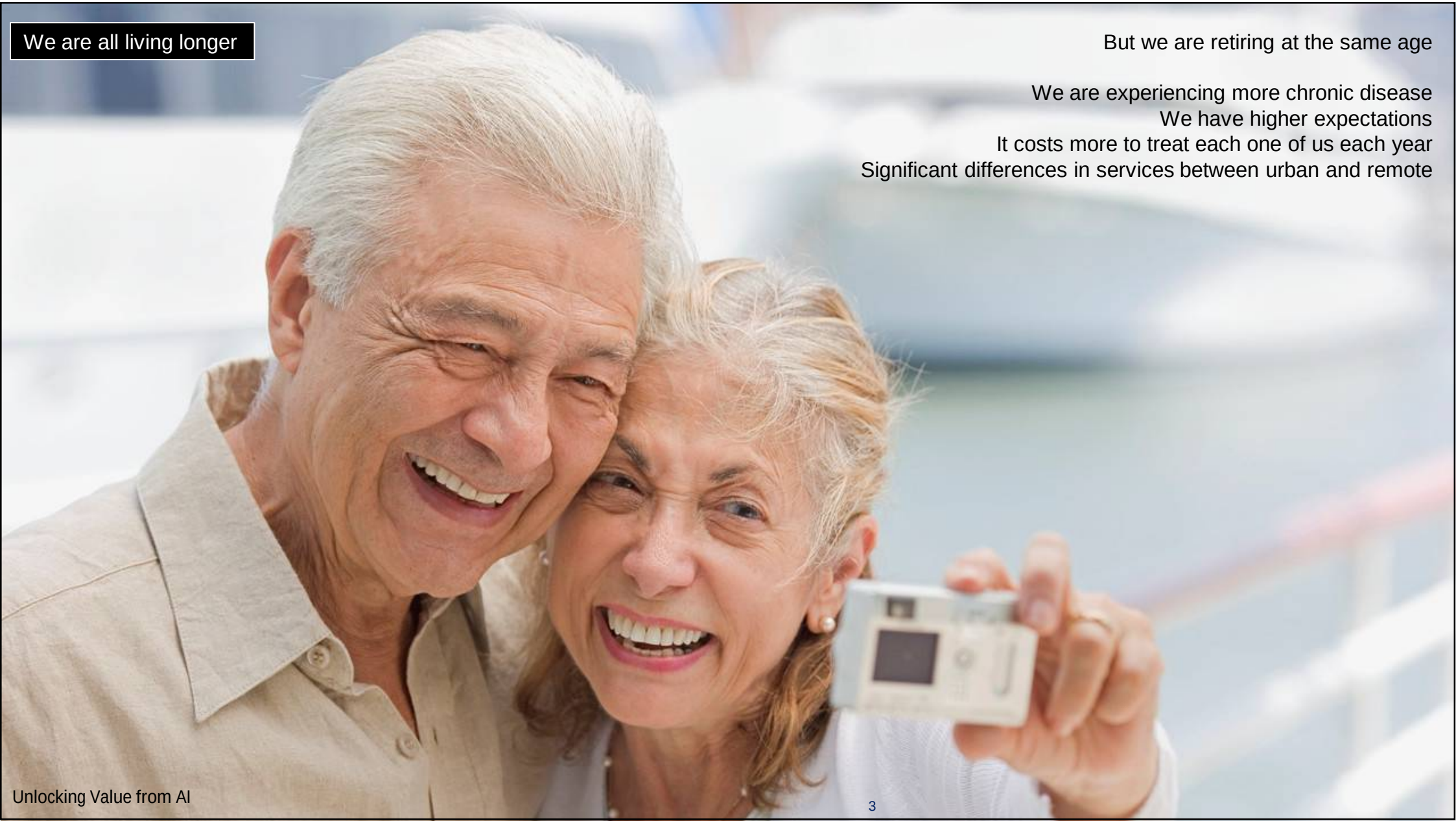
Some thoughts on Unlocking Value from AI

Dr. Ian Oppermann
NSW Chief Data Scientist
UTS Industry Professor
May 2022

Most problems we face today are complex

2 http://www.theamazingpics.com/page/2/


We are all living longer
But we are retiring at the same age
We are experiencing more chronic disease
We have higher expectations
It costs more to treat each one of us each year
Significant differences in services between urban and remote

