Topic 7: Governance for Responsible AI and Ethical AI Program Development

7.1 Introduction

In the first six weeks of the trimester, you have explored **ethical issues, challenges** and risks of AI development and deployment.

We're in the early stages of AI technology, with applications across industries striving to create a better world, including efforts to reduce gender bias

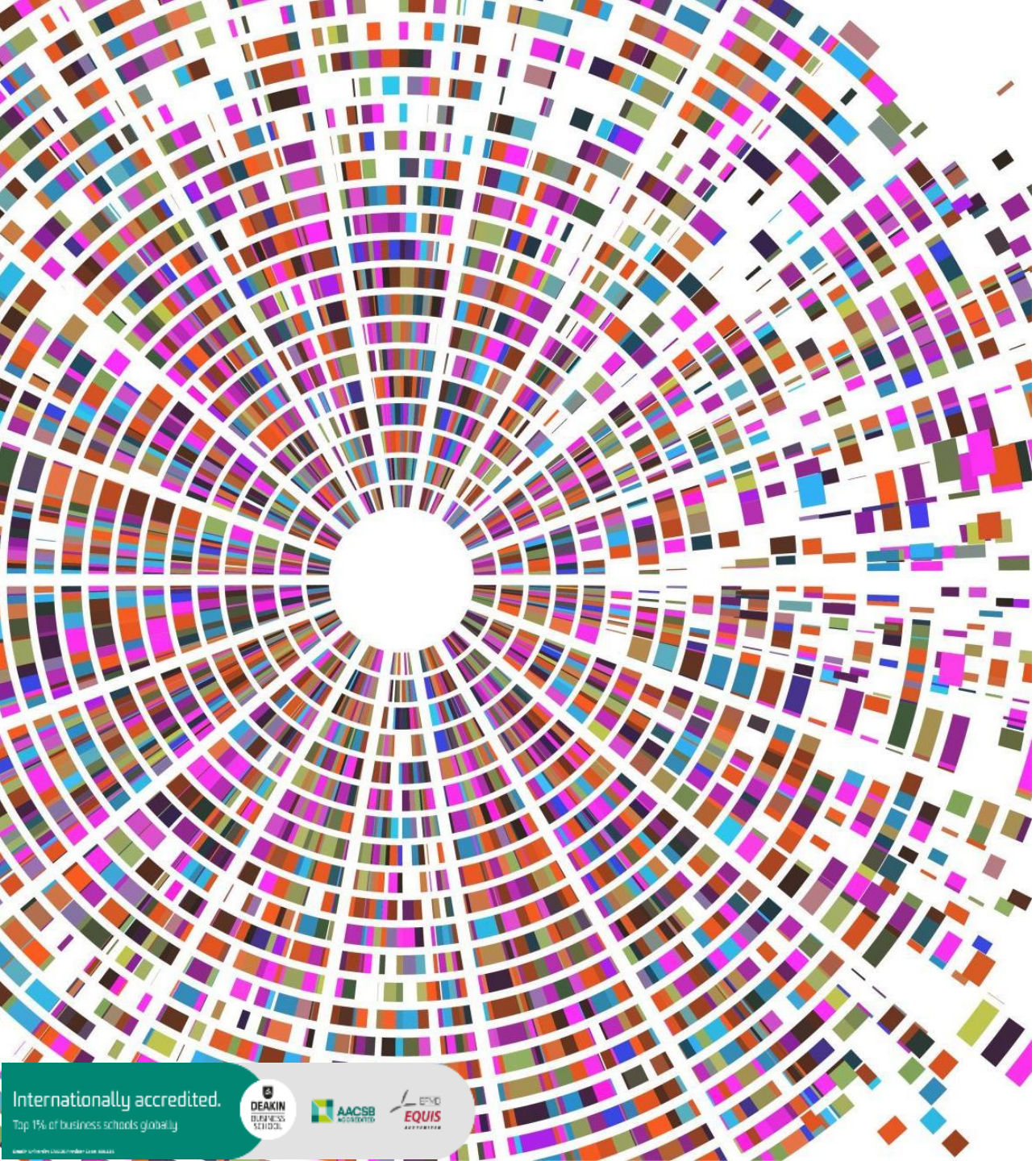
This week, you'll explore **approaches and frameworks for managing AI risks and bias** across organizational governance, implementation, and individual responsibilities

