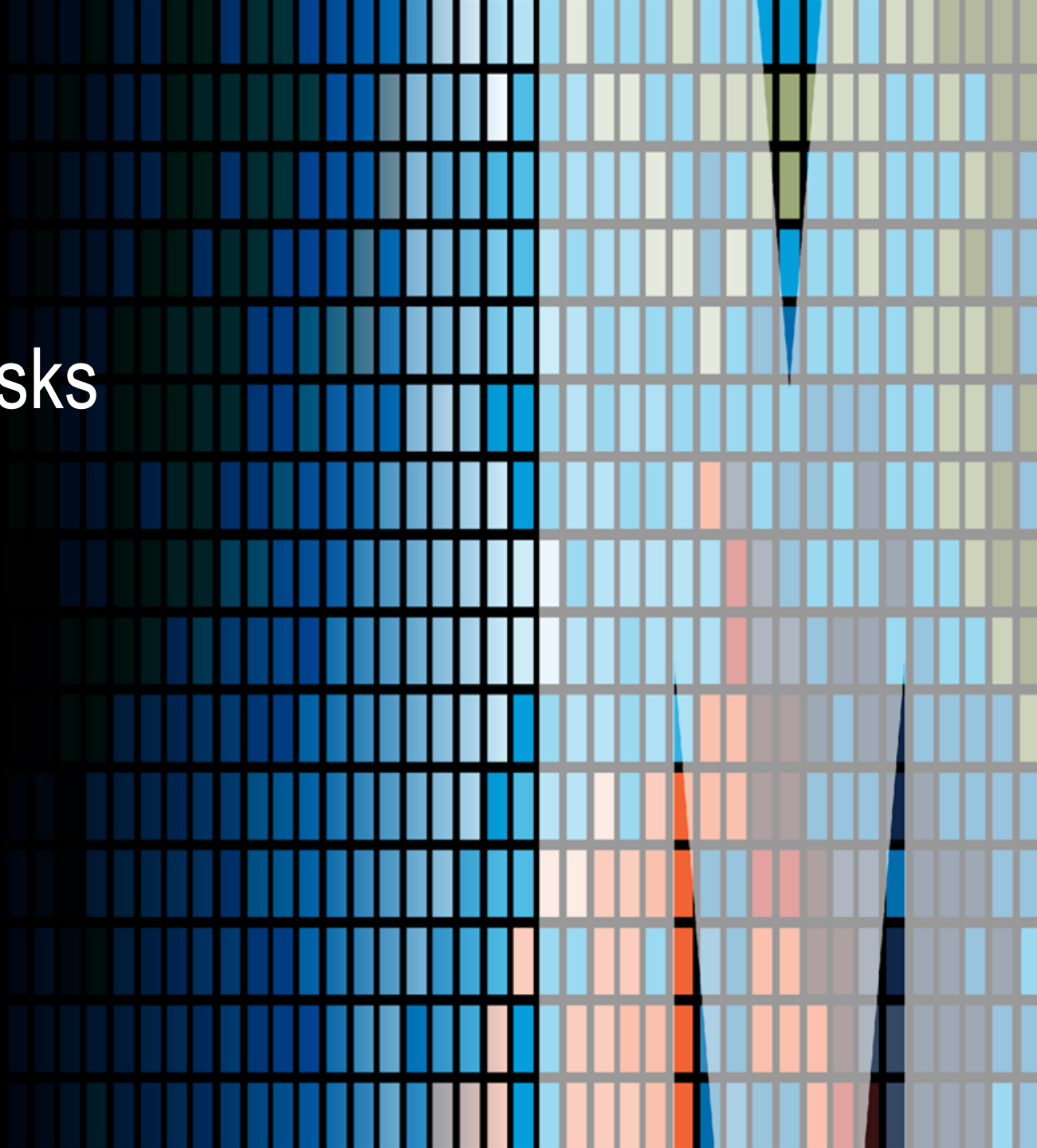
MONASH University

The Changing Face of Data Risks for Research in the Age of AI

Research Data Strategy
Office of PVC-RI

October 2025

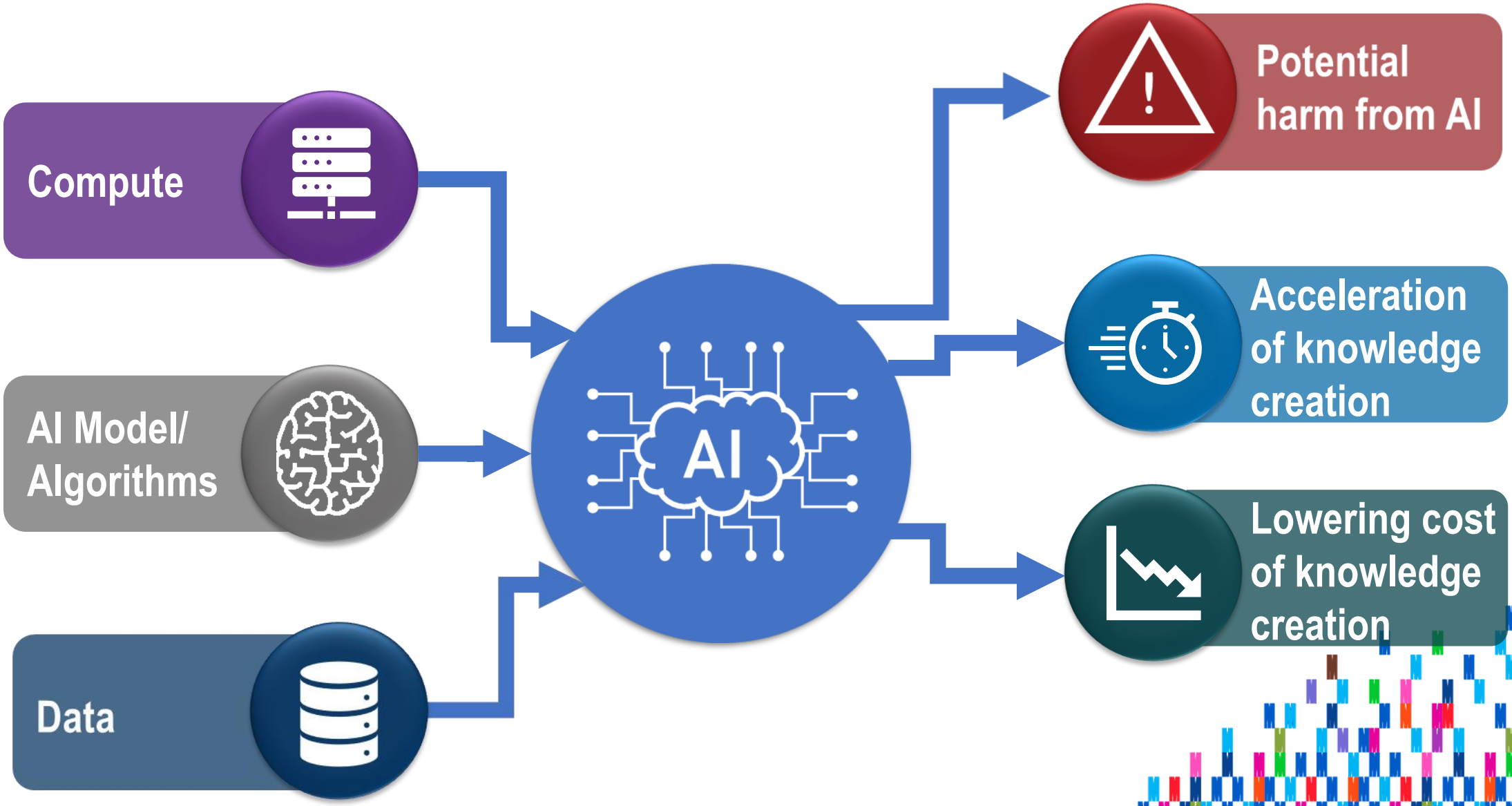
Dianne Brown
Komathy Padmanbhan



MONASH University

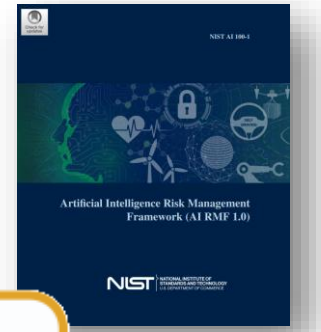
MONASH UNIVERSITY recognises that its Australian campuses are located on the unceded lands of the people of the Kulin nations, and pays its respects to their Elders, past and present.

