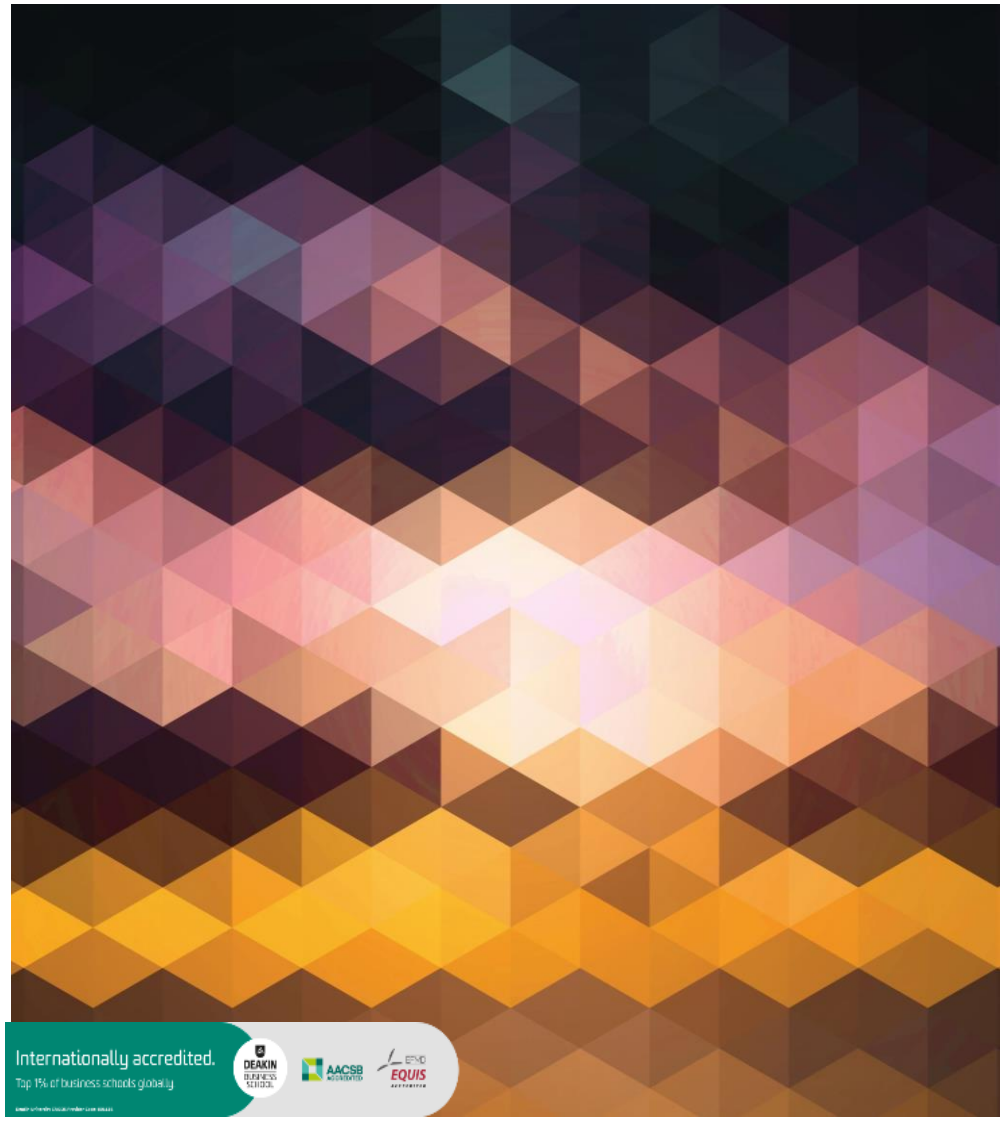
Topic 4: Taking Responsibility

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EQUIS



Topic 4. Taking responsibility: Accountability, Responsibility, and Transparency of AI

This week's focus is on the following core learning objectives

- Understanding the motivation of Accountability, Responsibility, & Transparency of AI
- Responsible Research and Innovation in the Development of AI Systems
- RRI in AI Use and Management; and the principles, definitions and importance of ART principles
- This week's topic is closely relevant to Assignment 2

