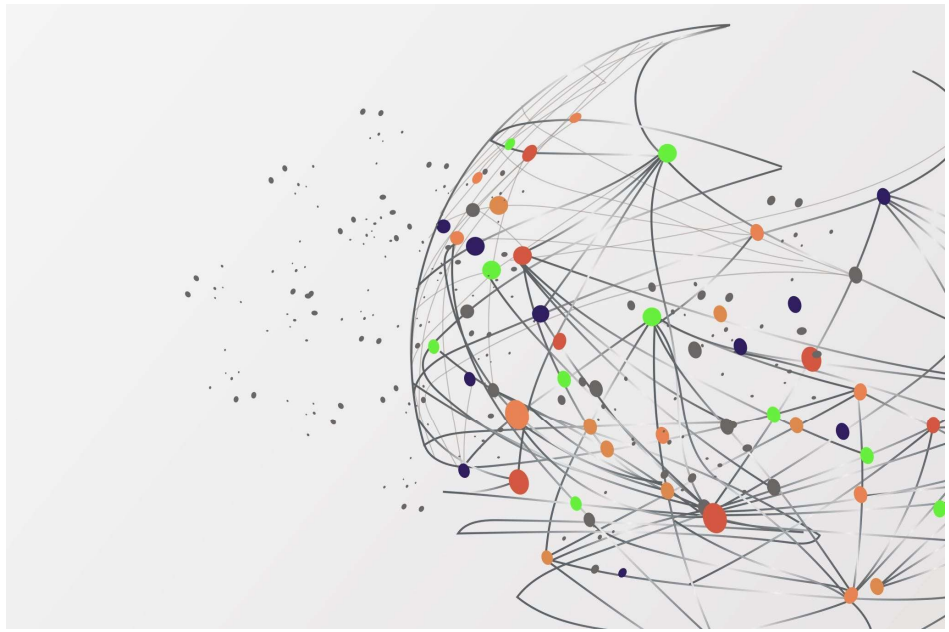
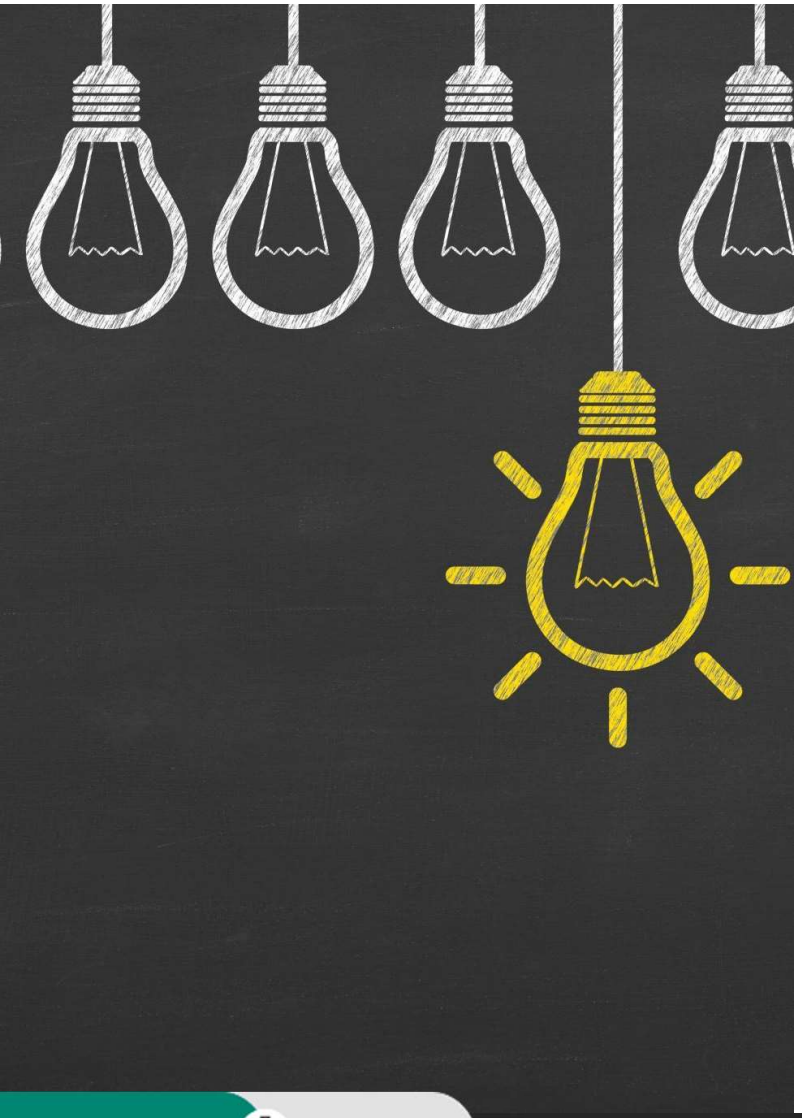