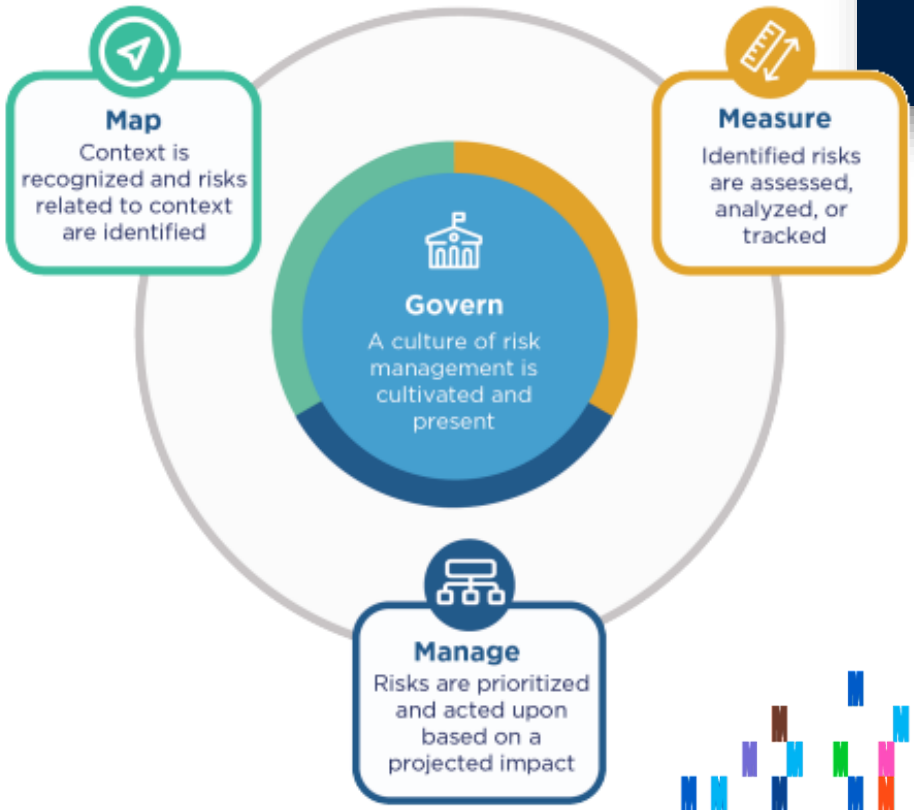
AI is reshaping the world and research

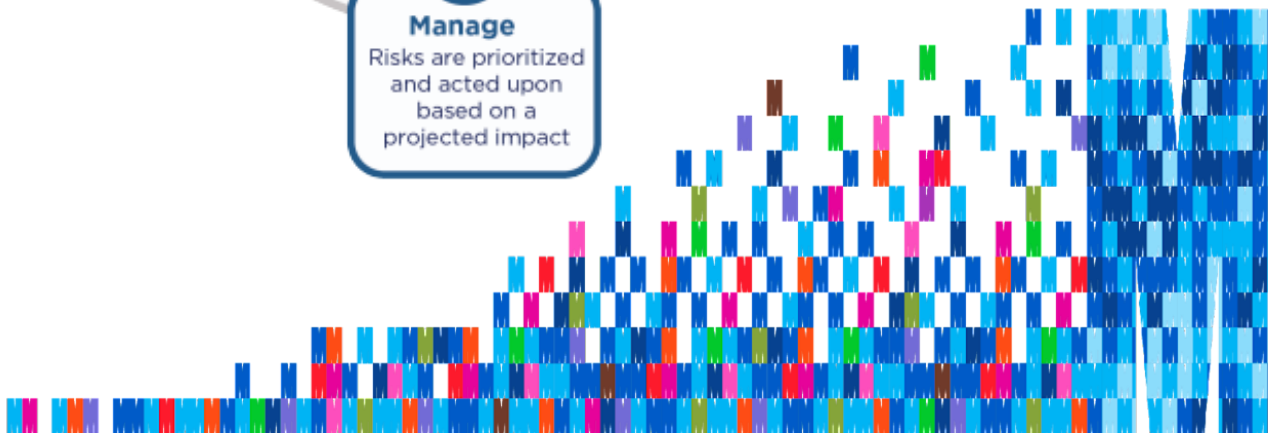


Governing requires mapping, measuring and managing the risk

NIST AI Risk Management Framework Core







Source: United States Government. Artificial Intelligence Risk Management Framework (AI RMF 1.0). In: National Institute of Standards and Technology (NIST), editor. U.S. Department of Commerce,2023.

