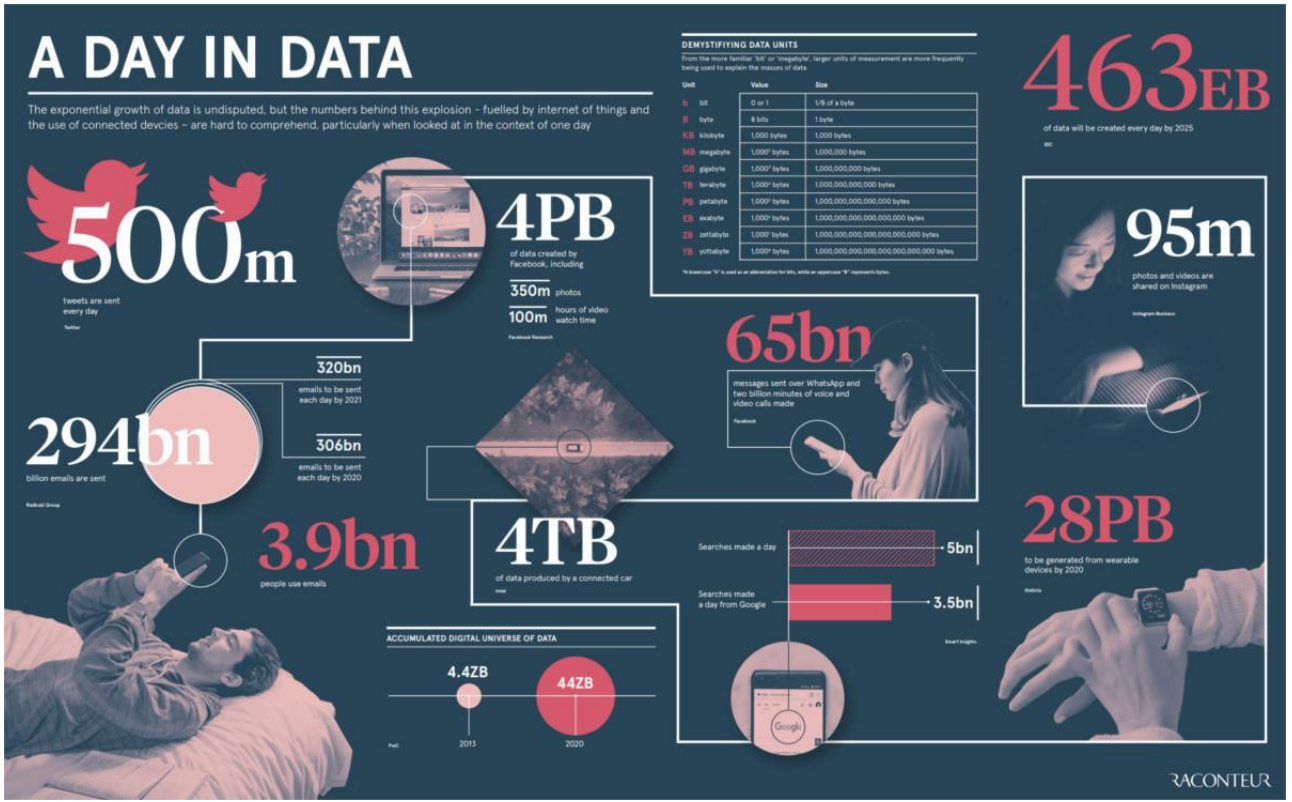
AI Timeline



Source: Paul Marsden
https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/

ISLab
Synthetic Intelligence Lab

Our World in Data!



Source: How much data is generated each day? World Economic Forum
https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/
Image: Raconteur

"By 2025, it's estimated that 463 exabytes of data will be created each day globally – that's the equivalent of 212,765,957 DVDs per day!"