Topic 2:
Responsible Artificial Intelligence



This week's focus is on the following Core Learning Objectives

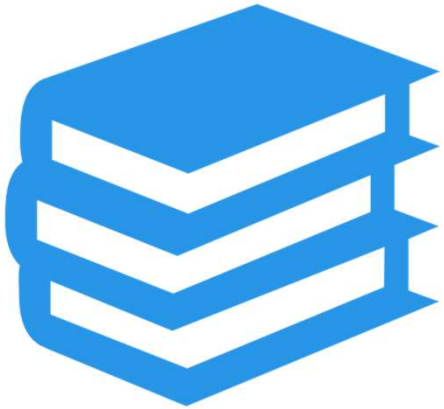
Understanding the notion of Responsible AI

Understanding why we need Responsible AI

Recognising basic principles of Responsible AI

Understanding current research and discussion of Responsible AI

This week's topic is relevant to Assignment 1, and Assignment 2



2.1 Introduction

“Can a system built using AI technology be trusted?”

Watch the following trailer from the 2004 Sci-Fi movie I,Robot starring Will Smith (See Section 2.1 on the unit site)

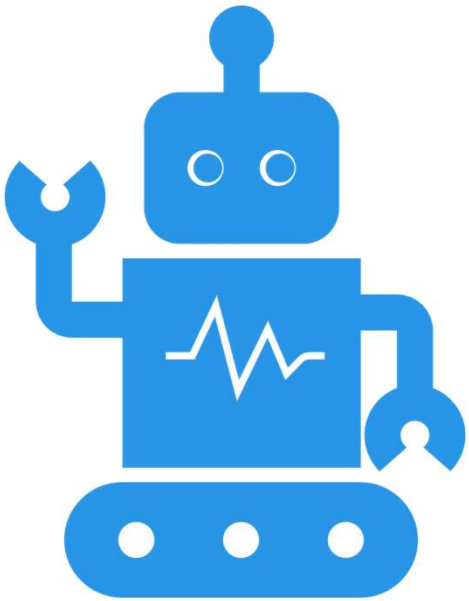
2.1. Introduction

This topic explores how responsible AI can engender trust and acceptance in those impacted by AI and the deployers of AI systems

The key to trust is accountability

If an AI system causes harm though physical damage or breaches of privacy or another action outside its defined behaviours, then those involved in the creation and deployers of the system should be held accountable

But some argue that the complexity of AI systems means we cannot expect those responsible for AI systems to be accountable