Data's Role

Training, Fine-tuning and Optimisation Data

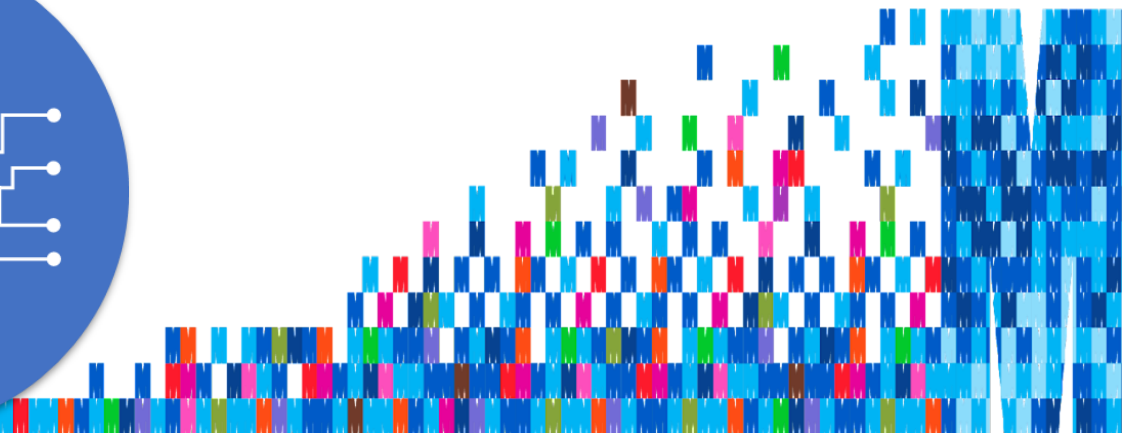
Input into the AI model for analysis

Output from the AI model

Data Fuels AI

AI

Potential harm from AI



DATA GOVERNANCE…

MONASH University

Governance Mechanisms

1. Acknowledges that data governance must sit across an organisation…

2. Views data as a strategic asset of the organisation…

3. Expects it will indicate who gets to set the rules…

4. Expects it will indicate what the rules are…

5. Expects it will indicate how the rules are articulated…

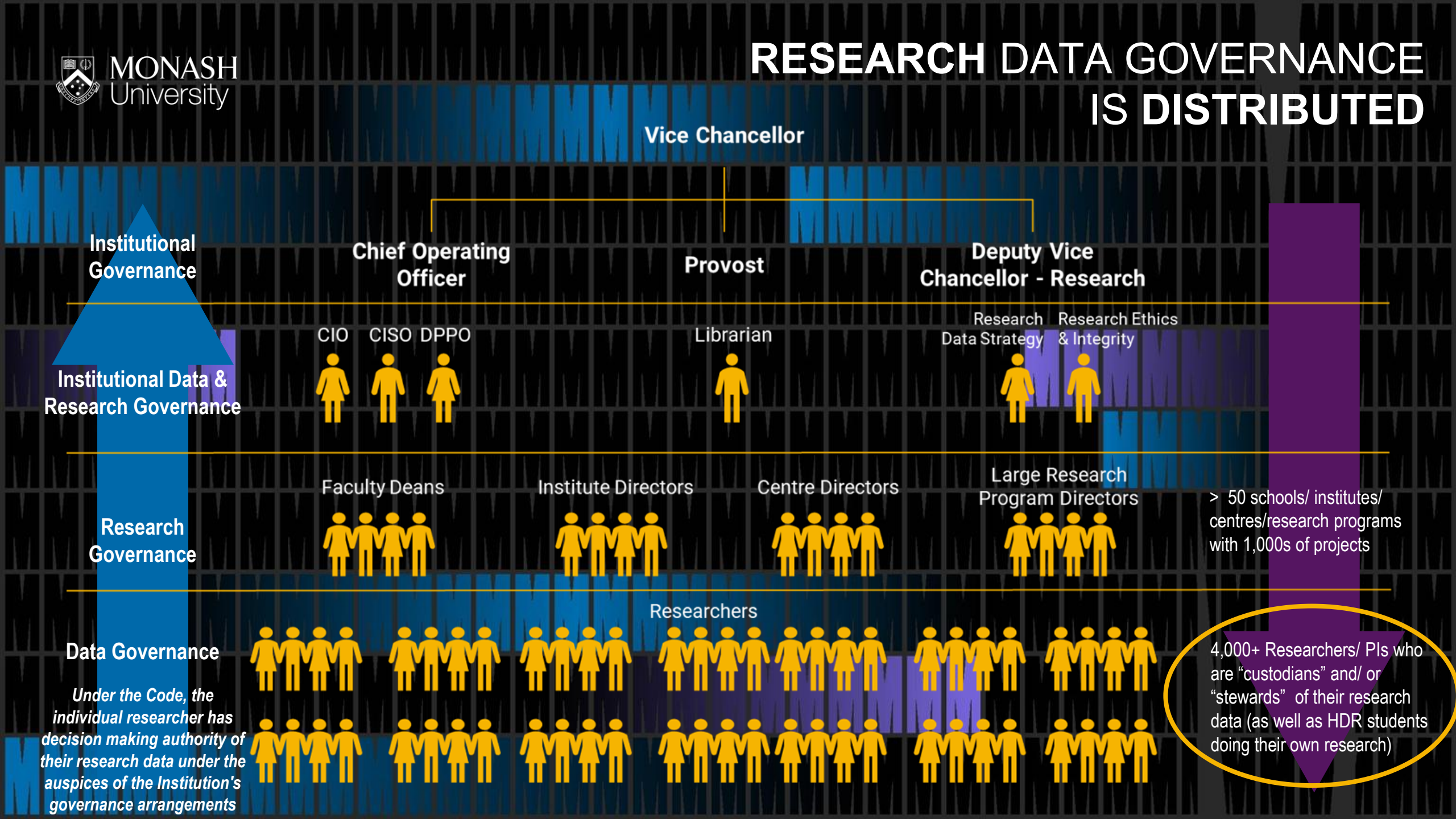
6. Expects it will indicate how the rules will be monitored…

*Data governance specifies a **cross-functional** framework for ma data as a **strategic enter** In doing so, d specifies **decision ri ities** for an organizat bout its data. Furtherm **malizes data po dures** and **monito***

