Research proposal writing

Ethical, privacy and fairness concerns

Blended IP Week - Utrecht | 28/01/2025

Lidia Marassi

hcaim human centred artificial intelligence masters

What is a research proposal

- The term «research proposal» refers to the act of writing a (sometimes brief, other times very long) "**plan of action**" describing a set of research activities intended to solve a research question.

- This may be academic oriented (e.g., thesis), industry oriented (e.g., test on the field) or a combination of both (the most common situation within several research domains).

Target

- In general, the aim is to **describe** the activities that the author wants to conduct, clarifying the motivations, the goals and the expected outcomes.

- This can be needed for several reasons:
  - Applying to founding
  - Graduation/dissertation
  - Getting grant to proceed from head of the research lab
  - …

- The emphasis is on presenting an organized and convincing view of the research, especially to those who have to evaluate/approve it.

Why a research proposal?

- The aim is to **clarify** aspects such as:
  - Why the research is important?
  - What is the expected impact?
  - What data will be used?
  - What is the timeline?
  - What methodology will be used?
  - What is the expected contribute to the literature?
  - What type of data is used?
  - Who are the people involved?
  - ….

Research proposal vs …

- …Dissemination (e.g., essay, research paper, etc.)
  - Dissemination activities usually happen at an advanced stage of a project (sometimes at the end) and are intended to spread the achievements across the scientific community .

- …Literature review
  - Is the process of (critically) reviewing the current literature to i) find competitors, ii) supporting works and iii) to frame the research itself
  - Literature review is often **part of** a research proposal, that can be done before (to support the proposal) or during the research activity (as part of it).

- **The research proposal is needed before the research activity beginning, as it is needed to described what the research is about.**

Pitch vs Defense

- In the very early stages of a research activity, you are often required to **pitch** your idea. A pitch is an ultra-simple version of the research proposal, designed to capture interest and communicate the key points: the problem, the importance, the methodology and the potential impact.

- In the late stage of a research activity, you are (often) required to present, disseminate and **defense** your work. This often implies describing the experimental scenario and the used methodologies, also motivating your choices. It is extremely important to have results (not necessary numbers) that support your claims.

Pitch

- A **pitch** is a short, concise presentation used to effectively and persuasively communicate an idea, project or proposal. The main aim is to capture the audience's interest by highlighting the key points and value of the proposal.

- Main characteristics of a pitch:
  1. **Short**: Usually a few minutes (often 1-3 minutes).
  2. **Clarity**: It must be understandable to people who are not experts in the field.
  3. **Engaging**: Must attract attention and arouse curiosity.
  4. **Impact-oriented**: Focuses on what makes the idea relevant and unique.

Typical pitch structure

- **Problem**: What is the problem or need you want to address? Explain why it is important to solve it.
- **Solution**: Present your idea or project as a solution to the problem. What does it consist of?
- **Added value**: What benefit or impact will your proposal have? Why is it relevant?
- **Next steps**: Briefly describe how you intend to implement the idea, with a concise roadmap.
- **Conclusion.**

Example

- **Problem:** Today, 60 per cent of companies lose customers due to poor user experience.
- **Solution:** We've developed an AI solution that analyses user behavior in real time and makes suggestions to improve the user experience.
- **Added value:** With our system, companies can increase customer loyalty by 40 per cent in six months.
- **Next steps:** The project will be developed in three phases: (e.g. prototyping, testing on a sample and large-scale implementation).
- **Conclusion:** With your support, we can change the way companies interact with their customers.

Structuring a research proposal

- The pitch easily allows for **feedback** on structure and content before committing to the thesis, thus improving its quality.
- It becomes a **roadmap** for the thesis, reducing uncertainty and the time needed to organize the content.

Structuring a research proposal

- The research proposal can be considered extremely useful precisely because, if well structured, it provides a solid and organized basis that can serve as a **skeleton for the thesis**.

- Key elements of the proposal can be expanded and enriched in the dissertation, rather than starting from scratch.

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:
  - **Introduction**: to give the reader the overall context. The aim is to make all interested readers being able to understand the main terms, needs and difficulties that the research will have to deal with.

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:
  - **Related works**: This is where the literature review is usually inserted, describing the current literature (if any) focusing on the same problem. A good approach is to structure the section to describe pros and cons of the "competitors", highlighting in the final paragraphs gaps that the research will try to fill (also motivating why this is important).

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:

  - **Research methodology**: this is the often the wider section, intended to detail the precise methodology (tools, procedures, data, visualization, etc.) that the research will use. Usually, also data collection is described. If the research activities are planned to be long running (e.g., year long), a good practice is to identify some **milestones** allowing the monitoring of progresses based on objective measurements (e.g., deliverables done, data collected, tests done, etc.).