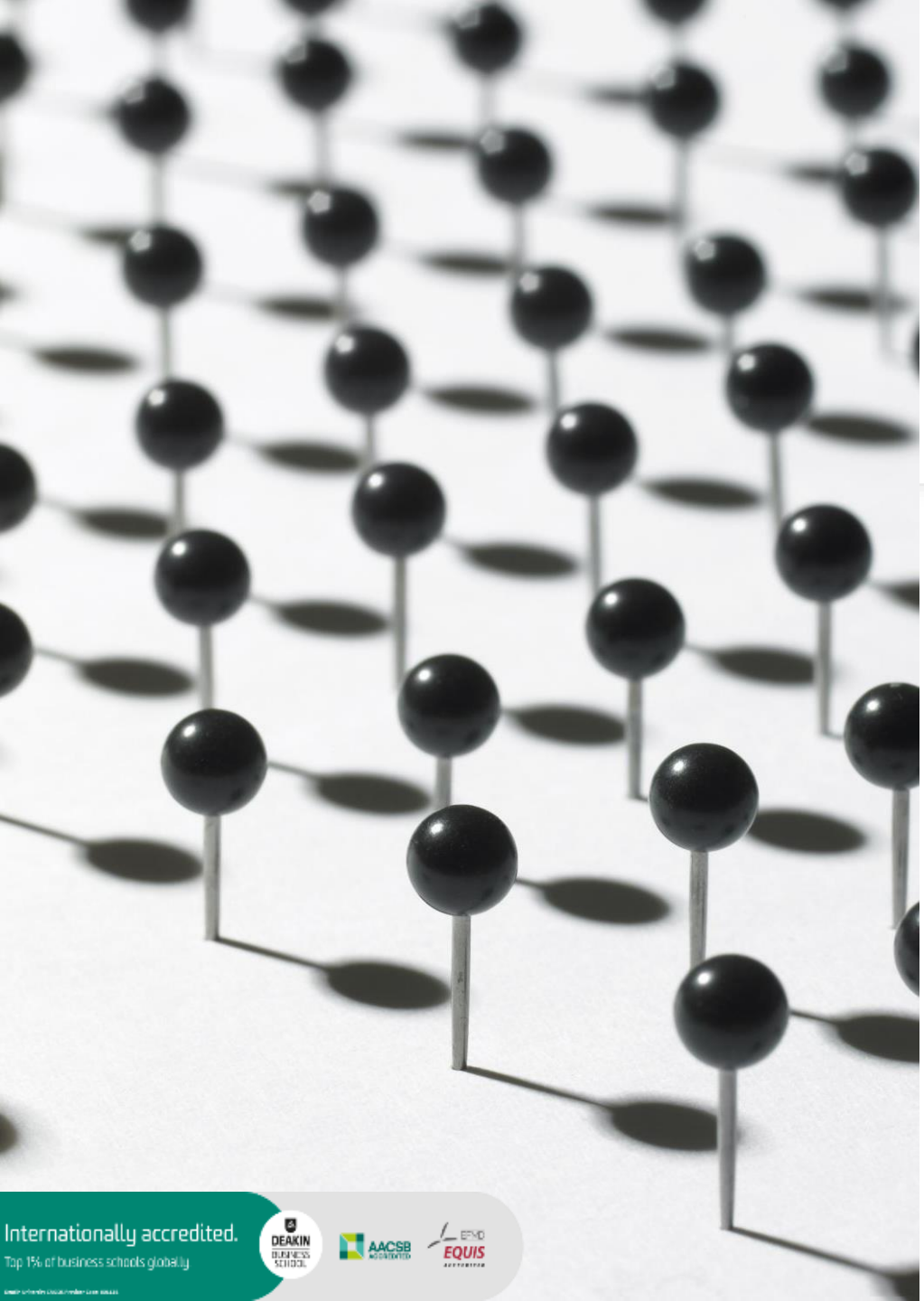
Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

4.1. Introduction

Let’s watch the trailer for the Netflix documentary "The Great Hack"

Discuss Cambridge Analytica's improper data acquisition from Facebook to influence US elections

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED



Reflection Points

Consider algorithm misuse in shaping public opinion

Discuss ethical concerns regarding autonomy, fairness, and justice

4.2 Underlying General Principles for Algorithmic Transparency and Accountability

Awareness: awareness means educating all involved parties about potential biases throughout the development stage and their potential harm to individuals and society

Access and Redress: Access and Redress means the existence of a process for looking into and revising incorrect judgments or decisions

Accountability: Accountability means institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results

Explanation: Explanation means the logic of the algorithm must be communicable in human terms no matter how complex it is