Source: Abraham, Rene, Schneider, Johannes, & vom Brocke, Jan. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

MONASH University

RESEARCH DATA GOVERNANCE IS **DISTRIBUTED**



Multiple Ways to Frame AI Risk

MONASH University

Who it impacts

- harm to individuals;
- harm to groups/ communities;
- harm to societies

Australian Government. Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings. Department of Industry Science and Resources,. Canberra 2024. Pg12

Cause

- by decision/ action of a human or an AI system itself
- intentional or unintentional
- pre-deployment or post-deployment

Impact domain

- discrimination & toxicity;
- privacy & security;
- misinformation;
- malicious actors & misuse;
- human-computer interaction;
- socioeconomic & environmental harms;
- AI system safety, failures and limitations

Slattery P, Saeri AK, Grundy EAC, Graham J, Noetel M, Uuk R, et al. The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence2024 August 01, 2024:[arXiv:2408.12622 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2024arXiv240812622S.

Org'n Specific

- affects accessibility & inclusivity;
- unfair discrimination;
- perpetuates stereotypes or demeans;
- causes harm;
- compromises privacy;
- causes concerns about security of data/ system;
- influences decision-making that causes harm
- poses reputational risk/ undermines public confidence

Australian Government. Pilot AI assurance framework guidance: Attachment Risk consequence rating advice: Digital.gov.au; 2024 [Available from: https://www.digital.gov.au/policy/ai/pilot-ai-assurance-framework/guidance/attachment]

Public Concern

- Spreads fake and harmful content
- Cyber attacks against govt/ org/ people
- Loss of data privacy
- AI doing jobs of humans
- AI-assisted surveillance that violates privacy & liberty
- Inaccurate decision making
- Bias and discrimination in decision making
- Failure of AI embedded in critical infrastructure
- AI being unsafe, untrustworthy, unaligned with human values

Saeri A, Noetel M, Graham J. Survey Assessing Risks from Artificial Intelligence (Technical Report) (March 7, 2024). Available at SSRN: https://ssrn.com/abstract=4750953. University of Queensland; 2024

Frame-work

- CBRN Information or Capabilities
- Confabulation.
- Dangerous, Violent, or Hateful Content
- Data Privacy.
- Environmental Impacts
- Harmful Bias or Homogenization
- Human-AI Configuration
- Information Integrity
- Information Security
- Intellectual Property
- Obscene, Degrading, and/or Abusive Content
- Value Chain and Component Integration

United States government. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. In: National Institute of Standards and Technology (NIST), editor.: U.S. Department of Commerce; 2024.

12 NIST GenAI Risks

MONASH University

CBRN Information or Capabilities

Eased access to or synthesis of materially nefarious information or design capabilities related to chemical, biological, radiological, or nuclear (CBRN) weapons or other dangerous materials or agents.

Confabulation

The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications") by which users may be misled or deceived.

Dangerous, Violent or Hateful Content

Eased production of and access to violent, inciting, radicalizing, or threatening content as well as recommendations to carry out self-harm or conduct illegal activities. Includes difficulty controlling public exposure to hateful and disparaging or stereotyping content.

Data Privacy

Impacts due to leakage and unauthorized use, disclosure, or de-anonymization of biometric, health, location, or other personally identifiable information or sensitive data.