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:

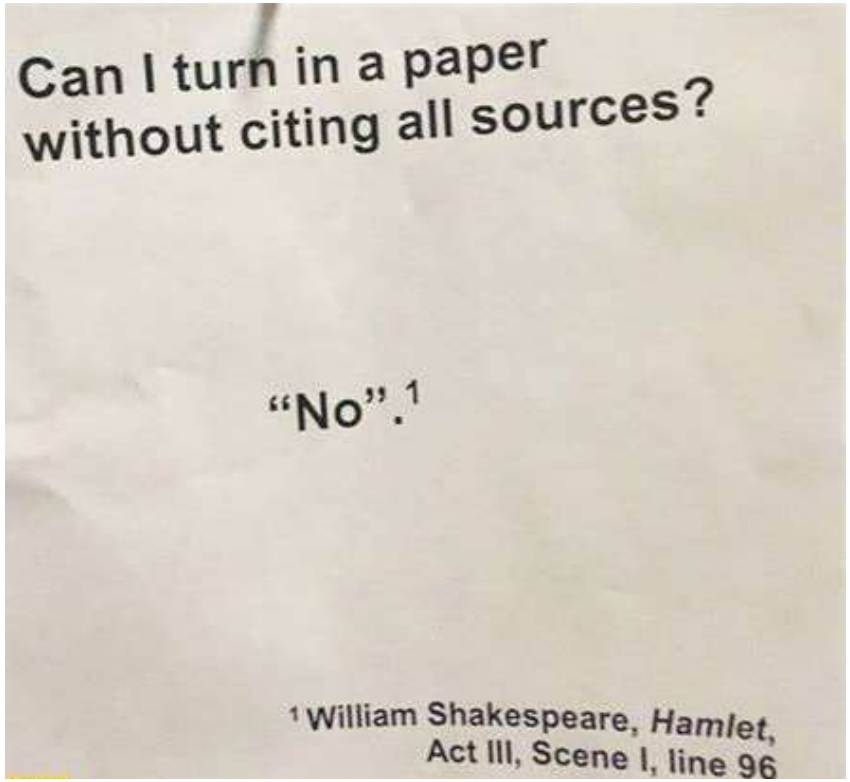
  - **Budget and timeline**: *In view of the nature of the dissertation, this section is not required in the same format as a traditional research proposal*. However, it is useful to consider the computational budget (i.e. the technological resources required for the work) and the timeline (i.e. how you plan to organize the time to complete the work).A common error is to design unrealistic plans (e.g., not enough time allocated for an activity that is known to require time).

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:
  - **Conclusions**: to sum up all the most important points, lacks in the current literature, expected contribution and timeline.

Structuring a research proposal

- In the scientific domains, a research proposal is often structured as a research paper:



Bibliography: to report the list of bibliographic sources supporting the claims made through the proposal, to cite datasets and third-party tools, previous researches on the same topic, etc.

Structuring a research proposal

- Starting with a well-defined proposal **reduces the time** needed to revise and reorganize the thesis.
- The clear structure of the research proposal facilitates the writing process and ensures that all sections are aligned and consistent.
- *In relation to your dissertation*, the research proposal is not only a first step, but also a **valuable guide to the dissertation** writing process, which simplifies and accelerates the overall work.

Common mistakes to avoid

1. **Unrealistic plans**

- Too tight a schedule for complex activities.
- Resources (budget, data, tools) inadequately planned or underestimated.

2. **Lack of clarity or synthesis**

- Long or overly technical descriptions without clear definitions.
- Failure to distinguish between priorities and secondary details.

Common mistakes to avoid

human centred artificial intelligence masters

3. Lack of bibliographical support

- Lack of reference to existing literature to justify the search.
- Outdated, non-authoritative or irrelevant sources.

4. Lack of internal consistency

- Outcomes not in line with the proposed methodology.
- Contradictions between different parts of the proposal.

Common mistakes to avoid

human centred artificial intelligence masters

5. Negligence in risk and impact analysis

- Failure to consider ethical, fairness or privacy implications.
- Lack of plans to deal with potential problems.

- **Why is this important?**

Implications

- Even today, excluding some domains in which research activities may directly harm the safety or security of people (e.g., medicine), *implications of the proposed research are* ***often not explicitly considered***.

- In the field of AI, it is important to consider also the ***ethical, fairness or privacy*** concerns associated with the research proposal.

- Being able to define this issues and plan accordingly is a pro.

Implications

- In the context of human-centred AI, the focus is usually on creating systems that respect human values, rights and well-being.

- Issues of design responsability come into play:
  - AI projects may have unforeseen **impacts on society.**
  - Ethical failures can lead to reputational and legal damage.

- **Researcher responsibility**: Integrate fairness, ethics and privacy analyses from the outset to avoid future problems.