ISLab
Synthetic Intelligence Lab

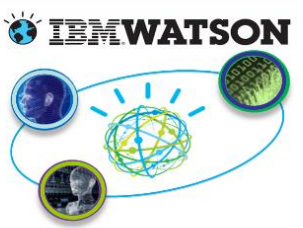
Deep Blue

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

Cognitive systems

- **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 evaluat learning.
  - Deep QA



AlphaGo

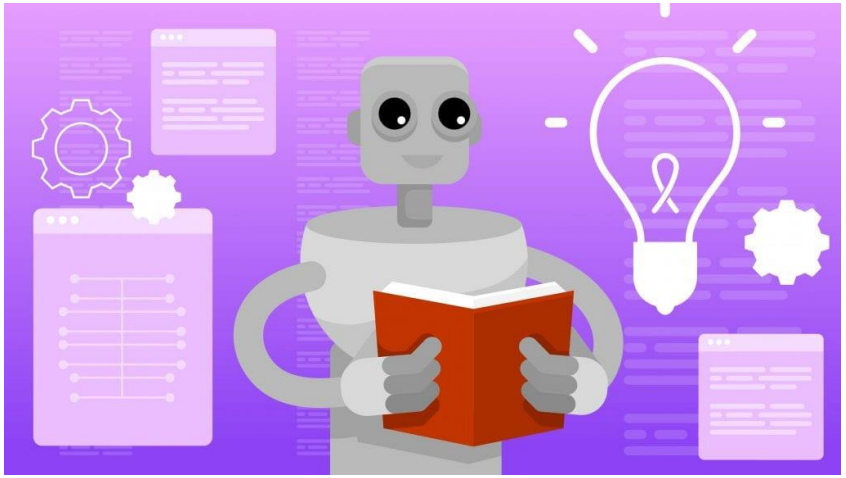
- **AlphaGo (Google DeepMind) - 2016**
  - The match between man and the machine did not go well for Lee Se-dol the Go game world champion;
  - Go is a board game for 2 players that is similar to chess but more complex in relation chess.
  - AlphaGo combines deep neural networks (evaluation) and Monte Carlo tree search (choice). With a combination of supervised learning and reinforcement.

ISLab
Synthetic Intelligence Lab

GPT-3
OpenAI

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

ISLab
Synthetic Intelligence Lab

AI Where?

- **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.**

ISLab
Synthetic Intelligence Lab

Robots (everywhere)

Google Self-Driving Car Project

Hiroshi Ishiguro

Industry 4.0

Hanson
Robotics

ISLab
Synthetic Intelligence Lab

Manufacturing

Brain Corp
Adaptable and flexible robots.
They can navigate unstructured environments like warehouses and store floors.

Rethink Robotics
Collaborative robots that can work in the same environment as humans.
Eg., Sawyer prepped materials for assembly, completed inspections and adapted to human workflow changes.

Datacolor – AI Tolerancing for Fabric Color Matching
To ensure that the original design colours match the colours in a finished textile product.

Source: Sam Daley (2018)
19 examples of artificial intelligence shaking up business as usual
https://builtin.com/artificial-intelligence/examples-ai-in-industry

ISLab
Synthetic Intelligence Lab

Finance

AlphaSense: AI-powered financial search engine
AI-powered financial search engine to help investment firms.

Using a combination of linguistic search and natural language processing, the program can analyze key data points across 35,000 financial institutions.

Source: Sam Daley (2018)
19 examples of artificial intelligence shaking up business as usual
https://builtin.com/artificial-intelligence/examples-ai-in-industry

Betterment: Robo-advisor pioneer
Automated financial investing platform and a pioneer of robo-advisor technology that uses AI to learn about an investor and build a personalized profile based on his or her financial plans.

Bank of America - Erica (Virtual Assistant)

ISLab
Synthetic Intelligence Lab

Customer Service - Virtual Assistant

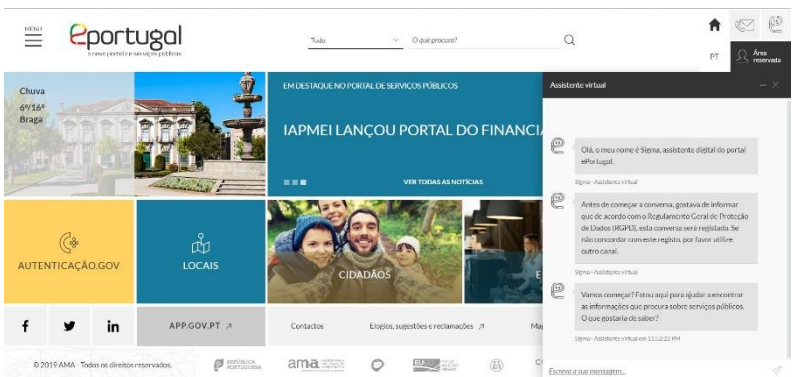
SNAPS - The Platform for Conversational AI

Online platform that empowers brands to provide personalized e-commerce, proactive support, and engagement, creating a wholly unique brand experience for each customer.

