



ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE

MASTER I SCIENCE DE DONNÉES ET INTELLIGENCE ARTIFICIELLE (SDIA)

DR ILHAM KITOUNI

2023-2024

LESSON 4 : ACCOUNTABILITY AND TRANSPARENCY IN AI



INTRODUCTION TO ACCOUNTABILITY AND TRANSPARENCY IN AI

- Accountability and transparency in AI refer to the ability of humans to understand how AI systems make decisions and to hold the developers of such systems responsible for their actions.
- AI can have a significant impact on people's lives, and it is essential that these systems are developed in a way that is fair, just, and ethical.
- Accountability and transparency are necessary to ensure that decisions made by AI systems are explainable and that AI developers are held responsible for any harmful outcomes of their systems.



INTRODUCTION TO ACCOUNTABILITY AND TRANSPARENCY IN AI

Challenge:

- One of the challenges of achieving accountability and transparency in AI is the technical complexity of these systems.
- AI algorithms are often opaque, making it difficult for humans to understand the decision-making process.
- Additionally, there are legal and ethical challenges in determining who is responsible for the actions of AI systems.

WHY IS ACCOUNTABILITY AND TRANSPARENCY IMPORTANT?

- AI can have a significant impact on people's lives, from things like who gets a loan or a job to who is targeted with advertising or law enforcement.
- It is important that AI systems are accountable and transparent so that we can ensure that they are used fairly and ethically.
- Accountability and transparency can help to:
 - Build trust in AI systems
 - Identify and address biases in AI systems
 - Improve the performance of AI systems



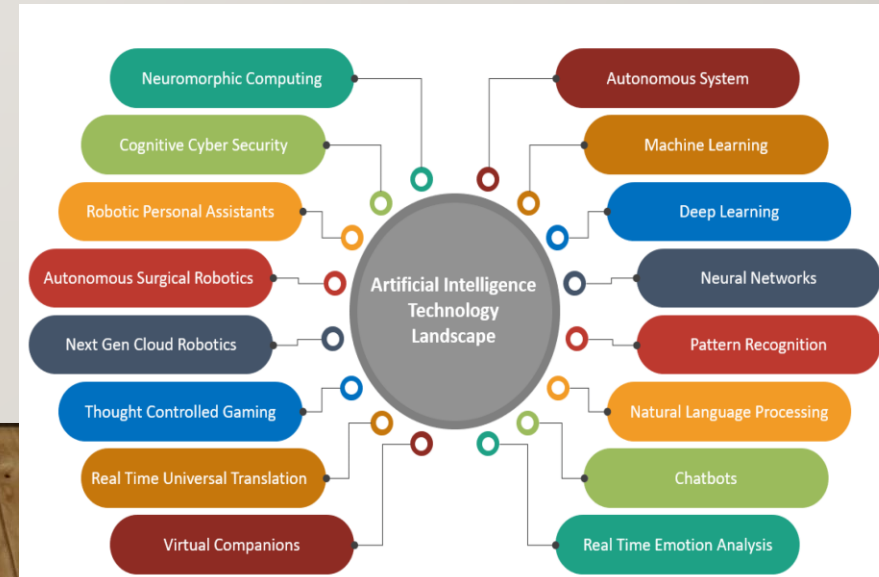
OVERVIEW OF THE PRESENTATION



- This presentation will discuss the challenges and importance of achieving accountability and transparency in AI.
- First we explore different perspectives on responsibility in AI, including the role of stakeholders such as AI developers, governments, regulators, users, and the public.
- Second we focus on ways to increase transparency in AI systems, discussing technical and ethical considerations and case studies of AI systems that have been made more transparent.