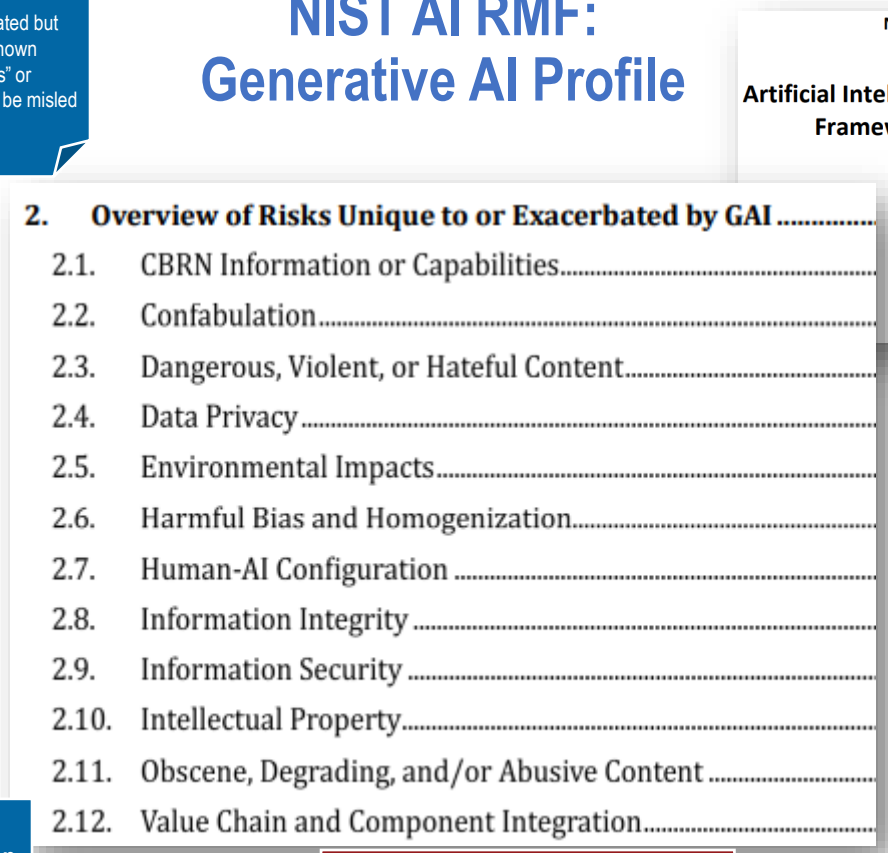
Environmental Impacts

Impacts due to high compute resource utilisation in training or operating AI models, and related outcomes that may adversely impact ecosystems

Harmful Bias or Homogenization

Amplification and exacerbation of historical, societal, and systemic biases

NIST AI RMF: Generative AI Profile



NIST Trustworthy and Responsible AI
NIST AI 600-1

Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

This publication is available free of charge from: https://doi.org/10.6028/NIST.AI.600-1

July 2024

2. Overview of Risks Unique to or Exacerbated by GAI .....
- 2.1. CBRN Information or Capabilities.....
- 2.2. Confabulation.....
- 2.3. Dangerous, Violent, or Hateful Content.....
- 2.4. Data Privacy.....
- 2.5. Environmental Impacts.....
- 2.6. Harmful Bias and Homogenization.....
- 2.7. Human-AI Configuration .....
- 2.8. Information Integrity .....
- 2.9. Information Security .....
- 2.10. Intellectual Property.....
- 2.11. Obscene, Degrading, and/or Abusive Content .....
- 2.12. Value Chain and Component Integration.....

Human-AI Configuration

Interactions between a human and an AI system which can result in the human anthropomorphizing GAI systems or experiencing algorithmic aversion, automation bias, over-reliance, or emotional entanglement

Information Integrity

Lowered barrier to entry to generate and support the exchange and consumption of content which may not distinguish fact from opinion or fiction or acknowledge uncertainties. Loss of data providence

Information Security

Lowered barriers for offensive cyber capabilities; increased attack surface for targeted cyberattacks

Intellectual Property

Eased production or replication of alleged copyrighted, trademarked, or licensed content without authorization (possibly in situations which do not fall under fair use); eased exposure of trade secrets; or plagiarism or illegal replication.

Obscene, Degrading, and/or Abusive Content

Eased production of and access to obscene, degrading, and/or abusive imagery which can cause harm, including synthetic child sexual abuse material (CSAM), and nonconsensual intimate images (NCII) of adults

Value Chain/ Component Integration

Non-transparent or untraceable integration of upstream components, including data that has been improperly obtained or not processed or other issues that diminish transparency or accountability for downstream users.

Source: United States Government. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. In: National Institute of Standards and Technology (NIST), editor: U.S. Department of Commerce: 2024.

Australian AI Ethical Principles

Human, societal and environmental wellbeing: AI systems should benefit individuals, society and the environment

Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals.

Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.

Privacy protection and security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data

Reliability and safety: AI systems should reliably operate in accordance with their intended purpose.

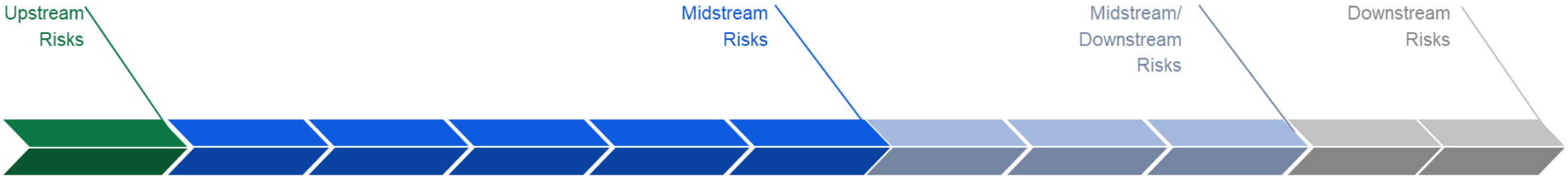
Transparency and explainability: There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, & can find out when an AI system is engaging with them

Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

Accountability: People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Australian Government. Australia's Artificial Intelligence Ethics Principle: Department of Industry Science and Resources,; 2019 [updated 11 October 2024. Available from: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles.

Examples of Temporal Nature of AI Use Risk



Value Chain/ Component Integration

Non-transparent or untraceable integration of upstream components, including data that has been improperly obtained or not processed or other issues that diminish transparency or accountability for downstream users.