Through chatbots, voice skills and social messaging.

Source: https://snaps.io/

E-Portugal - Sigma



Source: https://eportugal.gov.pt/
By GFI Portugal

17

ISLab
Synthetic Intelligence Lab

Expectation

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

ISLab
Synthetic Intelligence Lab

AI moment

- "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.

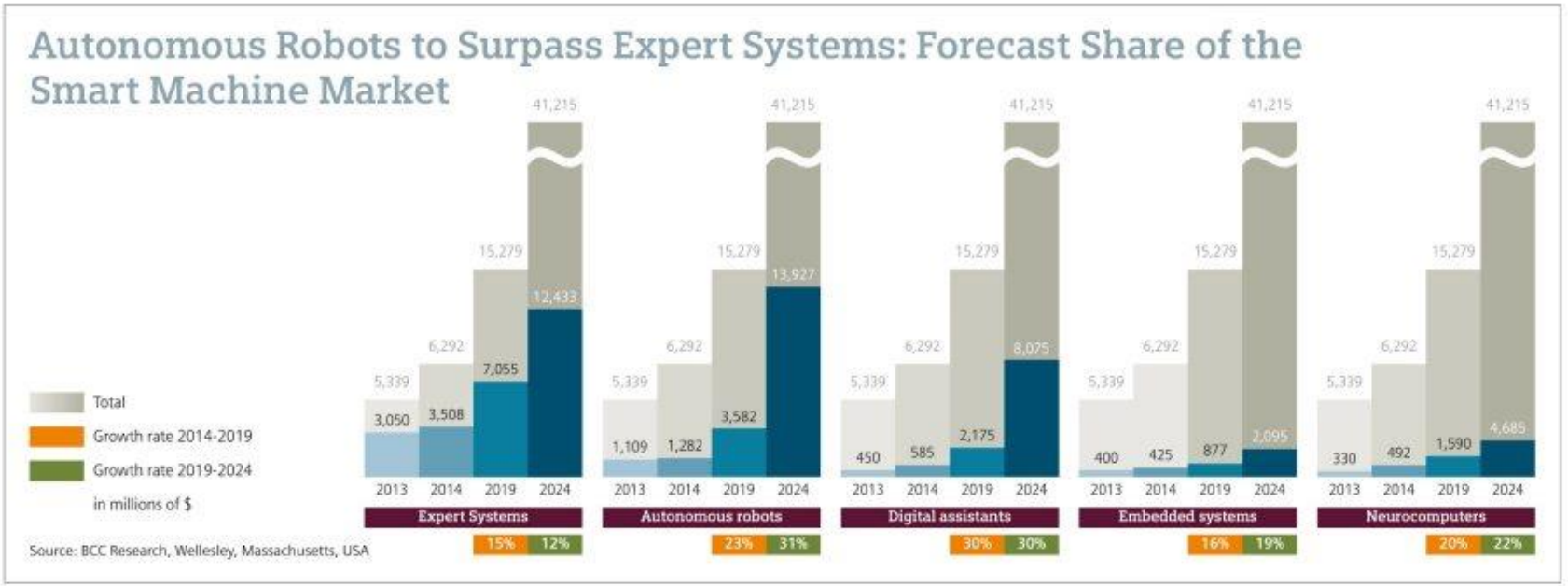
- We are automating tasks (we are not automating jobs !!!)
  - There is no evidence that there will be mass unemployment.
- What we are talking about and seeing is thinking in an absolutely different way.

- 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.