The world of 2030

Global Challenges

Trustworthy

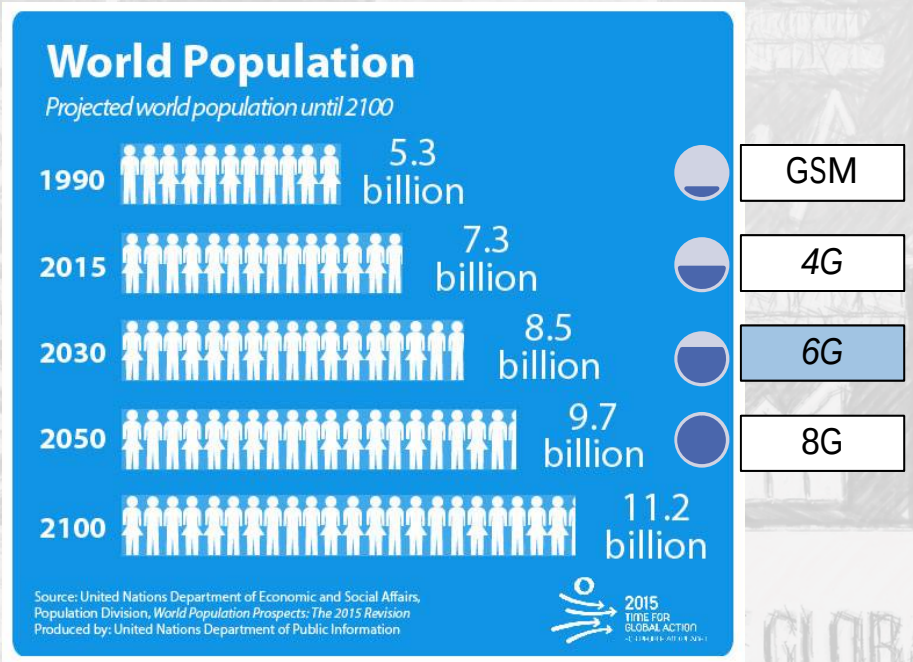
Green

Privacy Preserving

Guaranteed

Secure

Seamless



GSM 3G 4G 5G 6G 7G 8G

Some Considerations for use of AI

Artificially Intelligent Autonomous Devices

Quite Cool
Can augment humans
Can be more efficient
Uses in manufacturing, agriculture, hospitality

but what if

... the robot interacts directly with your personal networks
... local optimisation is based on robot “needs” not your needs
... services offered are based on what is learned from your data
... the robot harvests energy from your movement while engaging
... your data forms part of future robot knowledge base

When interacting with an AI

- Can you tell?
- Does it matter?
- How is it qualified?
- Who is responsible?
- Can you trust it?
- What is happening to my data?
- How informed is your consent?

