



Ethical AI microcredential

BOOKLET

CU4 | Transparency

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. Importance of transparency in AI systems

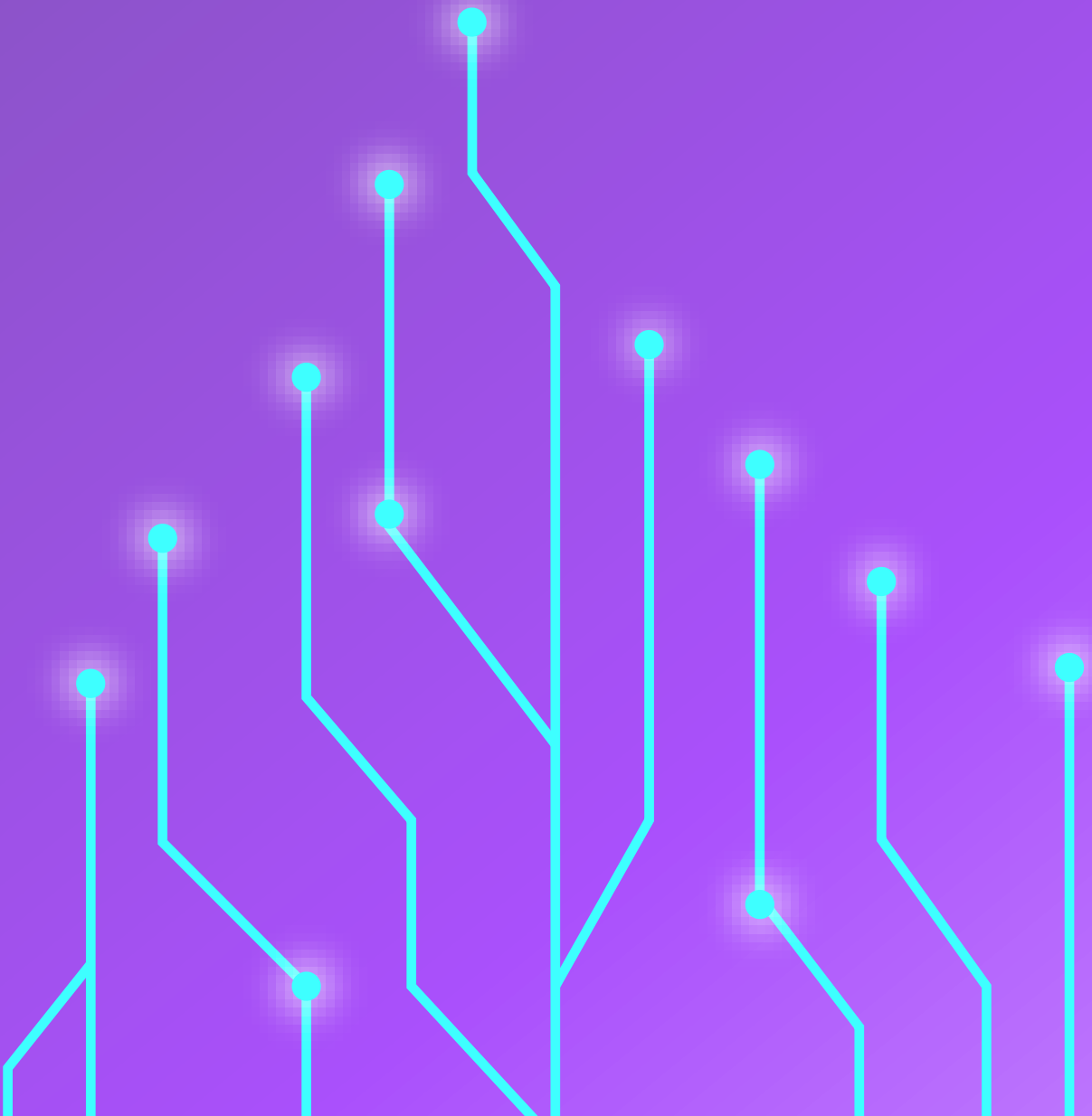
03. Relationship between transparency and Algorithmic Bias

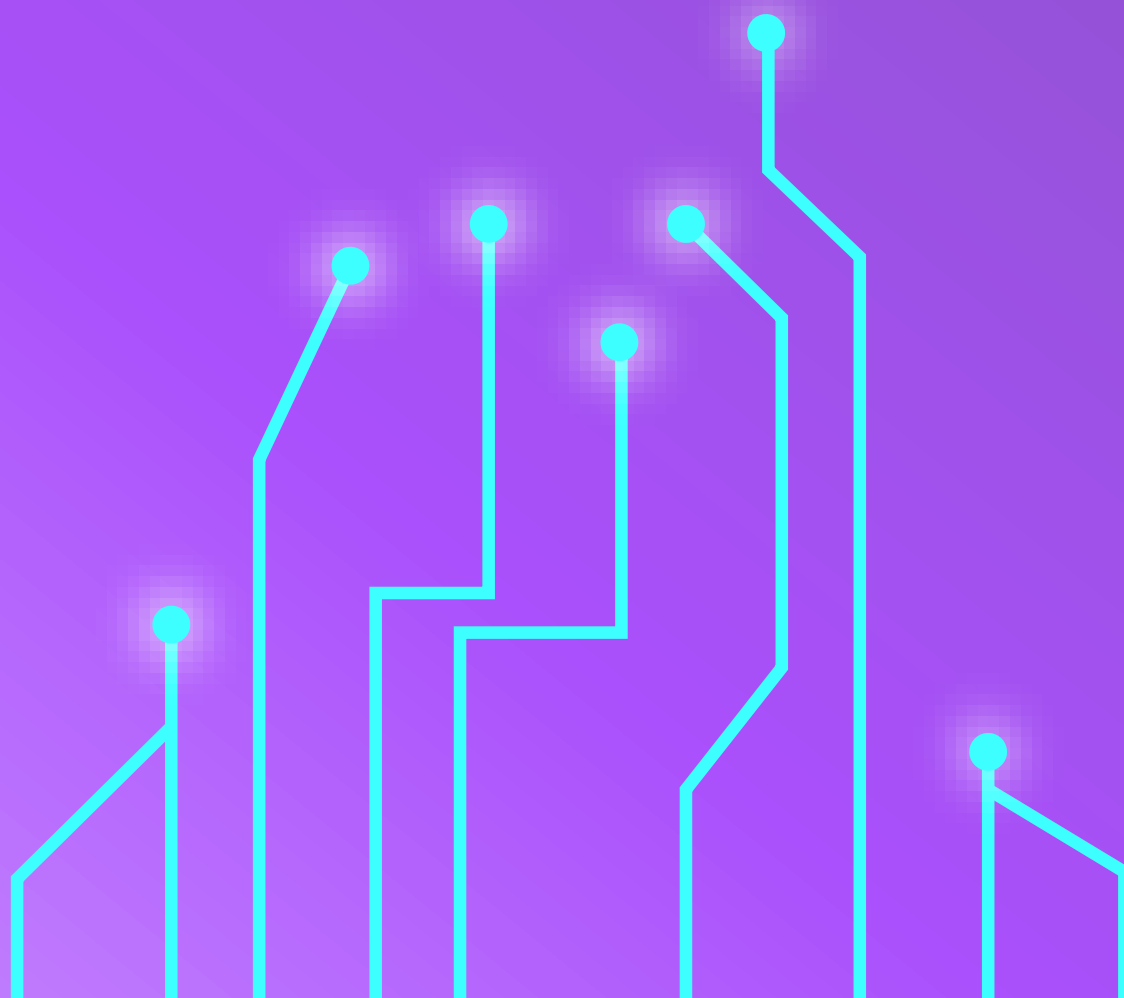
04. Strategies for promoting transparency in AI systems

