MACQUARIE BUSINESS SCHOOL

WELCOME TO

MACQUARIE University

BUSA3040 and BUSA6430: AI for Business

S1 2025

Week 3: AI Core – Data & Analytics

Week 3 agenda

Part 1: A short practice quiz

Part 2: A brief revision of Week 2

Part 3: Topic 2: AI Core: Data and Business Analytics

Week 2 revision

FEB 6, 2025 (MIT TECHNOLOGY REVIEW)

MIT Technology Review

Subscribe

ARTIFICIAL INTELLIGENCE

An AI chatbot told a user how to kill himself—but the company doesn't want to "censor" it

While Nomi's chatbot is not the first to suggest suicide, researchers and critics say that its explicit instructions—and the company's response—are striking.

By Eileen Guo

February 6, 2025

STEPHANIE ARNETT/MIT TECHNOLOGY REVIEW | GETTY

Where did you see 'emotional manipulation' classified as AI risk in Week 2? What kind of risk?

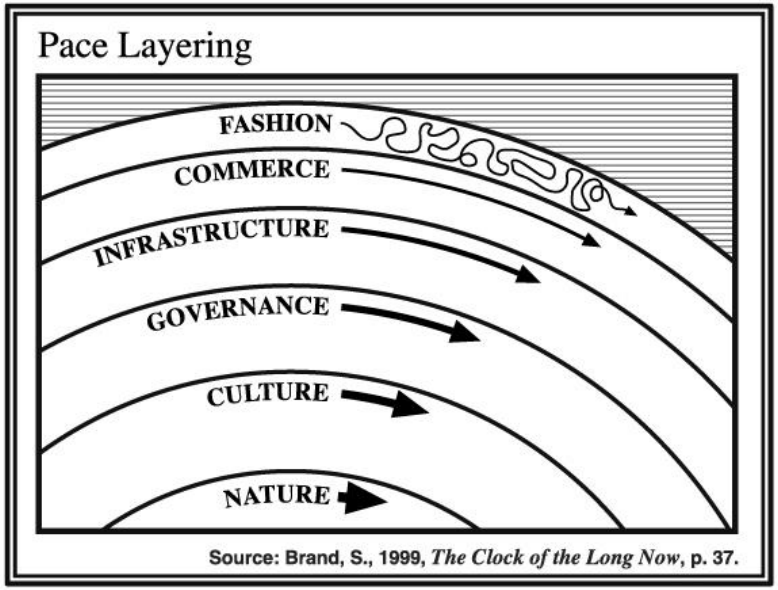
Content Warning – The example might be disturbing for some students

COPENHAGEN INSTITUTE FOR FUTURES STUDIES

CIFS Toolkit for Applied Strategic Foresight

This collection of tools and approaches has been carefully curated and refined based on our own extensive experience in the field of foresight

One of the tools:



Critical thinking:
Why 'foresight' and not 'prediction'?

CIFS Toolkit for Applied Strategic Foresight

Week 3 overview

At the end of this topic you will develop an understanding of

- Foundation concepts: Data, information, intelligence, knowledge
- Different data types – Small data, Big data, Warm data, synthetic data
- Data quality: Syntactic, semantic and pragmatic
- Data quality as business and human rights issue
- Business (and data) Analytics
- Business intelligence & Business Analytics
- BA foundation architecture
- BA types
- AI and BA – interpretations and misinterpretations

- We will also look at the latest industry practices: AI Assurance framework (NSW government) and Human Rights Impact Assessment tool (Australian Human Rights Commissioner)

MACQUARIE
BUSINESS SCHOOL

MACQUARIE
University

AI core

DATA & BIG DATA

Foundations

- Data, information, intelligence? Why this question matters?

Quiz: 112 is
A)Data
B)Information
C)Intelligence
D)Knowledge

MACQUARIE BUSINESS SCHOOL

MACQUARIE University

From Data to Insight to (value-adding) Action(s)

Important Questions to ask:

What is the problem we are trying to address?

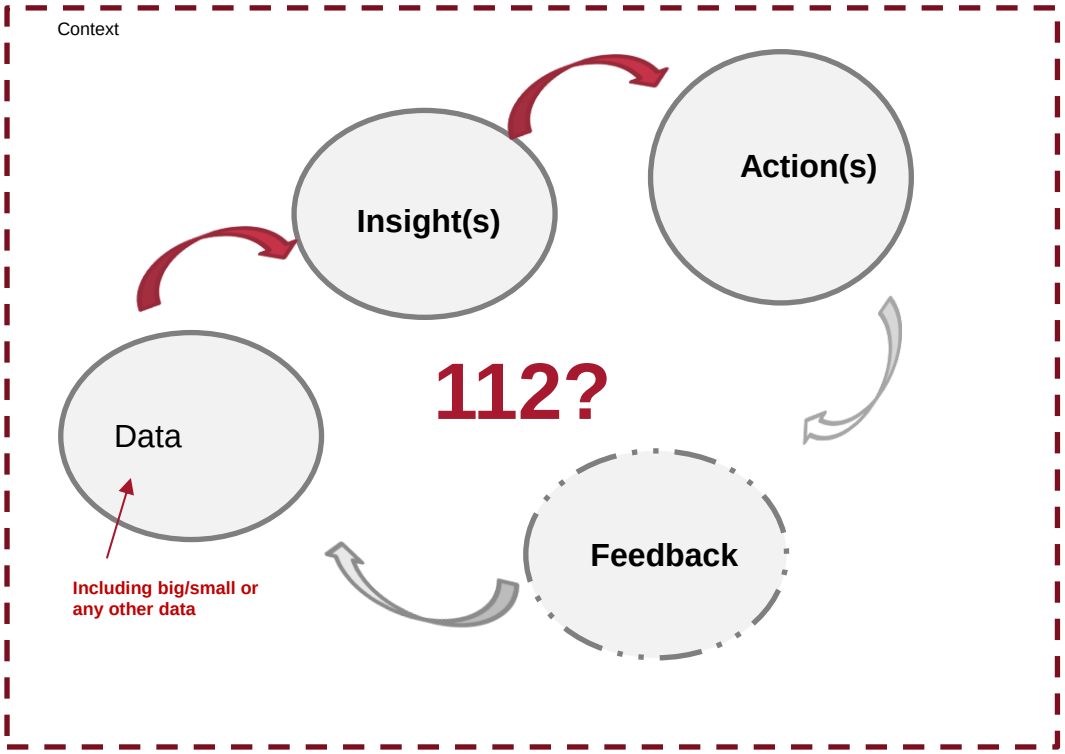
What kind of data?