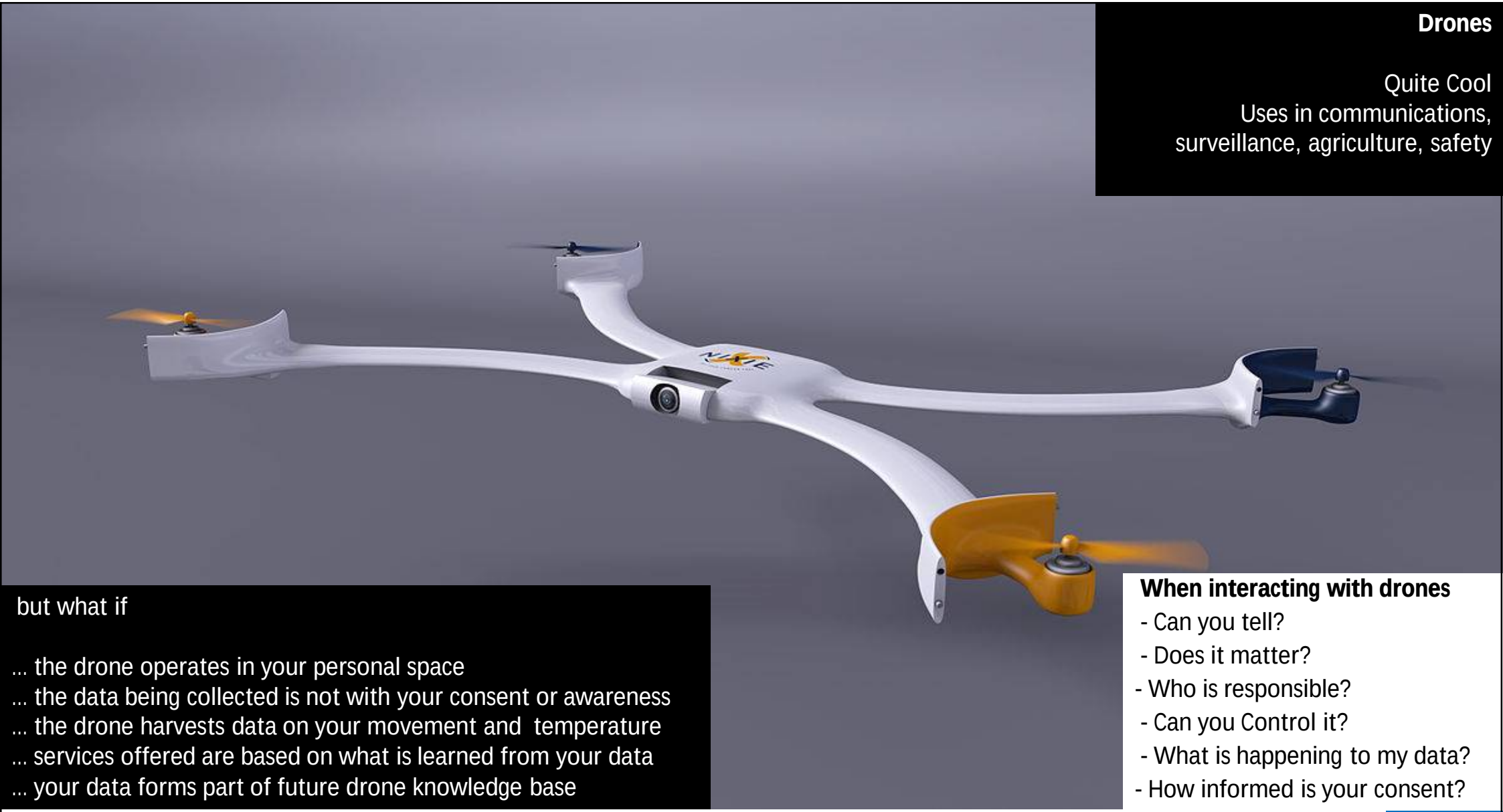
Shudu
Lil Miquela
Alexa
Siri

Image Source: Instragram

Can you tell?
Does it matter?
How is it qualified?
Who is responsible?
Can you trust it?
What is happening to my data?
How informed is your consent?

Drones

Quite Cool
Uses in communications, surveillance, agriculture, safety



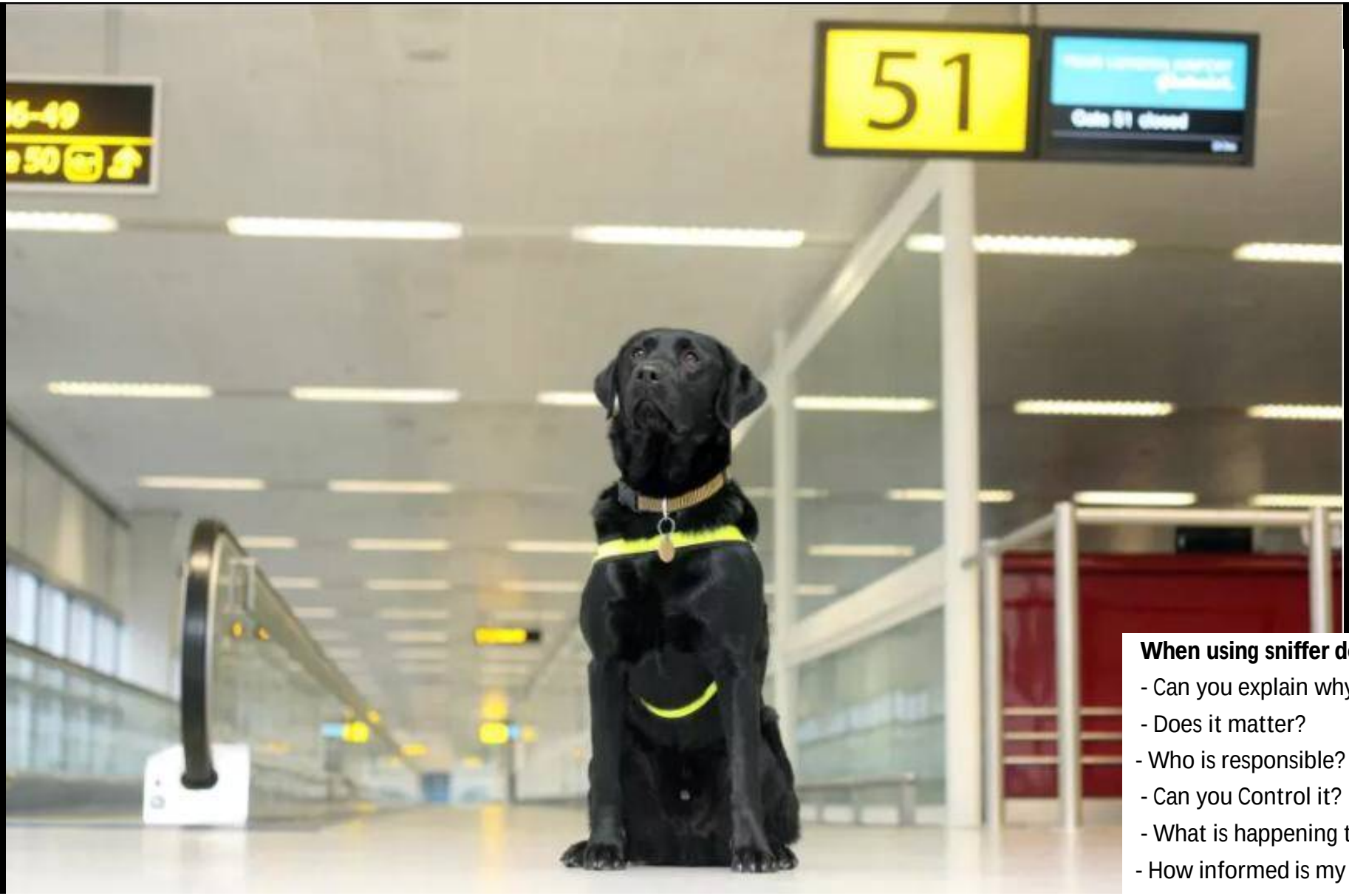
but what if

... the drone operates in your personal space
... the data being collected is not with your consent or awareness
... the drone harvests data on your movement and temperature
... services offered are based on what is learned from your data
... your data forms part of future drone knowledge base

When interacting with drones

- Can you tell?
- Does it matter?
- Who is responsible?
- Can you Control it?
- What is happening to my data?
- How informed is your consent?