05. Conclusion

01. Introduction

CU4 | Transparency





01. Introduction

In this competence unit, learners will gain knowledge regarding the importance of transparency in AI systems, focusing on understanding the basic concepts, the relationship between transparency and algorithmic bias and the relevance of the strategies to ensure that AI systems are understandable, explainable, and accessible to stakeholders, recognizing the real-world implications appreciating how important interpretable models, clear documentation, and effective communication can be in fostering a culture of transparency, mitigating algorithmic bias.

The knowledge outcomes for this course include:

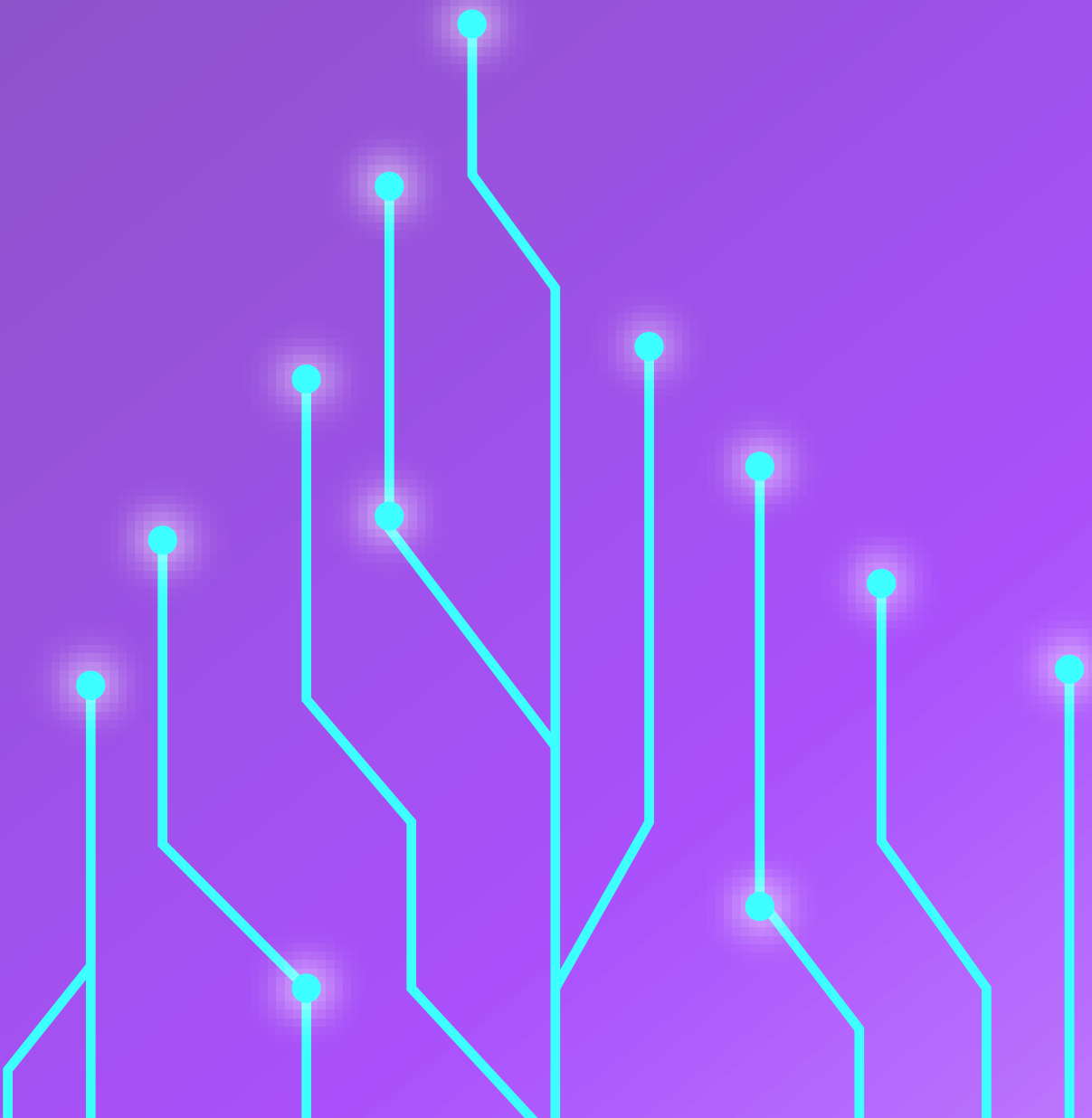
- **Importance of Transparency in AI Systems** and its relevance in ensuring that AI systems are understandable, explainable, and accessible to stakeholders. We will identify the benefits and value the importance of transparent AI systems for building trust and enabling stakeholder understanding. As an example: An AI model designed to detect cancer, even if it is only 1% wrong, could threaten a life. In cases like these, AI and humans need to work together, and the task becomes much easier when the AI model can explain how it reached a certain decision. Transparency makes AI a team player.



- **Relationship between Transparency and Algorithmic Bias** to find the connection between transparency and algorithmic bias, recognizing the dangers of opacity and how increased transparency can help identify, prevent, and mitigate biased outcomes in AI systems. We will recognize the significance of transparency in addressing and mitigating algorithmic bias. As an example: quite often, AI algorithms are opaque in the sense that such explanations are not available to all stakeholders. This opacity can have different sources. Sometimes institutions or corporations fail to communicate when they rely on AI systems or on how these systems work.
- **Strategies for Promoting Transparency in AI Systems**, such as using interpretable models, providing clear documentation, and communicating the decision-making processes of AI applications. We will explain how important these strategies are to promote a culture of transparency and mitigating algorithmic bias. As example, AI can affect various stakeholders, such as users, clients, employees, managers, regulators, or society. To ensure transparency and accountability, you need to engage and empower your AI stakeholders throughout the IS lifecycle.

02. Importance of transparency in AI systems

CU4 | Transparency





02. Importance of transparency in AI systems

Transparency stands as a foundational principle in the development and deployment of artificial intelligence (AI) systems.

Transparency in AI refers to the openness and accessibility of AI systems, allowing stakeholders to understand how algorithms operate, why certain decisions are made, and what factors influence their outputs. It encompasses various aspects, including the availability of information about data sources, algorithmic models, decision-making processes, and potential biases. Transparent AI systems enable stakeholders, including users, developers, policymakers, and the general public, to scrutinize and challenge algorithmic outcomes, fostering trust and accountability.

