



Ethical AI microcredential

BOOKLET

CU6 | AI Ethics, a practical approach

Project number:
2022-1-ES01-KA220-HED-000085257

How to use this Flipbook?

This document is interactive. Throughout the document, you will find links to additional information.



Button that takes you to the beginning of the document. This icon appears on the top right corner of the pages.



Whenever you see this arrow, it means that you have an **interactive color text** to click on, that has an external link associated to it.

DISCLAIMER: Please note that we cannot guarantee the continued availability of external content, such as videos, as they may be subject to change or removal by its authors or host platforms.

Index

Click on the menu

01. Introduction

02. Understanding the ethical implications of AI

03. Identification and mitigation of unethical practices in AI

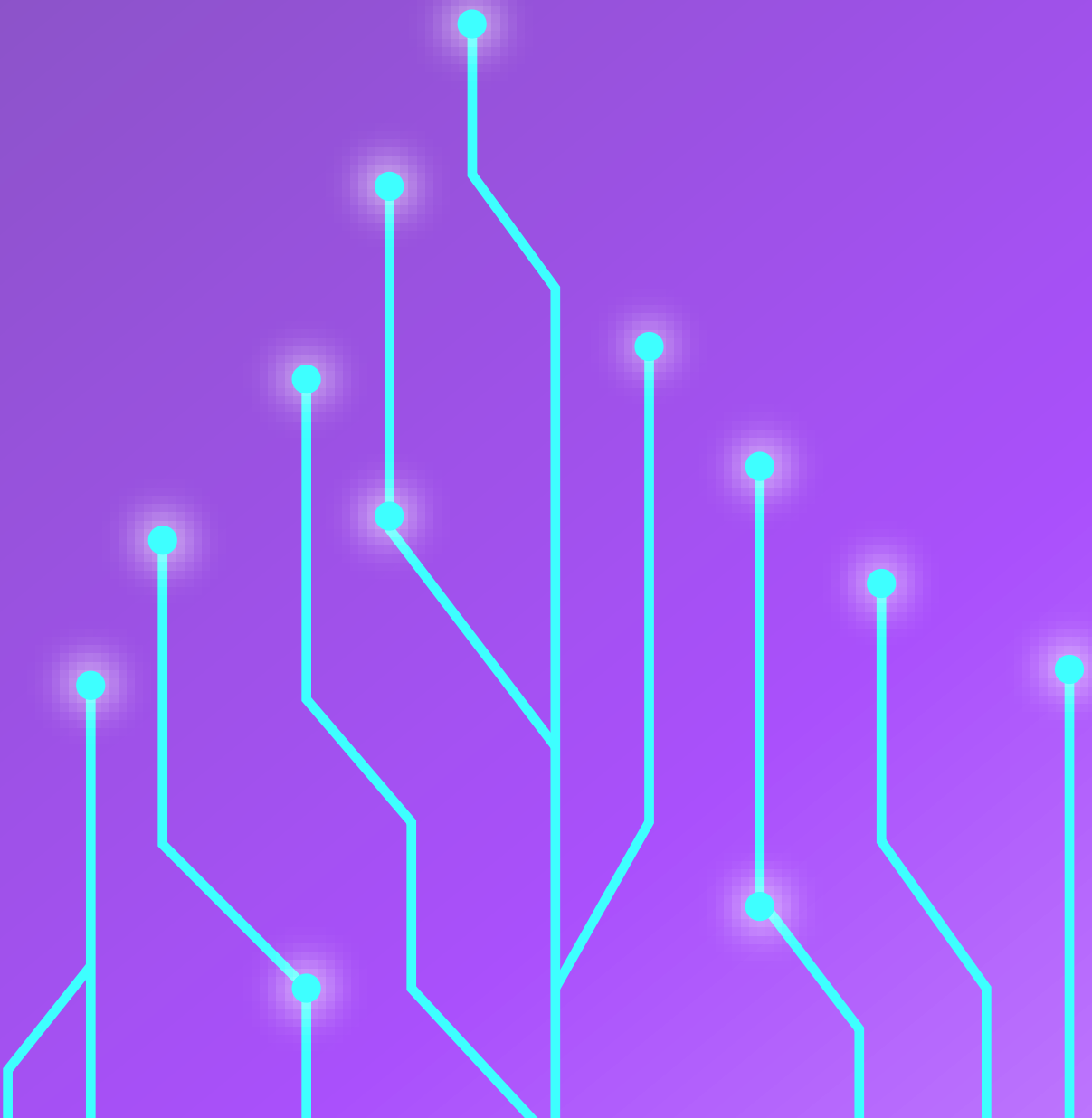
04. Practical application of ethical guidelines in AI

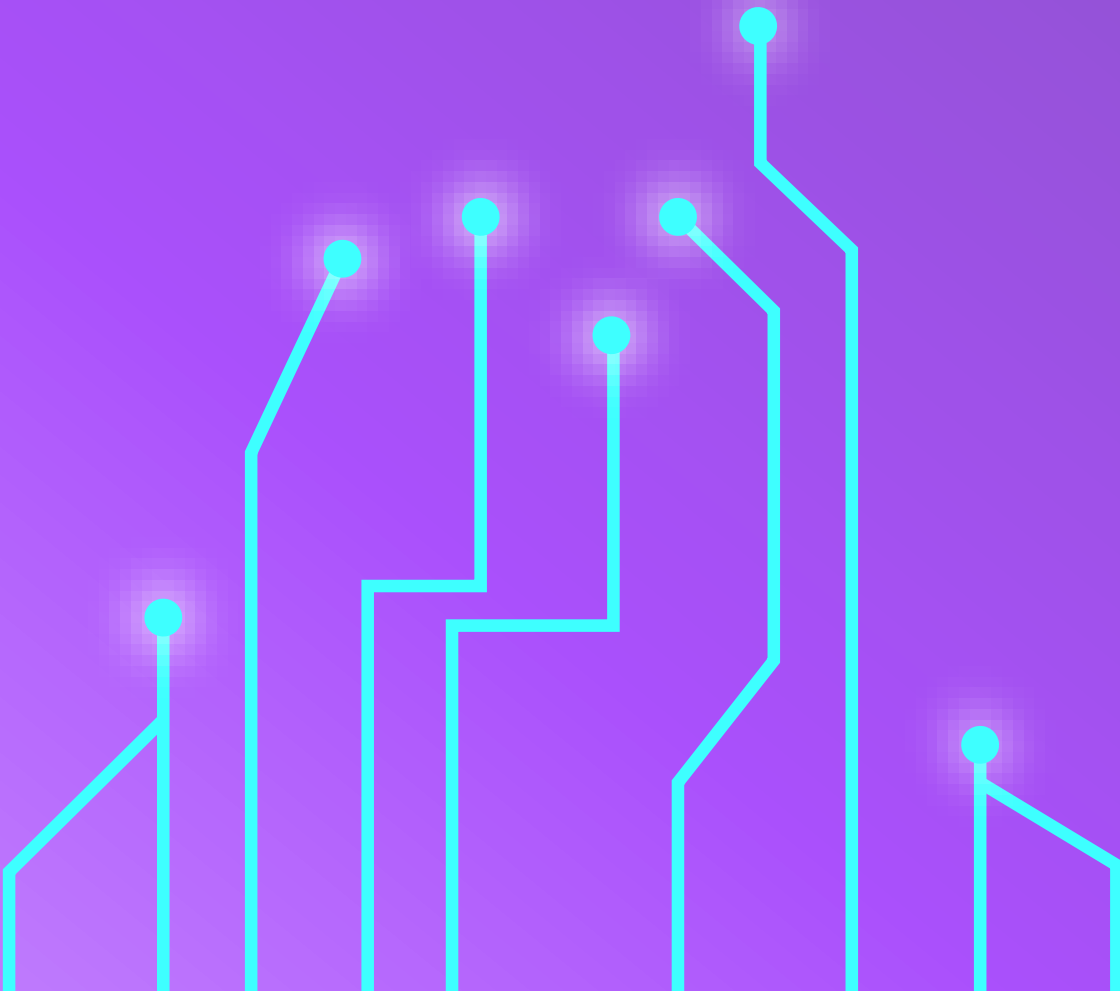
05. Fostering responsible AI development and deployment

06. Conclusion

01. Introduction

CU6 | AI Ethics, a practical approach





01. Introduction

This competence unit is designed for students to delve into the practical aspects of AI ethics, embedding principles of responsible AI development and usage in real-world scenarios. Students are encouraged to be proactive in implementing ethical guidelines in AI environments, drawing from theoretical knowledge and translating it into actionable insights.