Data Provenance: ensures data reliability and trustworthiness

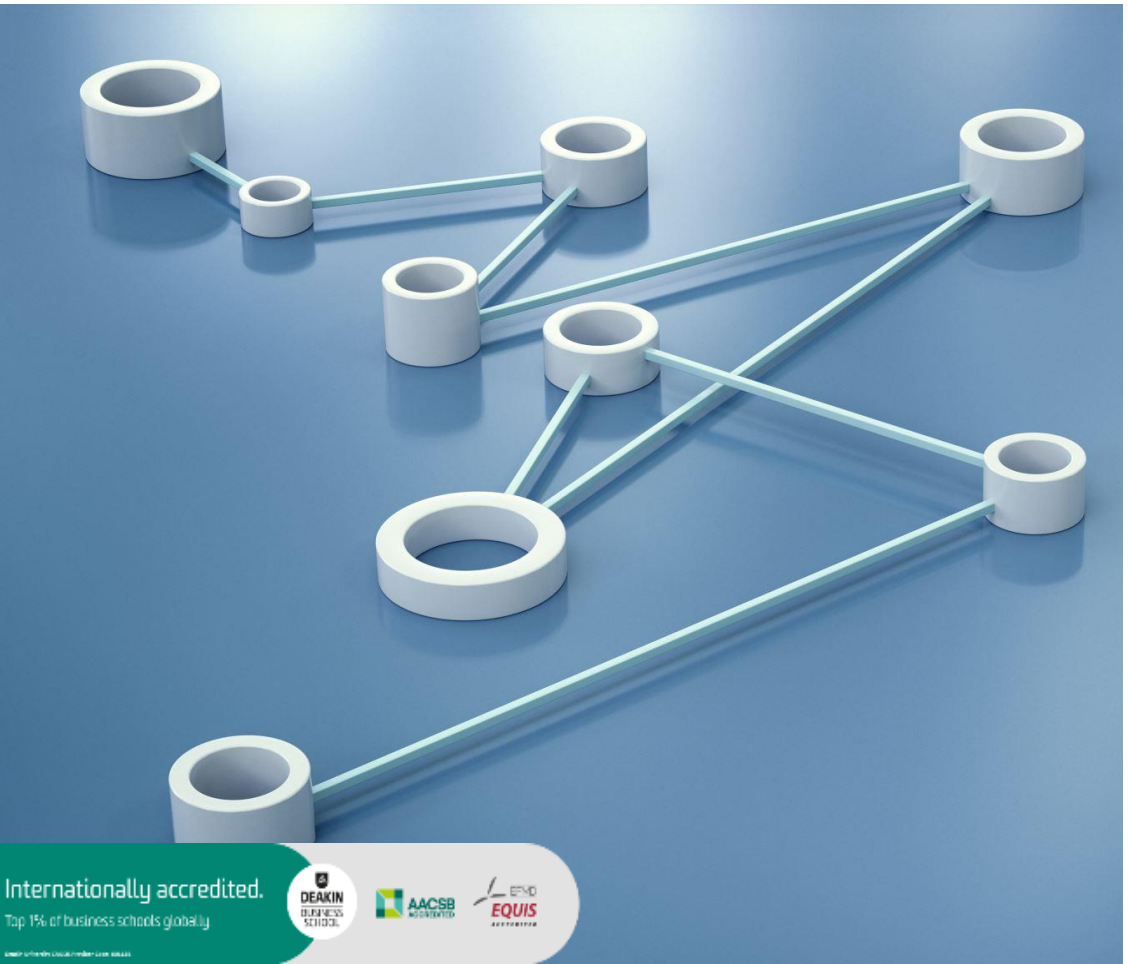
Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

4.2 Underlying General Principles for Algorithmic Transparency and Accountability

Auditability: Auditability involves auditing models, algorithms, data, and decisions when harm is suspected

Validation and Testing: Validation and testing of automated systems should be ongoing, utilizing techniques like regression tests, corner case analysis, and red-teaming for computer security to boost confidence in the systems

4.3 AI Accountability



- Organizations and individuals involved in AI systems must collect relevant data, ensure proper functioning throughout the system's lifecycle, and adhere to roles and regulatory frameworks, demonstrating compliance through actions and decision-making
- Ability of a system to explain and justify its actions to its users and other stakeholders
- Accountability should extend beyond technical artifacts to include their relationship with broader decision-making processes, such as development and implementation
- Accountability encompasses ethical, moral, or other expectations guiding individuals or organizations in designing and using AI systems

Watch

We need AI to be held accountable

Whose responsibility is it do you think to advocate for accountability?

Discuss

Do you think this lack of accountability in computerised systems has come about suddenly?

Or is it something that has slowly become the norm?