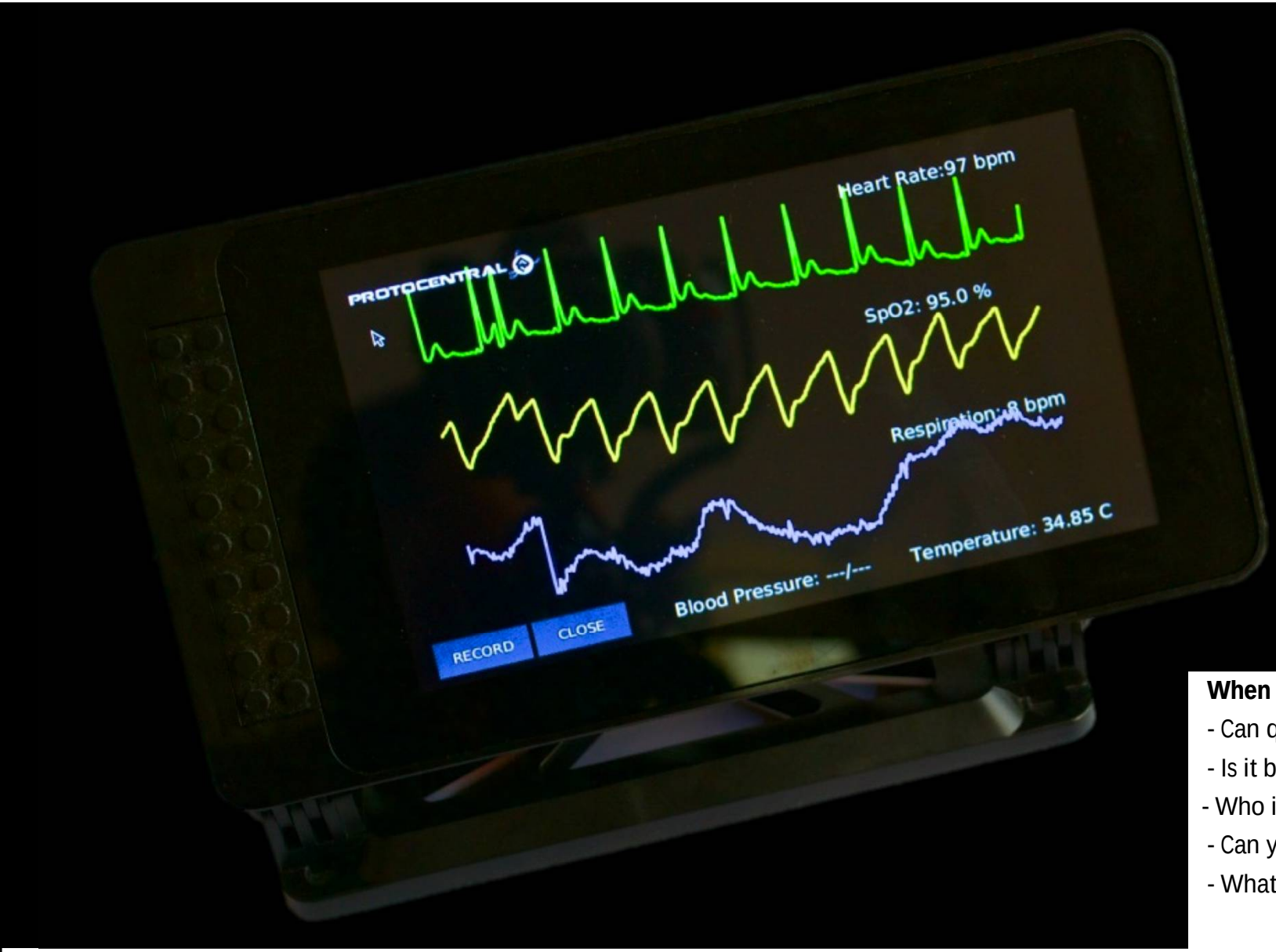
Explainability



When using sniffer dogs

- Can you explain why?
- Does it matter?
- Who is responsible?
- Can you Control it?
- What is happening to my data?
- How informed is my consent?