The outcomes of this competence unit encompass:

- **Understanding the ethical implications of AI:** students will begin with a solid grounding in the various ethical dimensions that AI encompasses. Following the insights provided by Floridi et al. (2018) concerning the ethical concerns surrounding AI, students will be introduced to the potential implications of AI on society, individuals, and the global community.
- **Identification and mitigation of unethical practices in AI:** addresses the need for recognizing and mitigating potential unethical practices in AI. Drawing from the reflections of Bostrom (2014), who explored the future of AI and its alignment with human values, students will learn how to identify and prevent unethical practices in AI development and deployment.

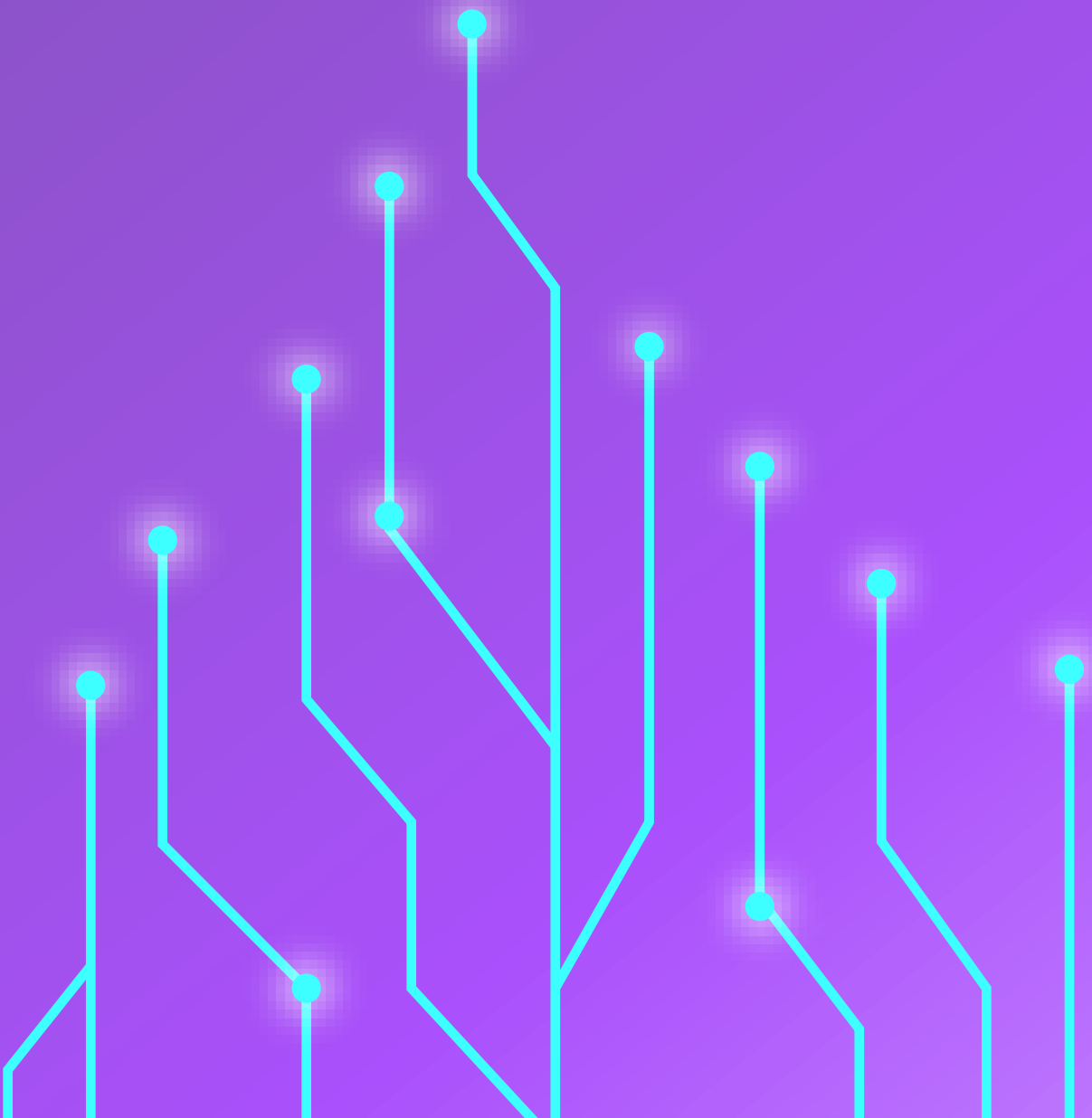


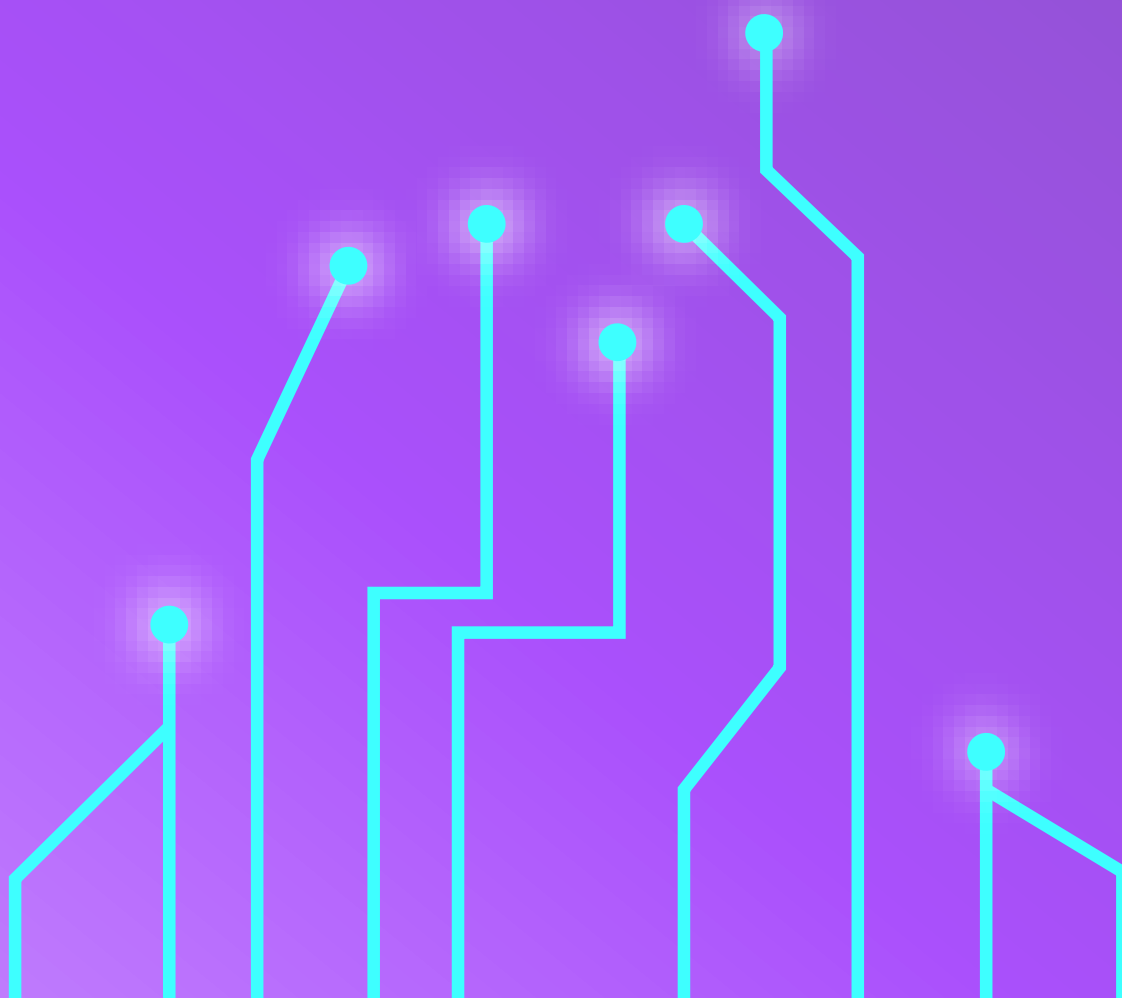
- **Practical application of ethical guidelines in AI:** students in this module are encouraged to translate theoretical understanding into practical action. With insights from the works of Ryan & Stahl (2020), focusing on the formulation of ethical guidelines, students will engage in activities that foster the practical application of these guidelines in AI scenarios.
- **Fostering responsible AI development and deployment:** will focus on fostering an environment that encourages responsible AI development and deployment. Based on the framework proposed by Jobin et al. (2019), which discussed global perspectives on AI ethics, students will explore methods to foster global collaboration and the creation of responsible AI.



02. Understanding the ethical implications of AI

CU6 | AI Ethics, a practical approach





02. Understanding the ethical implications of AI

Throughout this course we took a deep dive into the world of AI, exploring its capabilities and associated risks. By now, you're likely well aware of AI's transformative power and its growing presence across various sectors. Yet, with such immense power comes a profound ethical responsibility.