Core Learning Objectives

Recognise governance strategy guidelines and frameworks for responsible AI

Recognise the role of Corporate or Profession Codes of Conduct in AI development and deployment

Apply the governance strategy guidelines and frameworks to analyse different AI use contexts



7.2 Governance strategy to prevent and mitigate AI risks

Source: (Lucas, 2021)

AI taskforce and governance board

Establish an AI taskforce and governance board for developing ethical program/framework and overseeing AI strategy and implementation

Governance principles

Establish high-level AI ethics principles tailored to the organisation's technology and industry

Institutionalise these principles across various business levels through leadership communication, conduct guidelines, training, and rewards systems

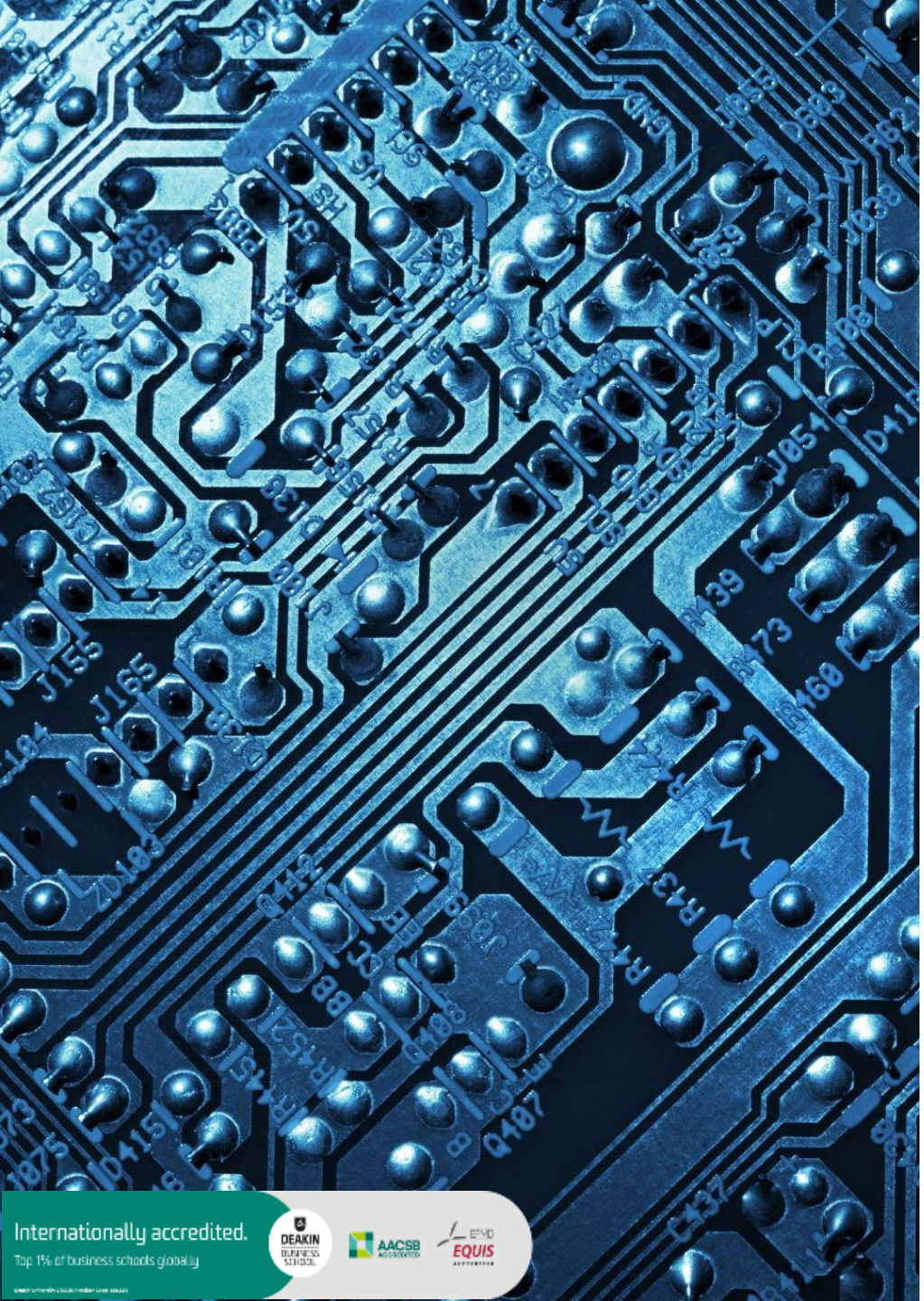
Encourage and empower staff to raise ethical concerns