One of the key benefits of transparent AI systems is their understandability. When AI algorithms are transparent, stakeholders can comprehend how they function and why they produce certain results. This understanding enables users to trust AI technologies and make informed decisions about their use. For example, in the context of a medical diagnosis AI model, transparency allows healthcare professionals to understand how the model arrived at its diagnosis, enabling them to validate its accuracy and reliability before making treatment decisions.





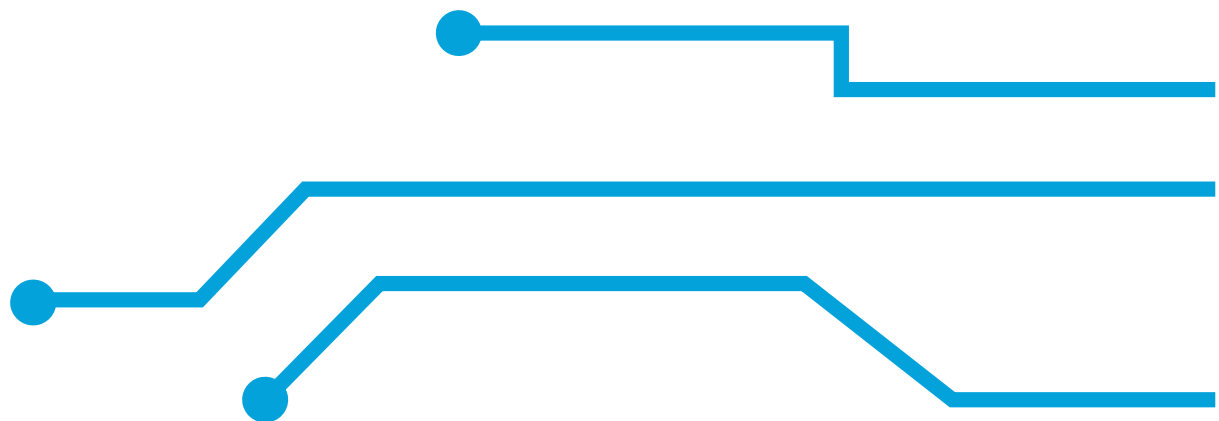
Moreover, transparency facilitates explainability, which is essential for ensuring that AI systems can provide interpretable explanations for their decisions and actions. Explainable AI empowers stakeholders to understand the rationale behind algorithmic outcomes and to identify and correct biases or errors. For instance, in the case of a loan approval AI system, transparency and explainability enable loan applicants to understand why their application was approved or denied, providing insights into the decision-making process and avenues for recourse if they believe the decision was biased or unfair.

Additionally, transparency enhances the accessibility of AI systems, making them more inclusive and equitable. When AI algorithms are transparent, stakeholders from diverse backgrounds and levels of expertise can access and interpret information about their functioning and outcomes. This accessibility ensures that AI technologies are not only understandable but also usable by a wide range of users, including those with disabilities or limited technical knowledge. For example, in the development of AI-powered accessibility tools for people with disabilities, transparency enables users to understand how the tools work and how they can benefit from them.



An illustrative example of the importance of transparency in AI systems is the development of AI models for medical diagnosis, such as detecting cancer. Even if an AI model is highly accurate, with a 99% success rate, the remaining 1% margin of error could have life-threatening consequences for patients. In such critical scenarios, transparency becomes essential for ensuring that healthcare professionals can understand how the AI model reached its diagnosis and can verify its accuracy before making treatment decisions. By providing transparent explanations of its decision-making process, the AI model becomes a valuable tool for healthcare professionals, enhancing their ability to diagnose and treat patients effectively.

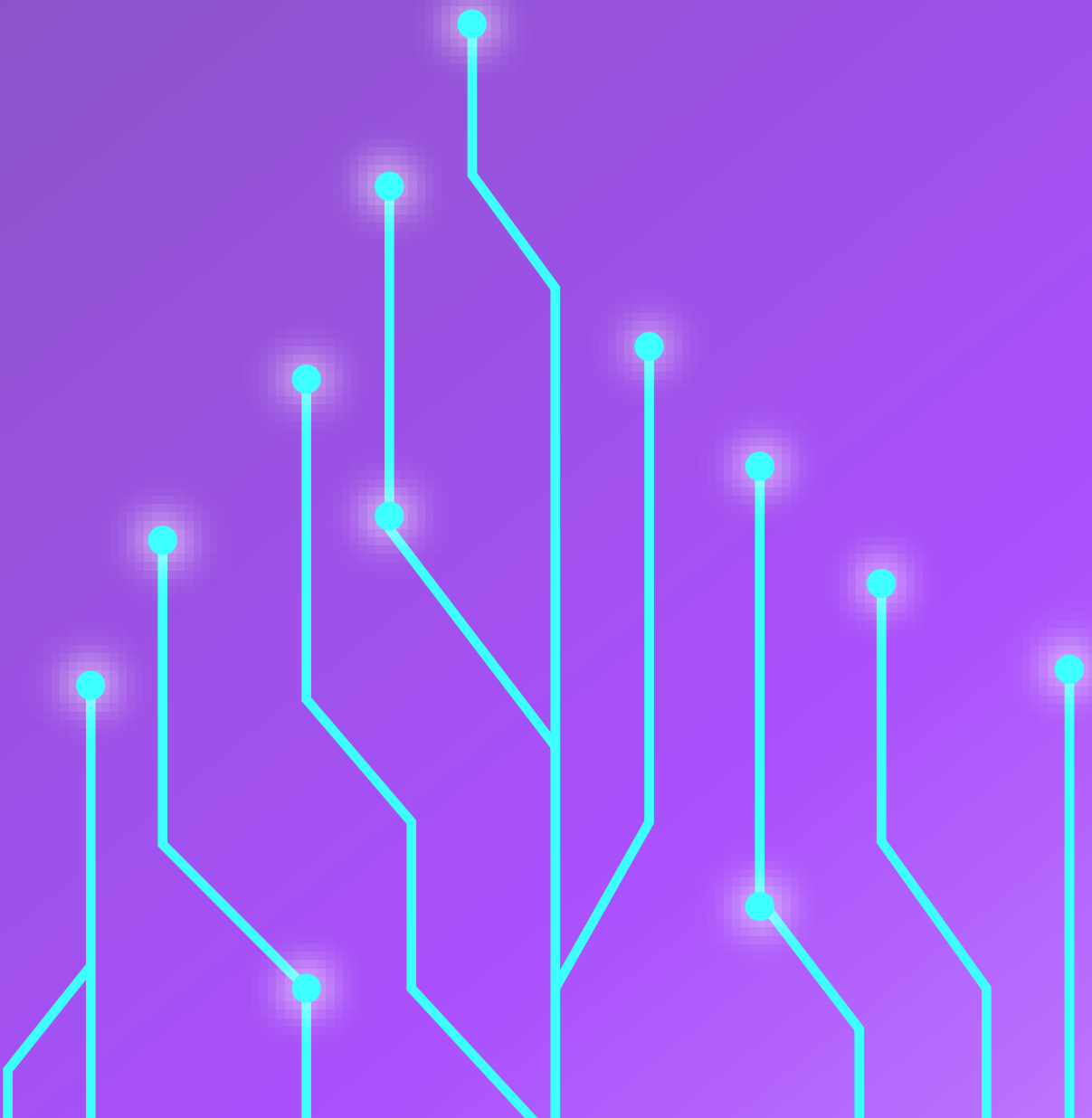
As we have already read in this booklet, algorithmic bias refers to systematic errors or unfairness in AI algorithms that result in discriminatory outcomes for certain individuals or groups. These biases can arise from various sources, including biased training data, flawed algorithmic design, or human bias encoded into the system. The consequences of algorithmic bias can be far-reaching, perpetuating inequalities, reinforcing stereotypes, and undermining trust in AI systems.