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

Accountability Erosion

Dissemination of Responsibility: Computerized systems involve multiple parties, complicating accountability

Bugs in Software: Software bugs are often used as an excuse, hindering prevention efforts

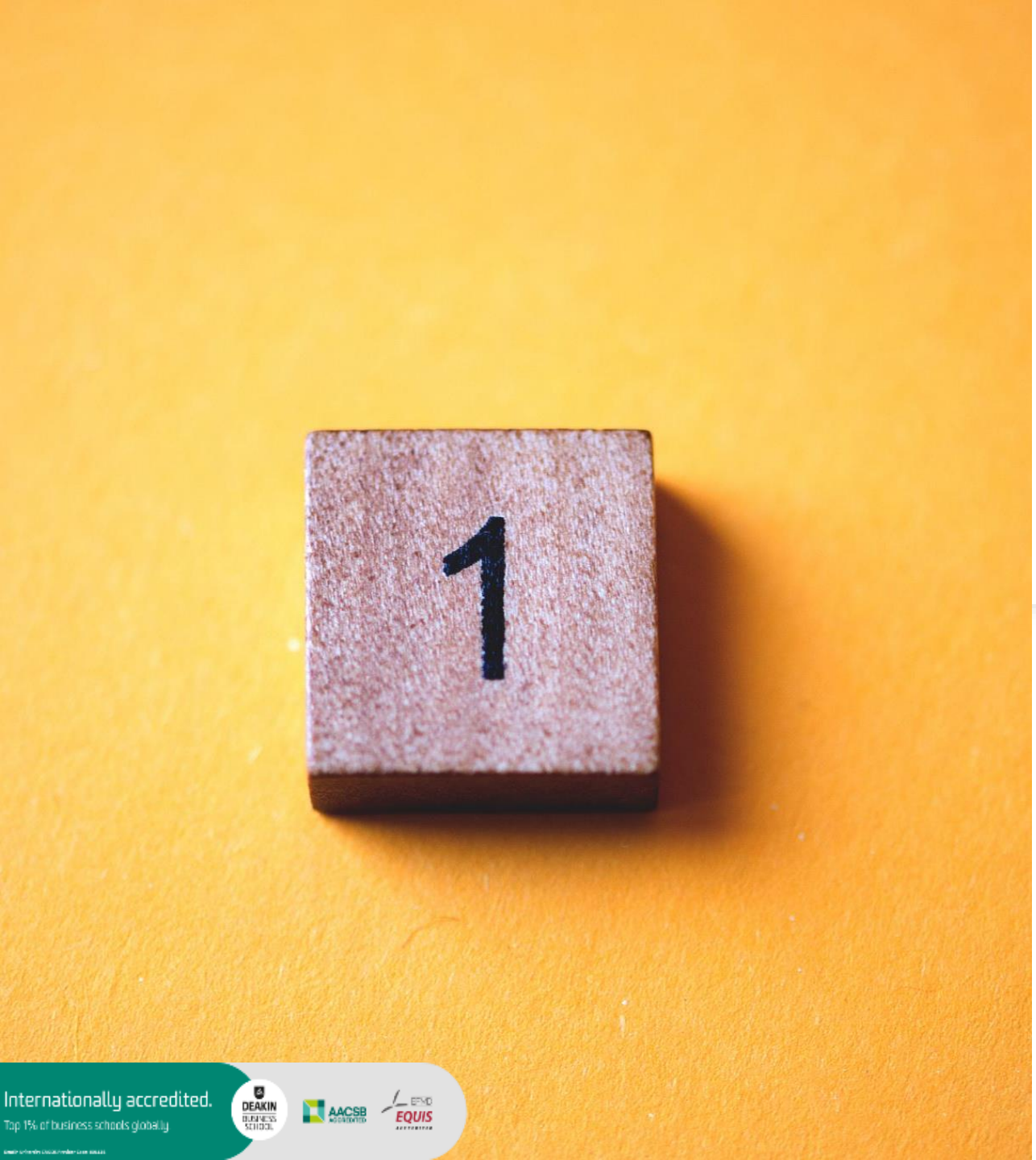
Blaming Computers over Humans: Tendency to blame computers rather than human involvement in errors

Lack of Developer Accountability: Absence of accountability among developers adds to challenges

What should we do with accountability?

Involvement of All Participants: All stakeholders should define moral values **guiding system operations**

Accountability Through Governance: Organizations can be held accountable through governance, despite algorithms' lack of moral or legal responsibility



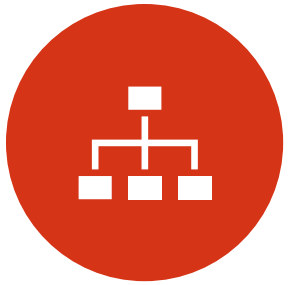
Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

4.4 AI Responsibility

Responsibility in AI

- Refers to humans involved in development, manufacture, selling, and use of AI systems
- AI systems as tools
- Created by humans with specific goals
- Not inherently responsible actors

Human responsibility



Irreplaceable, even with system accountability and openness

Responsible for interactions with AI systems

Incorporating ethical principles

Necessary at all stages of AI development lifecycle

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

Two possible outcomes when it comes to the responsibility of AI systems

Consider the following example from Dignum (2019, P.58):
"…**who will be liable if a medicine pump modifies the amount of medicine being administered? Or when a predictive policing system wrongly identifies a crime perpetrator?"**

- The builder of the software?
- The ones that have trained the system to its current context of use?
- The authorities that authorised the use of the system?
- The user that personalised the system's decision-making settings to meet her preferences?