Risk assessment

Conduct a comprehensive AI ethics risk assessment to identify potential issues and embed changes into organisational governance processes

Implement human oversight for AI decisions to enhance accountability

Develop an AI governance plan with short, medium, and long-term goals, regularly updated and overseen by qualified individuals



Monitor AI Impacts

Regularly monitor the impact of AI as ethical issues may arise unexpectedly

Assessing potential risks of AI before implementing the system is essential, as well as frequently monitoring the impact of the system once it is in use

Monitoring AI impacts is critical as it will enable organisations to respond quickly to identify the issues and alleviate the damage

Dataset management

Ensure that data sets used for training AI are complete, current, and representative to reduce the potential for bias and discrimination

Data protection and privacy procedures must be robust and transparent

Like-minded partners

Check with AI providers and future business partners to make sure they share similar ethical standard to your organisation

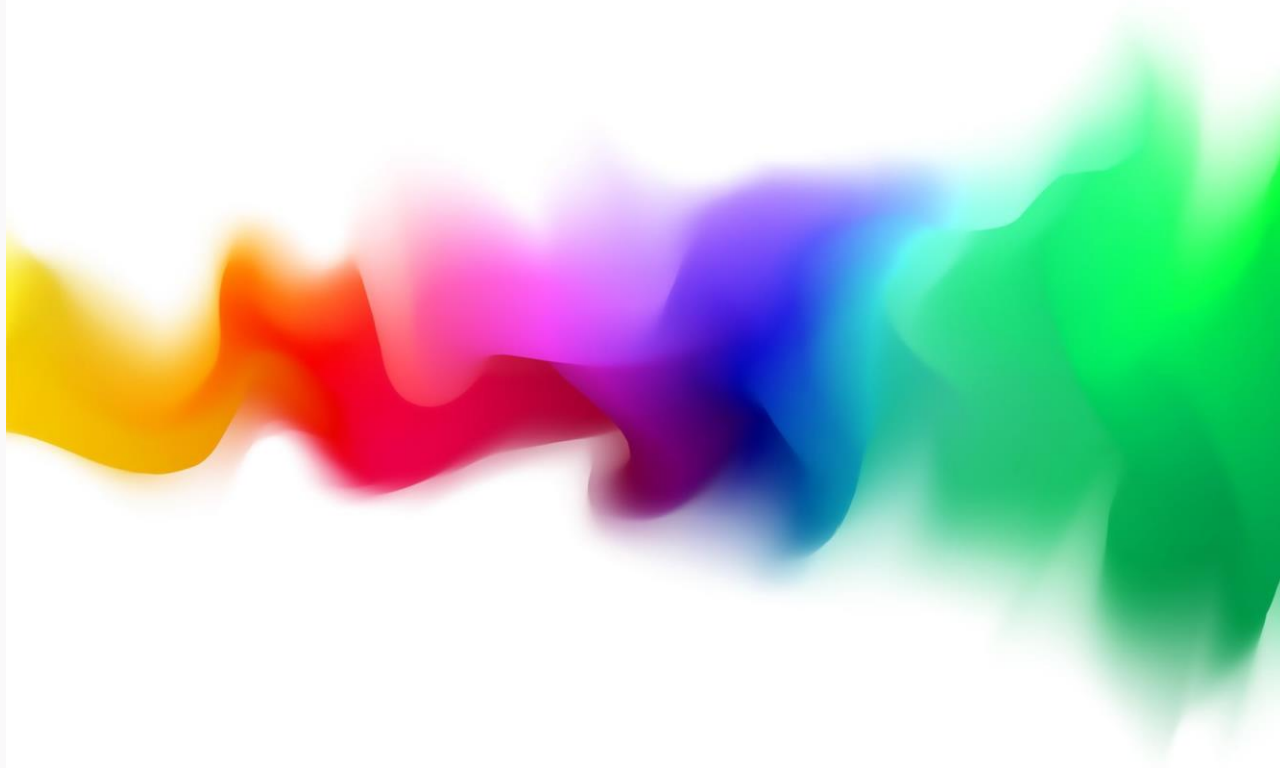
7.3. Guidelines to develop ethical AI program

Developing a data and AI ethics program is NOT a one-size-fits-all solution

Since companies across different industries have different values and requirements, it is important to customise a data and AI ethics program to address the specific business and regulatory needs of a company