2.2. AI Challenges: Should we trust AI?

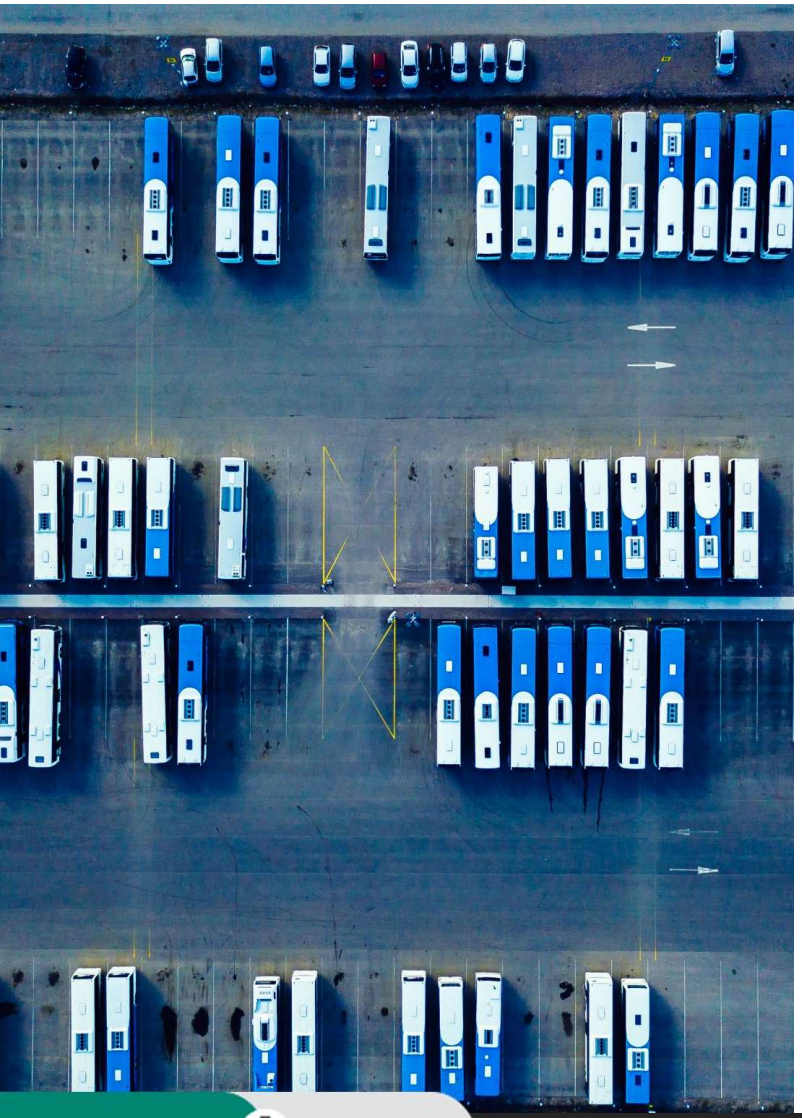
Trusting AI is problematic

In the article "AI & Global Governance: No One Should Trust AI" published by United Nations University Centre for Policy Research, Dr Joanna Bryson explains why trusting AI is problematic and why accountability of those responsible for AI systems is important

At present due to the lack of standards, there is no way to guarantee that an AI system will not step over the line

There are laws and standards for gathering, storage and use of personal information and, even then, the technology and uses are always ahead of policy makers and industry standards

These and other systems need to be well regulated



Trusting AI is problematic

A good example of a very well-regulated industry is vehicle manufacture

In Australia, any vehicle, imported or locally manufactured, must comply with Australian vehicle standards and is assigned a compliance plate

After sale, the use, and to some degree maintenance of, vehicles is controlled by legislation such as the "Road Safety Act 1986" and enforced by government bodies such as the Department of Transport and the Police Force

AI Challenges - What people think about AI?

- Can current regulations and laws make AI safe to use?
- How will AI impact jobs?
- Can the use and operation of AI be understood?
- What impacts will AI have on society?

Activity 1: Section 2.2 on the unit site

01

EXPLORING A MODEL OF THE KEY DRIVERS OF TRUST AND ACCEPTANCE OF AI SYSTEMS

02

DO YOU THINK THE MODEL IS APPROPRIATE AND A GOOD BASIS FOR UNDERSTANDING AND INCREASING TRUST AND ACCEPTANCE OF AI?

