MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

University of Cyprus

MAI613: Research Methodologies and Professional Practices in AI

Dr Kalia Orphanou

Fall Semester 2022

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Lecture Outline

1. Purpose of EC regulatory framework
2. Classification of AI applications based on risk categories
3. Obligations for providers of high-risk AI systems
4. Examples of AI systems of each category
5. Bias in AI systems
6. Transparency in AI Systems

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Commissioner for Internal Market Thierry **Breton** said:

"AI is a means, not an end. It has been around for decades but has reached new capacities fueled by computing power. This offers immense potential in areas as diverse as health, transport, energy, agriculture, tourism or cyber security. It also presents a number of risks. Today's proposals aim to strengthen Europe's position as a global hub of excellence in AI from the lab to the market, ensure that AI in Europe respects our values and rules, and harness the potential of AI for industrial use."

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Purpose of Regulatory Framework on AI

- There are multiple benefits of AI in the society **but many** AI systems create risks that need to be addressed to avoid undesirable outcomes.
- A clear and predictable legal framework that address the technological challenges is necessary.
- EU proposes the regulatory framework on AI to ensure **user safety**, as well as increasing **user trust** in emerging technologies.
- **Biometric identification** systems are included in this framework
  - AI decisions related on important personal interests such as in the area of education, healthcare, recruitment.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

AI Systems Characteristics

- Safety
  - Product security
  - Mental safety (user safety)

- Connectivity
  - Indirectly the product can be hacked leading to security threats and affecting the safety of users i.e. children smartwatch
  - Cyber-threads of industrial applications

Source: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, 2020

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

AI Systems Characteristics

- Autonomy
  - Self-learning feature of AI products and systems enable the machine to take decisions that are different from what the user expects.
  - May harm mental health i.e. AI humanoid robots
- Opacity
  - Black-box models
  - Decision-making process of the system is difficult to trace
  - Especially in critical domains, humans should understand how AI reaches a decision
  - Transparency, robustness, accountability, unbiased outcome to build trust

Source: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, 2020

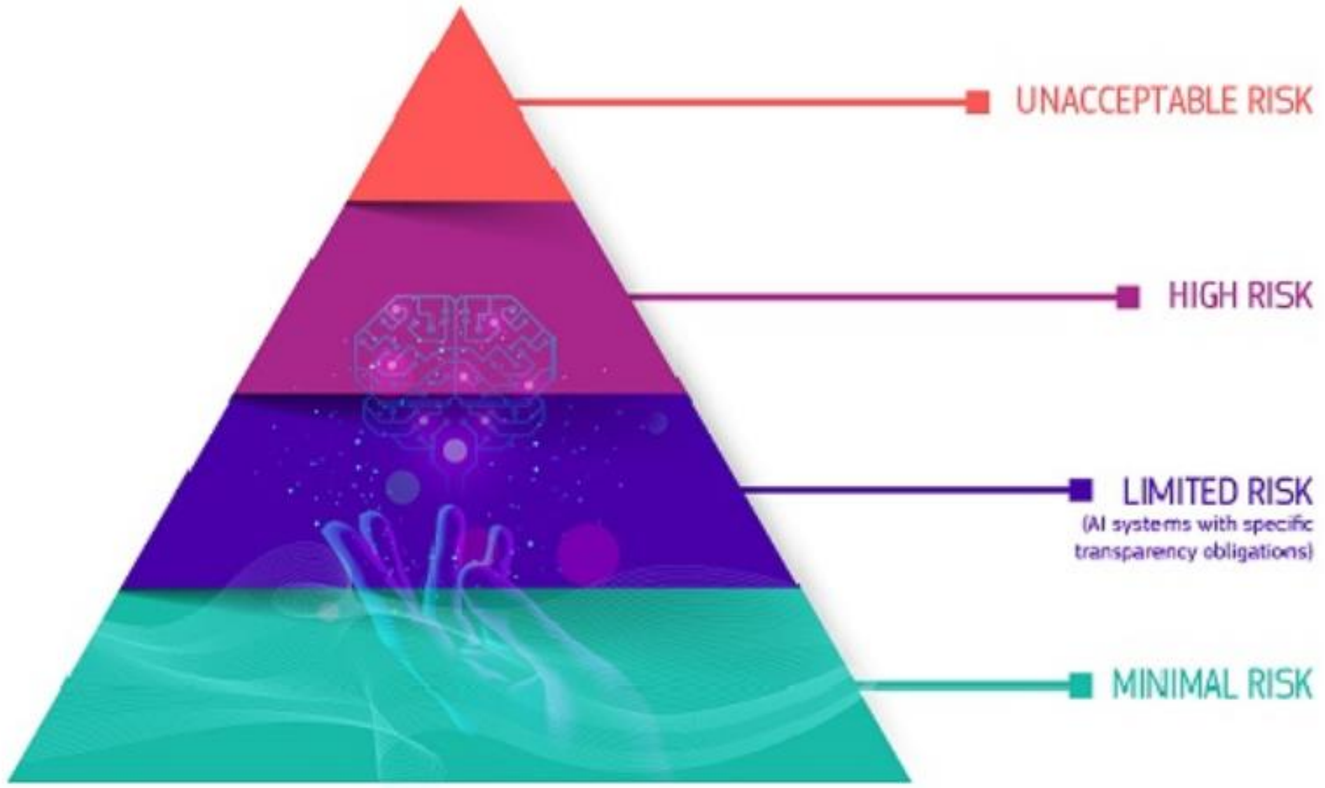
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The proposed rules will:

- address risks specifically created by AI applications;
- propose a list of high-risk applications;
- set clear requirements for AI systems for high-risk applications;
- define specific obligations for AI users and providers of high-risk applications;
- propose a conformity assessment before the AI system is put into service or placed on the market;
- propose enforcement after such an AI system is placed in the market;
- propose a governance structure at European and national level.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Risk Categories of AI Systems

- **Unacceptable Risk:** A very limited set of harmful AI applications that violate the fundamental rights i.e. exploitation of vulnerabilities of children, live remote biometric identification systems in publicly accessible spaces.

- **High Risk:** A limited number of AI systems that creates an impact on people's safety or their fundamental rights.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Risk Categories of AI Systems

- **Limited Risk:** In AI applications where there is a clear risk of manipulation. In limited risk applications, transparency requirements are proposed such as users should be aware that they are interacting with machines.

- **Minimal Risk**: All the rest of AI systems/applications can be used based on the existing legislation without additional legal obligations. The vast majority of AI systems belong to this category.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Unacceptable Risk AI Systems - Examples

Social scoring by government systems. Source by: https://www.eupoliticalreport.eu/artificial-intelligence-and-social-scoring/

- Ban in social scoring systems by public authorities in Europe.
- In China, the governments use social scoring to deny people access to public services.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Unacceptable Risk AI Systems - Examples

- Smart toys using voice assistance encourages dangerous behavior.
- Toys such as smart dolls apps i.e. Cayla, that use AI to collect data about the child so that they will personalize the gaming/learning activities.