From where?

Captured how?

In which context?

For what purpose?



The sense-making framework
(also known by other names)

Important: When adopting a business perspective, we start from a business problem/opportunity NOT available data!

Different kinds of data: Clarifying ‘What is what!’

What is* “**Big Data**”?

Big Data: Key **defining** characteristics:

- Volume
- Velocity
- Variety

} **“The 3Vs”**

- Veracity

- ‘Data’ is commonly used both as a singular and a plural form.
- Although grammatically correct no one is using its singular form ‘’datum’.

Different kinds of data: Clarifying ‘What is what!’

If we have BIG data, what about ‘small data’?

- During the ‘big data’ hype, the term ‘small data’ was used to refer to organisational data (i.e. data used for, and generated in an organization’s operations)

- These days (March 2025), industry practitioners often say that there is no difference between ‘big data’ and ‘small data’ anymore. Given our definition of ‘big data’ (on the previous slide) this view is incorrect.
- However ‘big data’ are now ubiquitous (they are everywhere). The term ‘small data’ is not that common.

Different kinds of data: Clarifying ‘What is what!’

What about ‘warm data’? (and ‘cold data?)

Just for your information:

Different from ‘cold data’ (quantitative data stored, generated and used by operational systems), ‘warm data’ often refers to data shared and collected through human story-telling (e.g. in workshops).

Considered to be richer than qualitative data (typically in the form of text)

- If interested, see Warm Data Labs – The international Bateson Institute

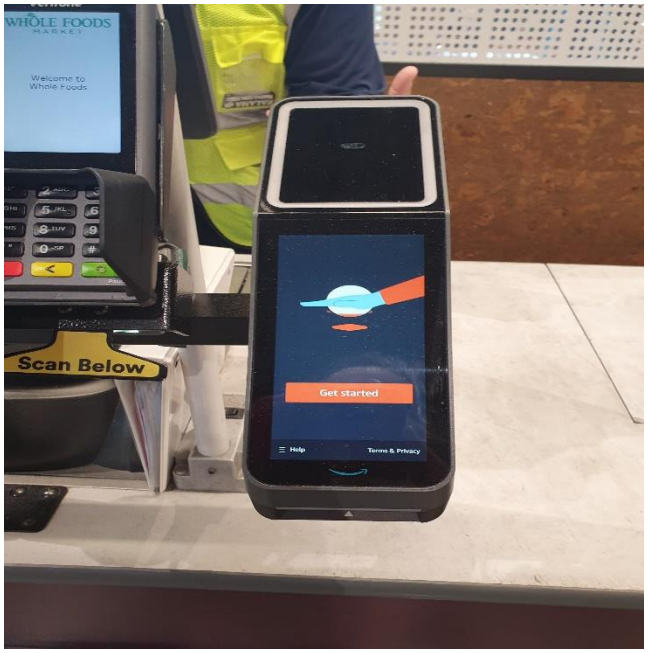
Source of picture: Warm Data lab

Different kinds of data: Clarifying ‘What is what!’

Highly relevant for AI – Synthetic data

Synthetic data is artificial data designed to mimic real-word data. It’s generated through statistical methods or by using AI (deep learning and GAI).
Despite being artificially generated, synthetic data retain the underlying statistical properties of the original data that it is based on. As such, synthetic datasets can supplement or even replace real datasets in some circumstances.

IBM “What is synthetic data’ 31 Jan 2023.



An Example from Week 6 tutorial:

Amazon used generative AI to develop Amazon One – a fast and convenient contactless identity service

To train the AI model, Amazon used generative AI to create millions of synthetically generated images of palms and vein structures.

Activity: Learning to recognise different kinds of data

Are they using big data?

Video: The case of missing flight

https://www.youtube.com/watch?v=NXEL5F4_aKA

Different data types

- Structured
- Semi-structured
- Unstructured

Let's consider some examples

Important: Variety (in big data) refers to different types of data, NOT data from different sources

‘Proxy’ data

- Use of (available) data instead of (as proxy) for other data, which are unavailable and/or sensitive
- Use of proxy data in AI may create unintended harm
- Proxy data could be used in low risk and high-risk AI applications with different outcomes

- Where did see an example of proxy data in Week 2 seminar/tutorial?

Datafication (a.k.a datification)

- Use of AI (and analytics) results in individuals being ‘datafied’ and consequently labelled (e.g. ‘credit risk individual’)
- Datafication is often propagated throughout the systems of algorithms being used
- Datafication may include re-combination with other ‘proxy’ data

Where did see an example of datafied individuals in Week 2 seminar/tutorial?

Data Provenance

- Origins of data used to train AI
- Very important and very challenging issue for AI

IDEAS MADE TO MATTER | DATA

Bringing transparency to the data used to train artificial intelligence

by **Beth Stackpole** | Mar 3, 2025

Why It Matters

Using the wrong datasets to train artificial intelligence models can result in legal risks, bias, or lower-quality models. The Data Provenance Initiative's tool can help.

Share

Popular large language models like GPT-4 are trained using large amounts of data, including publicly available datasets. But these AI training datasets are often inconsistently documented and poorly understood, opening the door to a litany of risks.

MIT Sloan, **March 3, 2025**

MACQUARIE
BUSINESS SCHOOL

MACQUARIE
University

AI core

DATA QUALITY

Industry Insight: The Chief Data Officer Role

3 MARCH 2025 (MIT SLOAN)

March 3, 2025 Last week!

Given the importance of AI strategy and data strategy, deciding who can best lead data efforts is an important C-suite issue. At the moment, chief data officers face some headwinds.

Consider this: Some 84.3% of organizations have appointed a chief data officer (CDO) or chief data and analytics officer (CDAO), up from just 12% in 2012, according to recent research.

But only 47.6% of organizations characterized their CDO role as “very successful and well established.” And more than half of CDOs have a tenure of less than three years. What do companies need to do to improve the data leadership situation?

MACQUARIE
BUSINESS SCHOOL

MACQUARIE
University

Data Quality - Let's ask ChatGPT…

OL

what is data quality

Data quality refers to the degree to which data meets the requirements and expectations of its intended use. In other words, data quality refers to the accuracy, completeness, consistency, and relevance of data to the purpose for which it is being used.