Data Privacy

Impacts due to leakage and unauthorized use, disclosure, or de-anonymization of biometric, health, location, or other personally identifiable information or sensitive data.

Harmful Bias or Homogenization

Amplification and exacerbation of historical, societal, and systemic biases

Confabulation

The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications") by which users may be misled or deceived.

Intellectual Property

Eased production or replication of alleged copyrighted, trademarked, or licensed content without authorization (possibly in situations which do not fall under fair use); eased exposure of trade secrets; or plagiarism or illegal replication.

Information Security

Lowered barriers for offensive cyber capabilities; increased attack surface for targeted cyberattacks

Information Integrity

Lowered barrier to entry to generate and support the exchange and consumption of content which may not distinguish fact from opinion or fiction or acknowledge uncertainties. Loss of data providence

Human-AI Configuration

Interactions between a human and an AI system which can result in the human anthropomorphizing GAI systems or experiencing algorithmic aversion, automation bias, over-reliance, or emotional entanglement

Environmental Impacts

Impacts due to high compute resource utilisation in training or operating AI models, and related outcomes that may adversely impact ecosystems

Obscene, Dangerous or Abusive Content

Eased production of violent, inciting, radicalizing, threatening, obscene, degrading, and/or abusive imagery or content

CBRN Info or Capabilities

Eased access to or synthesis of materially nefarious information or design capabilities related to chemical, biological, radiological, or nuclear (CBRN) weapons or other dangerous materials or agents.

NB: These are a summary of the risks (with the addition of "research ethics" and "contractual") as described by United States government. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). In: National Institute of Standards and Technology (NIST), editor.: U.S. Department of Commerce; 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

Adding in research specific risks…

2.2 Protection of patient information

2.2.1 Respect the patient's right to know what information is held about them, their right to access their medical records and their right to have control over its use and disclosure, with limited exceptions.

2.2.2 Maintain the confidentiality of the patient's personal information including their medical records, disclosing their information to others only with the patient's express up-to-date consent or as required or authorised by law. This applies to both identified and de-identified patient data.

Australian Medical Association. Code of Ethics 2004. Editorially Revised 2006. Revised 2016. 17 March 2017.

Upstream Risks

Midstream/ Downstream Risks

Downstream Risks

Value Chain/ Component Integration

Non-transparent or untraceable integration of upstream components, including data that has been improperly obtained or not processed or other issues that diminish transparency or accountability for downstream users.

Research Ethics or Data Privacy

Participants have not provided their consent and/ or were not informed that their data would be analysed or used to train AI

Harmful Bias or Homogenization

Amplification and exacerbation of historical, societal, and systemic biases

Confabulation

The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications") by which users may be misled or deceived.

Intellectual Property

Eased production or replication of alleged copyrighted, trademarked, or licensed content without authorization (possibly in situations which do not fall under fair use); eased exposure of trade secrets; or plagiarism or illegal replication.

Information Security

Lowered barriers for offensive cyber capabilities; increased attack surface for targeted cyberattacks

Contractual

AI is used in a manner that breaches the contracts entered by the University for their use

Information Integrity

Lowered barrier to entry to generate and support the exchange and consumption of content which may not distinguish fact from opinion or fiction or acknowledge uncertainties. Loss of data providence

Human-AI Configuration

Interactions between a human and an AI system which can result in the human anthropomorphizing GAI systems or experiencing algorithmic aversion, automation bias, over-reliance, or emotional entanglement

Environmental Impacts

Impacts due to high compute resource utilisation in training or operating AI models, and related outcomes that may adversely impact ecosystems