Eric Schmidt (Alphabet/Google)

ISLab
Synthetic Intelligence Lab

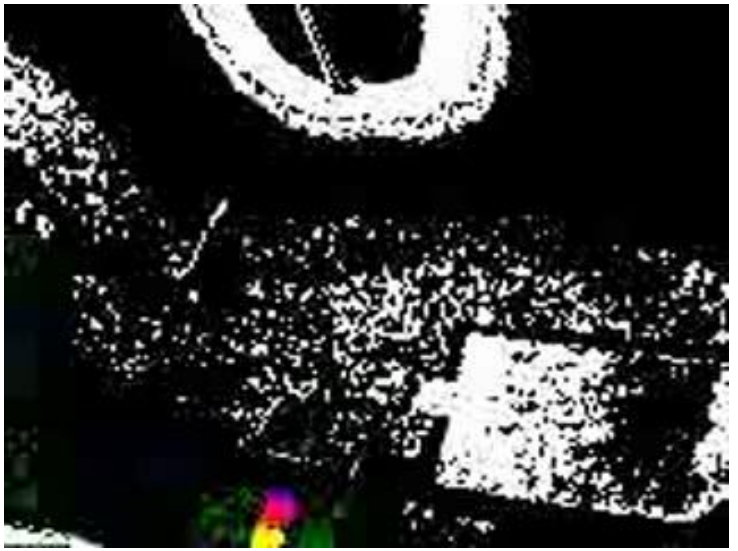
Forecasts for 2019-2024



Source: BCC Research forecasts for 2019-2024 (in millions)

ISLab
Synthetic Intelligence Lab

Put into context!



ISLab
Synthetic Intelligence Lab

Practical risks

Bad Data

- In the heart of the (actual) AI systems, we have Data (big data);
- The quality is naturally a key success factor;
- Bad data causing bad outcomes.

Malicious data

- Malicious data could cause malicious outcomes.

Transparency

- The 'black box' problem. The main difficulty is to understand (explanation and interpretation) how they work.

Misuse

- AI system is not transparent, it is (somewhat) unpredictable;
- In some cases, we can not say whether it is right or not, even go wrong; Predict the adverse effects failure might have.

Source: Chris Holder and Vikram Khurana, Artificial Intelligence: some practical risks, challenges and limitations in applying AI https://www.bristowscookiejar.com/trends/artificial-intelligence-some-practical-risks-challenges-and-limitations-in-applying-ai/

ISLab
Synthetic Intelligence Lab

New threats

ISLab
Synthetic Intelligence Lab

Autonomous weapons

Vox RECODE THE GOODS FUTURE PERFECT THE HIGHLIGHT FIRST PERSON PODCASTS VIDEO MORE

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

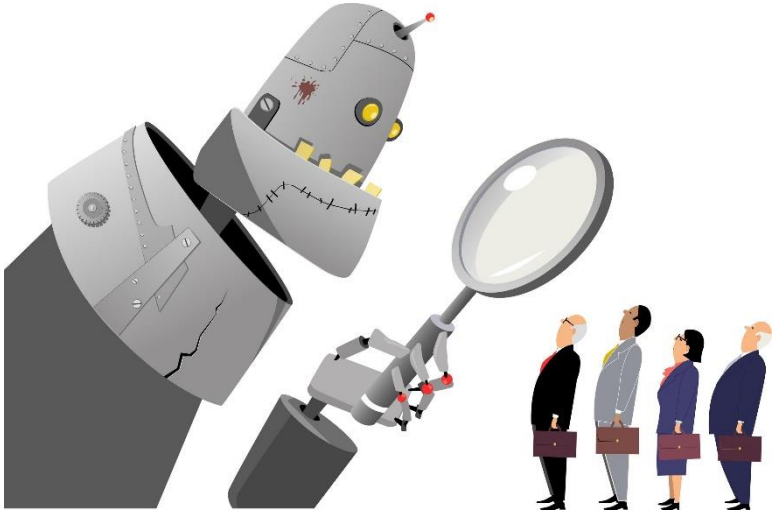
SHARE

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://www.vox.com/2019/6/21/18691459/killer-robots-lethal-autonomous-weapons-ai-war