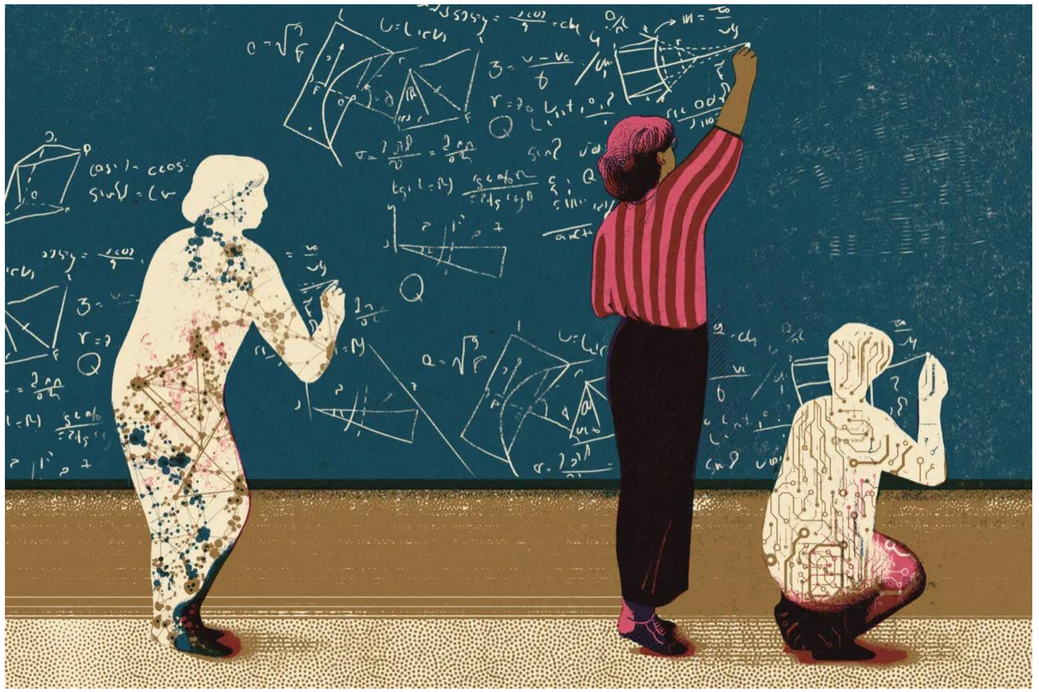
  The company that builds the toy can sell these data.
  - ➢ Makes the child extremely vulnerable
  - ➢ Someone can hack the device and communicate directly with the child.
  - ➢ Germany banned Cayla dolls and other similar games.

Source: https://myfriendcayla.co.uk/

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

High Risk AI Systems Examples

Tesla self-driving car. Source: https://www.bbc.com/news/business-52703767



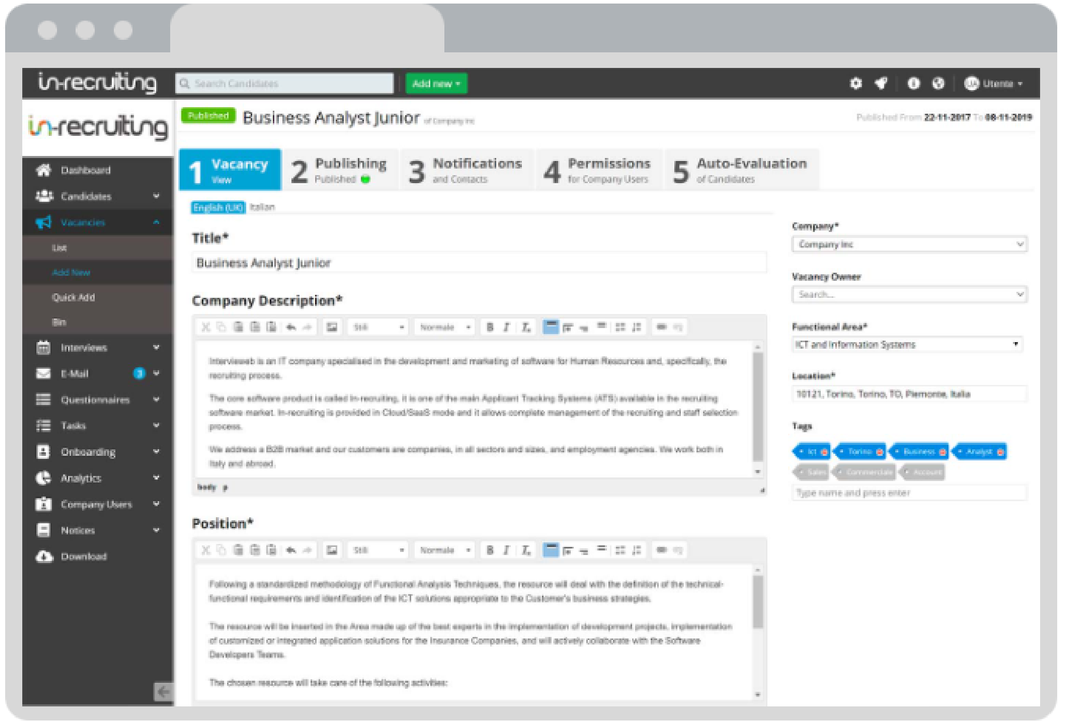
AI (Neural network) provide automatic feedback to students. Source: New York times (https://www.nytimes.com/2021/07/20/technology/ai-education-neural-networks.html)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