As we navigate the complex landscape of Ethical AI, various ethical schools of thought offer valuable guidance. Utilitarianism, for instance, emphasises maximising overall well-being, prompting us to consider how AI can be designed to benefit the greatest number of people. Deontological ethics, on the other hand, focuses on the inherent rightness or wrongness of actions, urging us to ensure AI systems operate within established ethical frameworks. Finally, virtue ethics highlights the importance of character and moral development, encouraging us to foster a culture of responsible AI development that prioritises human values.

In this final unit we will explore these core schools of thought, translating them into practical considerations for building ethical AI. We'll examine issues surrounding fairness, transparency, and accountability, and discuss how to develop robust frameworks that ensure AI serves humanity in a responsible and beneficial way.





By critically evaluating the ethical implications of AI, we can pave the way for a future where these powerful technologies contribute to a more just and equitable society.

> Overview of Ethical Theories and their Application in AI

Utilitarianism - a prominent ethical theory, it emphasises maximising overall well-being or "utility" for the greatest number of people. Applied to AI, this means prioritising outcomes that benefit society as a whole, while minimising potential harms.

- **Individual Impact:** Imagine an AI-powered security camera system in a neighbourhood. While it might enhance public safety for residents, it could also lead to increased surveillance and a feeling of being constantly monitored. A utilitarian perspective would weigh the individual's right to privacy against the potential reduction in crime for the entire community.
- **Community Impact:** AI-powered algorithms are increasingly used in social media platforms to optimise user engagement. However, these algorithms can contribute to the spread of misinformation, impacting public discourse and potentially harming communities. Utilitarianism would advocate for safeguards to minimise the spread of misinformation while maintaining user engagement on the platform.

- **Societal Impact:** The development of autonomous weapons systems raises complex ethical issues. While AI could potentially minimise human casualties in warfare, it also introduces questions about accountability and the morality of delegating life-or-death decisions to machines. Utilitarianism would urge careful consideration of the societal implications and potential long-term consequences of deploying such technologies.

Deontological Ethics - another influential theory, it prioritises the inherent rightness or wrongness of actions, independent of their outcomes. This approach emphasises established moral rules and principles such as human dignity, user consent, privacy protection, and ethical standards. These principles should guide developers in designing and deploying AI systems.

- **Individual Impact:** Facial recognition technology used in workplaces can enhance security, but it also raises concerns about user consent and potential bias in the algorithms. Deontological ethics would require ensuring clear user consent for facial recognition and ongoing monitoring for potential biases within the system.
- **Community Impact:** Predictive policing algorithms that aim to reduce crime by identifying high-risk areas raise concerns about fairness and potential discrimination. Deontological ethics would advocate for ensuring these algorithms are unbiased and don't disproportionately target marginalised communities.



- **Societal Impact:** The increasing use of AI in autonomous weapons systems presents a moral dilemma. Deontological ethics would likely raise questions about the ethical implications of delegating life-or-death decisions to machines and the importance of maintaining human accountability in warfare.

Virtue Ethics - inspired by Aristotle, it emphasises the cultivation of moral character and virtues like honesty, integrity, and empathy. Applied to AI development, this theory encourages promoting ethical behaviour and decision-making within the field. This means building AI systems that prioritise traits like empathy, transparency, and integrity.

- **Individual Impact:** Developers of social media platforms often prioritise user engagement to drive revenue. However, a virtue ethics approach would require them to consider the broader societal impacts of their platforms, such as the potential for addiction or the spread of misinformation.
- **Community Impact:** Job automation technologies powered by AI can increase productivity and efficiency. However, they also raise concerns about job displacement, leading to unemployment and economic inequality. A virtue ethics perspective would encourage developers to consider these potential consequences and explore ways to mitigate negative impacts on communities.

- **Societal Impact:** AI is increasingly used in healthcare to improve diagnostics and treatment plans. However, patient-centered care relies heavily on empathy and compassion. Virtue ethics would advocate for ensuring AI systems in healthcare complement human judgement and prioritise ethical considerations alongside efficiency and accuracy.