Replacing Existing Systems

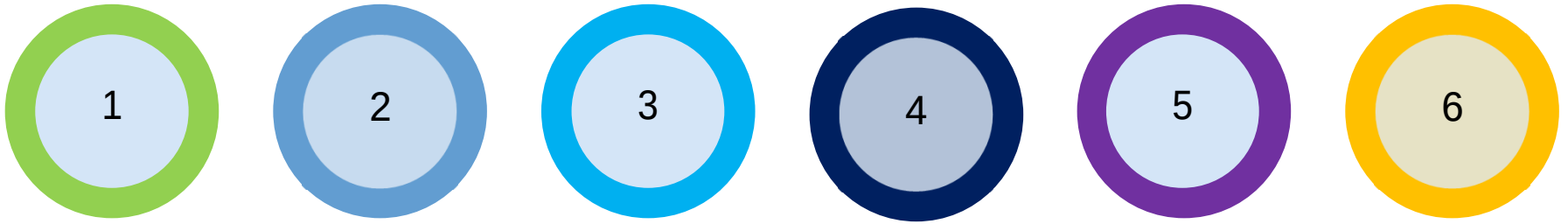


When comparing to existing systems

- Can quantify changes in risk?
- Is it better than the way we do things now?
- Who is responsible?
- Can you Control it?
- What is the risk of over (or under) reliance?

Considerations for Use of AI

The NSW AI Strategy is committed to Appropriate Use, Building Trust and Delivering on Outcomes



Appropriate Use | Co-Design | Data Considerations | Possible Harms and Appeals | Governance Considerations | Human Attention & Cognition

Building public trust: Delivering and showcasing positive outcomes for the community will contribute to public trust, and trust will be strengthened by openly acknowledging the strengths and weaknesses of AI and managing potential risks.

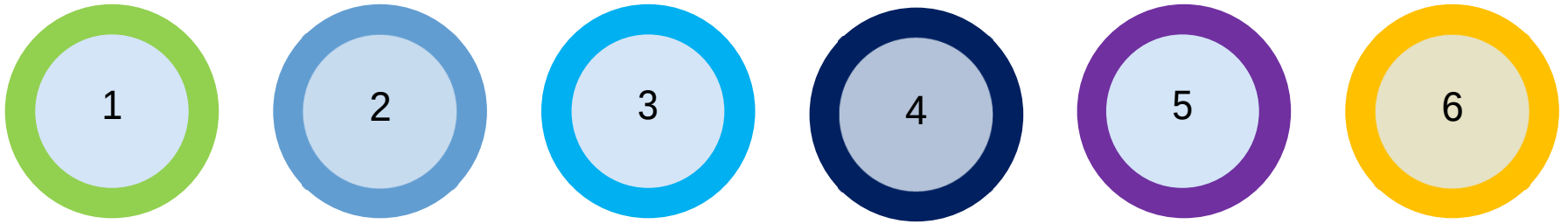
Digital uplift: Building public sector digital skills: There should be a broad program of digital capability uplift for the NSW public sector which also incorporates upskilling on emerging technologies.

Building data capability: NSW citizens must have confidence that data for AI projects is used safely and securely, and in a way that is consistent with privacy and data sharing frameworks, and community expectations.

Procurement: We must leverage opportunities to update procurement frameworks through innovation and proofs of concepts to take more timely advantage of emerging technologies.

Innovation and collaboration: We must explore working closer with industry and academia to drive better service delivery and solve some of our most complex problems.

Some dynamic tensions – not opposites but not always aligned



Appropriate Use | Co-Design | Data Considerations | Possible Harms and Appeals | Governance Considerations | Human Attention & Cognition

Complex Enough to be Useful — **Simple Enough to be Useable**

Identification of Risks — **Identification of Benefits Compared to Current Practices**

Explainable AI — **Powerful non-Explainable AI**

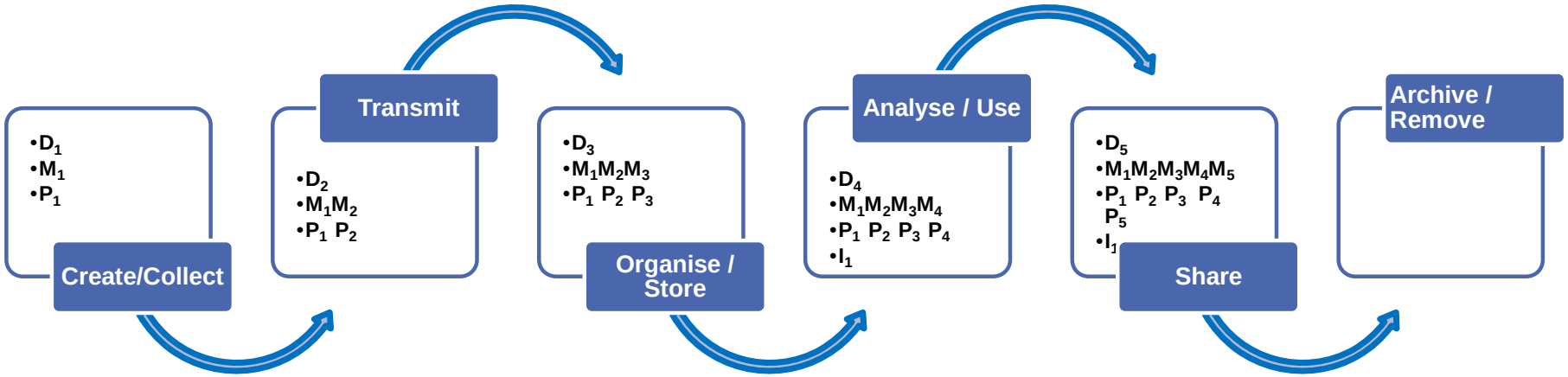
Assurance of "use of AI" — **Impinging on Authorising Environments**