High Risk AI Systems Examples

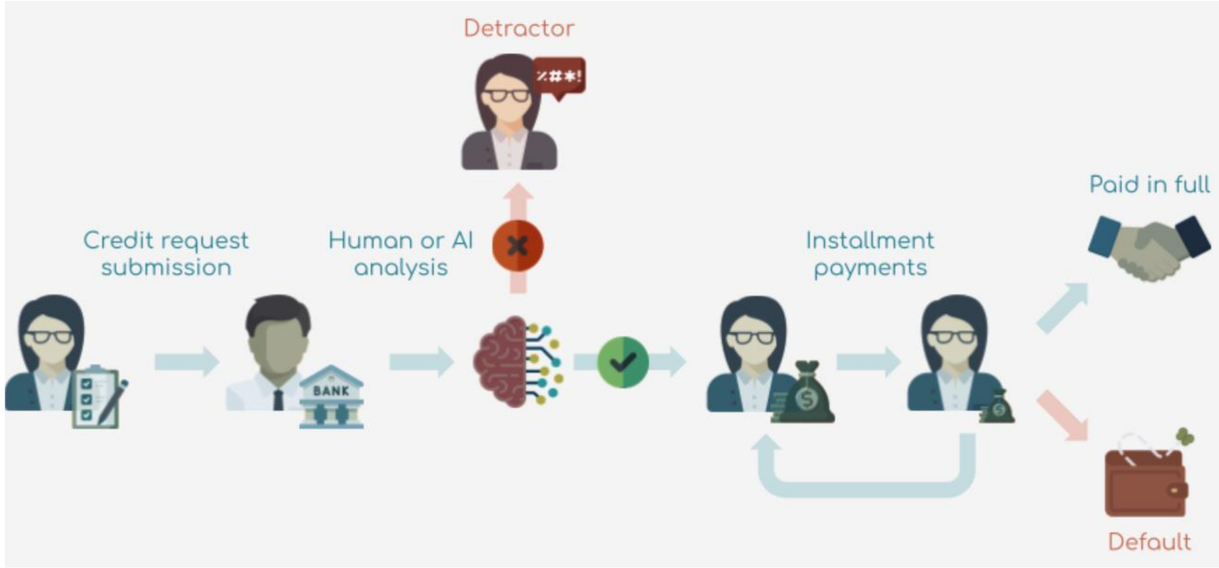
Surgeons operating on someone's arm using a robot Source: MIT Technology Review



CV-sorting software for recruitment using AI. Source: https://www.in-recruiting.com/en/solutions/companies/

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

High Risk AI Systems Examples



Credit score using AI. Source: https://nilg.ai/blog/202107/insights-in-ai-applied-to-credit-scoring/

Biometric identification. Your selfie, your new password. Source: https://www.global-imi.com/index.php/blog/biometrics-your-selfie-your-next-password

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Other High-Risk AI Systems

- Law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);
- Migration, asylum and border control management (e.g. verification of authenticity of travel documents);
- Administration of justice and democratic processes (e.g. applying the law to a concrete set of facts).