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

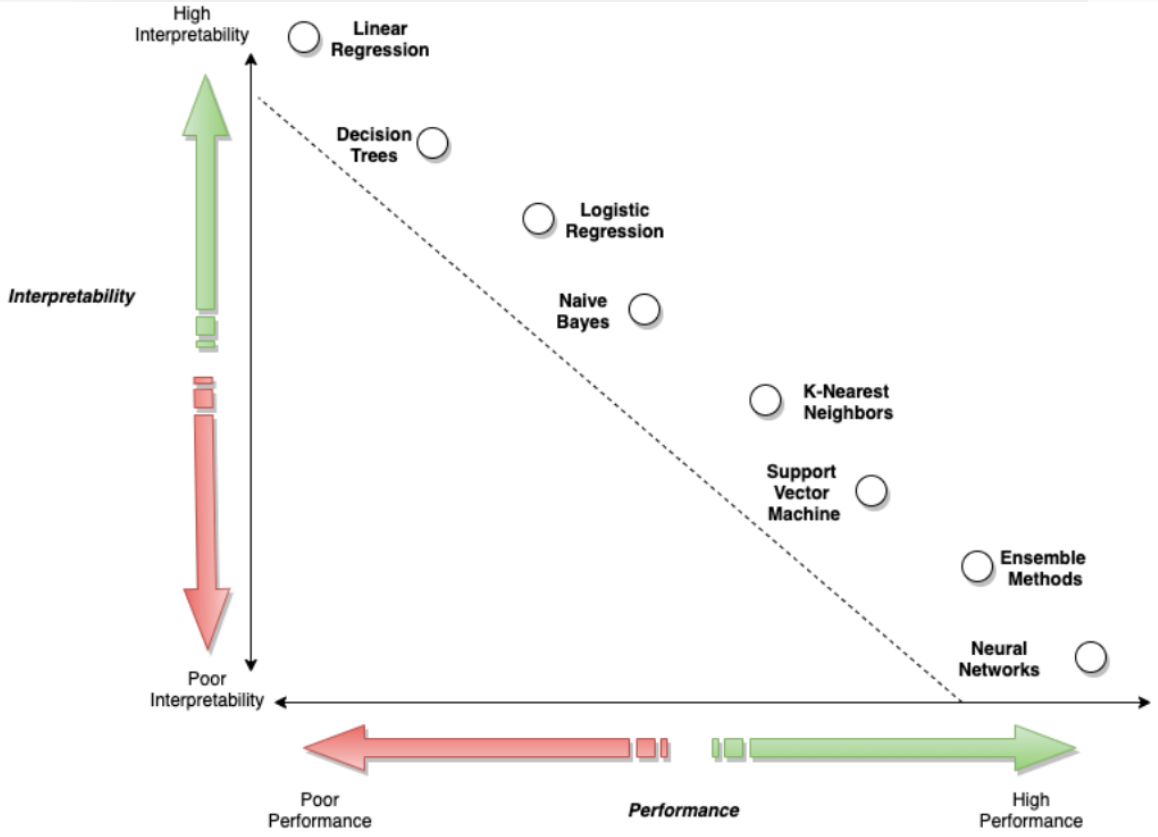
Watch

Background: Sophia the social humanoid robot developed by Hanson Robotics was activated on February 14th February

Watch the video from that event where Sophia has a short interview

As you watch the video consider what you would feel like living in a world where robots and humans are "equal"?