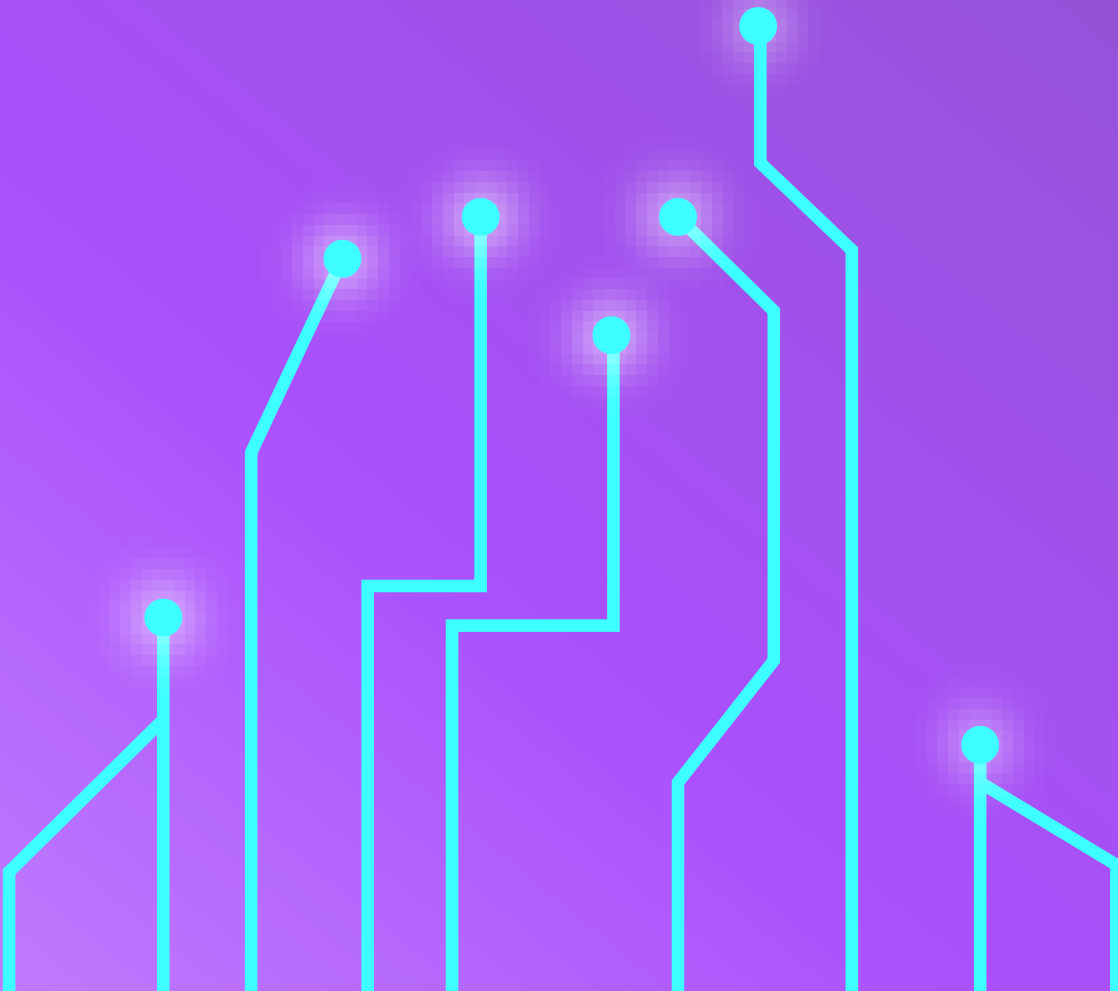




03. Relationship between Transparency and Algorithmic Bias

CU4 | Transparency





03. Relationship between Transparency and Algorithmic Bias

Opacity, or the lack of transparency, exacerbates the risks associated with algorithmic bias.

Quite often, AI algorithms are opaque, meaning that explanations of their decisions and actions are not readily available to all stakeholders. This opacity can stem from various sources, including institutional secrecy, corporate confidentiality, or technical complexity. When stakeholders lack access to information about AI systems, they are unable to assess the fairness, reliability, or ethical implications of algorithmic outcomes, leading to a lack of accountability and potential harm.

Transparency serves as a crucial antidote to opacity in AI systems, enabling stakeholders to scrutinise and challenge algorithmic decisions, thereby mitigating the risks of algorithmic bias. By increasing transparency, AI developers and practitioners can provide stakeholders with insights into how AI systems operate, why certain decisions are made, and what factors influence their outputs. Transparent AI systems empower stakeholders to identify and address biases, validate algorithmic accuracy, and hold developers accountable for the ethical and equitable use of AI technologies.