Seven key steps to follow to develop a personalized, functional, expandable, and long-lasting data AI ethics program that minimises risk, according to Blackman

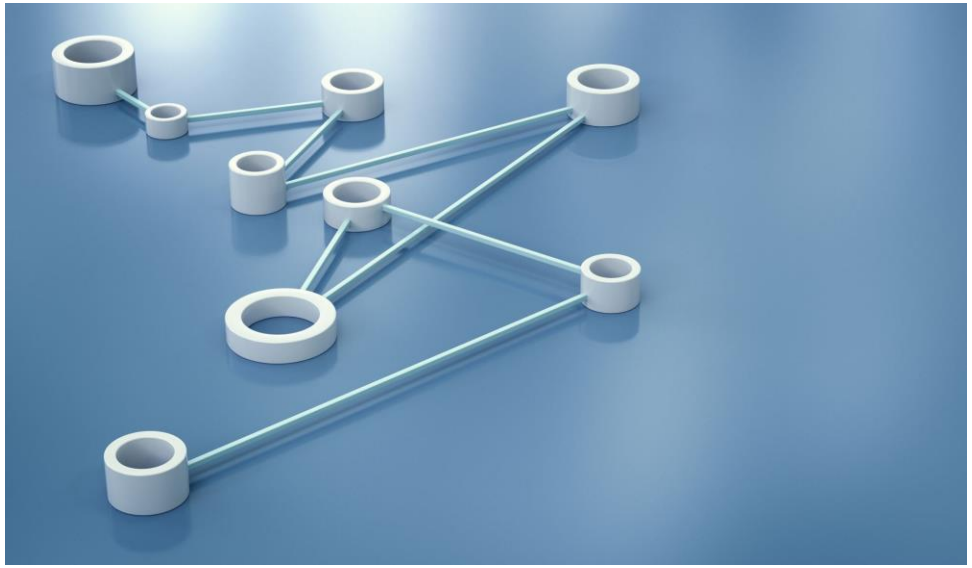
7.3. Guidelines to develop ethical AI program (Step by step)



Step 1. Organisations should leverage existing infrastructure, such as a data governance board

- The support and engagement of senior leaders within the company demonstrates their commitment to the program and their involvement also sets precedence to their employees about the seriousness of the program.

- Board members should be involved in making important decisions which will include AI strategy and will provide support and resources to minimise reputational, regulatory, and legal risk, protect the organisation's interest and ultimately make the program successful.



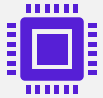
Step 2. Develop a data and AI ethical risk framework that is specific to your industry

- The framework should include clear statements of the organisation's ethical standards and potential ethical issues that could arise.

- It should also identify relevant internal and external stakeholders, recommend a governance structure, and determine how that structure will adapt to changes in personnel and other situations.

- Key performance indicators (KPI) and an assurance program need to be established. By defining and monitoring KPIs, organisations can ensure that their program is maintaining their effectiveness and achieving the intended outcome.

- Organisations need to ensure a quality assurance program to comply with relevant ethical standards, regulations and laws.



- As you saw in Topic 6, organisations need to have processes in place that mitigate biased algorithms, to ensure privacy and to be able to explain unusual outputs.

Activity

Discuss: Why is it critical for a company to tailor the guidelines to fit their context?

Step 3. Learn from industries that have had success

Executives can adopt healthcare industry approaches for risk mitigation, which have been refined over decades to systematically identify, assess, and mitigate ethical risks.

The issues of privacy, self-determination, and informed consent have been extensively examined by various experts in the medical and healthcare industry, as well as legal and regulatory professionals.

It is critical to ensuring that patients' rights and interests are protected. Patients are treated once they have given their informed consent. For instance, just as patients have the right to determine how their medical data is used, consumers should have control over how their personal data is collected, processed, and shared by organisations.

Applying similar principles used in the health care industry to the handling of personal data, the rights and interests of individuals can be protected, and the individual would have control over how their data is collected, processed, and shared.

Discussion

In healthcare before a procedure takes place full disclosure and explanation of a procedure is communicated to the patient before they give consent

Do you think these requirements could be transferred to AI products?

Could companies demonstrate this level of respect to the users of their product by providing them with full information on how their data is being collected, used and shared and would this make a difference to the user experience?