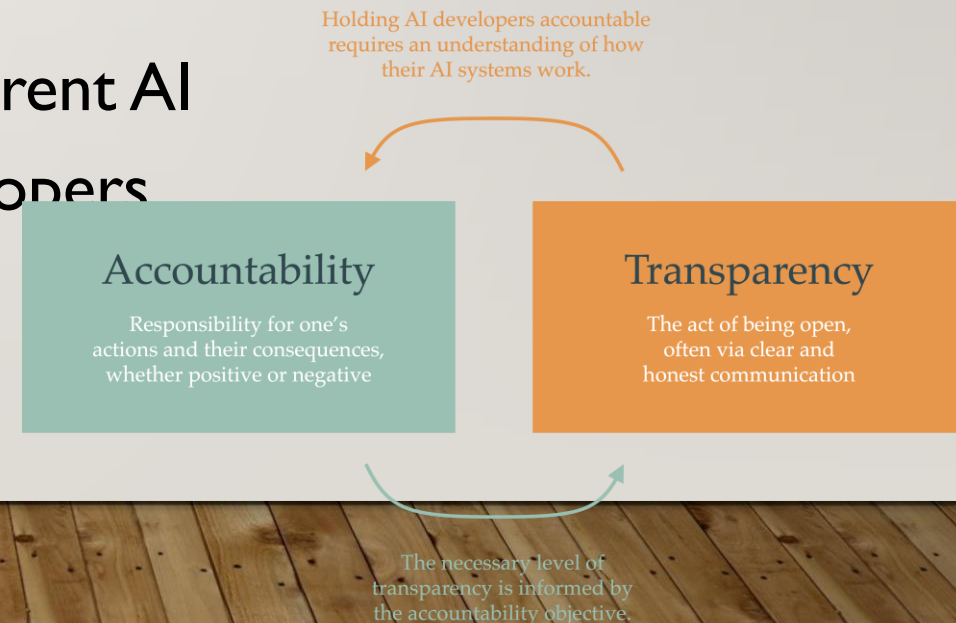
WHO IS RESPONSIBLE FOR THE ACTIONS OF AI SYSTEMS?

- There are several perspectives on who should be responsible for the actions of AI systems.
- Some argue that AI developers and deployers should be held liable for any harm caused by their systems.
- Others believe that governments and regulators should monitor AI systems to ensure their safety and ethical use.
- Additionally, users and the public can play a role in advocating for more ethical AI systems.