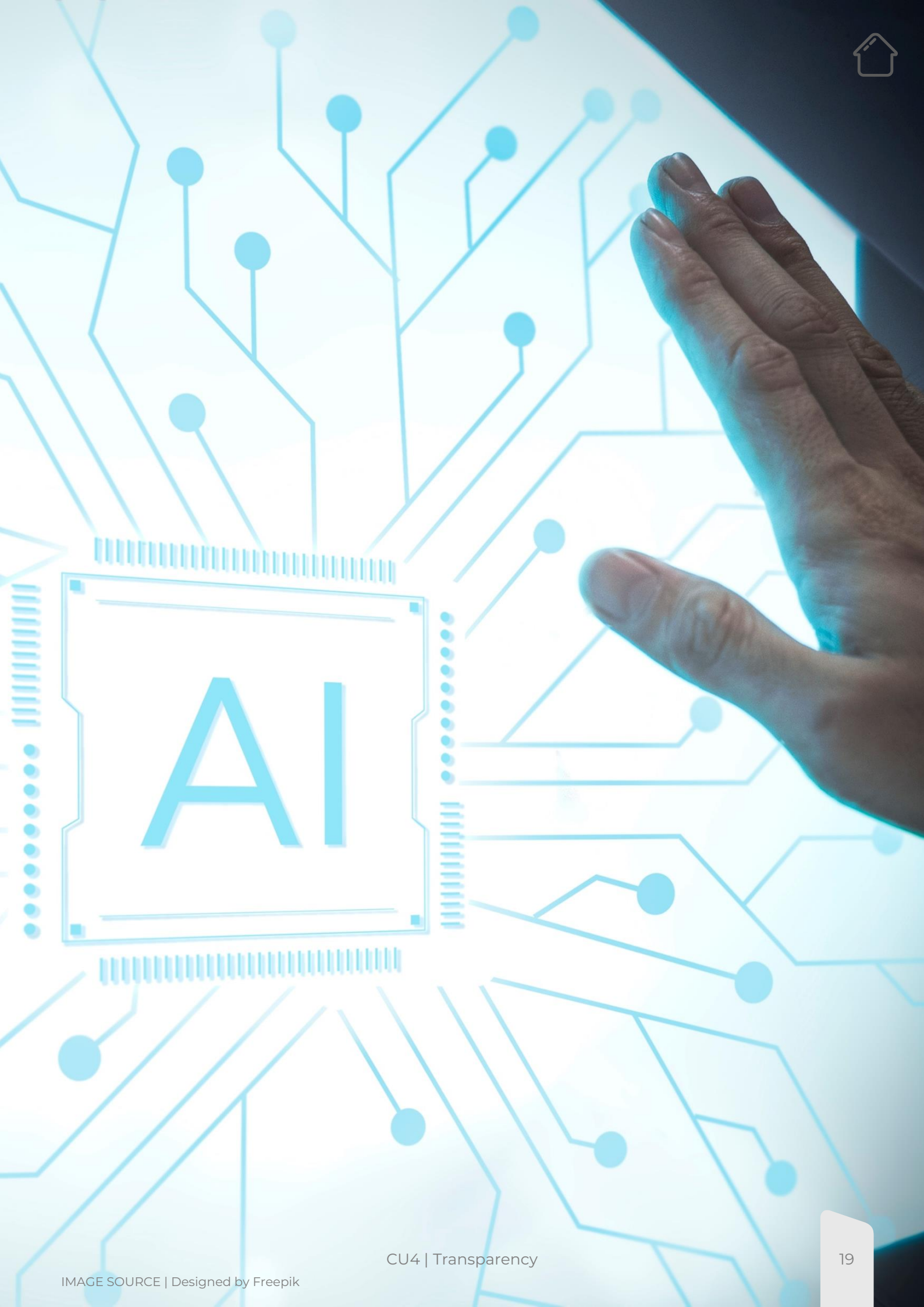
One of the key benefits of transparency in addressing algorithmic bias is the ability to detect and mitigate biased outcomes. When AI algorithms are transparent, stakeholders can examine the decision-making process and identify instances where bias may be present. For example, in the context of a hiring AI system, transparency enables stakeholders to assess whether the system unfairly discriminates against certain demographic groups in the selection process. By identifying biased outcomes, stakeholders can take corrective action to mitigate the harm caused by algorithmic bias and promote fairness and equity.

Moreover, transparency facilitates accountability and trust in AI systems. When stakeholders have access to information about AI algorithms, they can hold developers and practitioners accountable for the ethical and equitable use of AI technologies. Transparent AI systems build trust among users, regulators, and the general public, fostering confidence in the reliability and fairness of algorithmic outcomes. For instance, in the deployment of AI systems for criminal justice or healthcare, transparency enables stakeholders to understand how decisions are made and to ensure that these decisions align with ethical principles and legal standards.



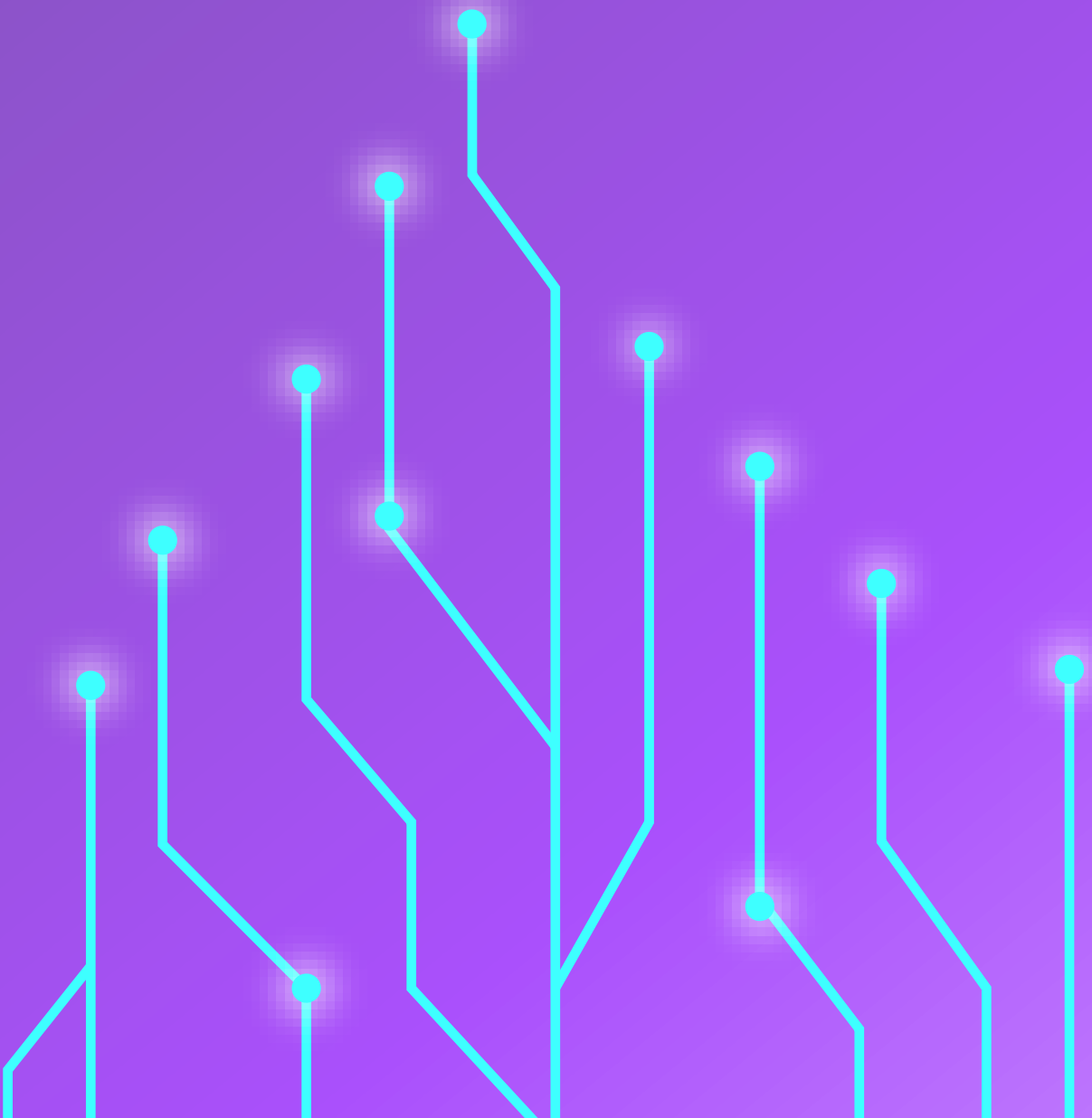
Transparency plays a pivotal role in addressing and mitigating algorithmic bias within AI systems. By increasing transparency, stakeholders can detect and mitigate biased outcomes, promote accountability, and build trust in AI technologies. As AI continues to evolve and become more integrated into various aspects of society, transparency remains essential for ensuring that AI systems are developed and deployed in a manner that upholds ethical standards and promotes fairness and equity. Through a comprehensive understanding of the relationship between transparency and algorithmic bias, learners can contribute to the responsible and ethical development of AI technologies, thereby creating a more equitable and inclusive future.