Step 4. Provide product managers with guidance and tools that are specifically tailored to their needs

While high-level guidance is helpful, it is often not enough to address the specific challenges that product managers face when creating and managing products

Product managers need to make a balanced decision considering the importance of making outputs ‘explainable’ and ‘accurate’

If, however, an organisation falls under certain regulations where the output requires an explanation, then this would be favoured over accuracy

Step 5. Develop an organizational culture that promotes awareness of data and AI ethics strategy

To successfully implement and maintain a data and AI ethics strategy, it is important to create a culture that values and prioritizes ethical decision-making

These days data and AI ethics are important in all areas of an organisation, not just restricted to certain departments

Anyone that interacts with data or AI products needs to understand the ethical considerations involved

This may mean further education and upskilling employees on the organisation's ethical framework, enabling them to ask important questions and concerns at critical junctures

Step 6. Get buy-in from employees

It is essential to encourage employees, both formally and informally, to participate in identifying ethical risks related to AI

Offering financial rewards to employees for behaving unethically can undermine the ethical standards of the organization.

At the same time, an organisation failing to financially reward ethical conduct can result in the employee feeling overlooked

Furthermore, if employees don't see that the organisation is investing in a robust data and AI ethics program, they may not see it as important and shift their focus on what will benefit their career advancement

Step 7. Engage stakeholders and monitor impacts following AI deployment

Even if all the steps in guidelines from the ethics framework were followed by employees, things can go wrong

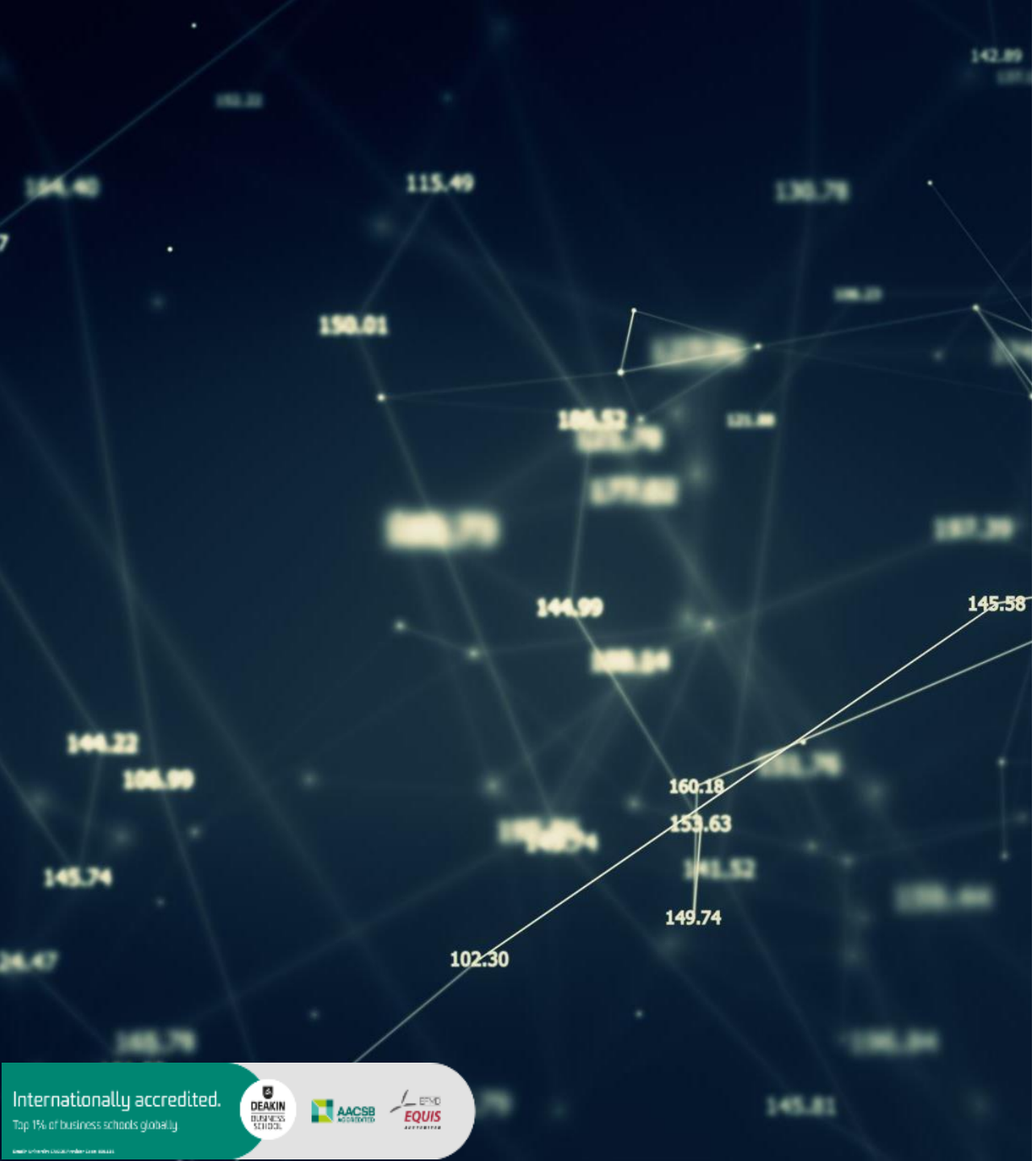
It is therefore necessary to monitor AI products and/or services once it is in the public domain to ensure that it is being used ethically