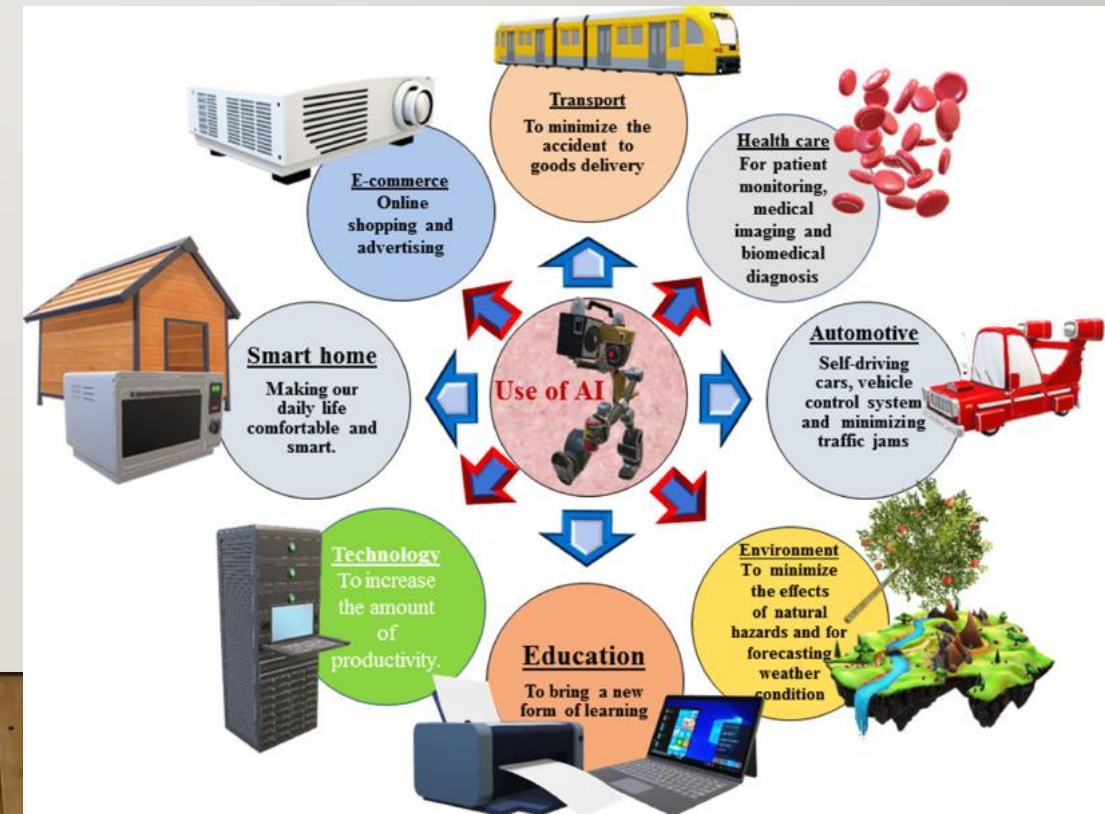
THE ROLE OF DIFFERENT STAKEHOLDERS IN ENSURING ACCOUNTABILITY

- AI developers and deployers have a responsibility to develop and deploy systems that are safe, ethical, and transparent.
- Governments and regulators can play a role in regulating AI systems to ensure their safety and ethical use.
- Users and the public can demand ethical and transparent AI systems and hold governments, regulators, and developers accountable for any harm caused by AI systems.



CASE STUDIES AND CHALLENGES OF ASSIGNING RESPONSIBILITY

- Examples of AI incidents and accidents include autonomous vehicle accidents, facial recognition technology errors, and biased language models.
- In some cases, responsibility for harm caused by AI systems has been assigned to developers or companies.
- In other cases, no clear responsibility has been assigned, highlighting the challenges of assigning responsibility for AI systems.



CHALLENGES OF ASSIGNING RESPONSIBILITY IN AI SYSTEMS

- Technical complexity: AI systems are often complex and difficult to understand, making it difficult to identify who is responsible for any harm caused by the system.
- Legal and ethical considerations: There are legal and ethical challenges in determining who is responsible for the actions of AI systems. For example, is the developer responsible for the actions of an AI system that learns and adapts over time?
- Difficulty of holding multiple parties accountable: AI systems are often developed and deployed by multiple parties, making it difficult to hold any one party accountable for harm caused by the system.



HOW CAN AI SYSTEMS BE MADE MORE TRANSPARENT?

- Transparency in AI refers to the ability of humans to understand how AI systems make decisions.
- There are different approaches to transparency in AI, including:
 - Providing access to training data
 - Making algorithmic decisions explainable
 - Enabling human oversight of AI systems

TECHNICAL CHALLENGES AND ETHICAL CONSIDERATIONS

- One technical challenge of making AI systems more transparent is operationalizing transparency techniques without sacrificing accuracy or performance.
- Additionally, some AI models may be difficult to interpret, and providing explanations may be complex or require significant resources.
- There are also ethical considerations in AI transparency, such as privacy concerns and potential discrimination. Transparency techniques must be designed to minimize the risk of exposing sensitive information or perpetuating biases.



EXAMPLES OF AI SYSTEMS THAT HAVE BEEN MADE MORE TRANSPARENT

- Examples of AI systems that have been made more transparent include:
 - Medical diagnosis systems: AI systems are being used to diagnose diseases more accurately and efficiently. By making these systems more transparent, doctors can better understand how the AI arrived at a diagnosis and make more informed decisions.
 - Predictive policing: AI systems are being used to predict crime rates and identify potential suspects. By making these systems more transparent, police departments can build trust with the community and ensure that AI is used fairly and ethically.
 - Credit scoring: AI systems are being used to assess creditworthiness and predict loan risk. By making these systems more transparent, consumers can better understand why they were approved or denied for a loan and take steps to improve their credit scores.