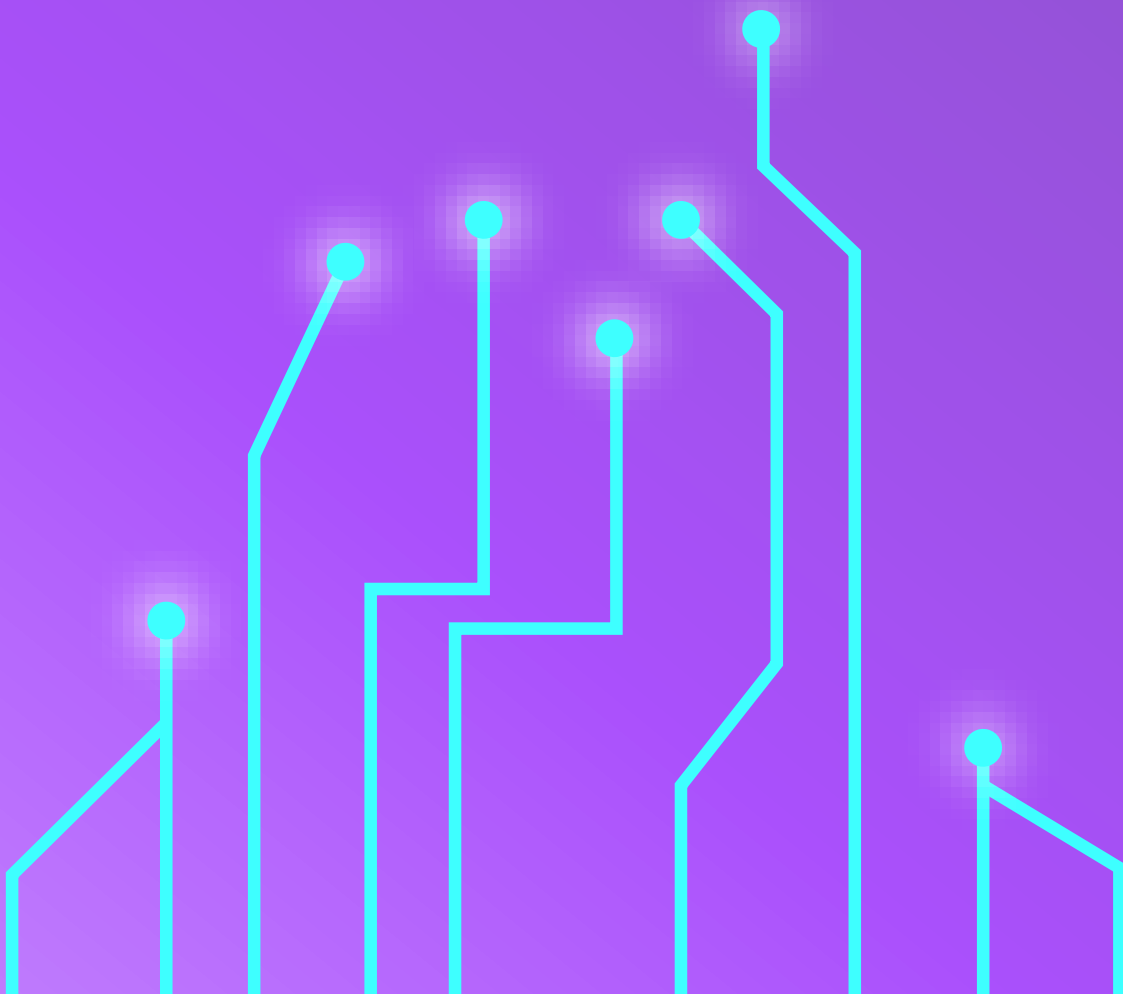




04. Strategies for promoting transparency in AI systems

CU4 | Transparency





04. Strategies for promoting transparency in AI systems

There are various strategies for promoting transparency in artificial intelligence (AI) systems such as using interpretable models, providing clear documentation, and communicating decision-making processes.

> Interpretable Models

Interpretable models represent a key strategy for promoting transparency in AI systems. They are machine learning models that produce results that are easy to understand and interpret by humans. Here are some examples:

- **Linear Regression:** Linear regression is a simple and interpretable model commonly used for predicting numerical outcomes. It works by fitting a straight line to the data points, making it easy to interpret the relationship between the input variables and the output.
- **Decision Trees:** Decision trees are hierarchical models that make decisions based on a series of if-then statements. Each node in the tree represents a decision based on a feature of the data, making it easy to follow the logic behind the model's predictions.



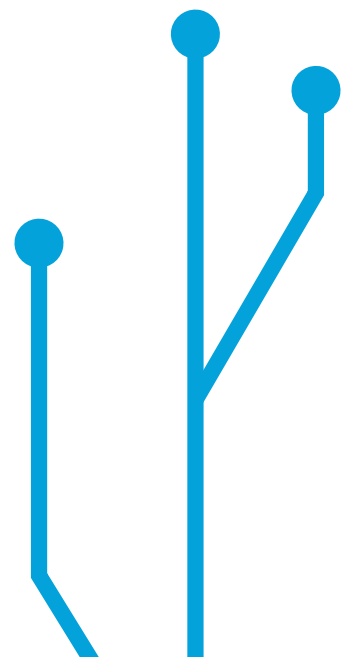
- **Logistic Regression:** Logistic regression is a statistical model used for binary classification tasks. It calculates the probability that an instance belongs to a certain class based on its input features, making it interpretable and easy to understand.
- **Rule-based Models:** Rule-based models, such as classification and regression trees (CART) or decision rules, directly translate input features into decision rules. These rules are easy to interpret and can provide insights into how the model makes predictions.
- **Generalised Additive Models (GAMs):** GAMs are flexible models that can capture complex relationships between input variables and the target variable while maintaining interpretability. They use smooth functions to represent the relationship between each input variable and the output, allowing for easy interpretation of the model's predictions.

Unlike complex black-box models, interpretable models allow stakeholders to understand how AI algorithms make decisions and what factors influence their outputs. By using interpretable models, AI developers can enhance transparency and accountability, enabling stakeholders to validate algorithmic outcomes and identify potential biases or errors. For example, in the context of a credit scoring AI system, using interpretable models allows stakeholders to understand the factors that contribute to credit decisions, such as income, credit history, and debt levels, thereby promoting transparency and fairness in lending practices.