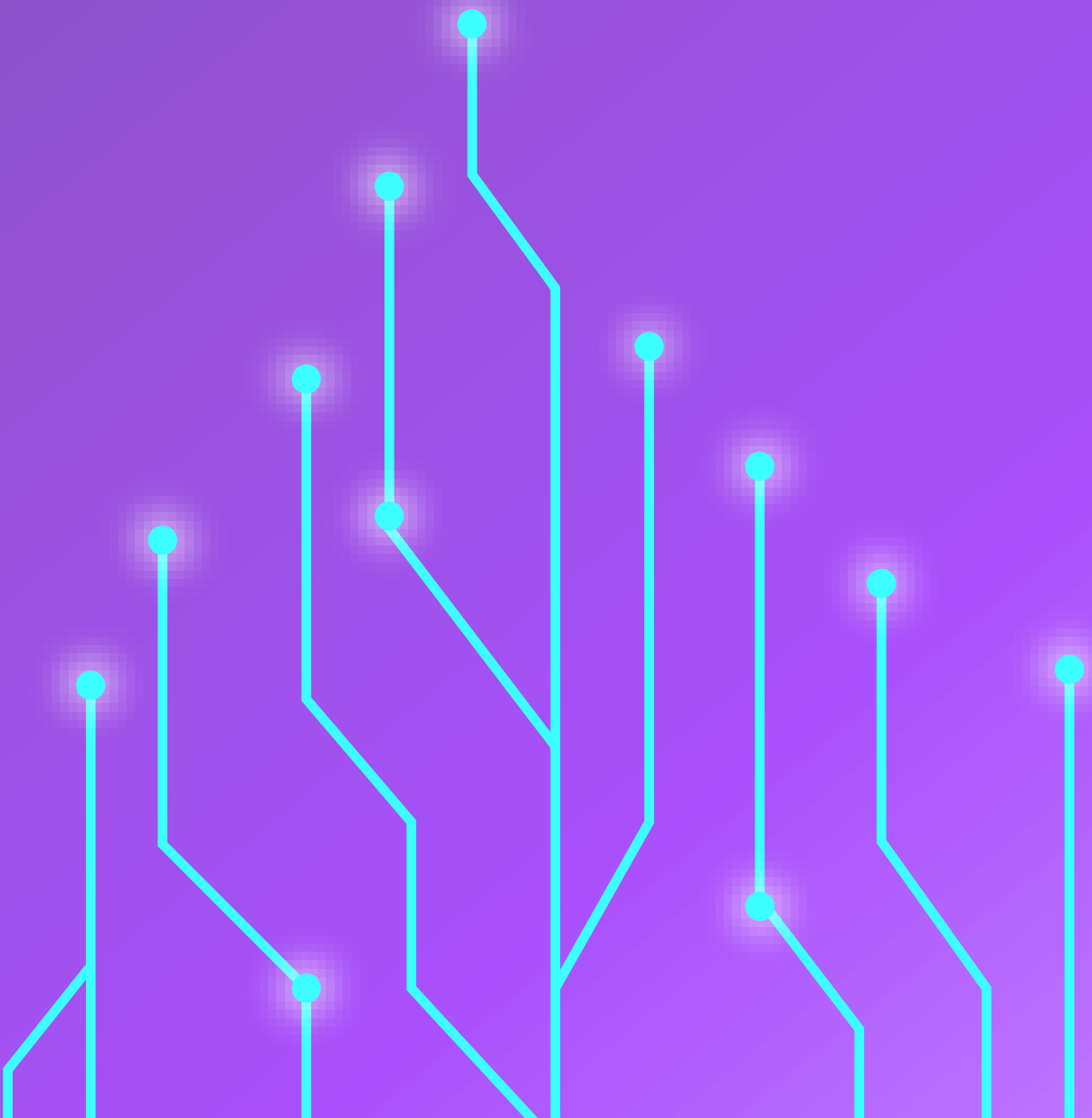


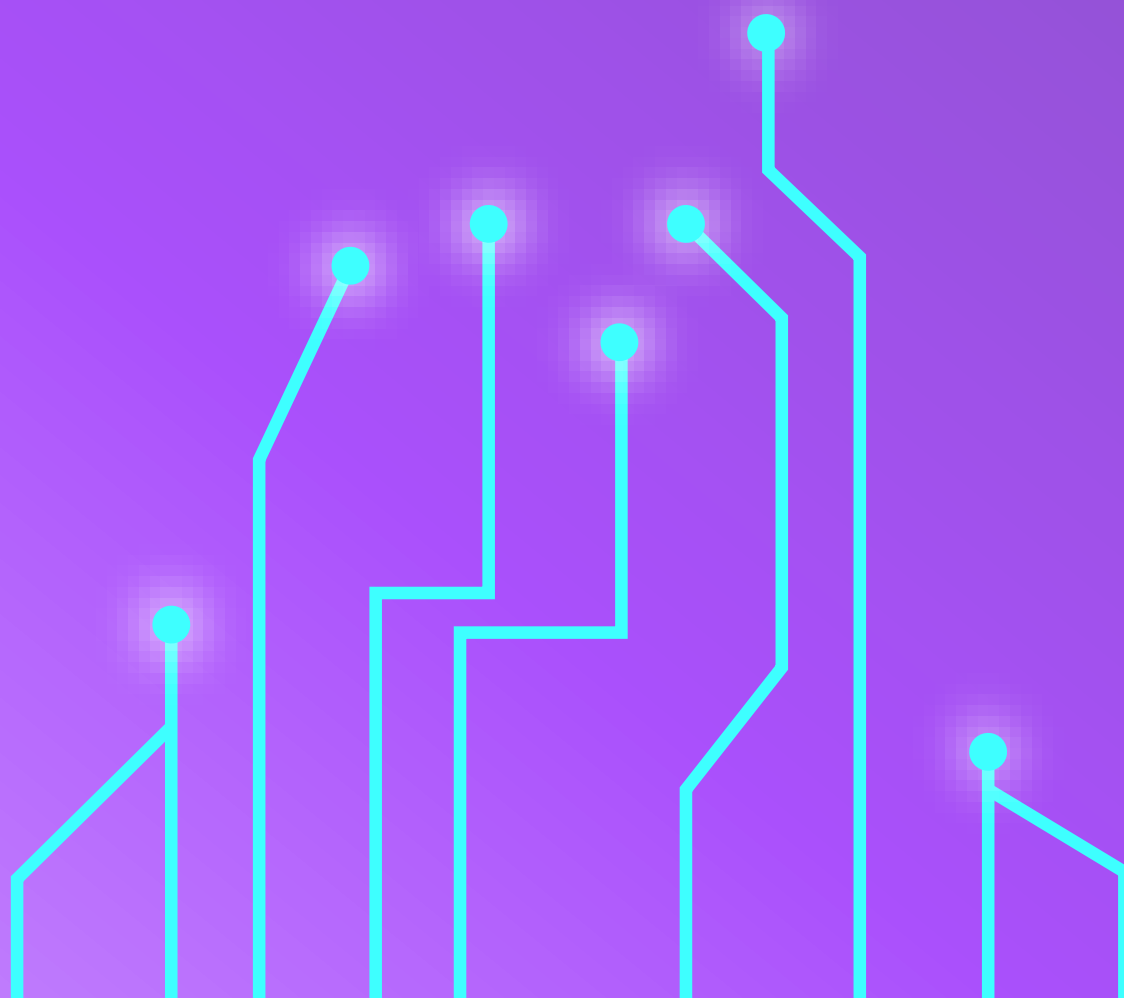


0101

03. Identification and mitigation of unethical practices in AI

CU6 | AI Ethics, a practical approach





03. Identification and mitigation of unethical practices in AI

As you know already, unethical practices in AI can manifest in various forms, including bias, privacy violations, and lack of transparency. In this section we will quickly review some of the main unethical practices covered so far, how to identify them, and then demonstrate how a potential mitigation strategy for each of them could work.

> Bias and Discrimination

- **Indicators:** Disproportionate impact on certain demographic groups. Systematic errors or disparities in decision-making outcomes. Lack of diversity in training data or underrepresentation of marginalised communities.
- **Example:** Facial recognition algorithms that exhibit higher error rates for people of colour, leading to discriminatory outcomes in law enforcement and surveillance.
- **Solution:** Diverse Datasets and Fairer Evaluation Metrics





- **Explanation:** AI models learn from the data they are trained on. If the data is biased, the model will perpetuate those biases. To mitigate this, developers are increasingly focusing on using diverse datasets that represent the real-world population. Additionally, fairer evaluation metrics are being developed to assess AI models for bias and ensure they don't unfairly disadvantage certain groups.

> Lack of Transparency

- **Indicators:** Black-box algorithms with opaque decision-making processes. Limited access to information about data sources, model architecture, and decision criteria. Absence of explanations or justifications for AI-generated outputs.
- **Example:** Financial algorithms that deny loan applications without providing clear reasons or criteria for rejection, exacerbating mistrust and frustration among applicants.
- **Solution:** Explainable AI (XAI) Initiatives
- **Explanation:** Explainable AI (XAI) is a growing field that focuses on developing AI models that are more transparent and interpretable. This allows humans to understand how the model arrives at its decisions, fostering trust and enabling us to identify and address potential biases. For instance, an XAI system might highlight which factors in a loan application data set most influenced the model's decision.

> Privacy Violations

- **Indicators:** Unauthorised access or misuse of personal data. Inadequate safeguards for protecting sensitive information. Failure to obtain informed consent for data collection and processing.
- **Example:** Social media platforms employing AI algorithms to analyse user behaviour and preferences without transparent disclosure or opt-out mechanisms, compromising user privacy and autonomy.
- **Solution:** Data Privacy Regulations and User Control
- **Explanation:** Data privacy regulations like GDPR (General Data Protection Regulation) and CCPA (California Consumer Privacy Act) are empowering users with more control over their personal information. These regulations require companies to be transparent about data collection practices, obtain informed consent, and provide users with the ability to access, correct, or delete their data.