BENEFITS AND LIMITATIONS OF DIFFERENT APPROACHES TO AI TRANSPARENCY

- **Providing access to training data:**
 - Benefits: Can help to identify and mitigate biases in AI systems.
 - Limitations: May compromise privacy if the training data contains sensitive information.
- **Algorithmic explainability:**
 - Benefits: Can increase trust and facilitate human oversight of AI systems.
 - Limitations: May not be possible in all contexts, such as for deep learning models.
- **Human oversight:**
 - Benefits: Can enable interventions in cases of malfunction.
 - Limitations: Requires significant resources and may involve harmful human biases.
- It is important to note that there is no one-size-fits-all solution to AI transparency. The best approach will vary depending on the specific AI system and its intended use. However, by being transparent about how AI systems work, we can build trust, identify and mitigate biases, and improve performance.

THE CASE OF THE SELF-DRIVING CAR ACCIDENT

- In 2018, an Uber self-driving car struck and killed a pedestrian in Arizona.
- The accident was caused by a combination of factors, including the self-driving system's inability to detect the pedestrian in time and the safety driver's failure to intervene.
- Who is responsible for the accident?
 - The self-driving car company is responsible for developing and deploying a safe system.
 - The person who programmed the self-driving car is responsible for ensuring that the system is programmed correctly.
 - The pedestrian may also be held partially responsible, depending on the circumstances of the accident.
- How could the accident have been prevented?
 - The accident could have been prevented if the self-driving system had been able to detect the pedestrian in time and take evasive action.
 - The accident could also have been prevented if the safety driver had been more attentive and intervened to prevent the collision.

THE CASE OF THE AI SYSTEM THAT PREDICTED CUSTOMER CHURN

- ChurnZero is an AI system that is used to predict customer churn for businesses.
- ChurnZero analyzes various data sources such as customer feedback, engagement data, and revenue data to predict the likelihood of a customer leaving and provides actionable insights to prevent churn.
- (successfully implemented by several large corporations, including Intuit and Indeed)
- What was the goal of the AI system? The goal of the AI system was to predict the likelihood of a customer leaving a company and to provide information on ways to prevent customer churn.
- How was it implemented? The AI system was implemented by analyzing various data points such as customer behavior, purchase history, and feedback which were then used to create personalized reports for businesses.
- What were the results? The results of implementing the AI system led to reduced customer churn rates and increased customer retention for the businesses that utilized it.



THE CASE OF THE AI SYSTEM THAT TARGETED PEOPLE WITH ADVERTISING

- In 2018, Facebook was embroiled in a scandal involving the use of personal data by the company Cambridge Analytica.
- Cambridge Analytica had used data from Facebook users to tailor political advertisements to specific audiences during the 2016 US presidential election, raising concerns about privacy and the use of personal data in targeted advertising.
- Facebook has since implemented stricter privacy policies for advertisers.
- What was the goal of the AI system?
 - The goal of the AI system was to improve the targeting of advertising by identifying specific audience demographics and behaviors.
- How was it implemented?
 - The AI system was implemented by analyzing large amounts of data concerning user demographics, search history, and past purchases, which were then used to tailor advertising content to specific user groups.
- What were the results?
 - The results of implementing the AI system led to increased advertising effectiveness and a better return on investment for companies. However, the use of personal data to target advertisements has also raised privacy concerns among users.

CONCLUSION

- AI systems are powerful tools that can be used for good or for bad.
- It is important to be aware of the potential risks and benefits of AI systems, and to take steps to mitigate the risks.
- We also need to develop ethical guidelines for the development and use of AI systems.

DISCUSSION QUESTIONS

- What are your thoughts on who should be responsible for the actions of AI systems?
- What are some ways to address the challenges of assigning responsibility in AI systems?

REFERENCES

- The case of the self-driving car that struck and killed a pedestrian : - "Uber self-driving car in fatal Arizona crash had software flaws: NTSB", Reuters, 2019. Disponible sur : <https://www.cbc.ca/news/business/uber-self-driving-car-2018-fatal-crash-software-flaws-1.5349581>
- The case of the AI system that was used to predict customer churn : - "How Indeed Uses AI to Predict and Prevent Customer Churn", AI Business, 2019. Disponible sur : <https://aibusiness.com/how-indeed-uses-ai-to-predict-and-prevent-customer-churn/>
- The case of the AI system that was used to target people with advertising : - "Cambridge Analytica scandal: What you need to know", BBC News, 2018. Disponible sur : <https://www.bbc.com/news/technology-43465968> and <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>