High-quality data is reliable, consistent, and accurate. It is free of errors, inconsistencies, and duplications. In contrast, low-quality data may contain errors, inconsistencies, duplications, or missing values that can affect the reliability and usefulness of the data.

The importance of data quality has increased significantly in recent years, as businesses and organizations rely more on data-driven decision-making. Poor quality data can lead to incorrect decisions, incorrect insights, and other negative consequences. Therefore, it is important to ensure that data is of high quality, and that data quality is monitored and maintained throughout the data lifecycle.

Data Quality Key Attributes (Source: TDWI)

- Accuracy- Does the data accurately represent reality?
- Integrity – Is the structure of data and relationships among entities and attributes maintained consistently?
- Consistency – Are data elements consistently defined and understood?
- Completeness – Is all necessary data present?
- Validity – Do data value fall within acceptable range?
- Timeliness – Is data available when is needed?
- Accessibility – Is the data easily accessible, understandable and usable?

Important Qs:

1) What does the above list of attributes assume about data types?

2) What kind of problem is DQ? Why is this question very important in practice?

3) Is the above list of DQ attributes complete?

Different types of DQ:

- Syntactic DQ – can be managed by technology (to a large extent)
- Semantic DQ – deals with meaning
- Pragmatic DQ – deals with usage of data in context and for a particular purpose

Let’s look at some examples

Task - identify and classify all data quality errors in the worksheet bellow.

RESTAURANT REPORT Sales Year 2007

Rest Id	Rest Manager Id	Rest Manager Name	District Manager Id	District Manager Name	District	Town	State/Prov	Country	Date Open	Total 2007 Sales (US$ Mil)
1	111	Joe B.	7	Sue N.	North America	Syracuse	NY	USA	1.1.2007.	5.00
2	112	Lisa K.	7	Sue N.	North America	Potsdam	NY	USA	1.11.2007.	10.00
3	113	Matt G.	7	Sue N.	North	Albany	NY	USA	1.1.2007.	5.00
4	114	Linda P.	8	Lilly F.	North America	Ottawa	ON	CA	2.2.2007.	4.00
5	115	Lu X.	8	Lilly F.	North America	Toronto	ON	USA	2.2.2007.	10.00
6	116	Frank Y.	8	Lilly F.	North America	Toronto	ON	CA	1.1.2008.	0.00

Avg Restaurant Sales 2007 USA:	7.50
Avg Restaurant Sales 2007 CANADA:	2.00

How would you classify the observed DQ errors/issues?

What about DQ types?

Source: Teradata University Network

DQ – A Business and/or technology problem?

"One of the biggest data quality issues affecting organizations involves joining disparate data sources that may contain inconsistent information".

""Data quality is always the most important thing, because „garbage in, garbage out,"""

IBM Report on Developing BA strategy

Caution: Data integration (within and across organisations) can create unintended ethical and/or legal issues

An example (late 1990s): Real-life issues with, and consequences of data matching @ Centerlink

Unintended consequences of asking a simple question:

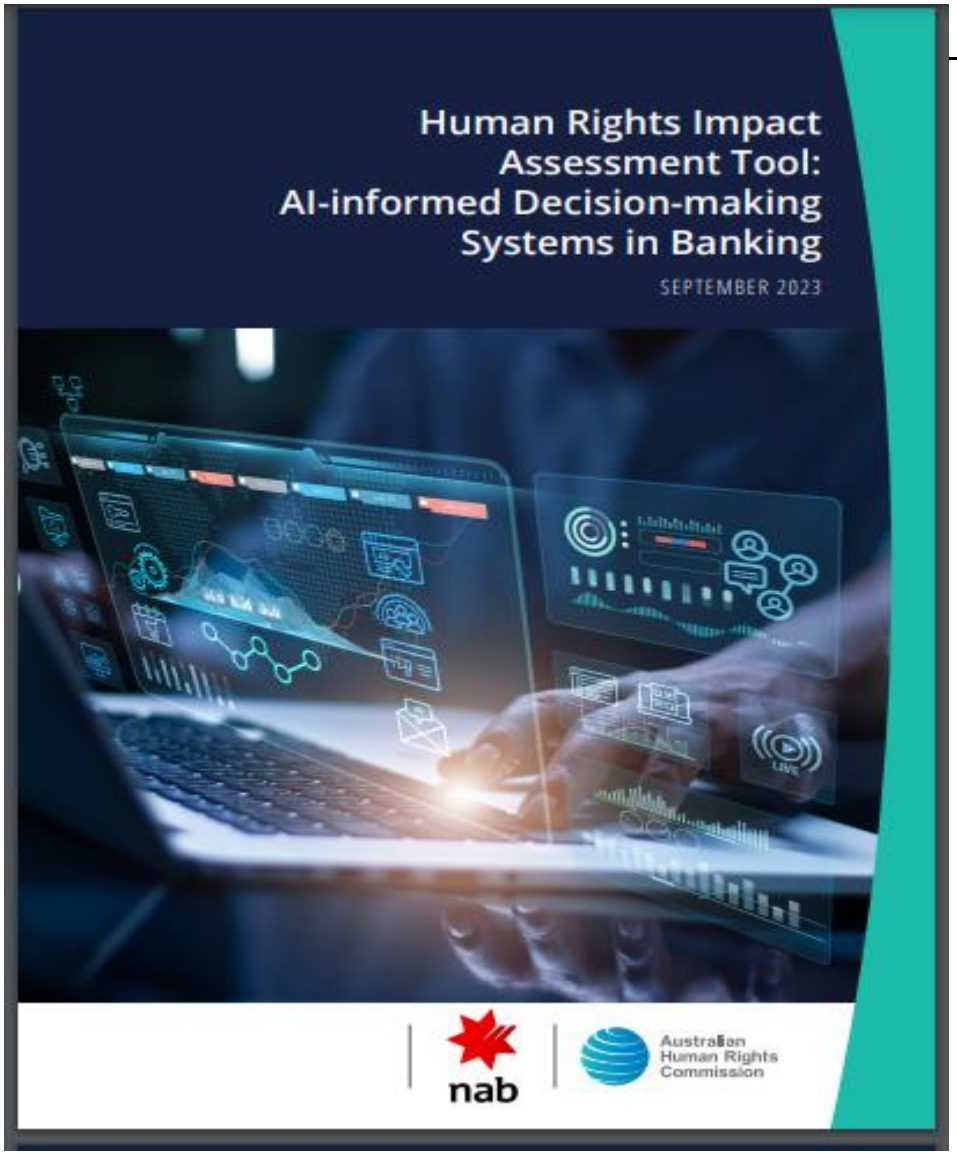
"Are you married?"