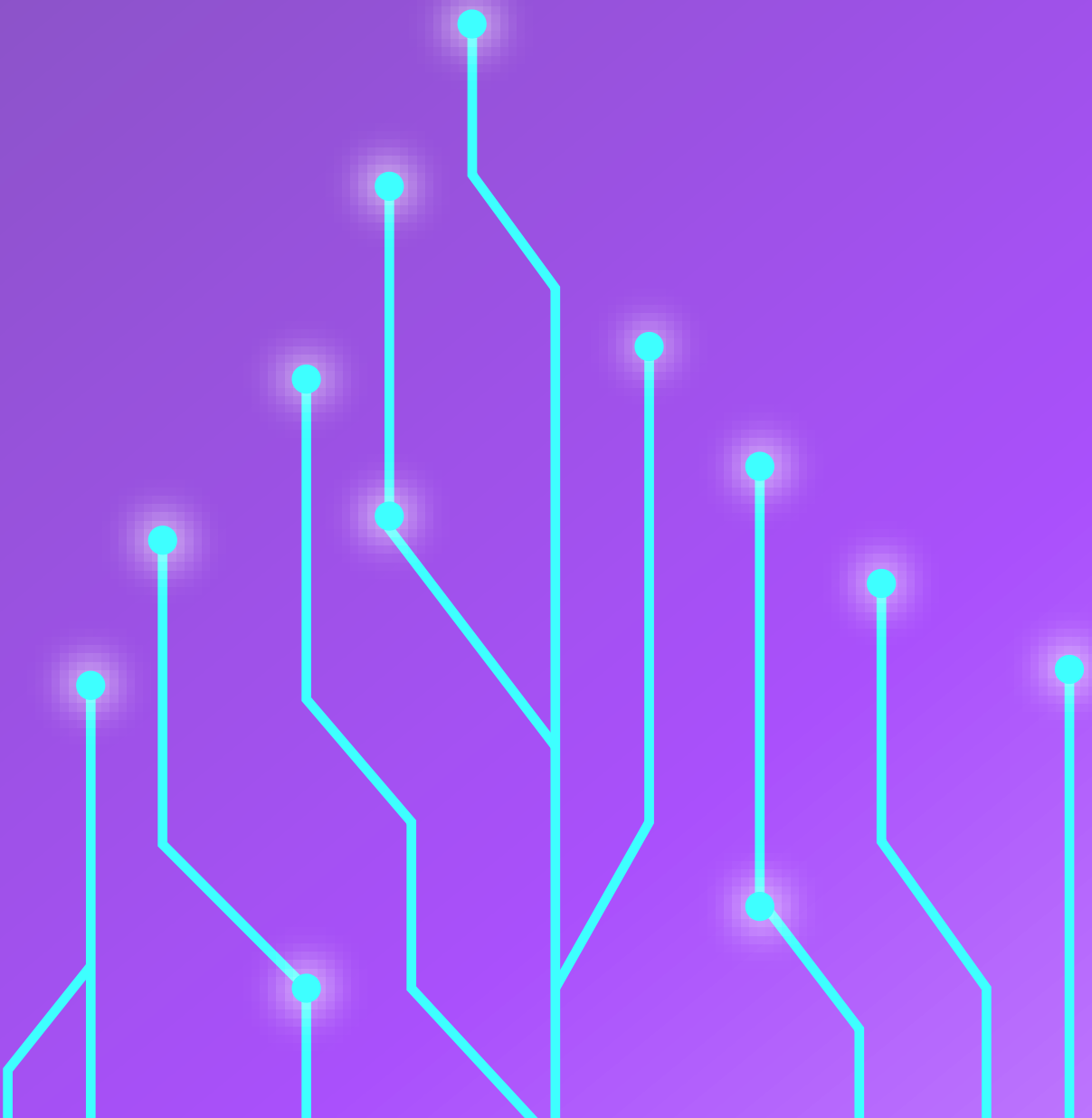


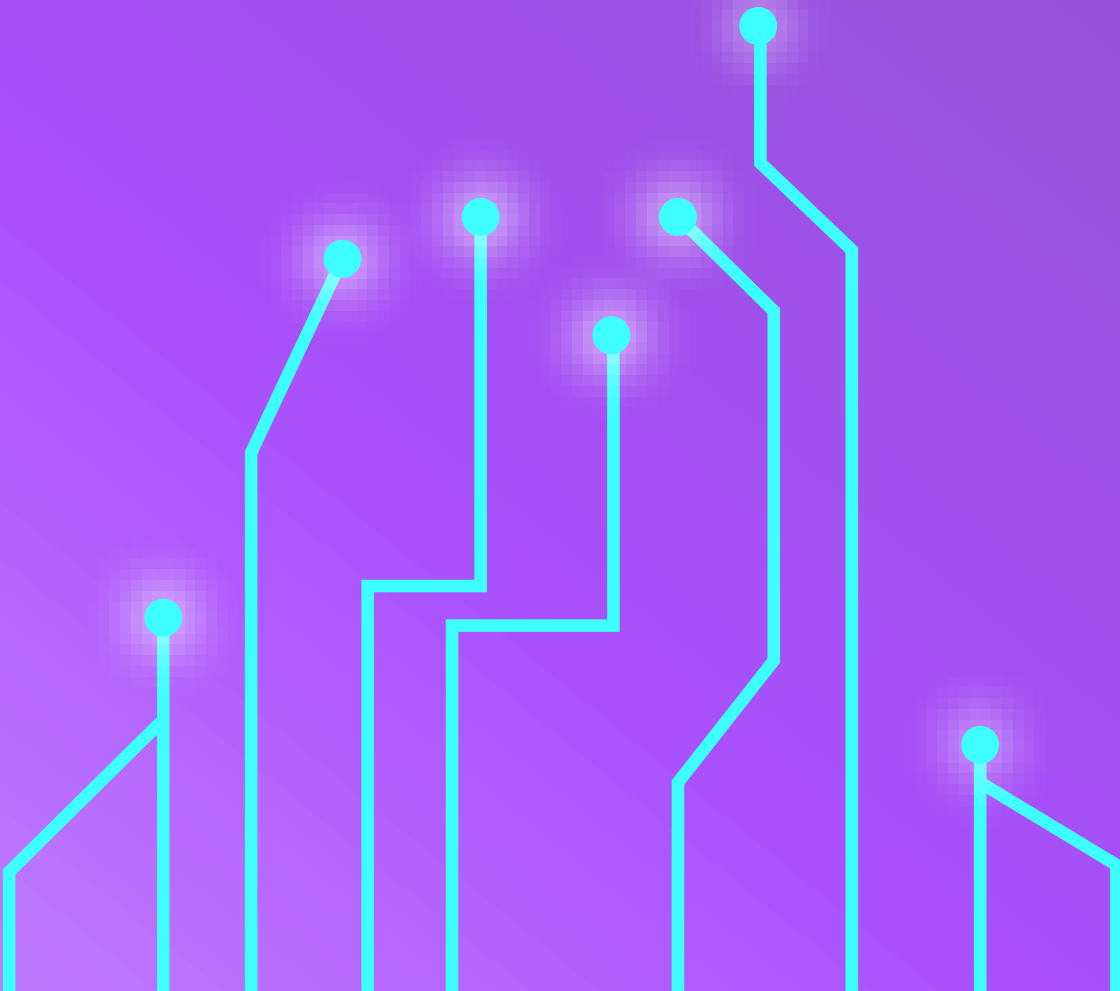
> Lack of Accountability

- **Indicators:** Absence of mechanisms for identifying responsible parties or holding them accountable for AI-related harms. Limited recourse or remedies for individuals affected by AI misconduct. Failure to conduct impact assessments or audits of AI systems.
- **Example:** Autonomous vehicles involved in accidents due to AI failures, where manufacturers evade liability or responsibility for damages, raising questions about legal and ethical accountability in AI-driven technologies.
- **Solution:** Human Oversight and Legal Frameworks
- **Explanation:** For high-risk AI applications, human oversight can be crucial. This means having humans involved in the decision-making process to ensure responsible use of AI and mitigate potential risks. Additionally, legal frameworks are being developed to establish clear lines of accountability for AI-related harms. This will help ensure that developers and companies are responsible for the ethical implications of their AI systems.

04. Practical application of ethical guidelines in AI

CU6 | AI Ethics, a practical approach





04. Practical application of ethical guidelines in AI

➤ **Building Your Ethical AI Toolkit – A Framework for Action**

Ethical guidelines provide a roadmap for ethical decision-making in AI development and deployment. Throughout this unit, we've explored the critical role of ethical considerations in AI development and deployment. We've examined various ethical schools of thought, each offering valuable frameworks for navigating the complex ethical landscape of AI. But how can we translate these abstract principles into practical action?

This section is a simple four-step framework to guide you in developing your own ethical AI guidelines.

1. Step 1: Delineating the AI System's Purpose

The first step involves clearly defining the intended use case for the AI system. What specific task is the AI designed to perform? Consider examples like facial recognition software for security purposes, an AI-powered hiring tool to automate resume screening, or a medical diagnosis assistant to support healthcare professionals. Understanding the system's purpose is crucial for identifying the stakeholders involved and the ethical considerations that will arise.



2. Step 2: Identifying Stakeholders and their Values

Once the use case is established, the second step involves pinpointing the various stakeholders who will be impacted by the AI system. Stakeholders can encompass a wide range of individuals and groups, including the system's users, developers, those affected by its decisions (e.g., loan applicants for an AI-powered lending platform), and society at large.

Each stakeholder group will have its own set of values and priorities that need to be considered when developing ethical guidelines. For instance, users might prioritize privacy and fairness in how their data is handled, while developers might focus on efficiency and accuracy in the system's operation. It's important to acknowledge these diverse values and find ways to balance them when formulating ethical principles.

3. Step 3: Enunciating Core Ethical Principles

Drawing on the understanding of the use case and the identified stakeholders, the third step involves defining the core ethical principles that will guide the development and deployment of the AI system. These principles serve as a foundation for ensuring the responsible and ethical use of AI technology. Some commonly considered principles include fairness, transparency, accountability, privacy, and safety. For example, in the case of an AI-powered hiring tool, fairness would necessitate ensuring the system doesn't discriminate against certain demographics based on biased data. Transparency would require clear communication about how the tool works and the factors it considers when evaluating candidates.