Fairness, ethics and privacy: Key definitions

human centred artificial intelligence masters

- **Fairness**: *Ensuring that AI does not discriminate or create unfair inequalities.*
  - Example: Fair access to opportunities regardless of gender, ethnicity or socio-economic status.

- **Ethics**: *Respecting moral principles in the use of technology.*
  - Example: Avoiding manipulation or harmful use.

- **Privacy**: *Protection of personal information and transparency in its use.*
  - Example: Compliance with regulations such as GDPR, minimising the risk of data breaches.

Proper section

- ***Thus, adding a section to take care of ethical, fairness and privacy concerns is strongly recommended.***

- Report, after the methodology, an analysis of possibly issues associated with ethics, fairness or privacy, together with the risk to happen and how to deal with them.
- This analysis not only demonstrates care and responsibility of the researcher, but also adds value to the proposal, making it more robust.

Different Approaches

- **Not all projects have the same level of risk**, but each proposal should analyze these aspects to be comprehensive and robust.

- .Projects of different types (technical, legal, interdisciplinary) may require different approaches.

- Consideration of fairness, ethics and privacy is not only good practice, but an essential element for the success of a research project.

What can go wrong?

- There are issues in AI that **cross the boundaries** between social, legal and technical disciplines.
- Some of these issues, once considered primarily of social interest, are now also central to developers, legislators and technical researchers (and viceversa).

What can go wrong?

- Malicious misuses of the technology
- Unwanted bias/problems with the data
- Unpredictable social impact
- Data/privacy leak
- Mass control
- Despicable applications
- ….

Failure examples

- **Staples** is an e-commerce web-site for office supplies, furniture, copy-print services and more. In 2012 it decided to use an algorithm to automatically set the items prices according to user information to operates a differential pricing strategy, based on user proximity to a competitor store.
- Staples seemingly rational decision led to higher prices for low-income customers, who (as it turned out) generally lived farther from competitor stores.
- *A lack of contextual analysis can exacerbate existing inequalities and take to inefficient technology solutions.*

Failure examples

- Tay was an artificial intelligence Twitter chatterbot, developed by **Microsoft,** that was originally released on March 23, 2016.
- Tay was designed to learn from user interaction, but was targeted by internet users.
- It caused controversy when the bot began to post inflammatory and offensive tweets, ending up spouting Nazi drivel, forcing Microsoft to shut down the service only 16 hours after its launch.

Failure examples

- In 2018, **Amazon** developed an AI-based system to automate the candidate selection process. The algorithm analyzed CVs sent by candidates to identify the most suitable ones, based on the company's historical hiring data.
- The system discriminated against female candidates, regularly rating them as less suitable than male candidates.
- The algorithm was trained on historical data dominated by male hires, particularly in technical roles.

Failure examples

- **COMPAS** is an algorithmic system used to help judges in the US assess risk of reoffending. It has been used to decide whether to grant probation or bail.
- A 2016 investigation by ProPublica found out that COMPAS had a significant problem with ethnic bias. The algorithm classified African-Americans as "high risk to reoffend" at a disproportionate rate compared to whites.
- The algorithm was based on historical data from the justice system, which already reflected systemic inequalities.

Failure examples

- In March 2018, an **Uber self-driving car** struck and killed a woman, Elaine Herzberg, in Tempe, Arizona, as she was crossing the street and pushing her bicycle out of a crosswalk.
- The vehicle was in autonomous mode at the time of the accident, with a human safety operator on board who failed to intervene in time.
- The vehicle's sensors detected the woman, but the algorithm failed to correctly classify the object as a pedestrian/bicycle.
- The AI system had not been properly trained to deal with complex scenarios such as pedestrians jaywalking.

Social Impact

- The complexity of the incident raised questions about the ethical and legal responsibilities of the various actors involved.
- Who was responsible?
  - The Uber company?
  - The algorithm development team?
  - The human operator who should have been supervising?
- *This incident has shaken public confidence in the safety of self-driving vehicles, and slowed progress in the testing and deployment of these technologies.*
- The Uber case is a real-life example of **how an automated ethical decision can lead to tragic consequences.**

Multidisciplinary approach

Technical issues: Algorithms need to be trained on quality and representative data to avoid bias. It is essential to integrate approaches such as human-in-the-loop and to conduct multiple tests in complex scenarios.

Regulatory issues: Autonomous vehicles need to comply with specific and evolving regulations for both autonomous driving and road testing. In addition, the management of collected data must comply with regulations such as GDPR.

Technical Issues	Legal Issues	Ethical concerns
Algorithm training	Compliance with autonomous vehicle regulations	Large-scale social consequences
Data quality and representativeness	Compliance in road tests	Trolley problem dilemma
Process automation	Regulations (such as the **GDPR**) for the protection of personal data collected by vehicles	...
Human-in-the-loop	comparison of regulations in different countries	
Diversified testing ...	...	