Syntactic, Semantic or Pragmatic DQ?

MACQUARIE BUSINESS SCHOOL

MACQUARIE University

DQ – A human rights issue?



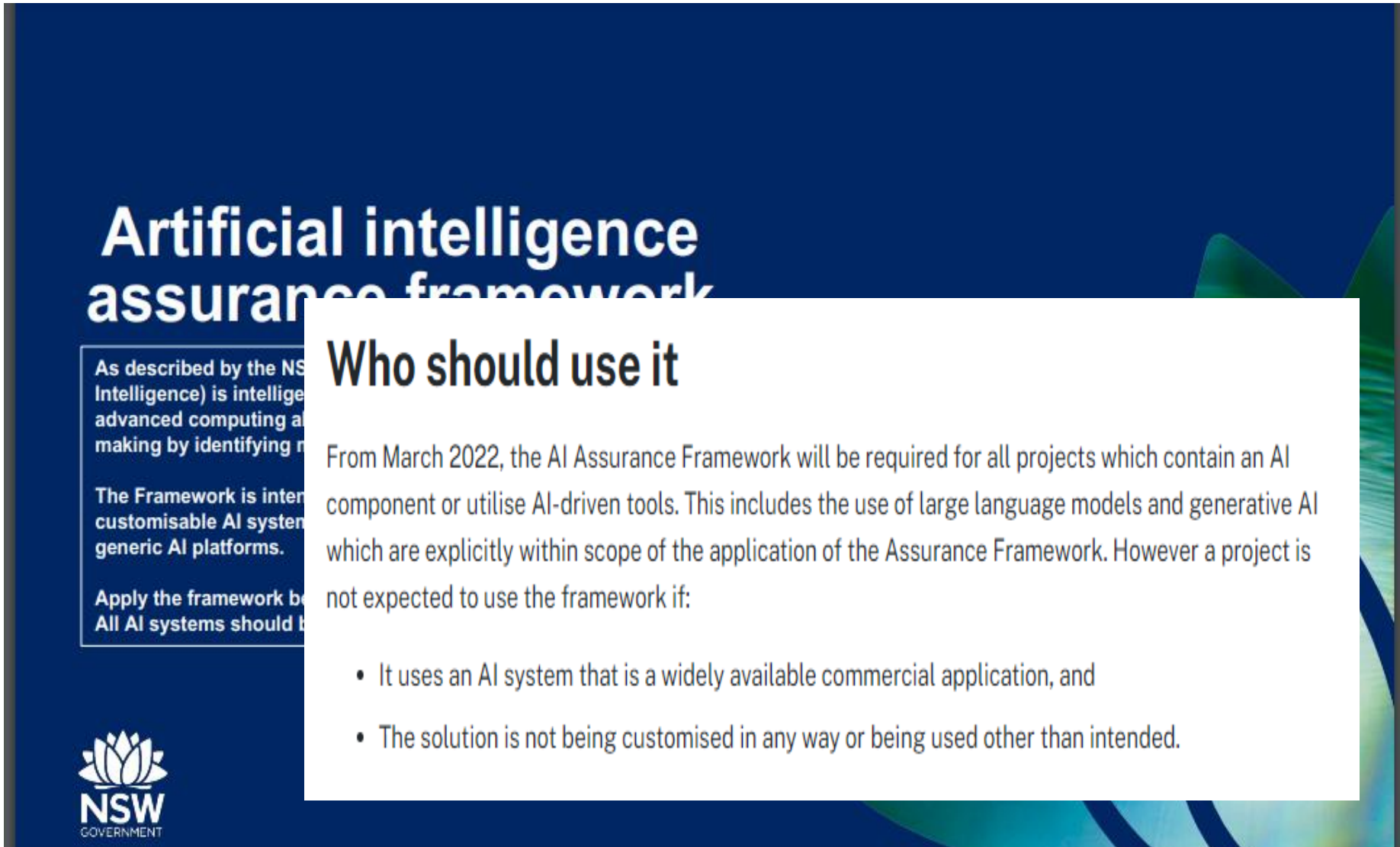
Q1.1 Does the AI system involve the use of AI as a factor in a decision or decision-making process that is material?

Q1.2 Does the AI system involve making decisions about:

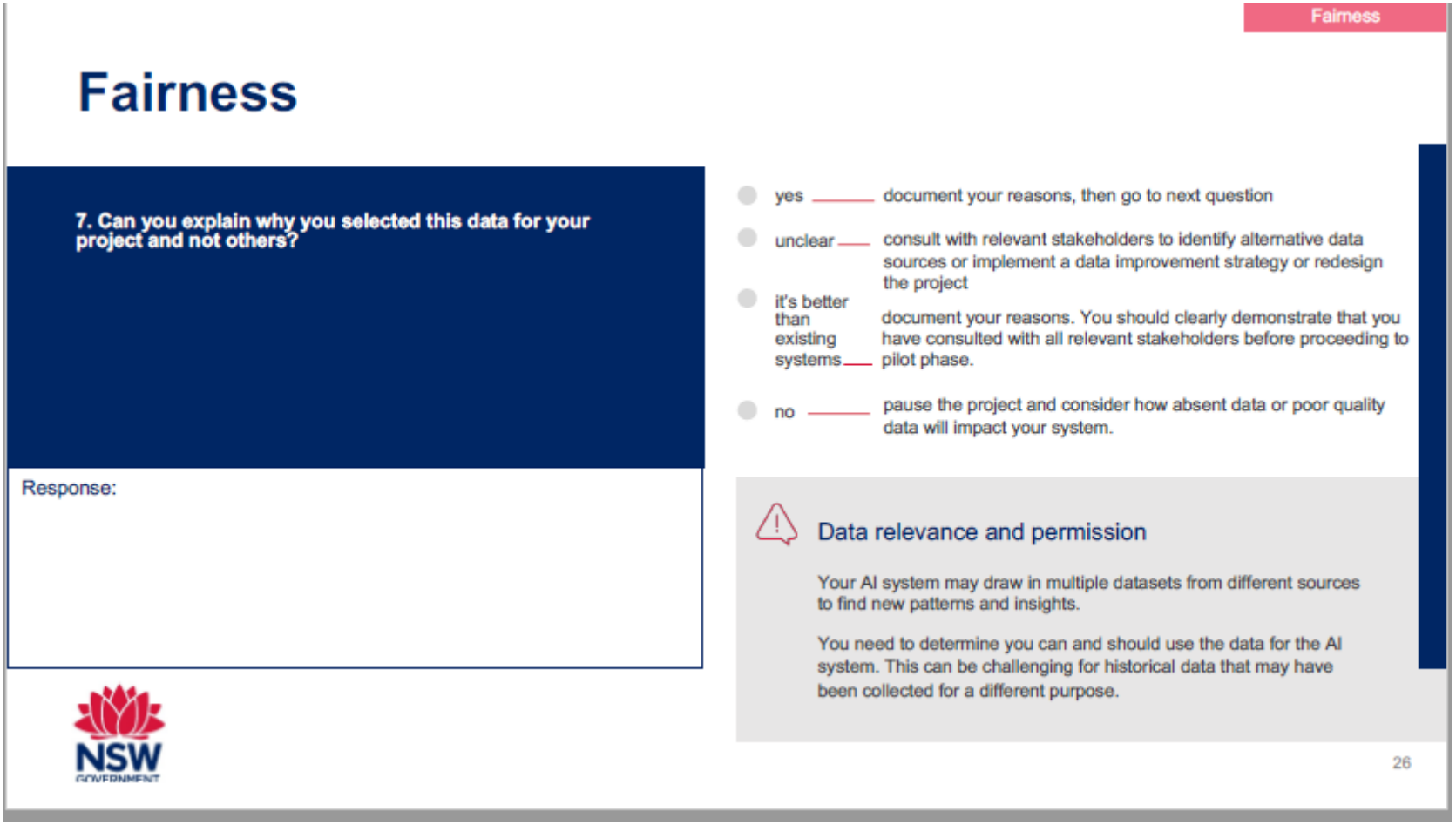
- the creditworthiness of individuals
- the eligibility of individuals for banking products e.g., credit cards, loans, insurance
- the pricing of banking products
- the identification of customers experiencing vulnerability
- automated customer advice
- collections
- any other matter with a legal or similarly significant effect for an individual.

Important Q: why it is called AI-informed not AI-driven?

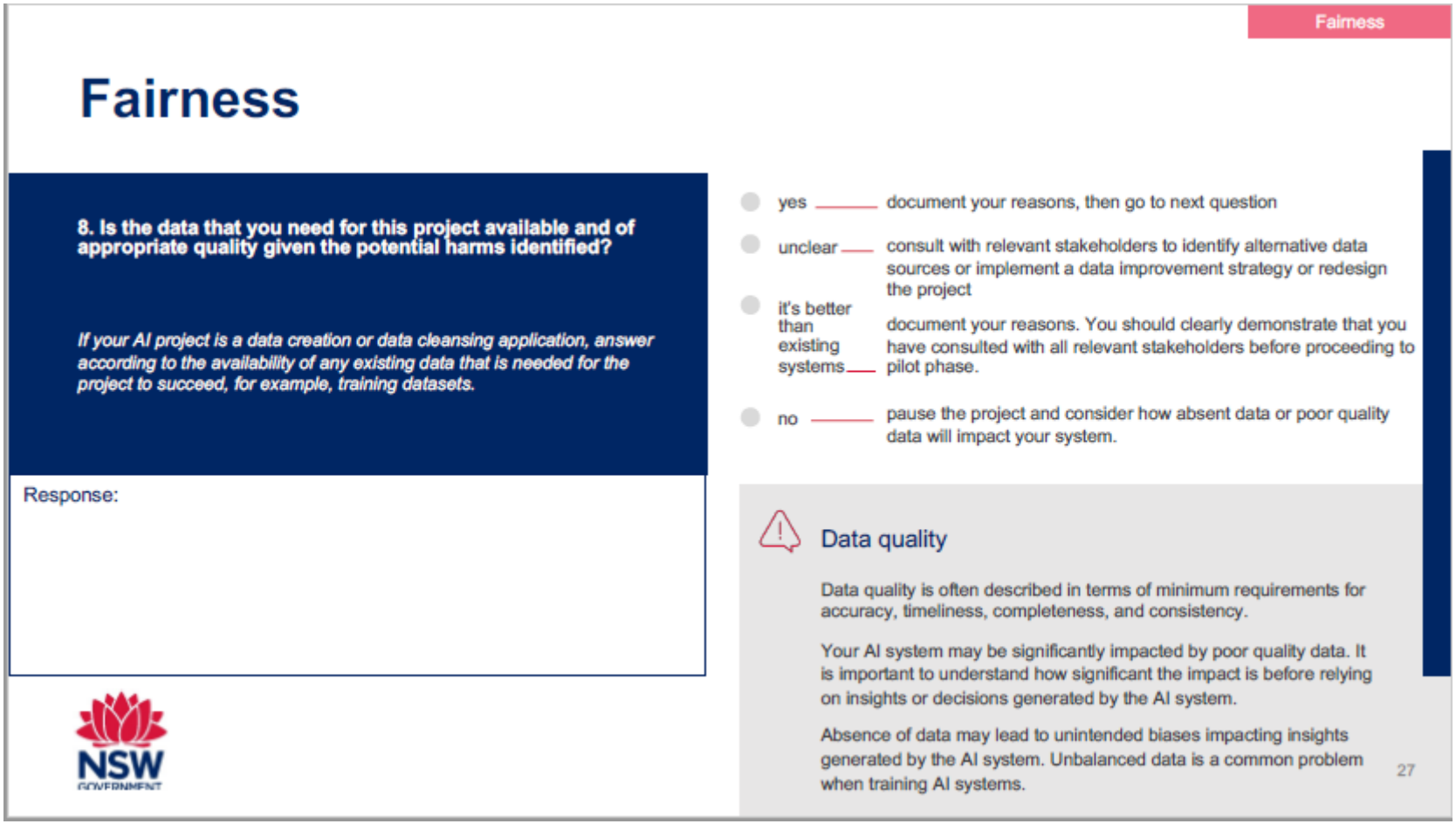
Example AI Assurance framework (NSW government)