4.5 AI Transparency-Explainability-Interpretability



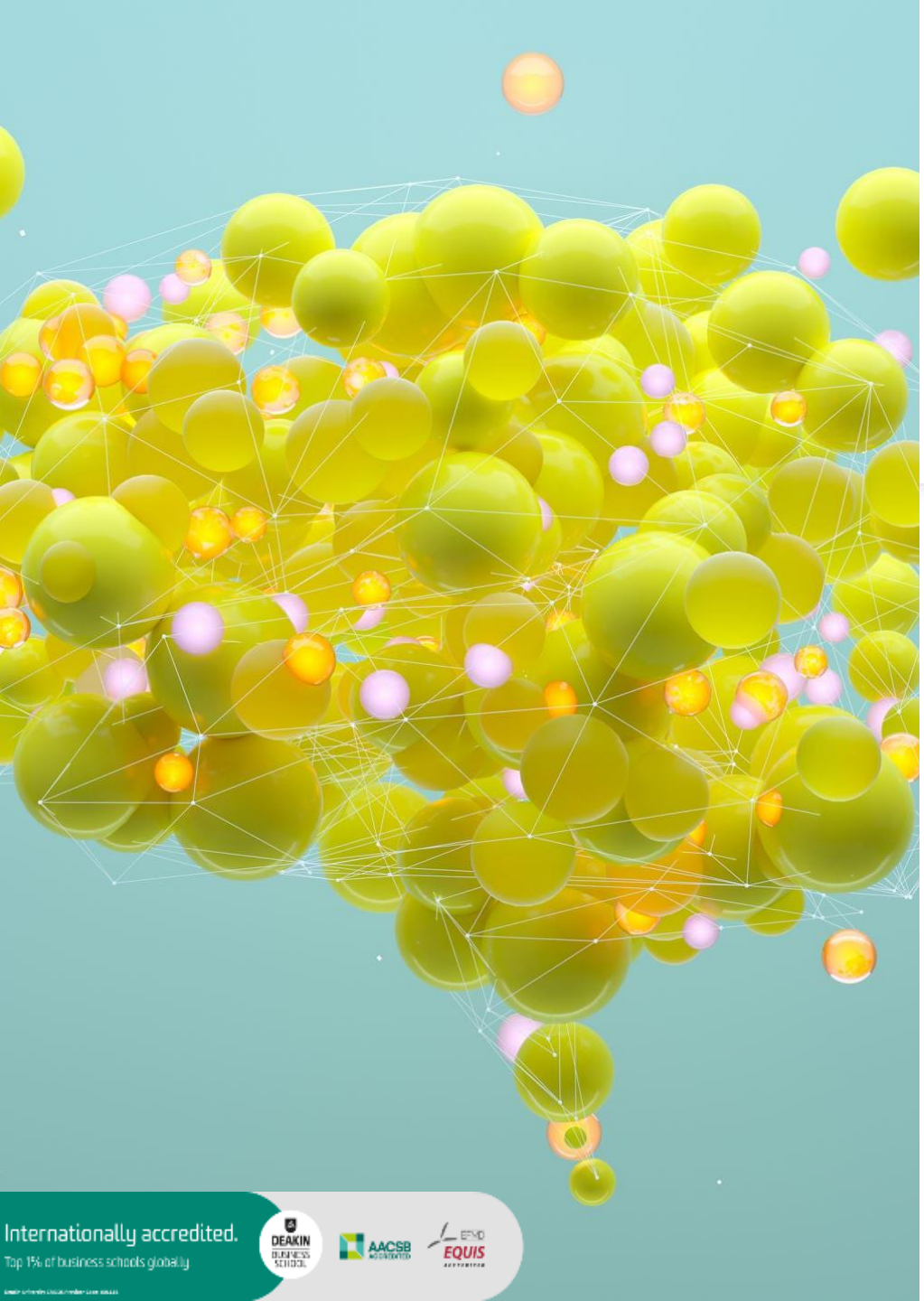
One of the reasons why people might be afraid of AI, is that AI technologies and algorithmic decision-making can be hard to explain.

When we can explain, justify, and interpret AI decision-making models, perhaps our fear of AI systems might reduce.

While some AI technologies are straightforward to explain, for instance semantic reasoning, planning algorithms and some optimisation methods, some other AI technologies especially data-driven models like Machine Learning, the relation between input and output of the models is much harder to explain

Source: AWS White paper

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EQUIS ACCREDITED



Transparent AI

Ensures outcomes of AI models can be explained, interpreted, justified, and communicated effectively

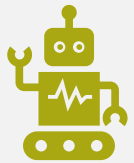
Also known as explainable, justifiable, and interpretable AI

Emphasizes explicitness and openness regarding data sources, development processes, and stakeholders

Facets of AI Transparency

Encompasses various dimensions

Crucial for institutional actors



Requires clarity in selection, implementation, and technical aspects of automated decision-making systems

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

Goal of AI Transparency

Essential for accountability and trust-building

The concept of AI transparency refers to the capability of people who use, regulate, and are impacted by AI systems to understand how AI reaches decisions

Transparency and explainability in Australia's Artificial Intelligence Ethics Framework