Ethical issues: Technologies such as autonomous vehicles raise questions about possible social consequences. In the case of unavoidable collisions, the trolley problem dilemma highlights the moral complexity of automated decisions in critical situations.

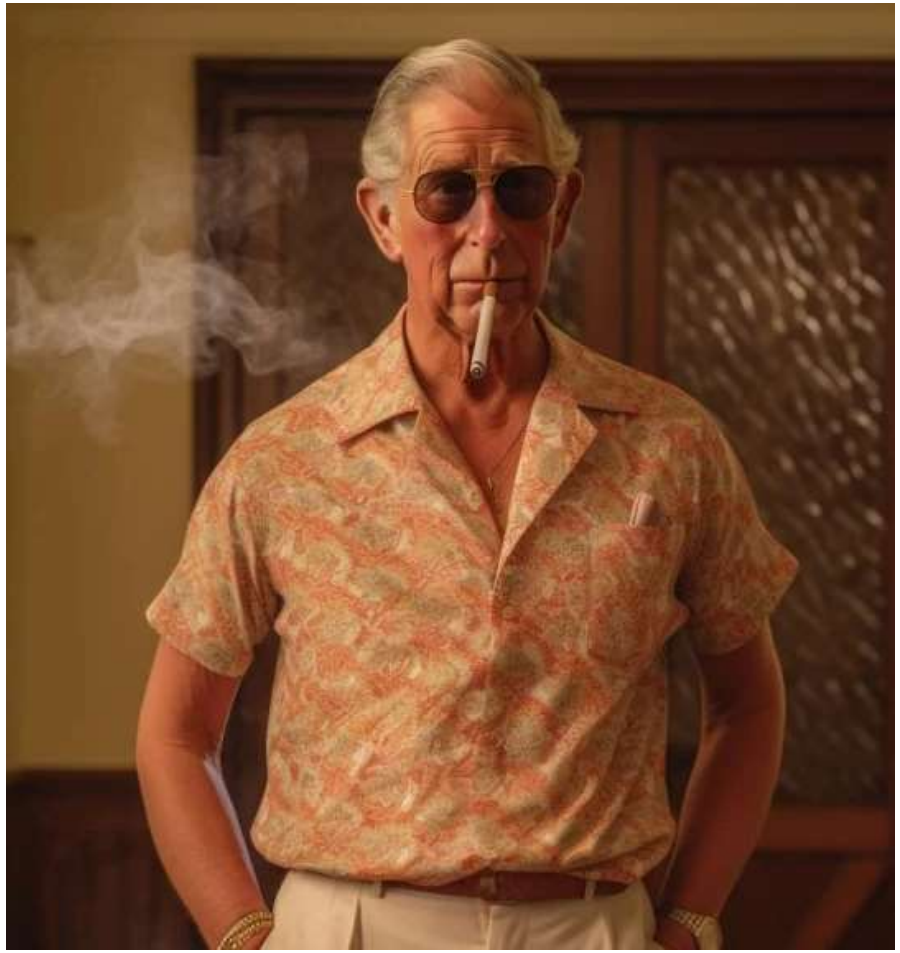
Ethical issues

- The *Trolley Problem* is a classic ethical dilemma in which you must choose between several options, each of which causes harm anyway (i.e., divert a train to save more people, but sacrifice one).
- In the case of IA applications, similar situations arise when there are **conflicting values.**
- When dealing with AI systems, it is crucial to anticipate these conflicts and analyze their **implications** on society, integrating technical, legal, ethical and social perspectives.

Failure examples

- ***Can we really expect technology to make a real difference to society?*** This is a debatable question, but surely what should be expected is that **technology should not make things worse.**
- AI, if well designed, should respect human values and hopefully help reduce inequalities rather than perpetuate them.
- The real challenge is not just to innovate, but to do so in such a way that technology makes a positive contribution without exacerbating existing problems.