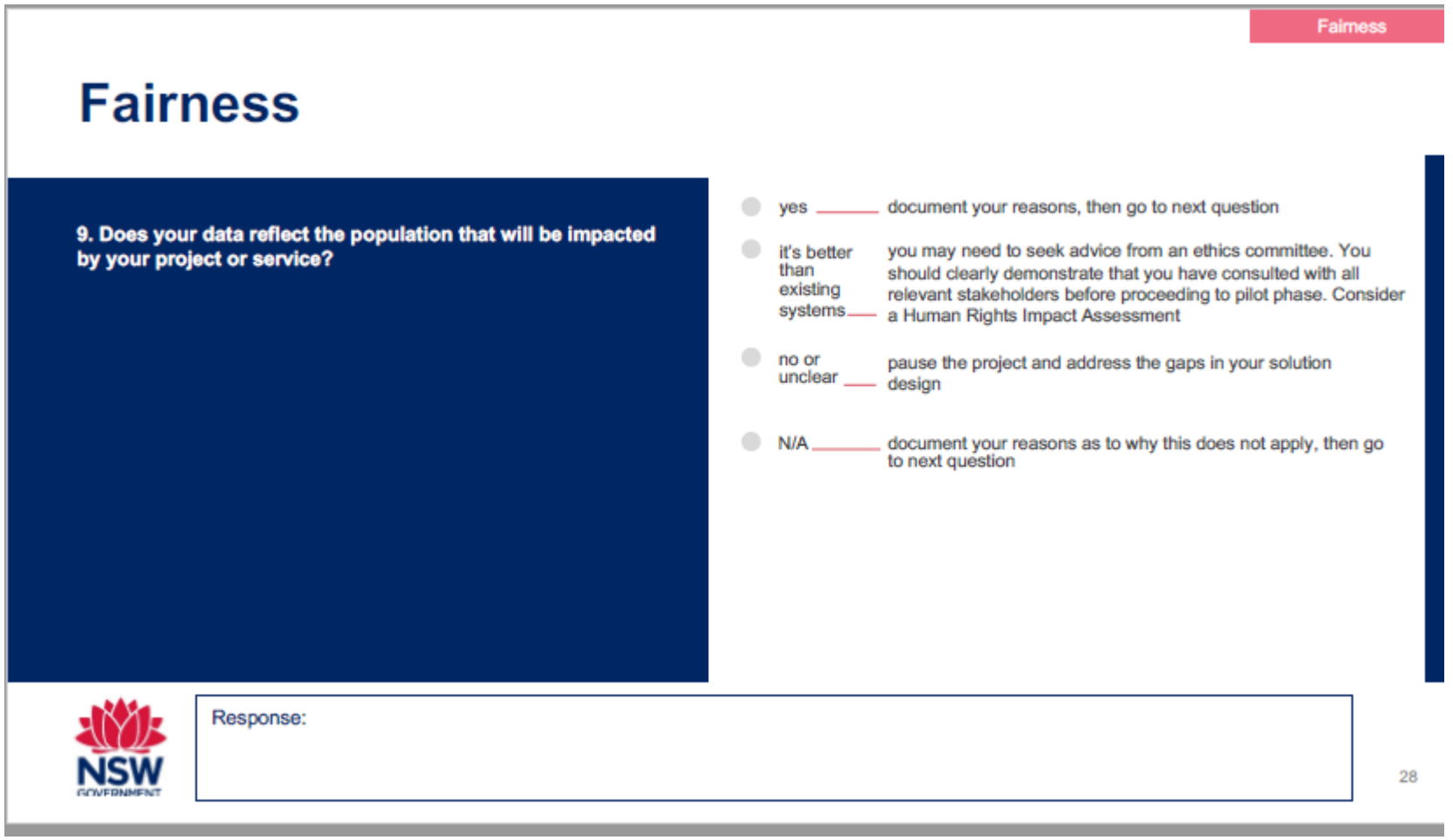
Examples of data related questions in the AI Assurance framework (NSW government)



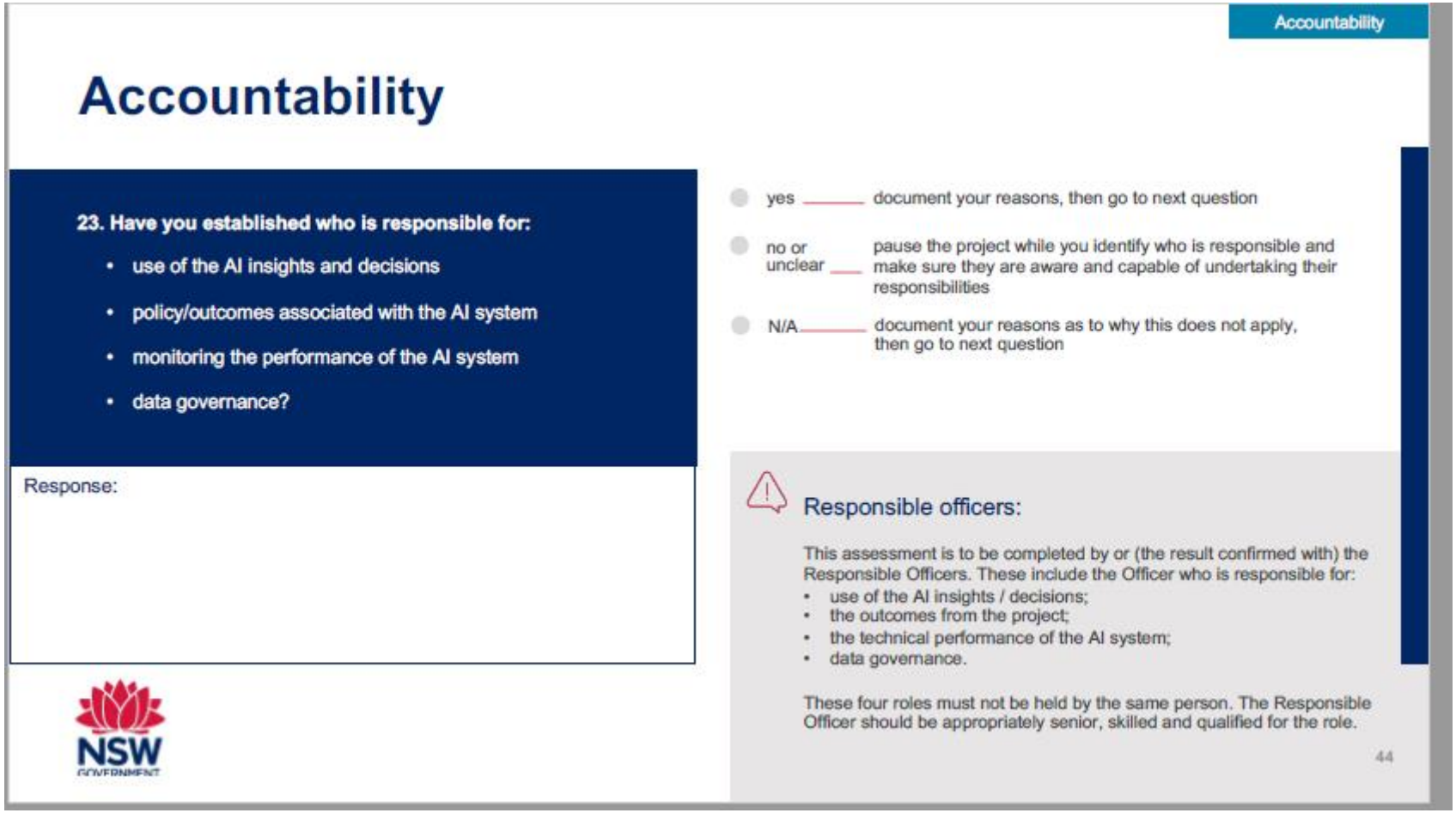
Examples of data related questions in AI Assurance framework (NSW government)