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Obligations for High-Risk AI Systems

- Risk management
- Data governance (managing the availability, usability, integrity and security of the data)
- Technical documentation
- Record keeping (traceability)
- Transparency and provision of information to users
- Human oversight (**the capability for human intervention in every decision cycle of the system)**
- Accuracy
- Cybersecurity robustness

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Obligations for Providers of High-Risk AI Systems

- To do a conformance testing before to place a high-risk AI system on the EU market.
- The system should comply with the mandatory requirements for trustworthy AI (e.g. data quality, documentation and traceability, transparency, human oversight, accuracy and robustness).
- In case the system itself or its purpose is substantially modified, the assessment will have to be repeated.
- For biometric identification systems, a third-party conformity assessment is always required.

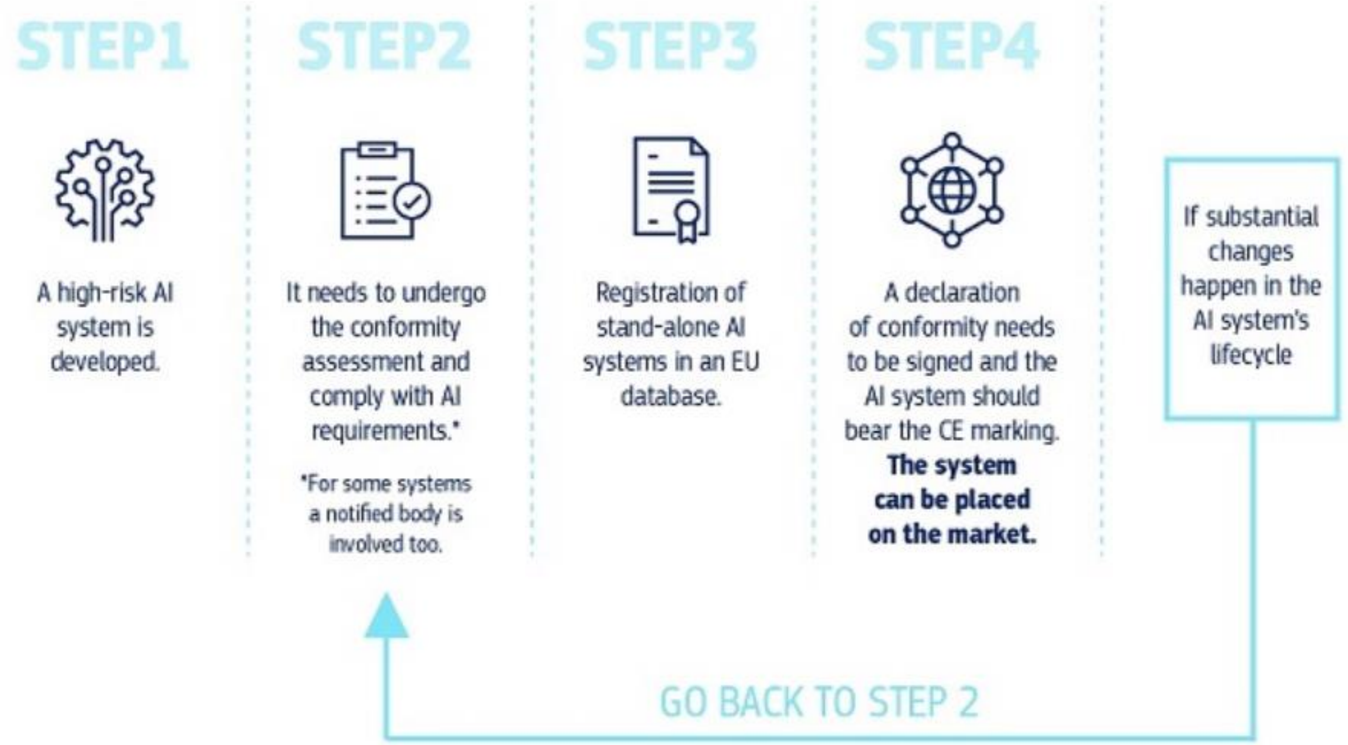
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Obligations for Providers of High-Risk AI Systems

- Providers of high-risk AI systems will also have to implement **quality and risk management in their systems** to ensure their compliance with the new requirements even after a product is placed on the market.
- Multiple audits from authorities will help on monitoring the high-risk AI systems after placing them on the market.