> **Clear documentation**

Clear documentation serves as another crucial strategy for promoting transparency in AI systems. Documentation provides stakeholders with insights into the design, development, and deployment of AI algorithms, including data sources, preprocessing techniques, model architectures, and evaluation metrics.

By documenting AI systems comprehensively, developers can enhance transparency and accountability, enabling stakeholders to understand the underlying processes and assumptions of AI technologies. For instance, in the development of a predictive maintenance AI system for industrial equipment, clear documentation allows stakeholders to assess the reliability and accuracy of predictive models, understand maintenance recommendations, and verify compliance with safety standards.





> **Effective communication of decision-making**

Effective communication of decision-making processes is essential for promoting transparency in AI systems. Communication ensures that stakeholders are informed about the rationale, logic, and implications of AI algorithmic decisions.

By communicating decision-making processes clearly and transparently, AI developers can build trust and confidence among users, regulators, and the general public. For example, in the deployment of AI systems for healthcare diagnosis, effective communication ensures that healthcare professionals and patients understand how diagnostic decisions are made, enabling them to trust and verify the accuracy of AI-generated diagnoses.



➤ Engagement and empowerment of AI stakeholders

Engagement and empowerment of AI stakeholders throughout the AI lifecycle are crucial for ensuring transparency and accountability. Stakeholder engagement involves users, clients, employees, managers, regulators, and society in the design, development, deployment, and evaluation of AI systems.

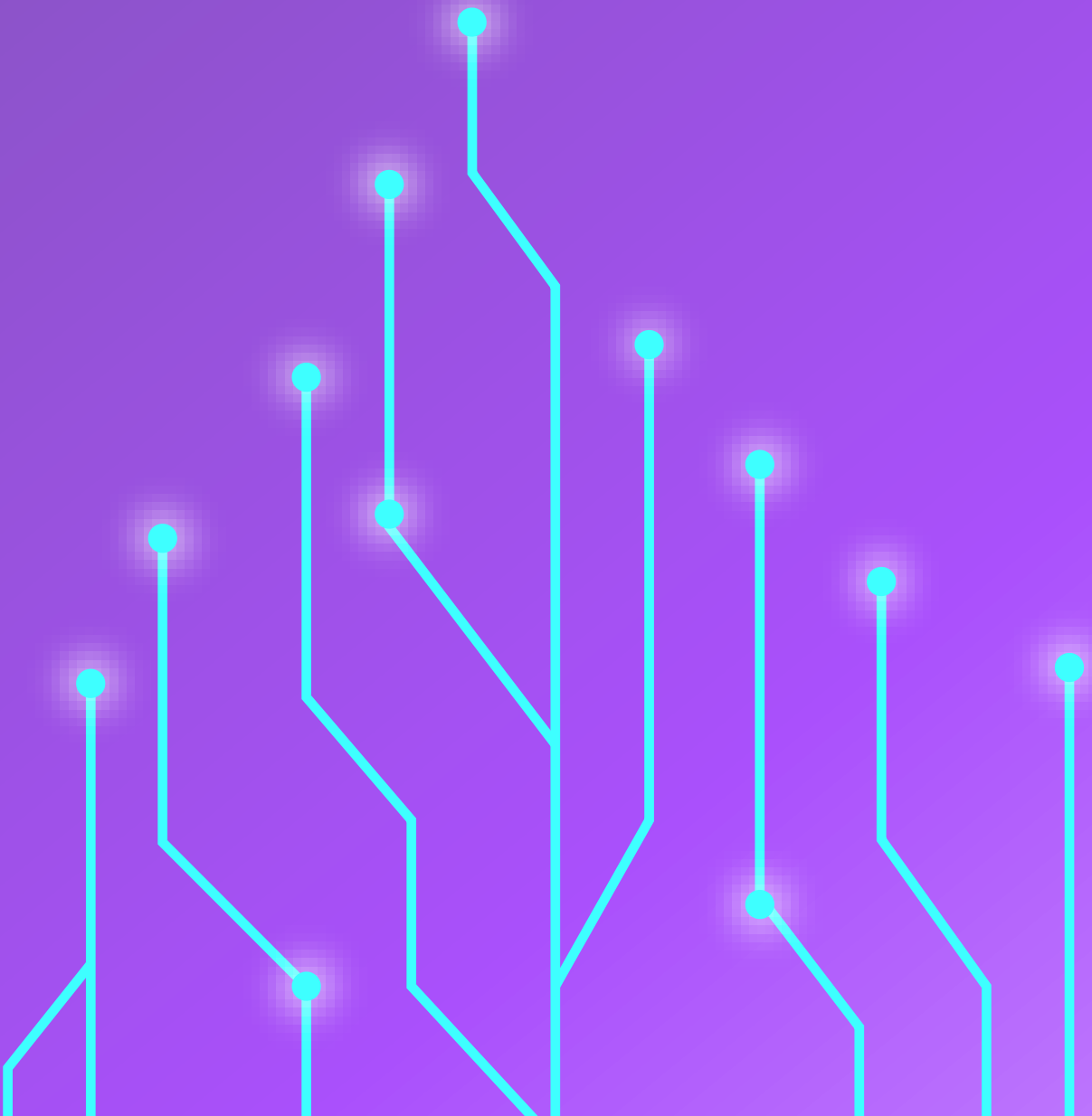
By engaging stakeholders, AI developers can gain valuable insights into their needs, preferences, and concerns, thereby promoting transparency, accountability, and ethical decision-making. For example, in the development of AI-powered autonomous vehicles, engaging with regulators and society ensures that safety, privacy, and ethical considerations are addressed, enhancing transparency and trust in the technology.