03

FOR MORE BACKGROUND SEE THE FULL KPMG REPORT ON THE UNIT SITE

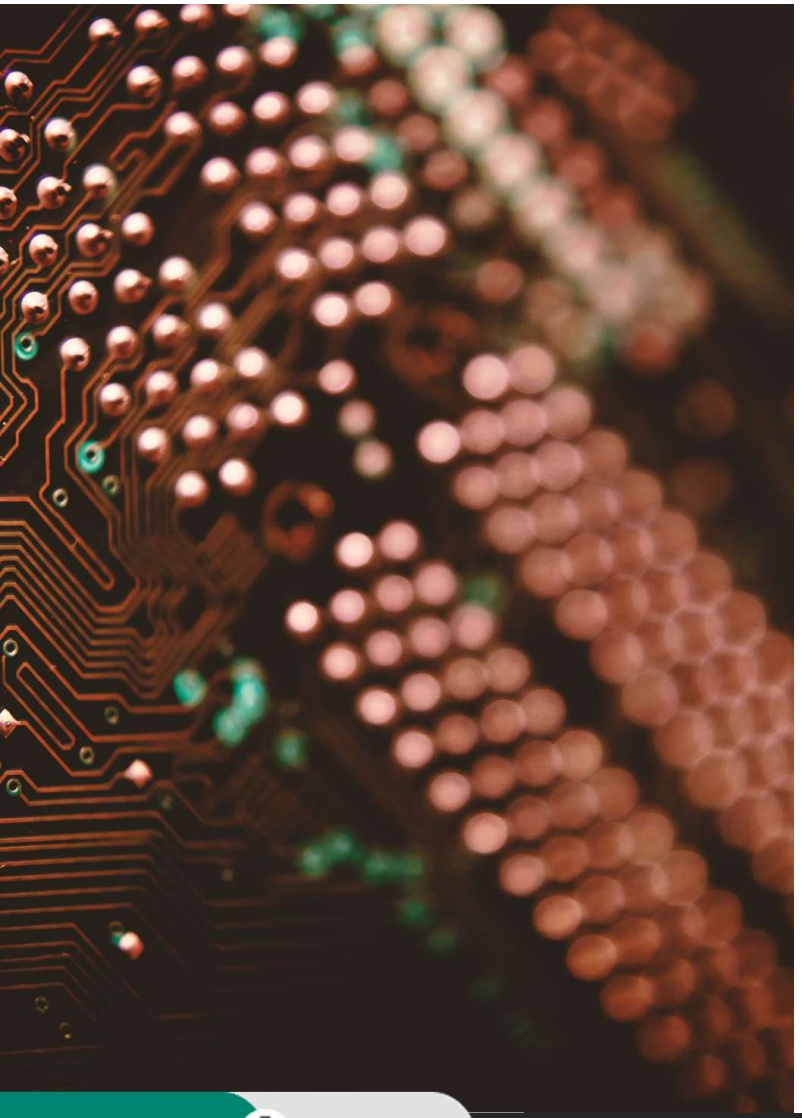
Activity 2: The Trolley Problem

Watch the video “The Trolley Problem” (see section 2.2 on the unit site)

What would you do?

Do you think an AI system could ever be sufficiently "human" to deal with the trolley problem?

Consider a self-driving vehicle that is faced with either running over a person or crashing into a tree

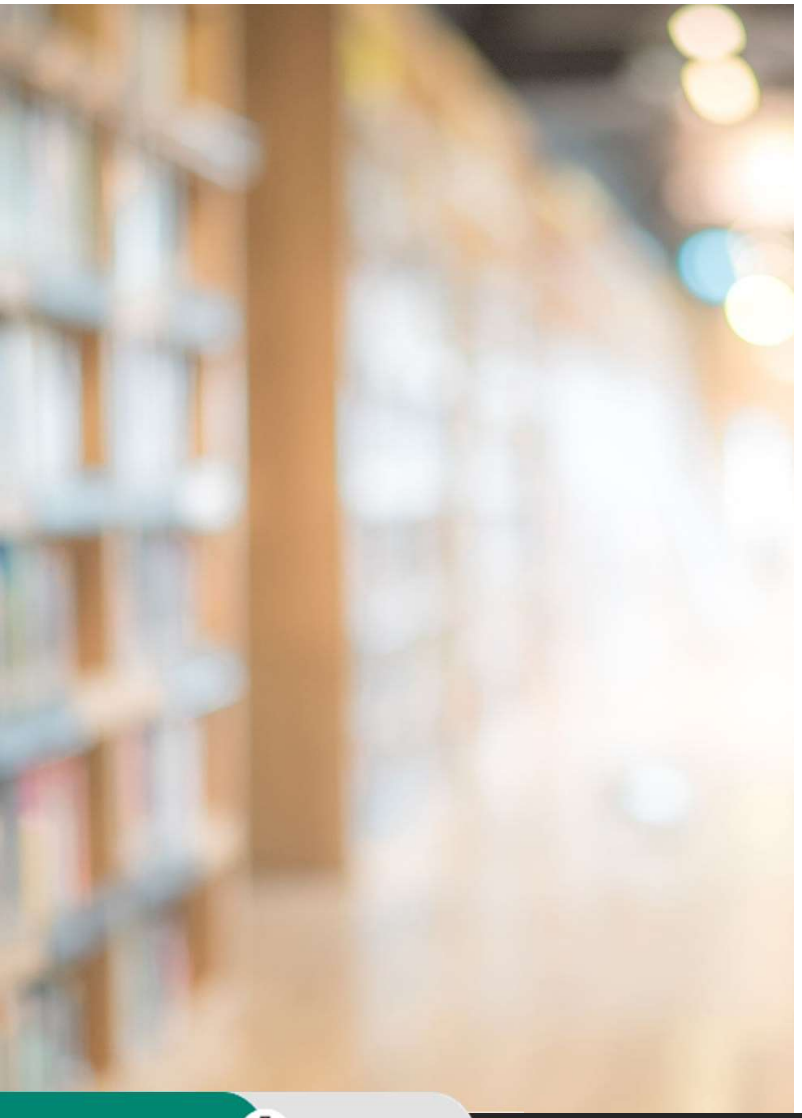


2.3 Why Responsible AI

The increasing use of AI has significant legal, ethical, societal, and economical implications for everyone

Concerns about how AI is used continue to rise with privacy breaches, bias in automated selection procedures, inappropriate generation of debt demands, and the dangerous behaviour of semi-autonomous vehicles

The influence of AI on jobs, people's well-being, healthcare, the distribution of wealth and other social considerations is unclear



What is Responsible AI?