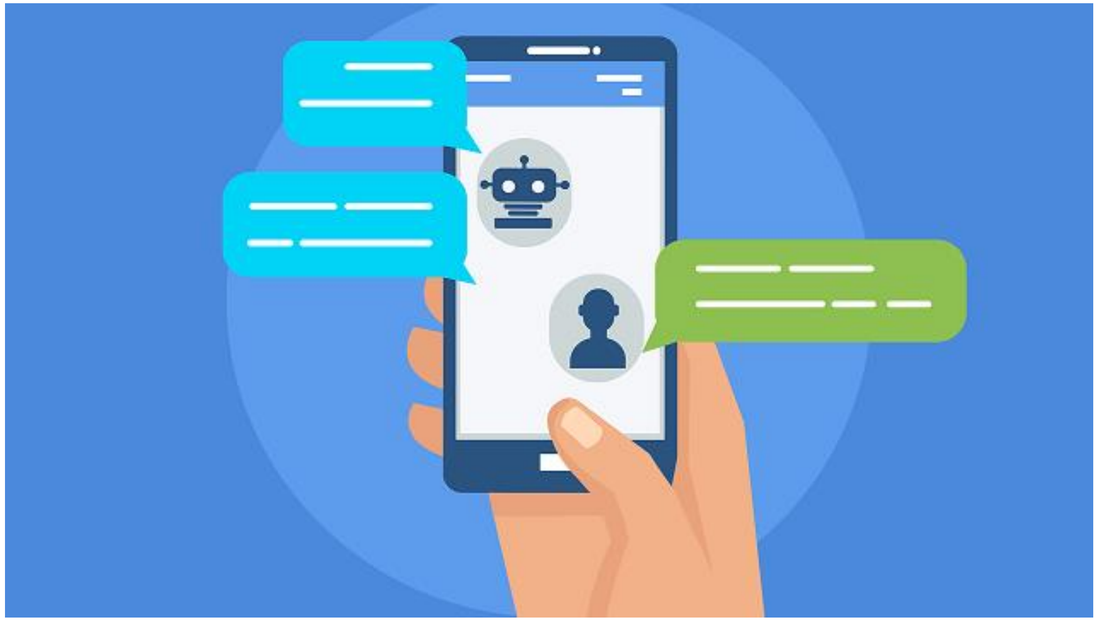
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How to Handle High-Risk AI Systems



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Limited Risk AI Systems Examples



Chatbots Source: Aalpha information systems (https://www.aalpha.net/articles/chatbot-app-development-advantages-and-disadvantages/)

Voice Assistants – Source: BBC news (https://www.bbc.com/news/technology-56602321)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Minimal Risk AI Systems Examples

AI video games: Virtual Reality Photo by Harsch Shivam

Source: github.com

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Voluntary Codes of Conduct

- Providers of non-high-risk applications can ensure that their AI system is trustworthy by developing their own voluntary codes of conduct.
- Or by following the codes of conduct adopted by other representative associations.