ISLab
Synthetic Intelligence Lab

AI and Bias



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

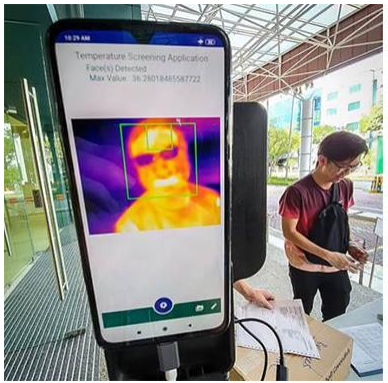
SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

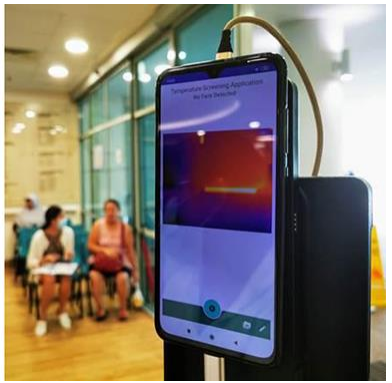
Source: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

ISLab
Synthetic Intelligence Lab

Social Control

- FACIAL RECOGNITION AND FEVER DETECTOR





Source: REUTERS

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

- INTELLIGENT DRONES & ROBOTS

Source: https://www.fierceelectronics.com/electronics/draganfly-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

ISLab
Synthetic Intelligence Lab

Big Brother?

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

ISLab
Synthetic Intelligence Lab

Deepfake

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.

ISLab
Synthetic Intelligence Lab

"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

ISLab
Synthetic Intelligence Lab

Responsible (society) AI

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 1st place guaranteed and in 2nd 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).

ISLab
Synthetic Intelligence Lab

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.

ISLab
Synthetic Intelligence Lab

Principles for responsible AI

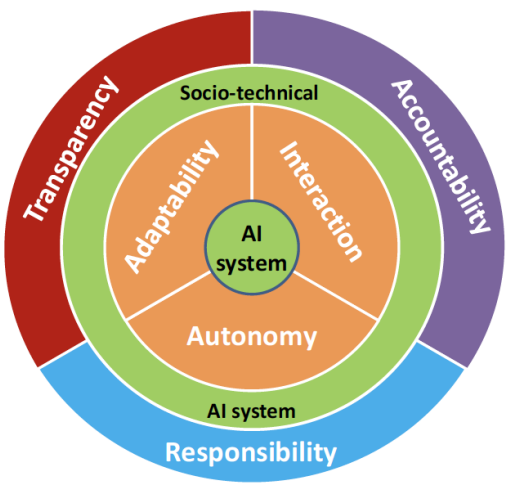
- **Accountability**
  - ability to explain and justify its decisions to users and other stakeholders
- **Responsibility**
  - the role of people themselves in their relation to AI systems
  - is not just about making rules to govern intelligent machines
  - whole socio-technical system englobes people, 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

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.

ISLab
Synthetic Intelligence Lab

AI systems

- Should be developed with responsibly 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.

ISLab
Synthetic Intelligence Lab

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

Transparent, auditable, explainable systems ...