4. Step 4: Translating Principles into Actionable Measures

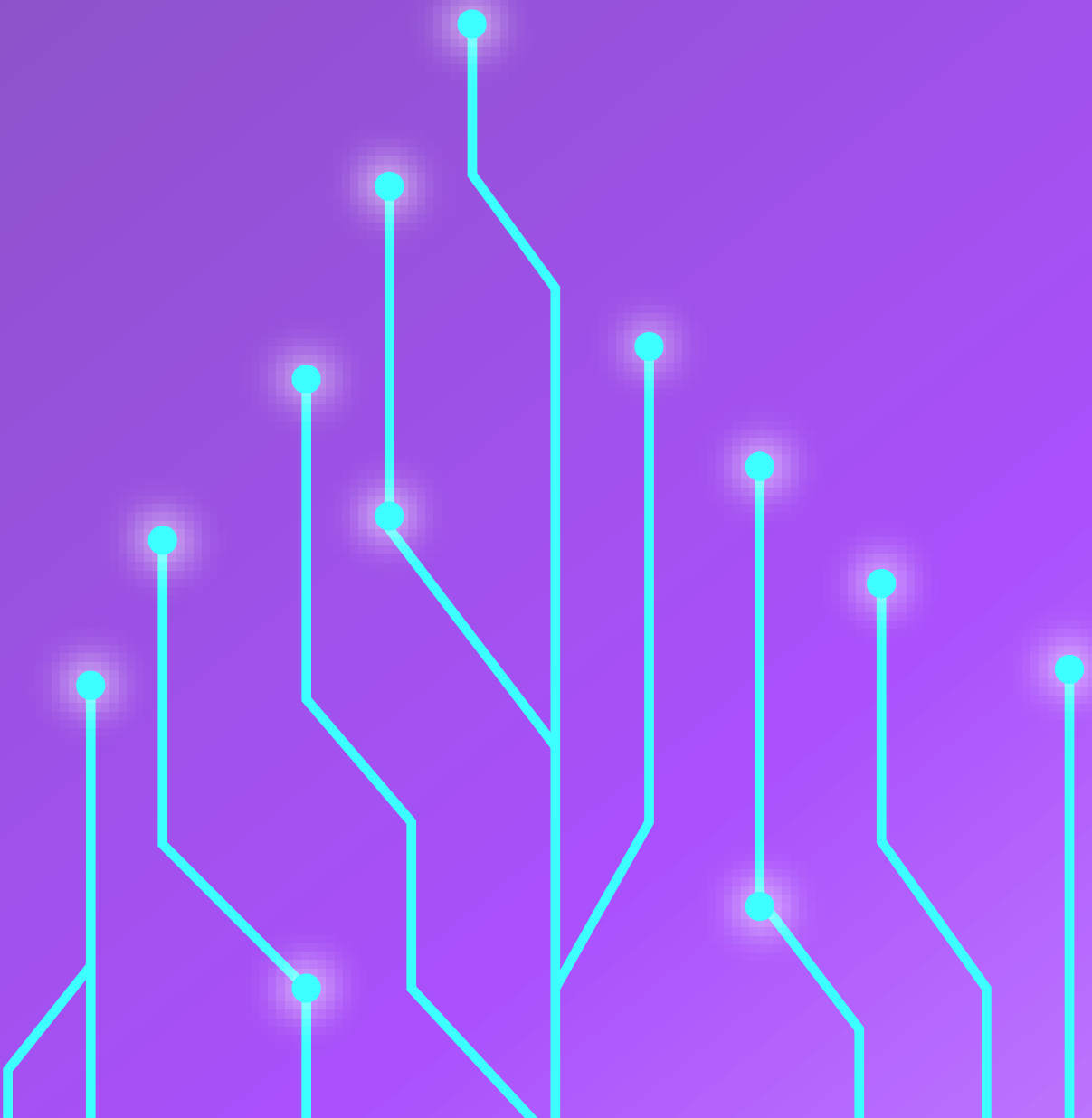
The final step translates the abstract ethical principles into concrete actions that can be implemented to ensure the AI system adheres to those principles. This might involve employing diverse datasets to mitigate bias in the system's decision-making. For transparency, Explainable AI (XAI) techniques can be incorporated to allow users to understand how the AI arrives at its outputs. Additionally, establishing clear lines of accountability for potential harms caused by the AI system is crucial. This could involve assigning responsibility to developers or creating mechanisms for redress for those negatively impacted by the AI's decisions.

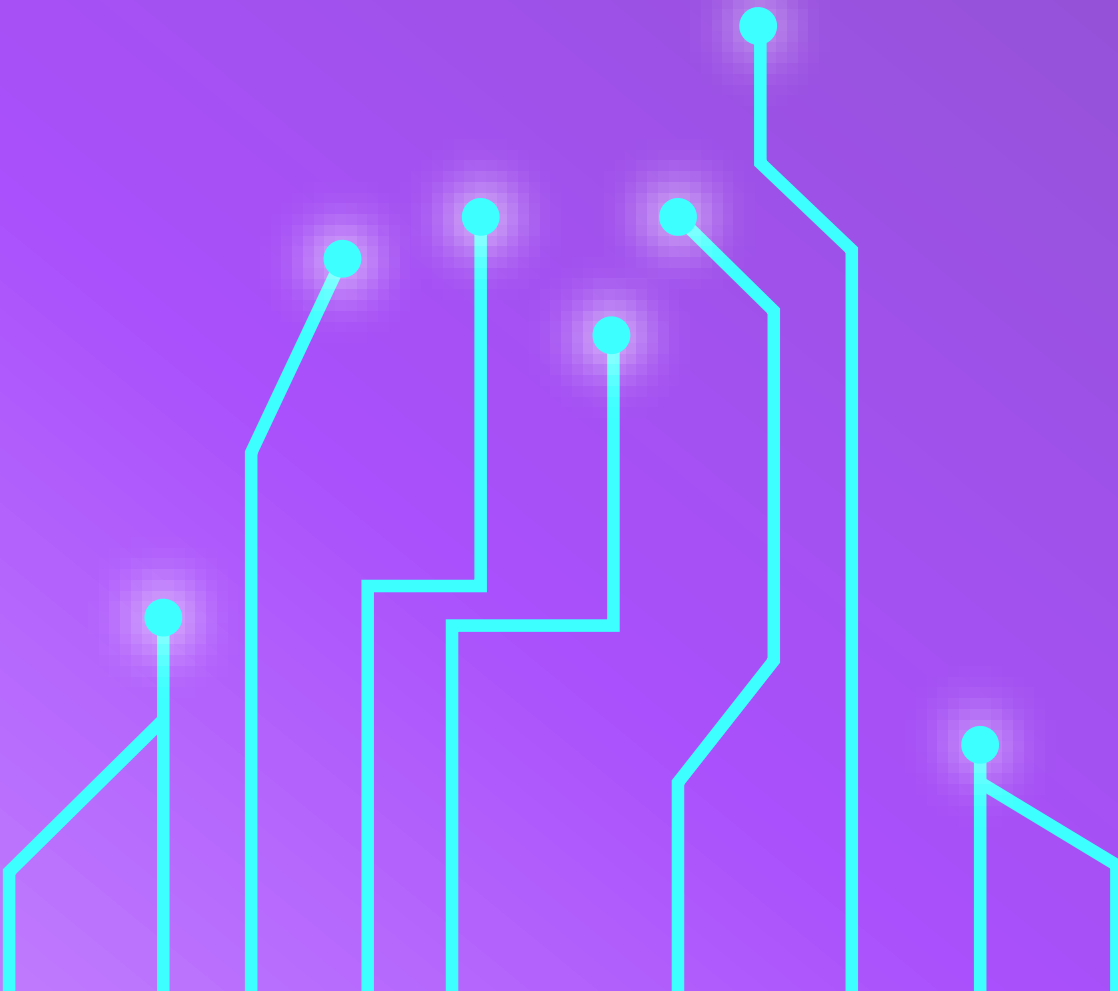


By following these steps, you'll be empowered to critically evaluate AI systems and contribute to the responsible development and use of these powerful technologies.

05. Fostering responsible AI development and deployment

CU6 | AI Ethics, a practical approach





05. Fostering responsible AI development and deployment

> Building a Better Future Together – Global Collaboration for Responsible AI

We've explored the importance of responsible AI development throughout this course. But ensuring ethical AI isn't just a local challenge – it requires a global effort. Here we will examine international initiatives and collaborative strategies that are working to foster responsible AI development and deployment on a worldwide scale.

- **The Global Partnership on Artificial Intelligence (GPAI):**

This is a multistakeholder initiative involving governments, industry leaders, and civil society organisations. The GPAI works to develop best practices and recommendations for responsible AI development and deployment across various sectors.

<https://gpai.ai/>



- **The OECD AI Policy Observatory:**

This initiative by the Organization for Economic Co-operation and Development (OECD) provides a platform for sharing information and best practices on AI policy development among member countries. It aims to foster international dialogue and collaboration on ethical AI governance. <https://oecd.ai/>





- **The UNESCO Recommendation on the Ethics of Artificial Intelligence:**

This is a non-binding international instrument adopted by UNESCO in 2021. It outlines key ethical principles for the development and use of AI, promoting human rights, fairness, transparency, and accountability.

<https://unesdoc.unesco.org/ark:/48223/pf0000380455#:~:text=AI%20actors%20and%20Member%20States,law%2C%20in%20particular%20Member%20States>



> The EU AI Act

The EU AI Act is the first attempt at establishing a comprehensive framework for regulating AI applications within the European Union. It was first proposed by the European Commission in 2021, having finally been approved by the EU Parliament and Council in May 2024.

The Act classifies AI systems into four risk categories—unacceptable, high, limited, and minimal—plus an additional category for general-purpose AI. AI applications deemed to have unacceptable risks, such as those that manipulate human behavior or exploit vulnerabilities, are outright banned.

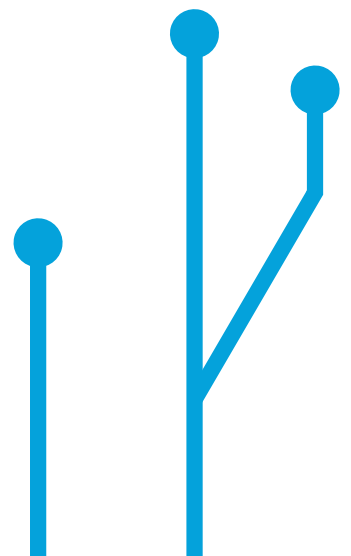
High-risk applications must adhere to stringent security, transparency, and quality requirements and undergo conformity assessments to ensure compliance. Limited-risk applications are subject to transparency obligations, while minimal-risk applications are not regulated. General-purpose AI, especially those with high capabilities, must meet transparency requirements and may be subject to additional evaluations.