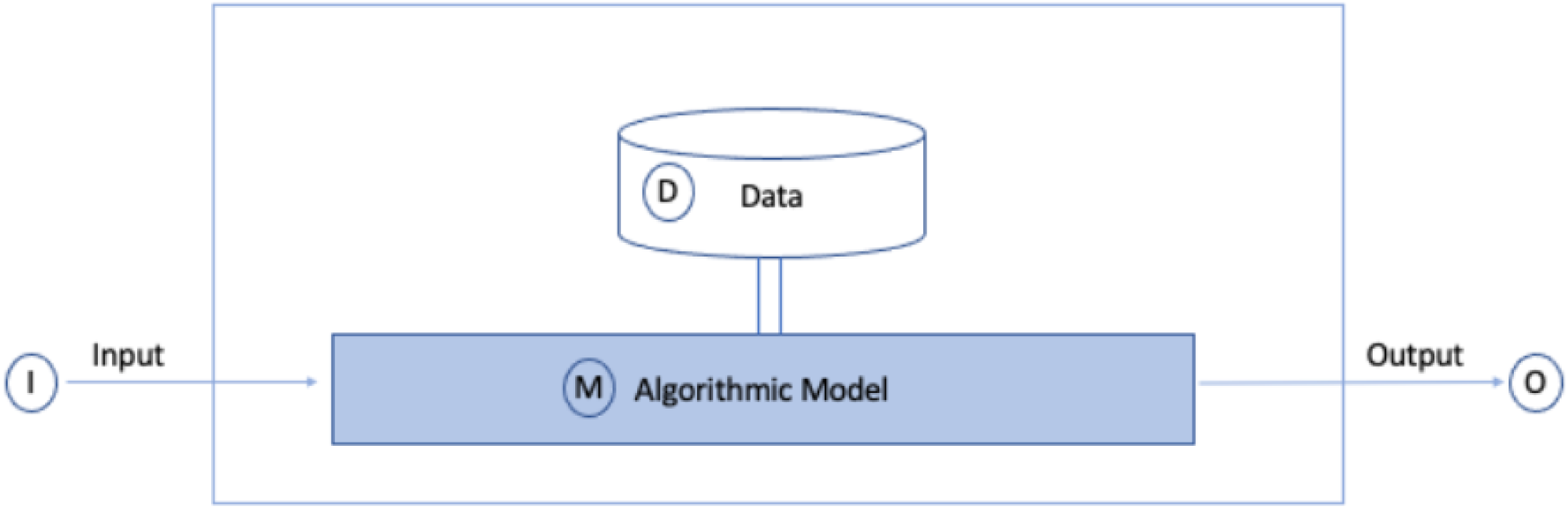
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Accountability and Transparency in AI Systems

- The complexity and opacity (“**black-box models**”) of some AI systems make their evaluation based on the fundamental rights legislation more difficult.
- A human-centric approach to AI means to ensure AI applications comply with fundamental rights legislation.
- Accountability and transparency requirements for the use of high-risk AI systems, combined with improved enforcement capacities, will ensure that legal compliance is factored **at the development stage**.

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

AI Decision-Making System



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Bias in AI Systems

- "a disproportionate weight *in favor of* or *against* an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair." Source: Wikipedia
- Multiple sources:
  - Input data
  - Training data
  - Algorithmic model
  - Output

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Mitigating Bias in AI Systems

- AI systems should not create or reproduce bias but instead to contribute to reduce bias and existing structural discrimination.
- The mandatory requirements for all the high-risk AI systems is to ensure that the output of the system is not disproportionately affecting protected groups (e.g. racial or ethnic origin, sex, age etc.)
- Auditing and other bias detection methods should be applied in high-risk AI systems
- Bias mitigation approaches will be used to reduce bias and discrimination based on the source of bias.
- Detailed documentation should be kept regarding the datasets used for the model training and testing (system transparency)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Transparency in AI Systems

- The decision process of the algorithmic models should be traceable
  - Use of interpretable ML models
  - Explainable AI techniques for black-box models
    - Black-box models: Deep learning, SVM, neural networks
- Be able to justify the particular decision outcome

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Regulatory Framework in AI vs GDPR

- Both set a global standard to respect the fundamental rights
- The requirements and obligations apply to providers and users of AI systems in the EU, regardless of whether AI systems are located in or outside the EU
- The penalty scheme is similar
- The methodology includes self-assessments (third-party assessments for biometric identification applications) to check if they conform the requirements and continuous monitoring.
- Accountability obligations require to keep a good documentation

Source: https://www.ey.com/en_es/law/european-draft-regulation-on-artificial-intelligence-key-questions-answered

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

References

European Union: European Commission, ***Commission Report on safety and liability implications of AI, the Internet of Things and Robotics,*** 19 February 2020, COM(2020) 64 final.

European Union: European Commission, ***New rules for Artificial Intelligence – Questions and Answers,*** Brussels, 21 April 2021

European Union: European Commission, ***Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence, Press Release,*** 21 April 2021

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423