Unwanted consequences



Unwanted consequences

Business Markets AI

How a fake AI photo of a Pentagon blast wiped billions off Wall Street

Davey Alba

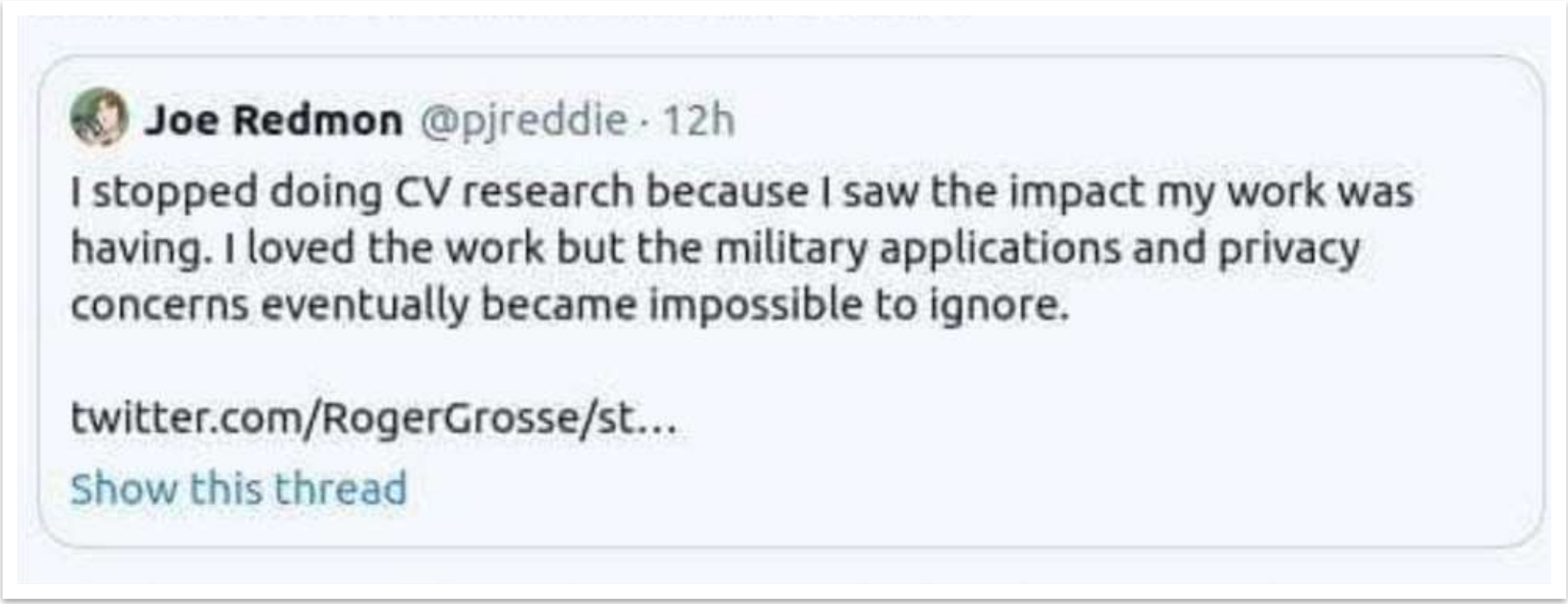
May 24, 2023 – 7.49am

Save Share A A A

Unwanted consequences

- AI applications have the potential to profoundly influence society, so it is crucial that researchers are aware of the **social consequences** of their work.
- Researchers should hopefully critically think about the purposes for which their technology will be used and carefully evaluate the effects that might follow (*of course, this is also a controversial issue*).

A viable solution?



Ethical, fairness and privacy concerns

Issue	Description
Consent from the participants	If the data is related to humans, have all the participants being informed about the research, data collection, storage, etc.? Is all the process GDPR compliant?
No harm	Are humans or animals involved? Does data collection harmed humans or animals or the environment in the process? Will data processing to the same?
No bias	Is the task, the data or the aim bias (gender, ethnicity, religion, etc.) free? Is it possible to demonstrate that also the results are bias free?
Privacy by default	If involved, privacy of the subjects must be guaranteed (i.e., impossible to perform re-identification)
Fair dissemination	All the results must be reproducible and all the outcomes made publicly available. No bias in the results (e.g., showing only successful attempts or setting up an unfair experimental setup).
…	

Legal relevance

- ***However, you do not necessarily have to identify and address these issues on your own.***
- In many cases, there are **established regulations** that can guide you through the process and help ensure that your project is compliant and responsible.
- **GDPR** (General Data Protection Regulation): In force since 2018, the European Data Protection Regulation sets out clear rules on how data should be collected, managed and used. It guarantees:
    - Transparency in the use of data.
    - Individual rights, such as access, rectification and deletion of their data.
    - Security measures to prevent data breaches.