This risk-based approach aims to mitigate potential harms while fostering innovation and ensuring AI development aligns with European values and principles.

https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf



<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>



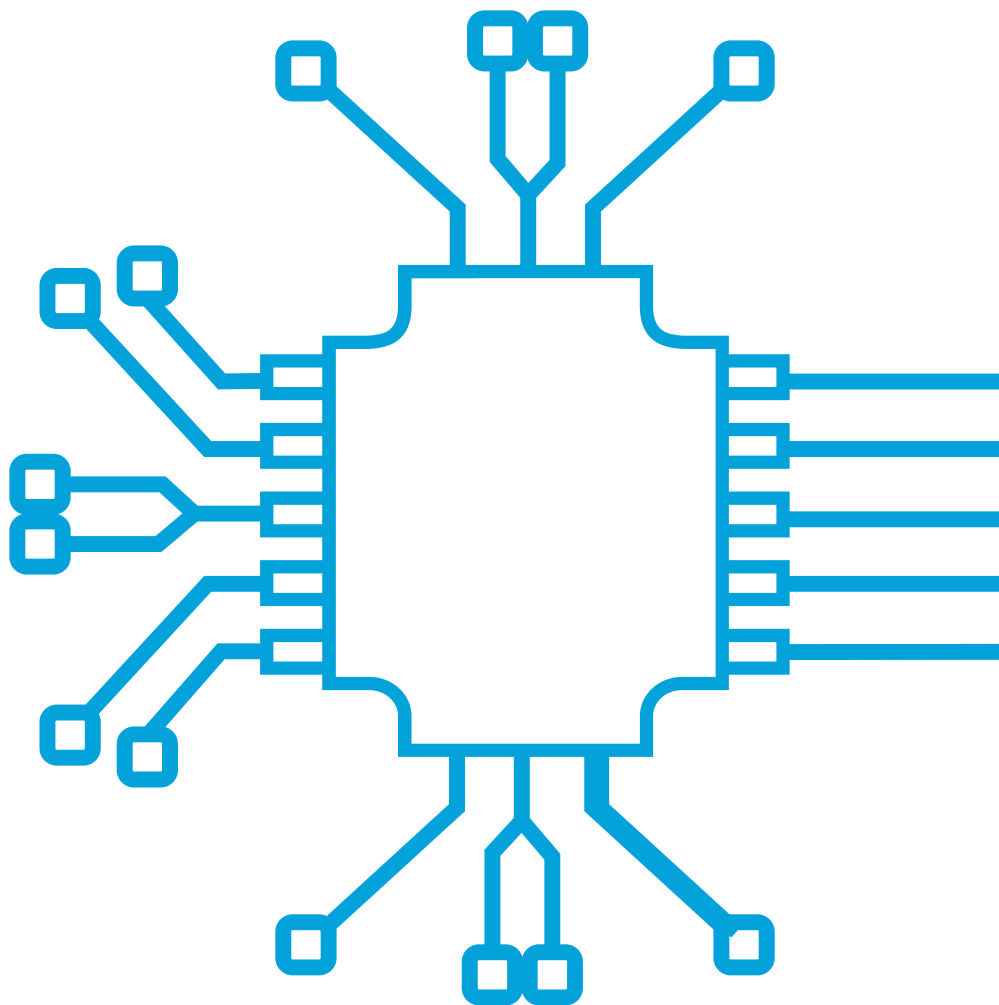


> You CAN Make a Difference - Strategies for Building a Responsible AI Future

You've now gained valuable knowledge about AI ethics and the importance of responsible AI development. But this journey doesn't end here. The future of AI is being shaped right now, and your voice matters. Below are some suggestions to equip you with practical strategies to get involved and contribute to a more responsible AI future in your community and beyond.

- **Raise Awareness:** Choose a specific ethical issue in AI that interests you (e.g., bias in facial recognition, privacy concerns with AI-powered assistants). Research the issue and prepare a short presentation or workshop for your school club, community centre, or even online platforms. Tailor your presentation to your audience and make it engaging (use visuals, interactive elements).
- **Advocate for Change:** Identify existing policy initiatives or organisations working on AI ethics (e.g., your local government representatives, advocacy groups). Research their work and identify areas where you can contribute your voice. Consider writing a letter to your local representative expressing your concerns about a specific AI application or advocating for transparency and accountability measures.

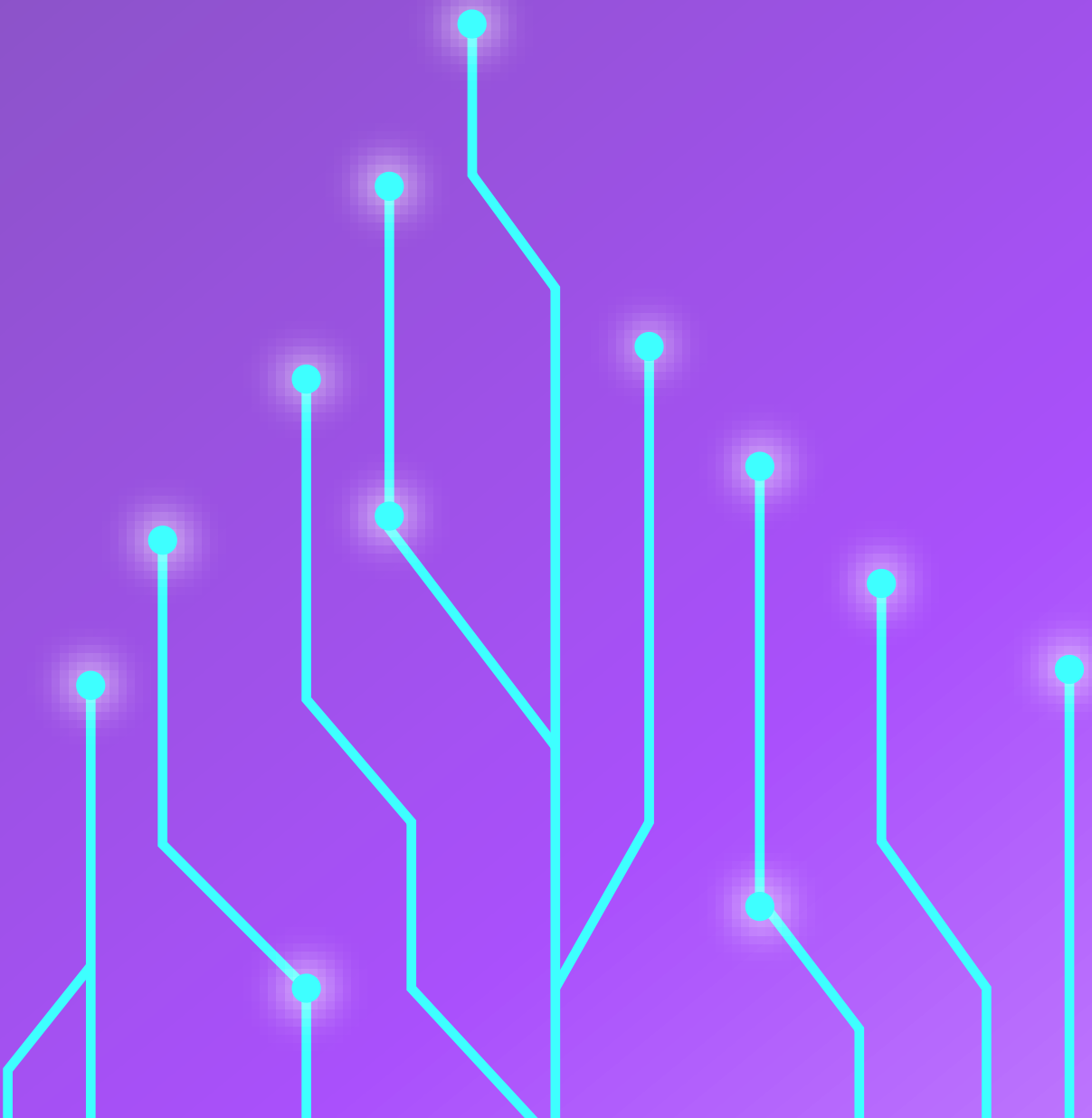
- **Join the Conversation:** Explore online forums, discussions, and social media groups focused on AI ethics. Participate in respectful discussions, share your learnings from this course, and learn from others' perspectives. Consider starting your own blog or online platform to share your thoughts and raise awareness about AI ethics.