Refers to the practice of designing, developing, and deploying AI with good intention to empower employees and businesses, and fairly impact customers and society—allowing companies to engender trust and scale AI with confidence

RAI is still a relatively new field that has rapidly developed over the past several years and has become a popular term in both business and the media

2.3 Why do we need it?

"*Some of AI's greatest failures to date have been the product of bias, whether it's recruiting tools that favor men over women or facial recognition programs that misidentify people of color. Embedding ethics and inclusion in the design and delivery of AI not only helps to mitigate bias — it also helps to increase the accuracy and relevancy of our models, and to increase their performance in all kinds of situations once deployed.*" (Salesforce's chief ethical and humane use officer cited in Renieris et al'. 2022)

Mature RAI programs will help to minimize AI system failures.

2.3 Why do we need it?

How AI systems amplify bias

Image recognition systems that use biased machine learning data sets will inadvertently magnify that bias. Researchers are examining ways to reduce the effects.

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	PASTA
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	FRUIT
HEAT	–
TOOL	KNIFE
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	MEAT
HEAT	GRILL
TOOL	TONGS
PLACE	OUTSIDE

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	VEGETABLES
HEAT	STOVE
TOOL	TONGS
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	MAN
FOOD	–
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN

In this example of gender bias, adapted from a report published by researchers from the University of Virginia and the University of Washington, a visual semantic role labeling system has learned to identify a person cooking as female, even when the image is male.

Source: Pratt (2020)

2.3 Why do we need it?

Arisen from a limited understanding of critical issues that emerged with the use and development of AI technologies

Recent studies and cases in practice have shown that AI can potentially create unintended consequences such as *biases, discrimination, errors, unexpected results*, and an overall lack of transparency regarding how outcomes are achieved

RAI is concerned with establishing ethical principles and human values to reduce biases, promote fairness and to ensure robustness and security in the use of AI technologies

Responsible AI as the top priority

Responsible AI requires informed participation of all stakeholders with education at the forefront to enable an understanding of the effects on society

Perhaps what is needed is a code of conduct for both AI systems, and the people that create and use these systems, which leads to AI systems conforming to our ethical expectations

Practitioners and researchers have called for AI regulation

RAI should be a top management concern and many organisations now have established responsible AI teams and units to ensure that AI is developed and used appropriatel

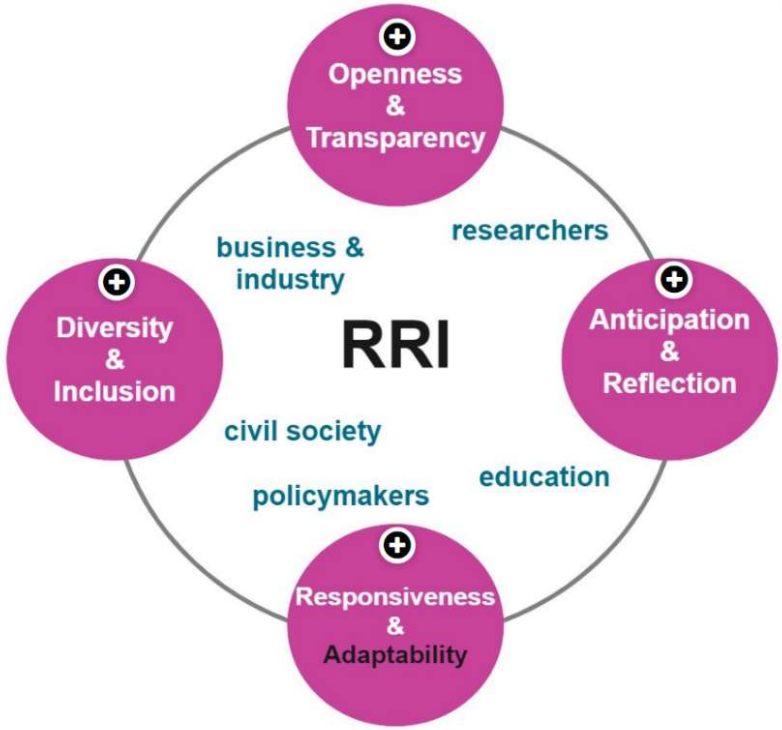
Current stage of RAI implementation in businesses

- The findings from the survey conducted by (Renieris et al., 2022) show that only 52% (11 out of 21 AI experts) believe that firms have an RAI program in place, and the result would explain why there have been an increase in AI failure cases.
- According to the study, mature RAI programs will help to minimize AI system failures.

2.4 Responsible AI Research and Discussion:

Basic principles of Responsible AI:

See section 2.4 on the unit site



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

THANK YOU!