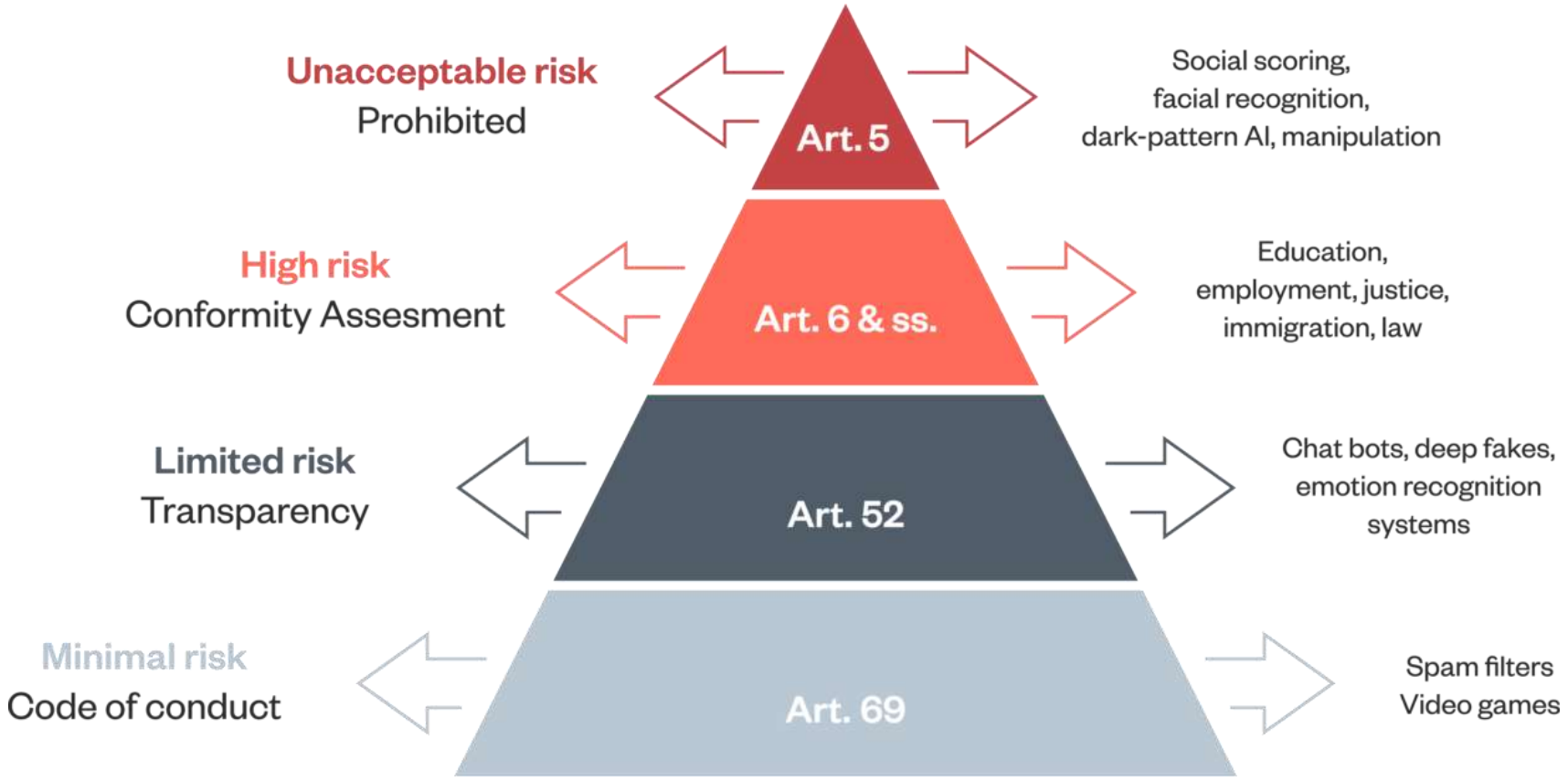
The AI Act

- In addition to the GDPR, the European Union has introduced a specific regulation for the use of AI.
- The **AI Act** (Artificial Intelligence Act) it is the world's first regulation on artificial intelligence.
- It is a benchmark for other countries, promoting an ethical and safe approach to technological innovation.

The AI Act

- It classifies AI systems according to their **level of risk** and sets specific requirements to ensure safety, transparency and fairness.
- The greater the risk, the greater the responsibilities and limits, right down to models that are too dangerous to be authorized.
- There are four levels: **minimum**, **limited**, **high** and **unacceptable**.

The AI Act



"Risk" is the combination of the probability of harm occurring and the severity of the harm.

Unacceptable

- *Subliminal manipulation*: Systems designed to influence people's behavior without their knowledge (e.g., inducing decisions without informed consent).
- *Exploitation of vulnerabilities*: Systems that exploit specific vulnerabilities of individuals/groups (e.g., children, the elderly, people with disabilities) to influence their decisions.
- *Social scoring systems*: Classifications of individuals based on behavior or personal characteristics (e.g., social credit) that could restrict rights or opportunities.
- *Real-time remote biometric identification in public spaces*: Generally prohibited, with strictly defined exceptions as a threat to public safety.

High risk

- Technologies that *can* significantly affect people's rights and safety.
- Require **mandatory conformity assessment** before use.
- Examples of applications:
  - *Healthcare*: Systems for AI-based medical diagnosis or treatment.
  - *Justice*: Algorithms to support legal decisions.
  - *Recruitment*: Systems for hiring decisions.
  - *Critical infrastructure*: AI systems for energy management, transport or telecommunications.
- High-risk AI systems are subject to several obligations, including model assessment, systemic risk mitigation and cybersecurity protections.