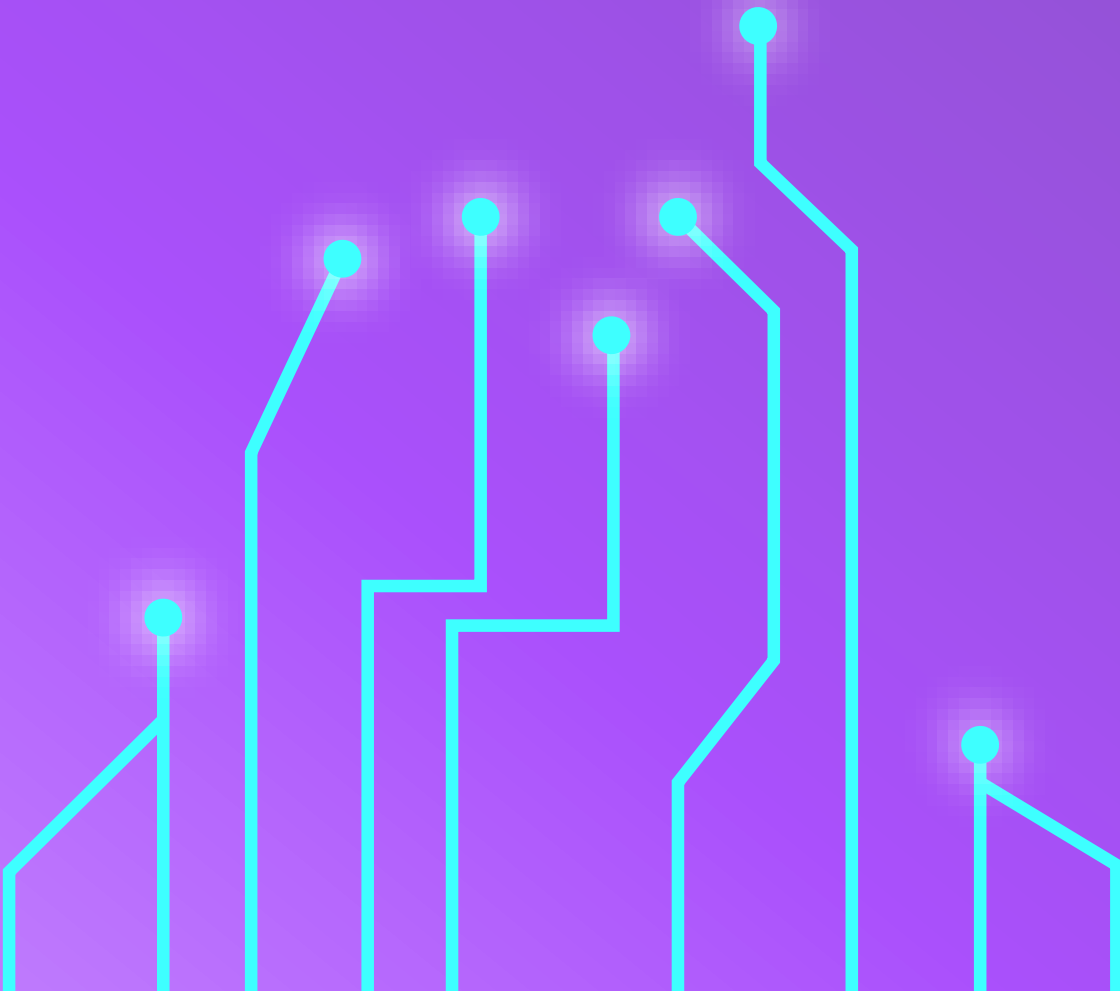




05. Conclusion

CU6 | AI Ethics, a practical approach





06. Conclusion

Throughout this course, we have explored the transformative power of AI.

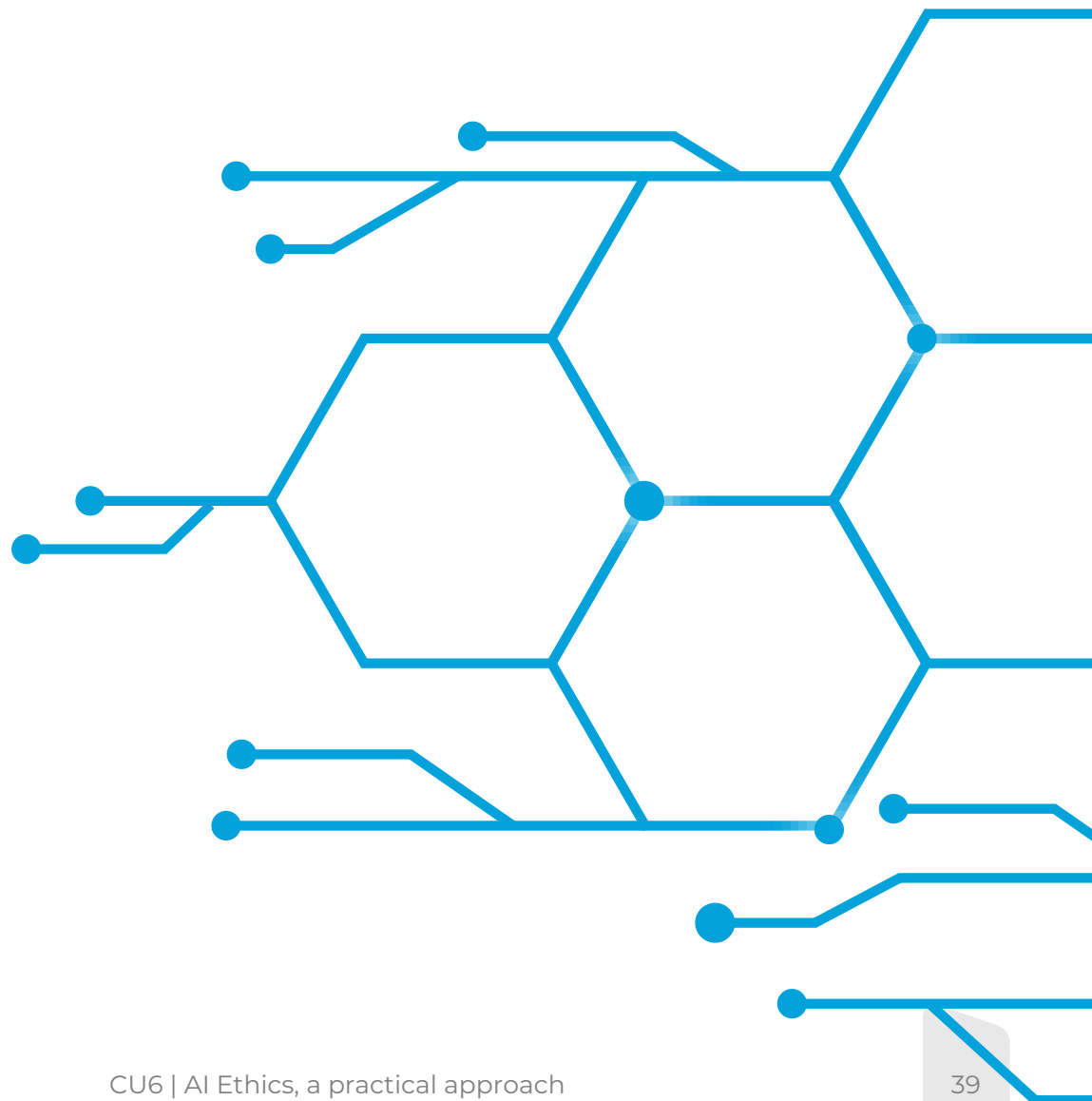
We started by examining the insidious nature of algorithmic bias, which is a critical lens through which all AI development should be viewed. In Unit 2, we delved into the concept of non-maleficence, which ensures AI systems do not cause harm and are designed with safety in mind. Unit 3 explored the crucial principle of accountability, emphasising the need for clear lines of responsibility in the development and deployment of AI. Transparency, which was the focus of Unit 4, ensures that we can understand how AI systems work and make informed decisions about their use. In Unit 5, we tackled the vital connection between AI and human rights, underscoring the importance of fairness and inclusivity in AI development. Finally, in Unit 6, we equipped you with a practical framework for developing your own ethical AI guidelines, empowering you to translate principles into action.

As we stand at the threshold of an AI-driven future, the ethical considerations we explored in this course are not mere academic exercises; they are the foundation for responsible AI development.



By actively engaging in discussions about AI ethics, advocating for responsible practices, and holding developers and policymakers accountable, we can ensure that AI serves humanity in a just, equitable, and beneficial way.

The future of AI is not predetermined; it is a future we will collectively create. Let us strive to build a future where AI becomes a powerful force for good, fostering progress, well-being, and a brighter tomorrow for all.





Charlie



Co-funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.



Universitat
de les Illes Balears



ENGAGING PEOPLE



INNOVATION TRAINING CENTER



AARHUS UNIVERSITY



VAMK UNIVERSITY OF APPLIED SCIENCES



2022-1-ES01-KA220-HED-000085257