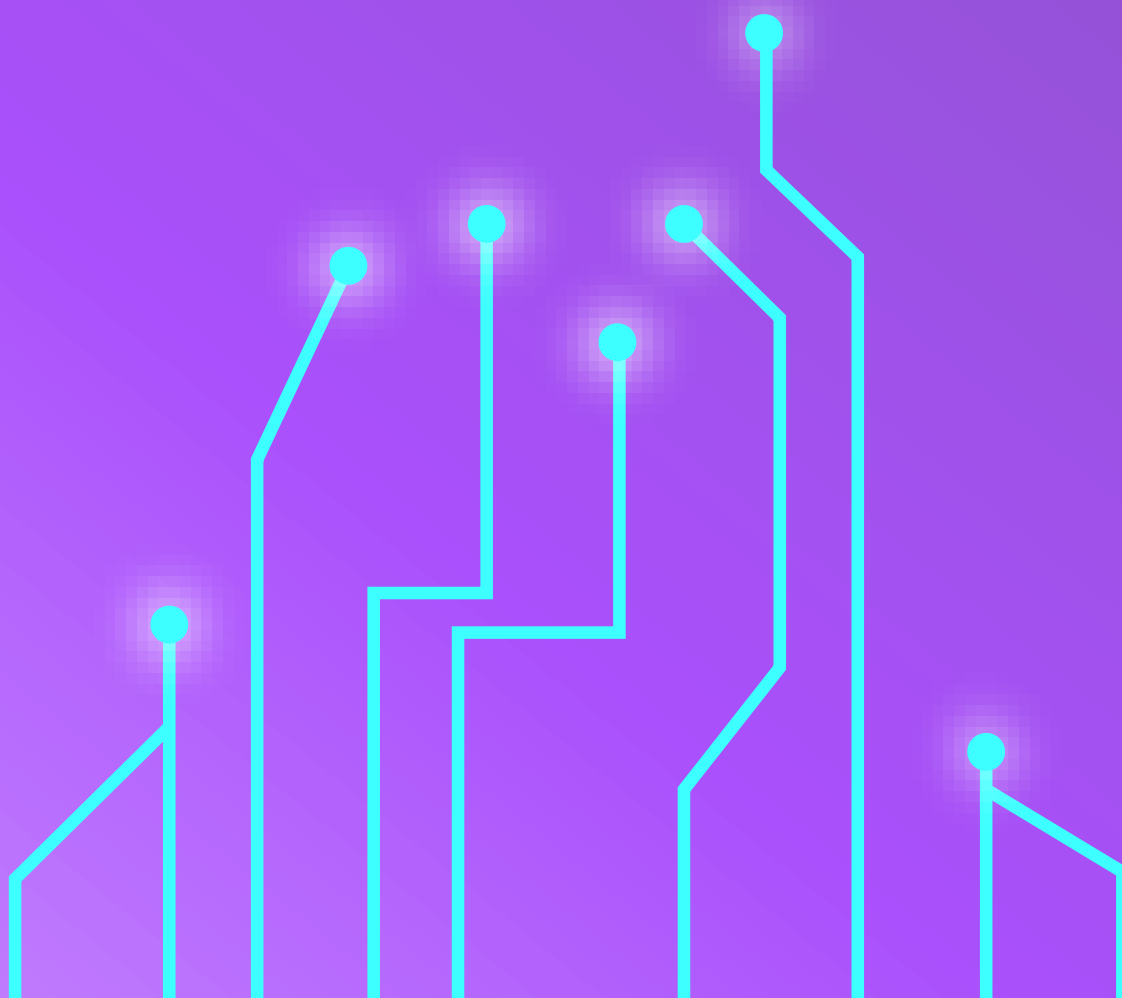




05. Conclusion

CU4 | Transparency

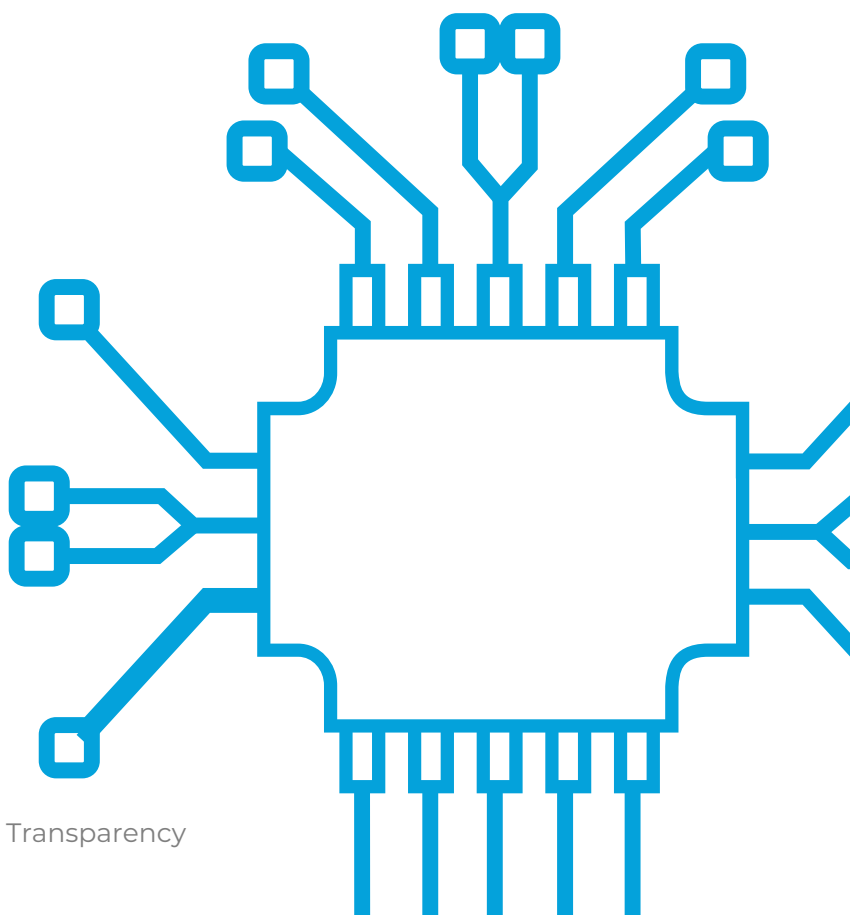


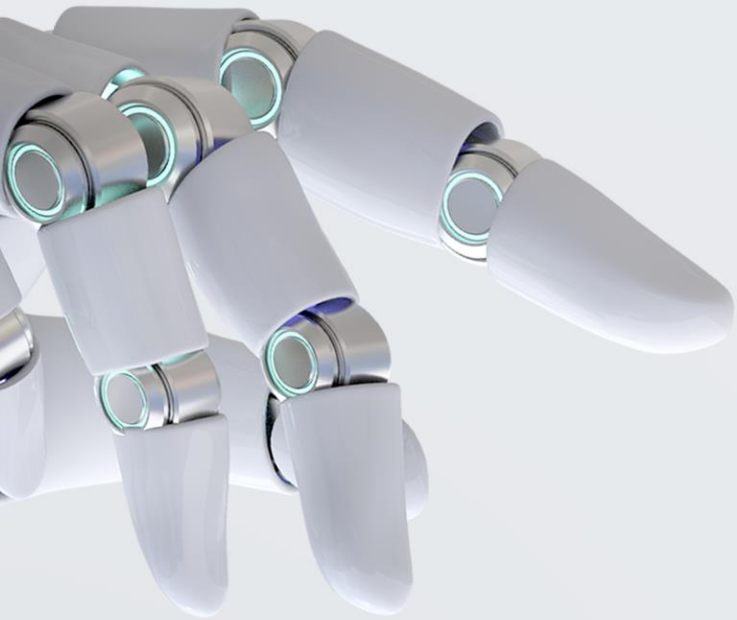


05. Conclusion

In conclusion, the importance of transparency in AI systems cannot be overstated, as it forms the foundation for building trust, accountability, and mitigating algorithmic bias. Moreover, understanding the relationship between transparency and algorithmic bias highlights the need to address opacity as a means to identify, prevent, and mitigate biased outcomes in AI systems.

Finally, the exploration of strategies for promoting transparency equips students with practical tools to enhance accountability and foster trust in AI technologies. Through a comprehensive understanding of these concepts, students are better prepared to navigate the ethical challenges of AI development and deployment, contributing to the advancement of responsible and equitable AI systems.







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



CC BY 4.0



BY