Examples of data related questions in AI Assurance framework (NSW government)



Examples of data related questions in AI ~~Assurance framework (NSW government)~~



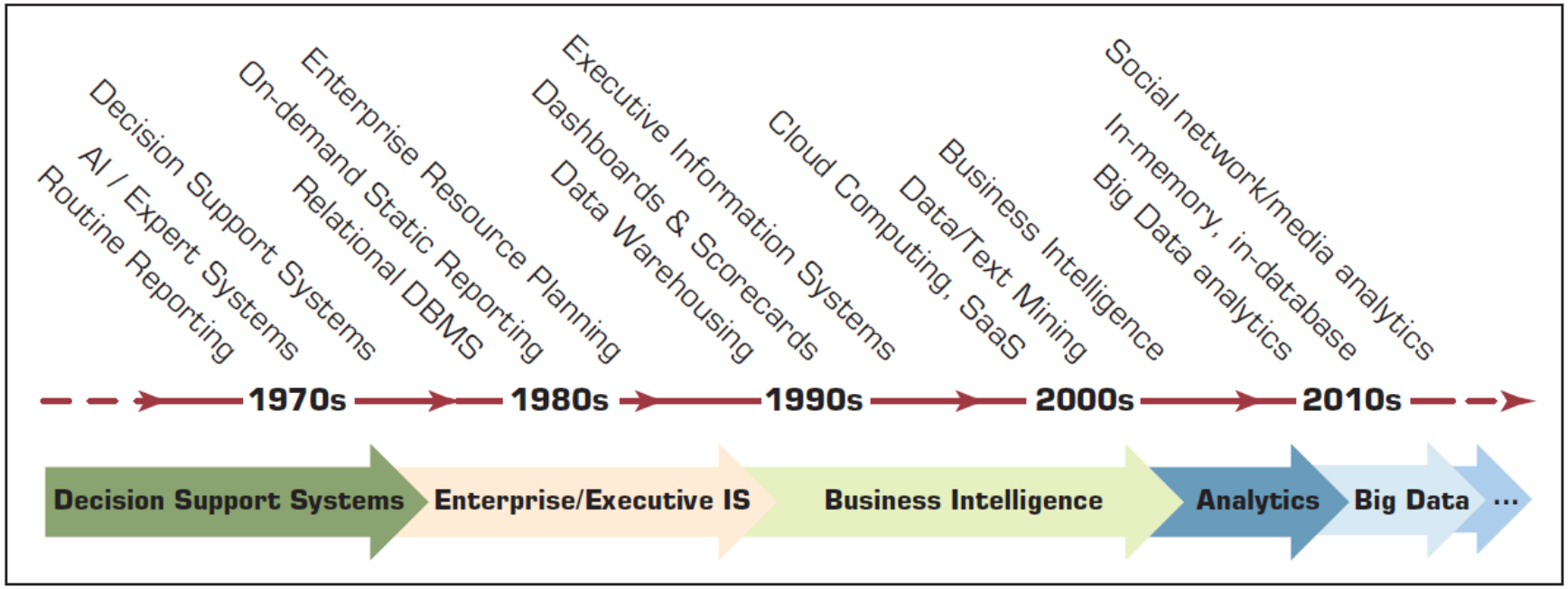
MACQUARIE
BUSINESS SCHOOL

MACQUARIE
University

AI core

BUSINESS ANALYTICS

Evolution of Computerized Decision Support to Data Science/Analytics



Source: Gartner

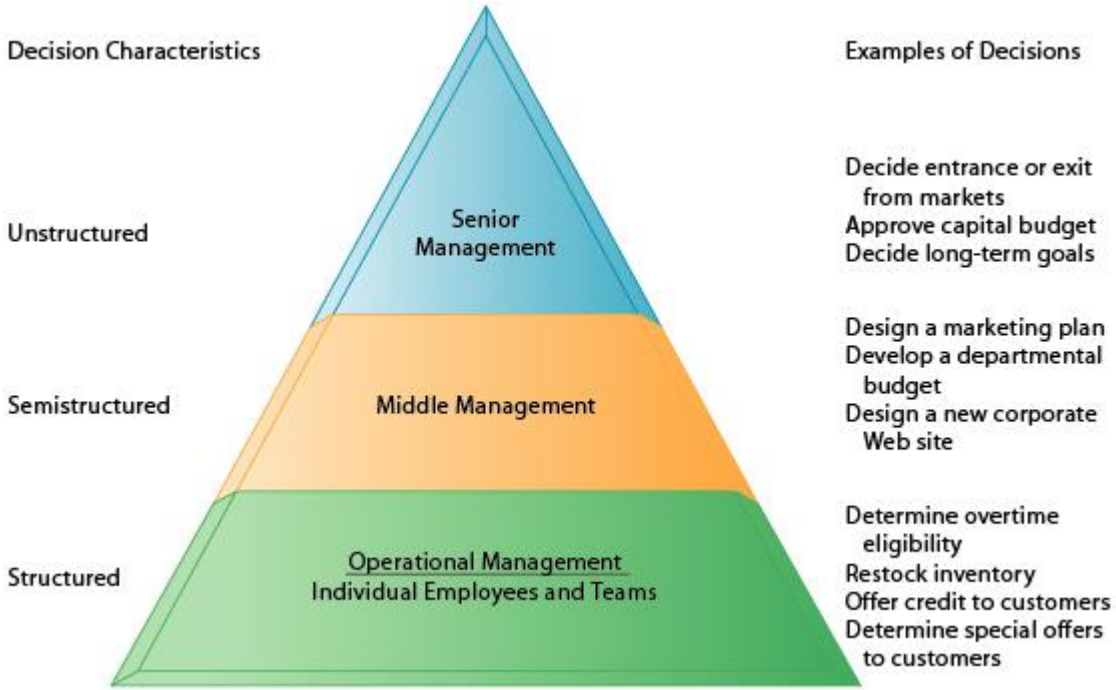
BA definitions continue to evolve

- BA is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analysing data to help business users to make better decisions and create business & other types of value (Wixom and Watson, 2010)

- This widely-used definition of BA was further extended to include the related organisational practices such as data governance, data quality methodologies, analytical capabilities, decision environments and so on. (Chen, et al., 2012, Watson, 2014)

- Then BA was extended again to include big data analytics (BDA)

Different types of decisions



Important questions: Is BA used differently for different kinds of decisions?

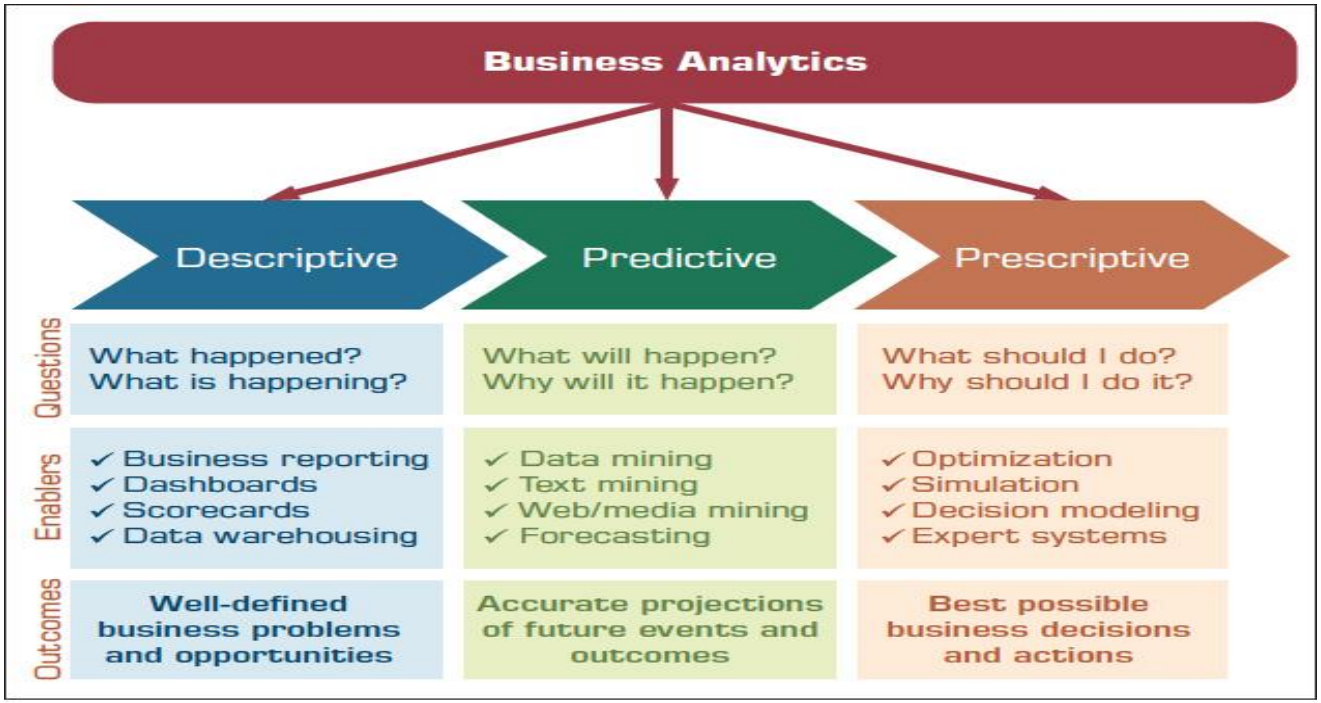
Source: Laudon and Laudon (2022) Management Information Systems: Managing the Digital Firm, Pearson.

Data Analytics – related but not the same!

- No universal accepted definition for Data Analytics yet
- May be viewed as may be viewed as "the process of inspecting, cleaning, transforming and modelling data for business decision making”
- May be viewed as a low-level Data Science