Limited risk

- AI systems intended to interact with individuals may pose specific risks but are unlikely to cause significant harm or violate fundamental rights.
- AI systems with transparency risks include, above all, chatbots and synthetic content generation: *Users must be informed that they are interacting with an AI system and not with a real person and systems that produce artificial content must clearly indicate that it has been generated by AI.*
- Transparency obligations for limited-risk AI systems are different for providers and deployers of such systems.

Minimal risk

- Technologies with impacts that are considered limited and do not require specific compliance obligations.

- Examples of applications:
  - Email spam filters.
  - Algorithms for improving image or video quality (also in videogames).
  - Non-critical recommendation systems (e.g. film or music recommendations).

- No specific requirements foreseen unless the use of the system changes the level of risk.

Referring to established regulations adds credibility

- Being aware of regulations such as the GDPR and the AI Act will help you to place your work within a clear regulatory framework, **defining project limits and responsibilities.**

- Regulations **help to identify potential project problems**, not only in the most legally oriented research proposals, as reference to regulations demonstrates awareness of the context *(for example, for projects that use personal data, the GDPR provides useful guidelines to ensure compliance and transparency).*

The role of education

- AI technologies will continue to evolve, with new applications and increasingly advanced models. It is essential to keep up with this evolution to ensure responsible use.
- **AI models, like any other technology, are just tools, not inherently good or bad.**
- What really matters is **how they are used** and for what purpose.
- AI has enormous potential to improve people's lives (from healthcare to creativity, from scientific research to education ...)

The role of education

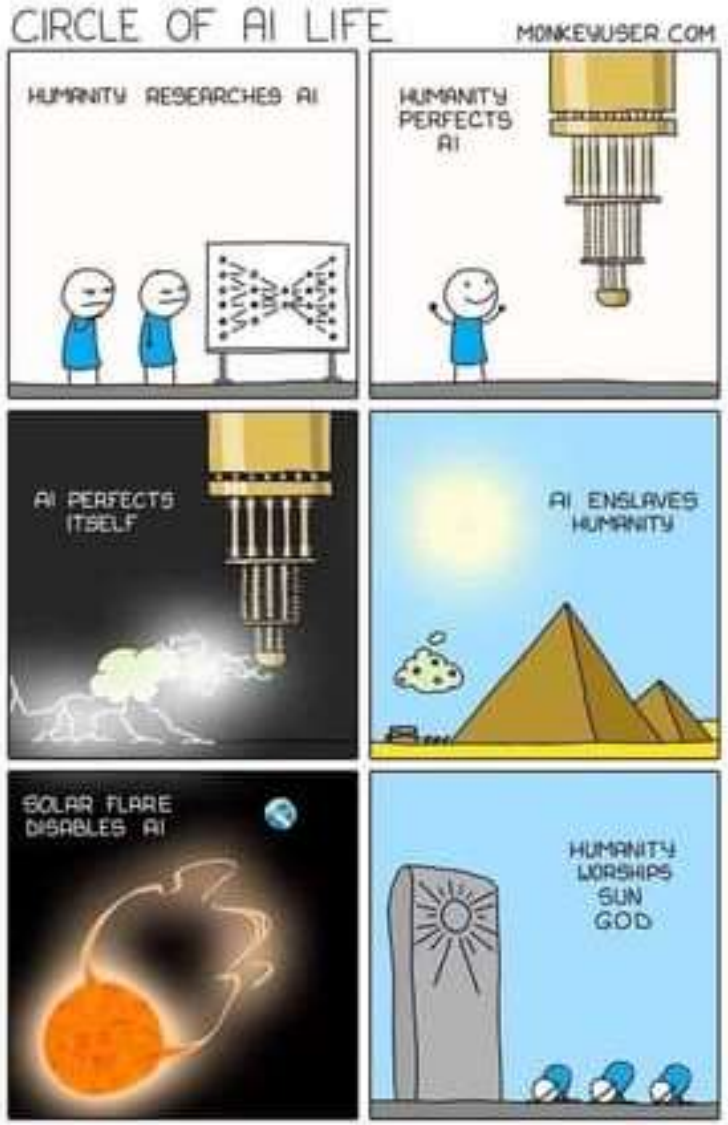
human centred artificial intelligence masters

- As the saying goes, 'with great power comes great responsibility'.
- It is therefore essential to develop methods and regulations that ensure its fair, safe and ethical use.