Human in the Loop — **Over or Under Reliance on AI**

Ensuring Individual Responsibility is Clear — **Unduly burdening Users and Deterring AI Uptake**

Data Problem, AI Problem or Policy Problem?

Data Lens 1 - Simplified Data lifecycles



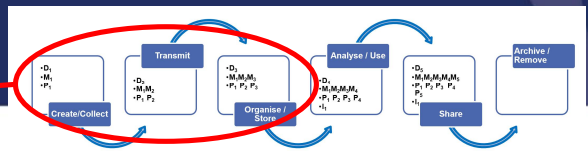
- Chain of governance
- Authorising framework
- Data provenance

- Data quality - fitness for purpose

- Guidance / prohibitions on use (and reuse)

Lens 2 - Cumulative Data Quality

Cumulative Impact on Data Quality



Create (Collect)

- Data completeness (e.g. spatial completeness, population completeness, system completeness)
- Timeliness
- Data precision (e.g. analogue-to-digital bits, precision of categories)
- Data accuracy (correct within precision)
- Data consistency (same phenomenon / feature leads to same data)
- Metadata re collection (e.g. time, date, location, temperature, collection device or method)

Transmit

- Impact on data accuracy (e.g. bit errors introduced)
- Impact on data precision (e.g. lossy compression)
- Impact on data completeness (e.g. down-sampling, averaging, interpolation)
- Impact on Timeliness (e.g. down-sampling, averaging, interpolation)
- Metadata re transmission (e.g. method, delay)

Store

- Impact on data accuracy (e.g. bit errors introduced)
- Impact on data precision (e.g. lossy compression)
- Impact on data completeness (e.g. down-sampling, averaging, interpolation)
- Impact on Timeliness (e.g. down-sampling, averaging, interpolation)
- Metadata re storage (e.g. method, data age)

Link / Analyse

- Impact on data completeness (e.g. down-sampling, averaging, interpolation)
- Impact on Timeliness (e.g. down-sampling, averaging, interpolation)
- Algorithm sensitivity assessment (e.g. to data precision, data accuracy, data completeness, Timeliness)
- Algorithm consistency assessment (same output from same input)
- Algorithm bias assessment (XXX)
- Metadata re analysis process (e.g. method, major factors affecting output)

Reuse (share)

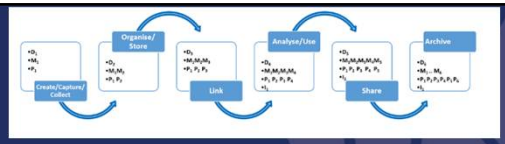
Archive

Delete

Data accuracy = minimum of (accuracy at collection, accuracy transmitted, accuracy stored, accuracy used for analysis)
Timeliness = minimum of (Timeliness at collection, Timeliness transmitted, Timeliness stored, Timeliness used for analysis)
Data completeness = minimum of (completeness at collection, completeness transmitted, completeness stored, completeness used for analysis)
Data consistency = data consistency at Create stage

Data quality includes four dimensions: accuracy, timeliness, completeness, and consistency. Data accuracy is defined as conformity between a recorded data value and the corresponding actual data value.