ISLab
Synthetic Intelligence Lab

Microsoft AI principles

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:primaryr

ISLab
Synthetic Intelligence Lab

ASILOMAR AI PRINCIPLES

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/

ISLab
Synthetic Intelligence Lab

ASILOMAR 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/

ISLab
Synthetic Intelligence Lab

Principles for AI development

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

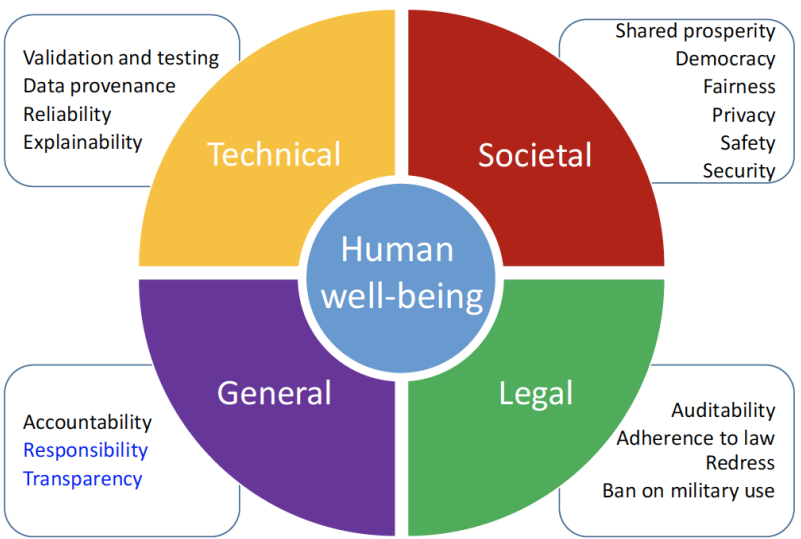
Source: The Barcelona declaration for the proper development and usage of artificial intelligence in Europe
https://content.iospress.com/articles/ai-communications/aic180607

ISLab
Synthetic Intelligence Lab

Responsible AI

"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



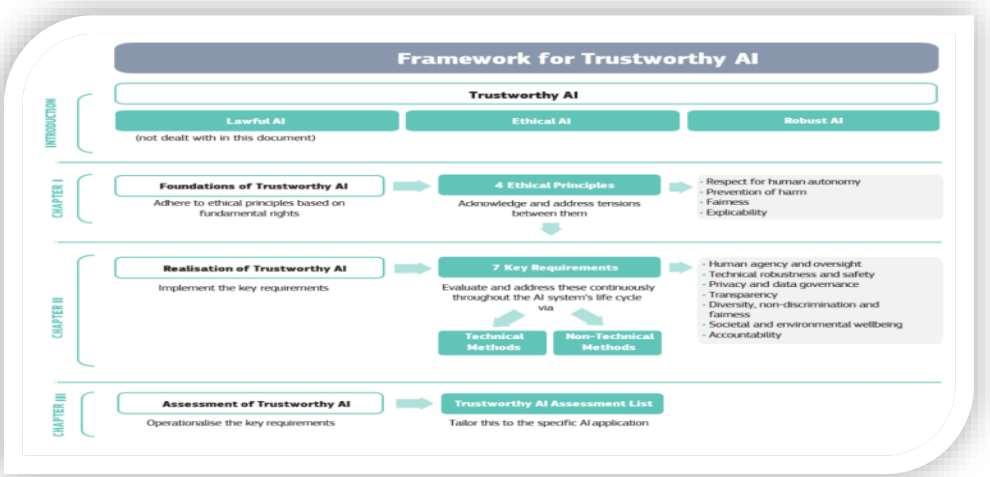
Source: Virginia Dignum (2019) Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer Artificial Intelligence: Foundations, Theory, and Algorithms, https://doi.org/10.1007/978-3-030-30371-6.