- For users, what the system is doing and why
- For creators, including those undertaking the validation and certification of AI, the systems' processes and input data
- For those deploying and operating the system, to understand processes and input data
- For an accident investigator, if accidents occur
- For regulators in the context of investigations
- For those in the legal process, to inform evidence and decision-making
- For the public, to build confidence in the technology

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

4.6 RRI in the Development, Use and Management of AI Systems

Responsible Research and Innovation is defined as: Transparent, interactive process by which **societal actors** and **innovators** become mutually responsive to each other with a view to the acceptability, sustainability and societal desirability of the innovation process and its marketable product

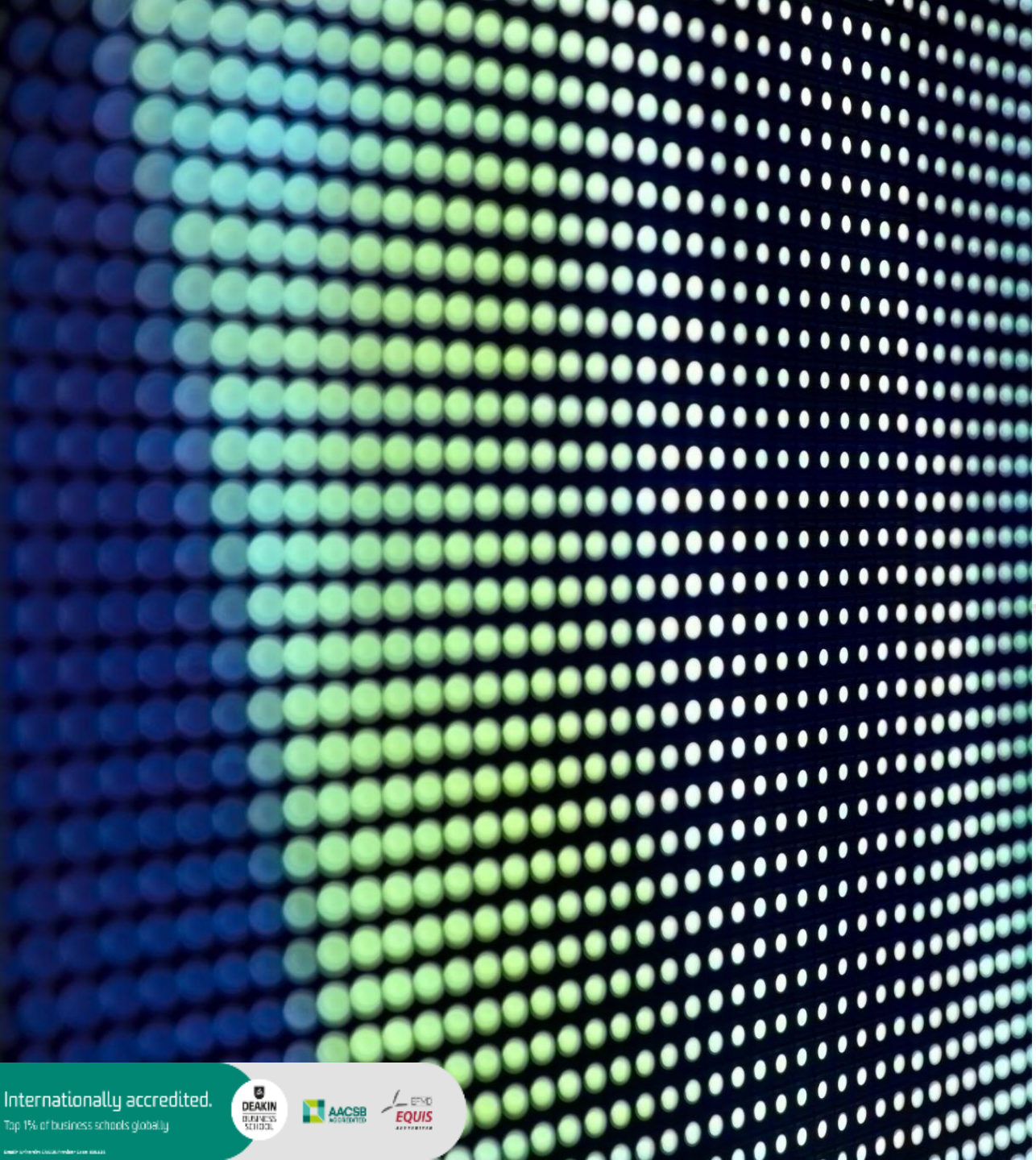
(Dignum, 2019, p. 50)

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

What are the Key Elements of RRI?