Data Lens 3
Control Required in Data Environment

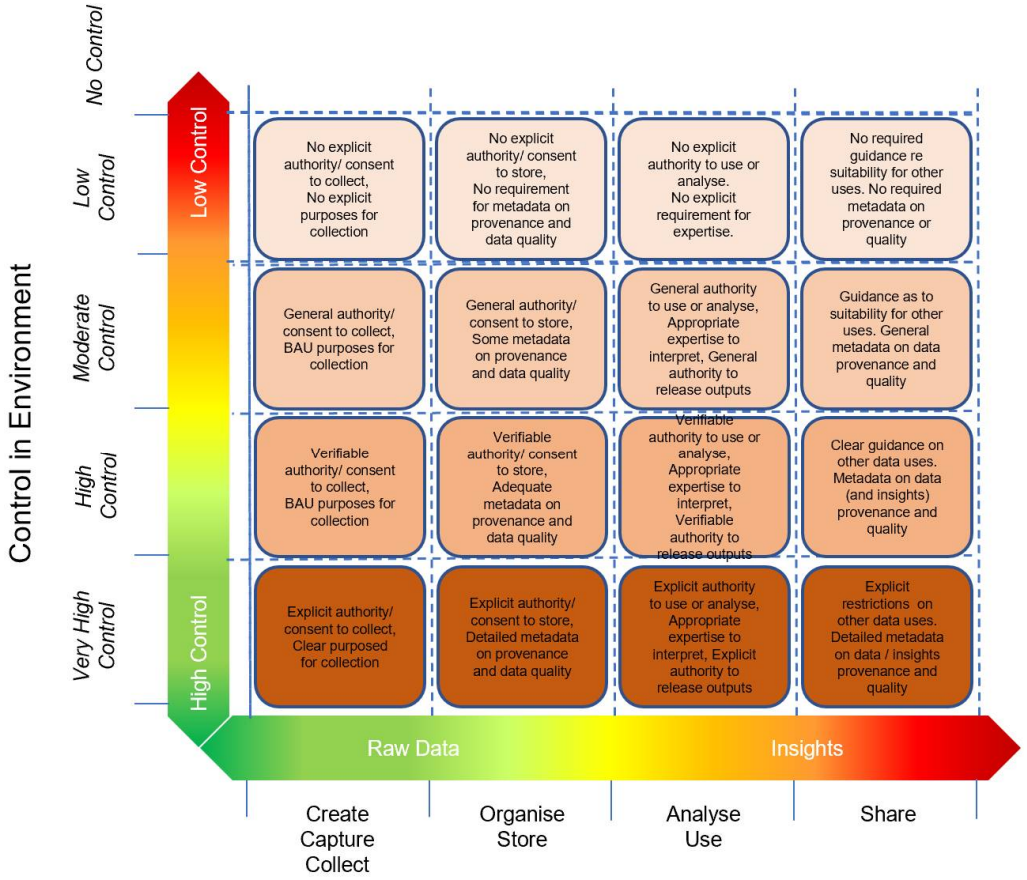


May have assumed authority to collect, use, and Use data. May have metadata on data provenance and quality. **Data** - low PIF.

Must have understanding of data quality and provenance, capable analysists and domain experts, adequate governance / security at each stage. **May have** broad authority to collect, use, and Use data. **Data -** moderately sensitive / moderate PIF.

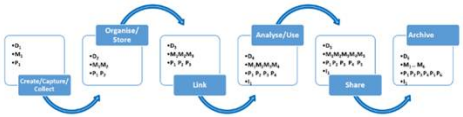
Must have understanding of data quality and provenance, highly skilled analysists and domain experts, strong governance / security at each stage. **May have** general authority to collect, use, and Use data. **Data -** high sensitivity / high PIF.

Must have explicit purpose and authority, high quality data and metadata, expert analysists and domain experts, strong governance / security at each stage. Explicit restrictions on secondary use of data and insights. **Data -** very high sensitivity and very high PIF



- Control = (proven) capability * (assessable) governance * (verifiable) purpose
- Capability includes skill in all stages of Data Lifecycle - data analysis, data provenance, governance, security
- High Control = skilled people working in strong governance environment with clearly authorised purpose
- No Control environment = no assessments or no restriction on people accessing or utilising data
- Requires an objective, repeatable, standardised assessment of
  - capability,
  - governance,
  - purpose,
  - data quality and provenance
  - sensitivity of data
  - degree of personal information contained in datasets

Working out which Trustworthiness / Control Environments



No Control environment

- Has data been approved for release as open data?
- Data is of sufficiently high quality for general use?

Answer No to any question

Low Control environment

- Data is ***not*** reasonably personally identifiable?
- Data does ***not*** contain sensitive subject matter?
- Data is of sufficiently high quality for general use?

Answer No to any question

Moderate Control environment

- Data is ***not*** reasonably personally identifiable?
- Data contains ***only*** moderately sensitive subject matter
- Data is of high quality for intended uses?

Answer No to any question

High Control environment

- Data is ***not*** reasonably personally identifiable?
- Data quality is well quantified?

Answer No to any question

Very high Control environment

- Data can only be accessed under an external instrument such as a PID,
- Data ***is*** reasonably personally identifiable
- Data may contain sensitive subject matter
- Data quality is well quantified

Data Problem, AI Problem or Policy Problem?

Lens 1 - Operational vs non-operational AI

Operational AI

Operational AI systems are those that have a real-world effect. The purpose is to generate an action, either prompting a human to act, or the system acting by itself. Operational AI systems often work in real time (or near real time) using a live environment for their source data.

Not all operational AI systems are high risk. An example of lower risk operational AI is the digital information boards that show the time of arrival of the next bus.

Operational AI that uses real-time data to recommend or make a decision that adversely impacts a human will likely be considered High or Very high risk.

Start assessment questions with

Community benefit

AI should deliver the best outcome for the citizen, and key insights into decision-making.

Non-operational AI

Non-operational AI systems do not use a live environment for their source data. Most frequently, they produce analysis and insight from historical data.

Non-operational AI typically represents a lower level of risk. However, the risk level needs to be carefully and consciously determined, especially where there is a possibility that AI insights and outputs may be used to influence important future policy positions.