An analogy to illustrate this point, is saying that just like a car can be built with all the safety features, it is still not safe to drive at high speeds on a side street

Both quantitative and qualitative research is needed to understand the impact of data and AI products

It is essential to consult stakeholders to determine how they are affected by the 'product' and to receive their feedback on what the product should and should not do

Engaging stakeholders early in the process ensures that the product is designed and developed in alignment to their values and needs



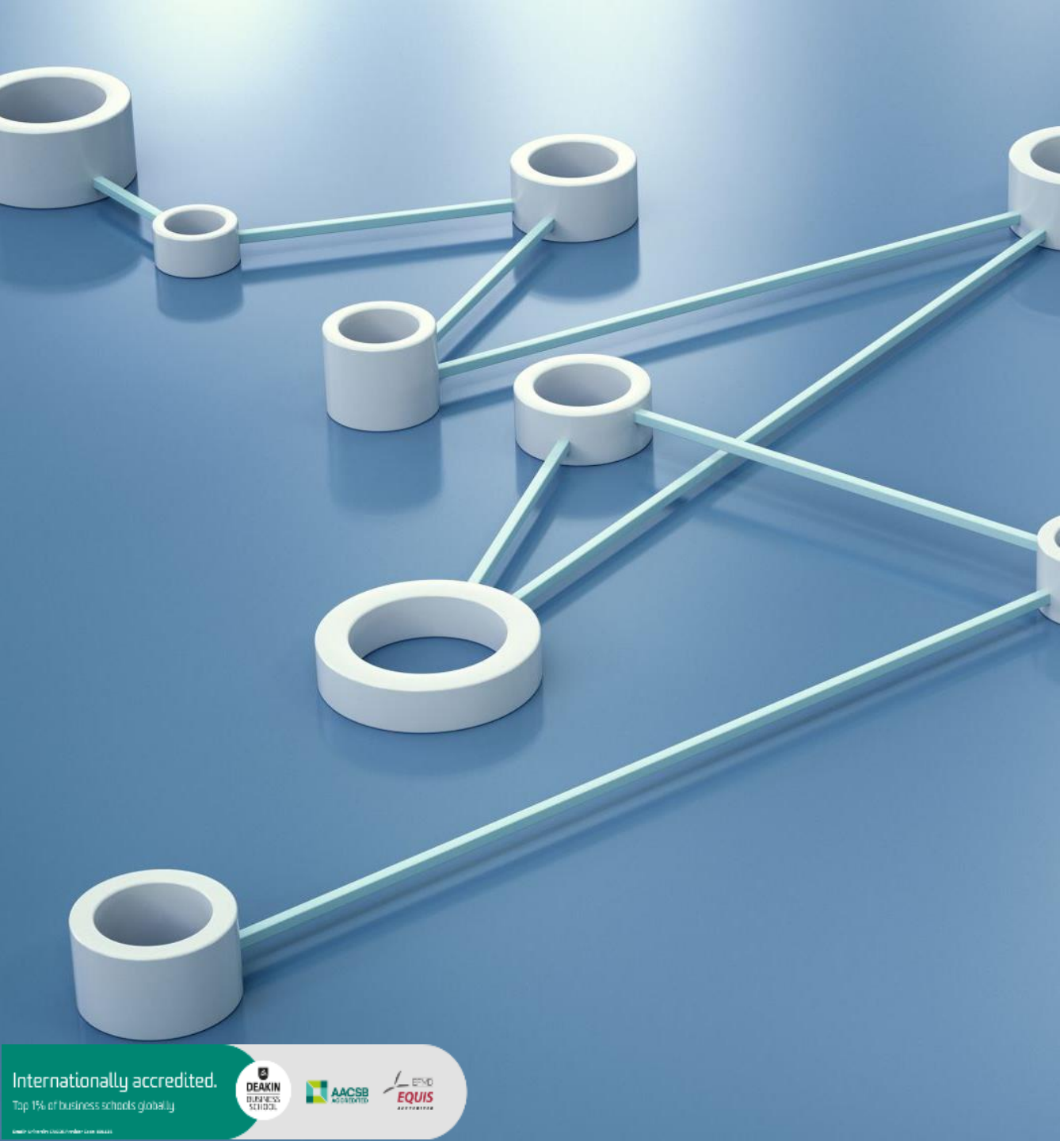
7.4 Framework to manage AI bias (Step by step)