- Doing the Right Thing
- Good and Reflexive Governance
- Creative Learning
- Choosing Together
- Unlocking the Full Potential of Technology
- Sharing Results
- Taking Care of our Planet

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED



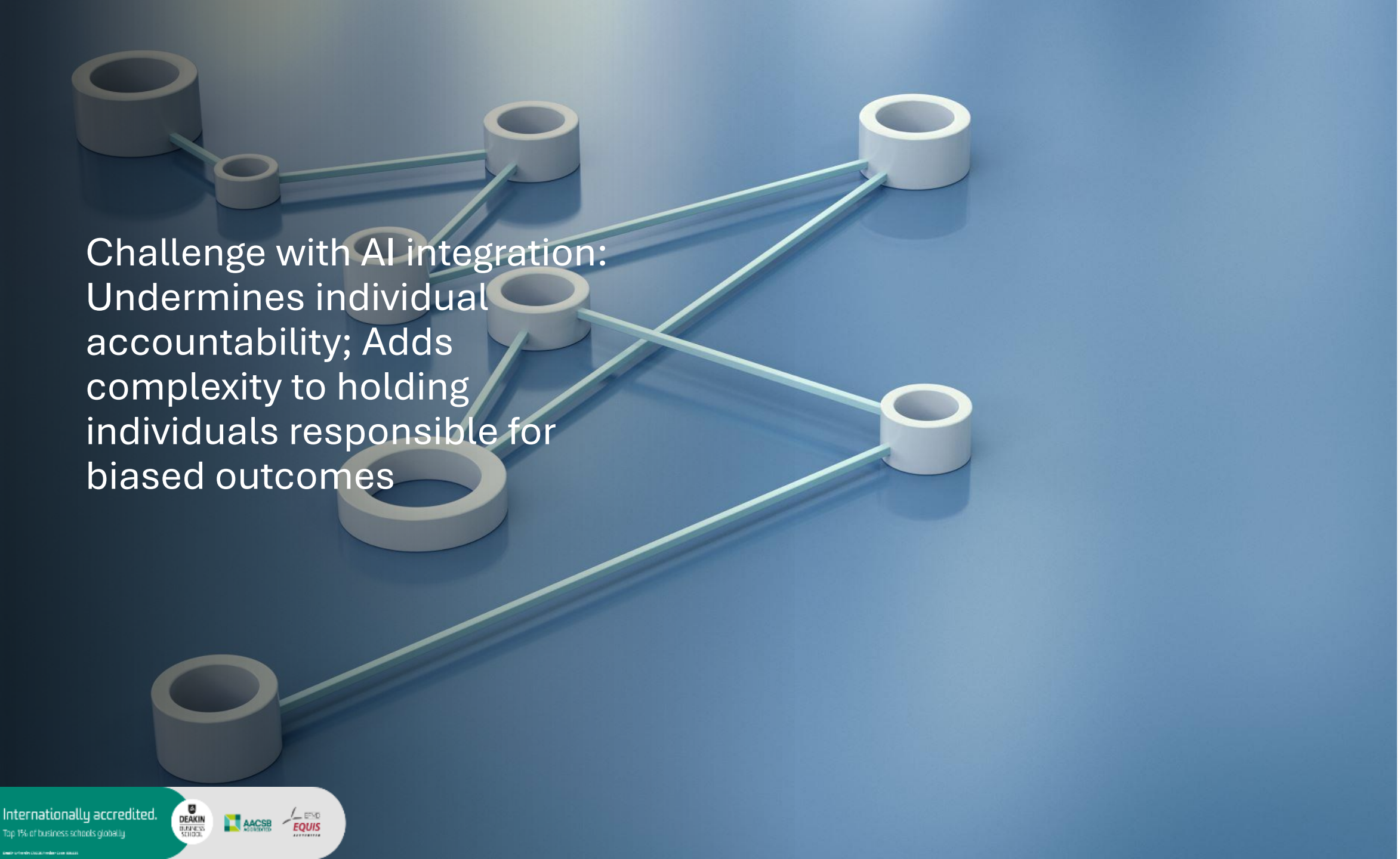
RRI in AI use and management

Organizations relying on human decision-makers: Need to control for unconscious bias among individuals; AI can assist in revealing such biases

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EQUIS ACCREDITED

Addressing biased outcomes:
Not solely reliant on standard antidiscrimination legislation;
Legislation effective only when individuals can be held responsible for decisions

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED



Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

What can executives do to head off such problems?

Explore impacts of outcomes, nature, and scope of decisions

Assess operational complexity and scalability limitations

Determine level of explanation required for decisions

Internationally accredited.
Top 1% of business schools globally
DEAKIN BUSINESS SCHOOL
AACSB ACCREDITED
EFMD EQUIS ACCREDITED

We will discuss this topic further in this week's seminars.

THANK YOU!