Start assessment questions with

Fairness

Use of AI will include safeguards to manage data bias or data quality risks, following best practice and Australian Standards

Lens 2 - AI risk factors exist on a spectrum

The key factor that determines risk is how the AI system is used, including whether it is operational or non-operational.

AI generates insights for **non-operational** human use from non-sensitive data

(example: analytics package reporting on historical non-sensitive data)

AI generates insights or alerts for **operational** human use with minimal potential for harm

(example: anomaly detection software; alarm system)

--

AI makes **operational** actions, decisions or recommendations with no routine human oversight with minimal potential for harm

(example: automated door; biometric login with alternative login methods; automated phone menu, smart sign showing driver speed)

AI generates **operational** insights / decisions / recommendations for human to action with some potential for harm

(example: public facing chatbot; red light camera, intruder alert system)

--

AI generates insights for **non-operational** human use analysing sensitive data

(example: analytics package operating on data of vulnerable individuals)

AI makes and implements **operational** decisions autonomously of human input in the interests of human safety and wellbeing

(example: anti-lock braking system)

--

AI makes and implements **operational** decisions within a specified range, and refers exceptions for human to review and action

(example: loan application system, autonomous tram)

AI makes and implements **operational** decisions that can negatively affect human wellbeing autonomously of human input

(example: autonomous benefits eligibility reviews; judicial custodial sentence recommendations, unconstrained autonomous system, self driving car)

Lens 3 - Risk factors for individuals or communities

Consider the risks of…	None, negligible, N/A	Reversible with negligible consequences	Reversible with moderate consequences	Reversible with significant consequences	Significant or irreversible
Physical harms	○	○	○	○	○
Psychological harms	○	○	○	○	○
Environmental harms or harms to the broader community	○	○	○	○	○
Unauthorised use of health or sensitive personal information (SIP)	○	○	○	○	○
Impact on right, privilege or entitlement	○	○	○	○	○
Unintended identification or misidentification of an individual	○	○	○	○	○
Misapplication of a fine or penalty	○	○	○	○	○
Other financial or commercial impact	○	○	○	○	○
Incorrect advice or guidance	○	○	○	○	○
Inconvenience or delay	○	○	○	○	○
Other harms	○	○	○	○	○
	Very low risk or N/A	**Low**	**Midrange**	**High**	**Very high risk**

Lens 4 - Fairness risk factors for AI projects

Consider the risks associated with…	Very low risk or N/A	Low	Midrange	High	Very high risk
Using incomplete or inaccurate data	○	○	○	○	○
Having poorly defined descriptions and indicators of “Fairness”	○	○	○	○	○
Not ensuring ongoing monitoring of “Fairness indicators”	○	○	○	○	○
Decisions to exclude outlier data	○	○	○	○	○
Informal or inconsistent data cleansing and repair protocols and processes	○	○	○	○	○
Using informal bias detection methods (best practice includes automated testing)	○	○	○	○	○
The likelihood that re-running scenarios could produce different results (reproducibility)	○	○	○	○	○
Inadvertently creating new associations when linking data and/or metadata	○	○	○	○	○
Differences in the data used for training compared to the data for intended use	○	○	○	○	○

And the Result?

Artificial intelligence assurance framework

As described by the NSW Government AI Strategy, AI (Artificial Intelligence) is intelligent technology, programs and the use of advanced computing algorithms that can augment decision making by identifying meaningful patterns in data.

The Framework is intended to be used for custom AI systems, customisable AI systems, and for projects developed using generic AI platforms.

Apply the framework before you use or deploy your AI system. All AI systems should be piloted before being scaled.

NSW GOVERNMENT

https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fedit.digital.nsw.gov.au%2Fsites%2Fdefault%2Ffiles%2F2021-12%2F22112021%2520-%2520%2520NSW%2520Government%2520Assurance%2520Framework.pptx&wdOrigin=BROWSELINK

ISO/IEC/JTC1 SC42 - Developing Standards for AI

SC 42 is developing an **AI Management System as a pathway to certification**, leveraging the work that has been conducted under all the working groups.

5 standards are now published, and 21 standards and projects are under development.

Including observers, currently 47 countries involved.

SC 42 work is complemented by

Decision to Share PAS 183 (SC40) ISO 37156	Frameworks for Data Sharing Agreements SC38 23751	Appropriate Use of Analytics
Data Quality		
Terminology – Use Cases		

Identified Standards Activities (and Gaps)

- Frameworks for Data Sharing Agreements - SC38 23751
- Decision to Share - PAS 183 (SC40) ISO 37156
- Data Quality - TC 184 (ISO 8000)
- Appropriate Use of Analytics
- Terminology – use cases
- Utility of Metadata (SC32)

Shorthand Legend :

SC27 (Cyber Security)
SC 32 (SQL)
SC 38 (Cloud Computing)
SC40 (Information Management)
SC42 (AI)

Standards Needed

Dr. Ian Oppermann
Chief Data Scientist, Department of Customer Service
UTS Industry Professor

e ian.oppermann@customerservice.nsw.gov.au |
2 – 24 Rawson Place, Sydney NSW 2000