Step 1: Decide on Data and Design

To minimise potential problems, an organisation must identify which class or group to protect for their specific application and determine the relevant issues to be mindful of

Following questions should be addressed:

- Is the issue about the difference in the sizes of groups, or different accuracy rates between the groups?
- For group sizes, does fairness demand an equal number from each group type or a proportional percentage?
- For differing accuracy rates, is the data accurately labelled, and, if so, which group needs the most predictive equity?



Step 2: Check Outputs

To prevent the possibility of the output of an AI system causing more harm than good, it is crucial to check its fairness and impact.

Algorithms tend to be literal and doesn't deal very well with intersectionality which could result in disparate impact for a group of the data

Intersectionality describes the interconnected nature of social identities and how they overlap

See the following example given by Townsen (2023)that highlights the issue

If we say a credit product needs to be equally available to men and women, disabled or not, an algorithm's solution could be to select male wheelchair users and only nondisabled women. This means that **equal numbers of men, women, disabled people, and nondisabled people are included in the data**, **but disabled women can never be selected**."

To counteract this issue Townson (2023) suggests using a 'generative adversarial networks, (GANS)' approach. **This framework uses two systems, the original model (or generator) and a second model that acts as an auditor (or discriminator).** The outputs of both models are compared to generate a fairer solution.

Step 3: Monitor for problems

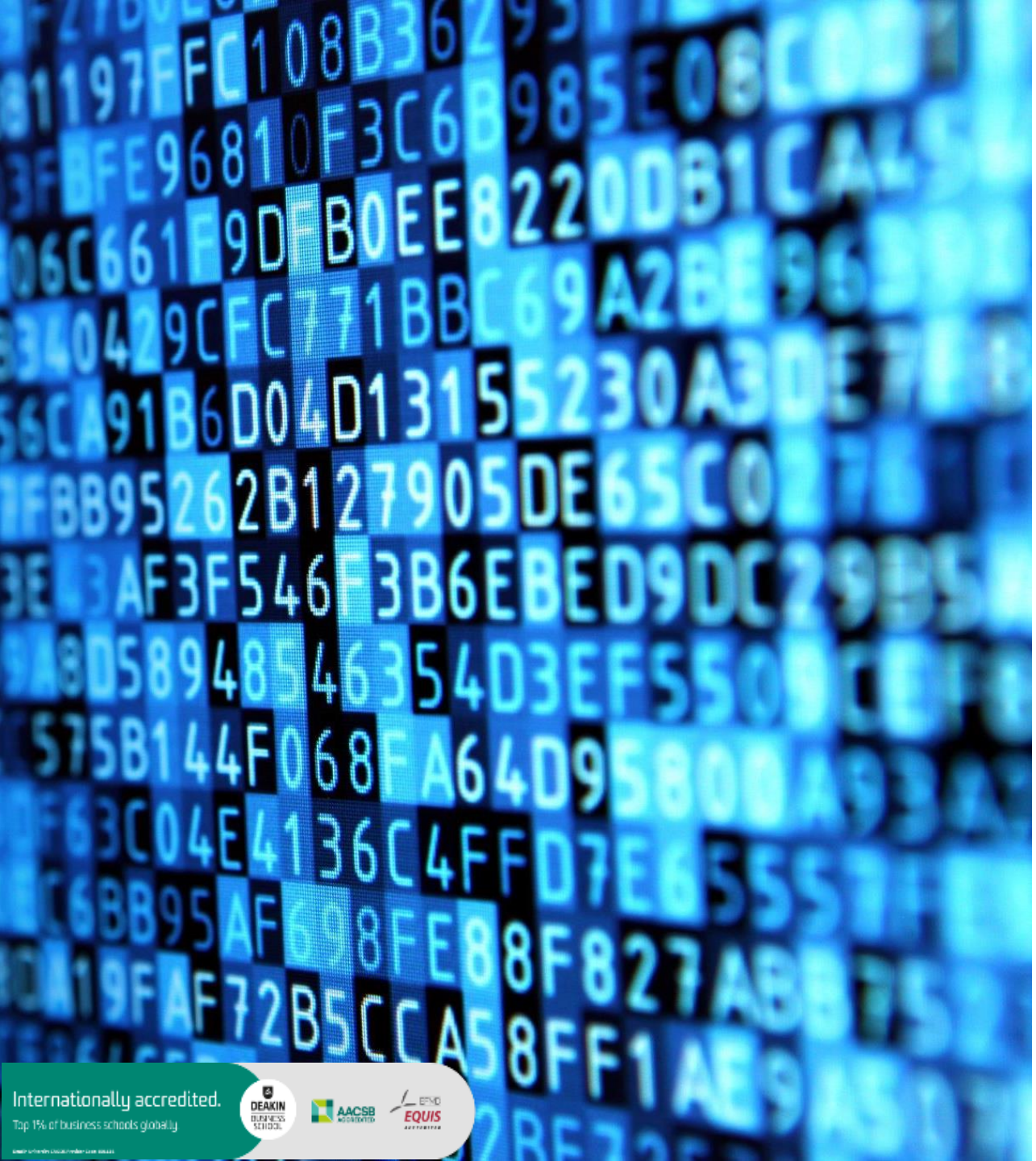
Monitoring the outputs over time is crucial and by frequently checking the AI model's output, suspicious patterns or errors could be identified.