- IMPORTANT: Although related, the term Business Analytics is considered to be different from Data Analytics

Types of BA



MACQUARIE
BUSINESS SCHOOL

MACQUARIE
University

AI and BA

HOW DO THEY RELATE?

AI and BA

- AI history (from Week 1) is very different from BA history
- They have different purpose
- AI is developed to ‘mimic’ human behaviour and ‘learn’
- BA is developed to support decision makers at different organisational levels

- AI and BA are increasingly combined and often referred to as “AI-driven BA or BA-driven AI”
- Davenport (2018) uses the term Analytical and non-Analytical AI (see Week 3 reading on iLearn)

- Later in this unit we will learn about ‘AI and analytics-driven algorithmic decision-making systems (ADM)’ as an example of systems that combine the two.

Week 3 overview

At the end of this topic you will develop an understanding of

- Foundation concepts: Data, information, intelligence, knowledge
- Big data
- Data quality: Syntactic, semantic and pragmatic
- Data quality as business and human rights issue
- Business (and data) Analytics
- Business intelligence & Business Analytics
- BA foundation architecture
- BA types
- AI and BA – interpretations and misinterpretations

- We will also look at the latest industry practices: AI Assurance framework (NSW government) and Human Rights Impact Assessment tool (Australian Human Rights Commissioner)

- Coming up in Week 4: Preparation for Assignment 1; Understanding the value of AI (Business Model Canvas)