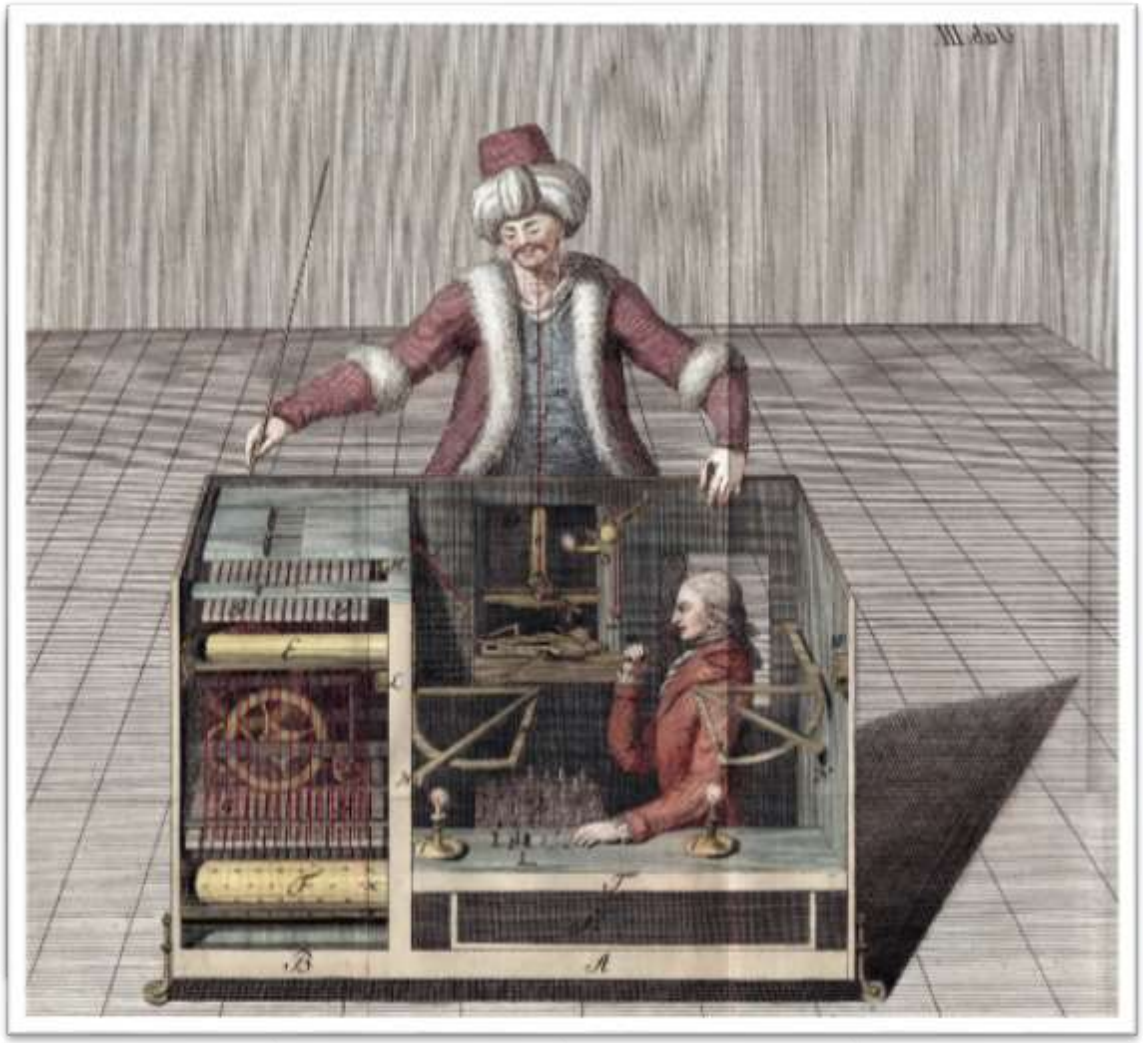
- **Education is the key!** Educating people about AI is essential to promote its conscious and responsible development and use, and to ensure shared benefits for all.

Conclusion

- The take home message is to pay attention to all these aspects during the research proposal writing, to be ready to identify them and deal with them as soon as they arise.



Conclusion



Further reading

- Onwuegbuzie, Anthony J. "Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety." Library & Information Science Research 19.1 (1997): 5-33.
- Connelly, Lynne M. "Ethical considerations in research studies." Medsurg Nursing 23.1 (2014): 54-56.
- Munhall, Patricia L. "Ethical considerations in qualitative research." Western Journal of Nursing Research 10.2 (1988): 150-162.
- Ketefian, Shaké. "Ethical considerations in research. Focus on vulnerable groups." Investigación y Educación en Enfermería 33.1 (2015): 164-172.
- Carter, Rickey E., et al. "Pragmatic considerations for fostering reproducible research in artificial intelligence." NPJ digital medicine 2.1 (2019): 1-3.
- Sallstrom, Laura, Olive Morris, and Halak Mehta. "Artificial intelligence in Africa's healthcare: Ethical considerations." (2019).

Thank you for your attention !