AI systems are designed to make predictions or decisions based on input data.

Accuracy of the output data is dependent on the quality of the input data and the assumptions made in the development stage of the AI system.

Once the AI system is released into the real world it may encounter new scenarios or input data that are present in the real world.

If input data changes due to seasonal or cyclical patterns such as weather patterns, further challenges are present, and the accuracy of predictions or outcomes need to be monitored for accuracy.



7.5. Corporate or Profession Codes of Conduct

We have learnt so far...

- Different approaches for organisations to implement both AI governance strategies and frameworks to implement these.
- A responsibility on the part of the individual employee to behave ethically when using/interacting with AI depends on their professions as well as their organisation's code of conduct.
- Existing law that can be applied to protect human and our society, e.g., antidiscrimination laws, and laws against killing and maiming pedestrians (regarding driverless cars).
- Organisations need to build a comprehensive code of conduct for their employees to follow that might help them to prevent and manage AI risks and users' unethical behaviours.

Organisation’s code of conduct

Refers to the human resource policy that outlines the standards of behaviour expected by your organisation from every employee including *ethical principles, values, accountability, standard of conduct, standard of practice, disciplinary actions*

The policy is particularly designed and customised for each specific organisation

Professional code of conduct

The set of principles which describe the professional conduct, personal conduct and professional competence expected of each profession

Ethics is at the core of any professional, and all professional bodies put so much emphasis on the need for members to abide by ethical codes of conduct

Australian Council of Professions states

"A profession is a disciplined group of individuals who adhere to ethical standards and who hold themselves out as and are accepted by the public as possessing special knowledge and skills in a widely recognised body of learning derived from research, education and training at a high level, and who are prepared to apply this knowledge and exercise these skills in the interest of others…"

The Code of Professional Conduct of the Australian Computer Society (ACS)

An essential characteristic of a profession is the need for its members to abide by a code of ethics. For the ACS, this code is established as the Code of Professional Conduct and is part of the Society's Regulations. These Regulations apply to any ACS member who works in the field of information and communications technology (ICT)" (ACS, 2014).

See section 7.5 on the unit site for the details.

Corporate or Profession Codes of Conduct: a summary

Corporate codes of conduct and Professional codes of conducts are needed to urge good judgement, integrity, and high ethical standards

Corporate codes of conduct govern people's behaviours so that they can be aware of what behaviours are acceptable and unacceptable in a professional setting

People know for most part, how not to do bad things

If they don't, training will be offered