Trends

- 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

ISLab
Synthetic Intelligence Lab

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

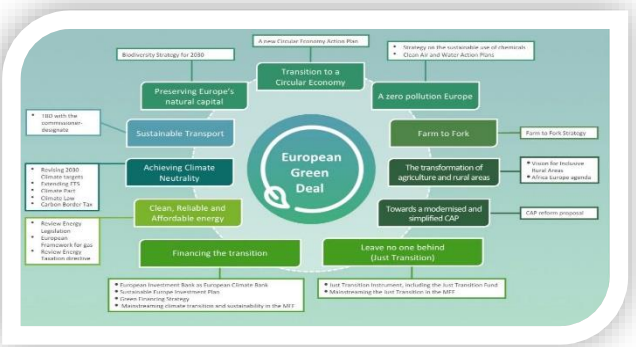
ISLab
Synthetic Intelligence Lab

Society 5.0

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

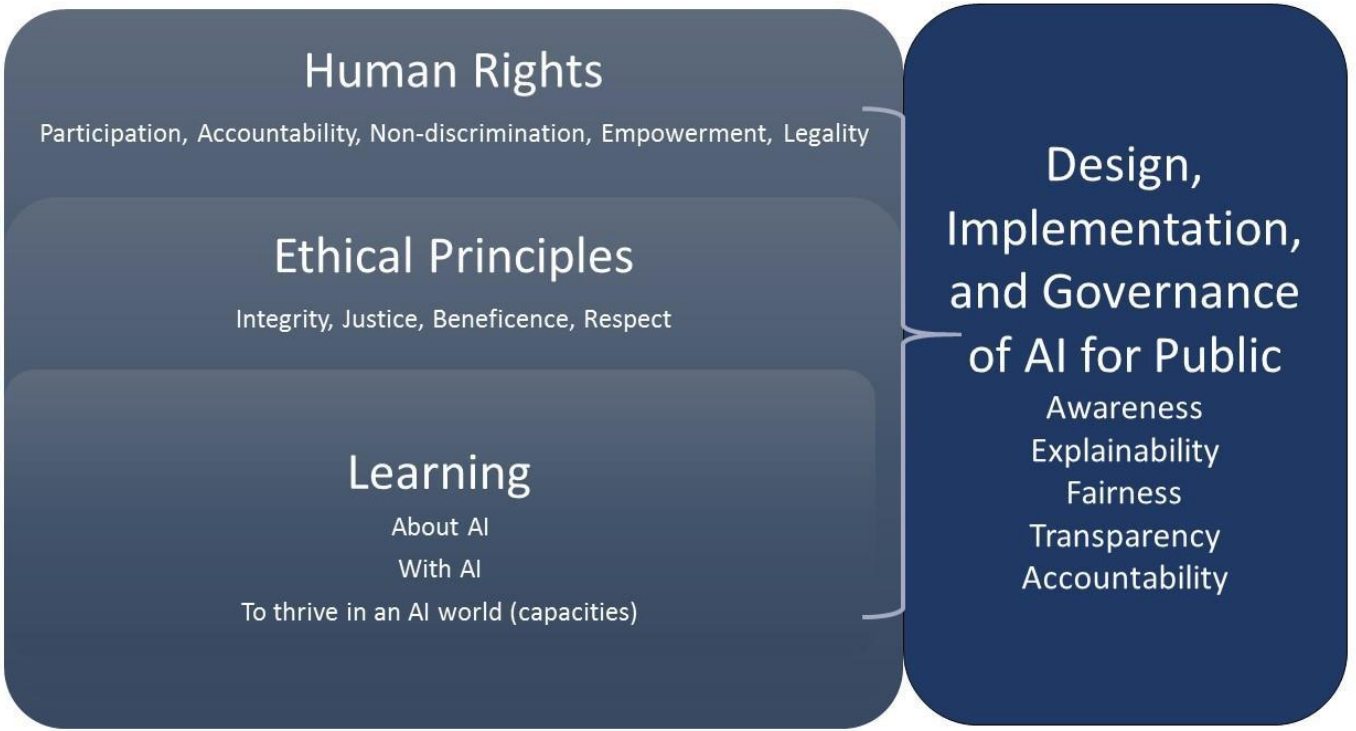
Example: Social problems identified by Japan for Society 5.0

- 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 inequality



ISLab
Synthetic Intelligence Lab

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.

ISLab
Synthetic Intelligence Lab

Main dimensions of the LASI AI strategy

LASI Laboratório Associado Sistemas Inteligentes

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.

ISLab
Synthetic Intelligence Lab

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

Skills:

- Ability to solve complex problems;
- Critical thinking;
- Creativity;
- Ability to interact and understand intelligent entities (human or virtual).

Basic competences:

- Science, technology, engineering and mathematics;
- Digital (Cloud, mobility, social, analytics).

Interdisciplinary knowledge.

Flexibility is the key to adaptation and survival

Environments with Intelligence

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.

Source: The future of AI? Helping human beings think smarter
Kevin Kelly, 2014.
https://www.wired.co.uk/article/brain-power

ISLab
Synthetic Intelligence Lab

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.

The New Yorker – October 23, 2017

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

ISLab
Synthetic Intelligence Lab

Readings

- 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.

Universidade do Minho
Escola de Engenharia
Departamento de Informática

A Responsible AI for Social Good

LICENCIATURA EM ENGENHARIA INFORMÁTICA
MESTRADO integrado EM ENGENHARIA INFORMÁTICA
Inteligência Artificial
2022/23