Obscene, Dangerous or Abusive Content

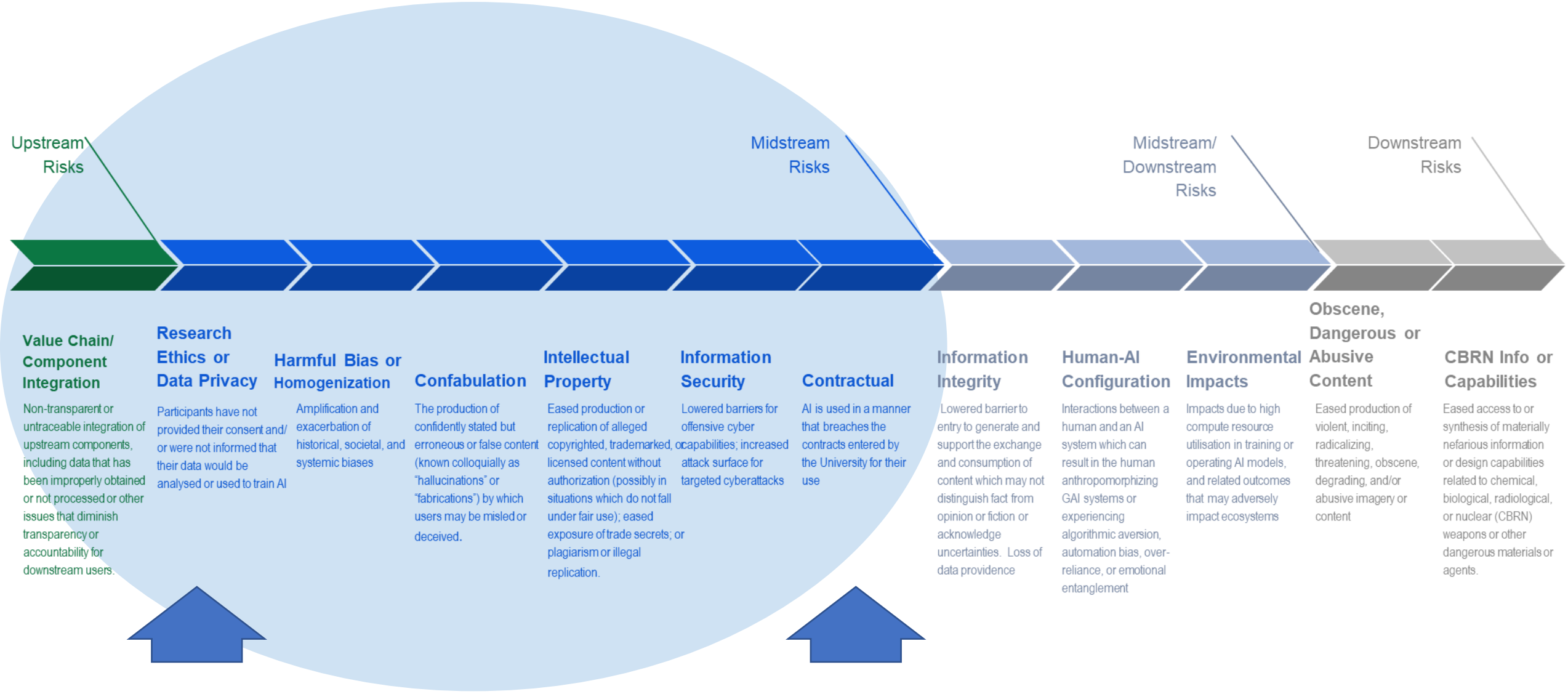
Eased production of violent, inciting, radicalizing, threatening, obscene, degrading, and/or abusive imagery or content

CBRN Info or Capabilities

Eased access to or synthesis of materially nefarious information or design capabilities related to chemical, biological, radiological, or nuclear (CBRN) weapons or other dangerous materials or agents.

NB: These are a summary of the risks (with the addition of "research ethics" and "contractual") as described by United States government. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). In: National Institute of Standards and Technology (NIST), editor.: U.S. Department of Commerce; 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

….then focussing on the key risks



NB: These are a summary of the risks (with the addition of "research ethics" and "contractual") as described by United States government. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). In: National Institute of Standards and Technology (NIST), editor.: U.S. Department of Commerce; 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

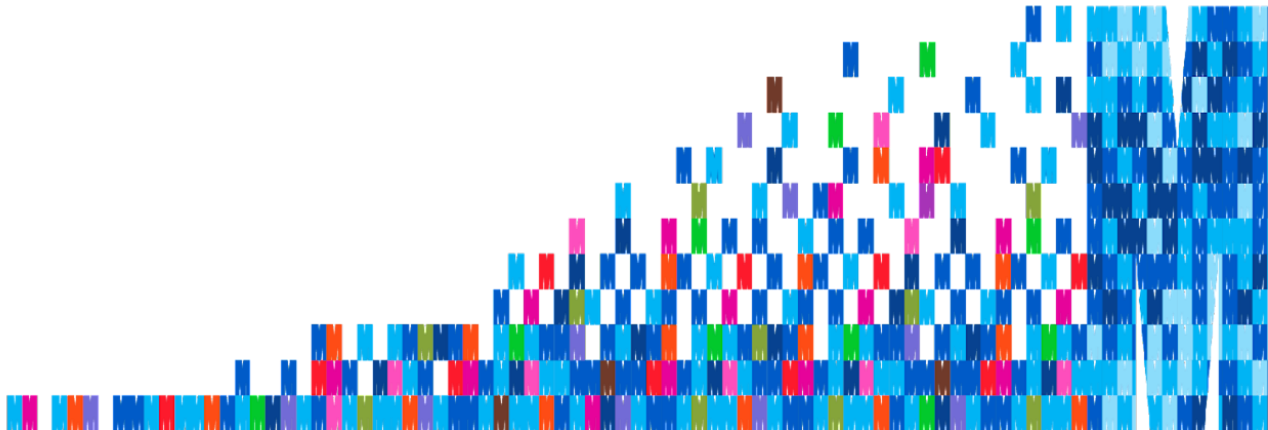
Risks identification is contingent on four key questions

Will you be a developer, deployer or end-user of AI

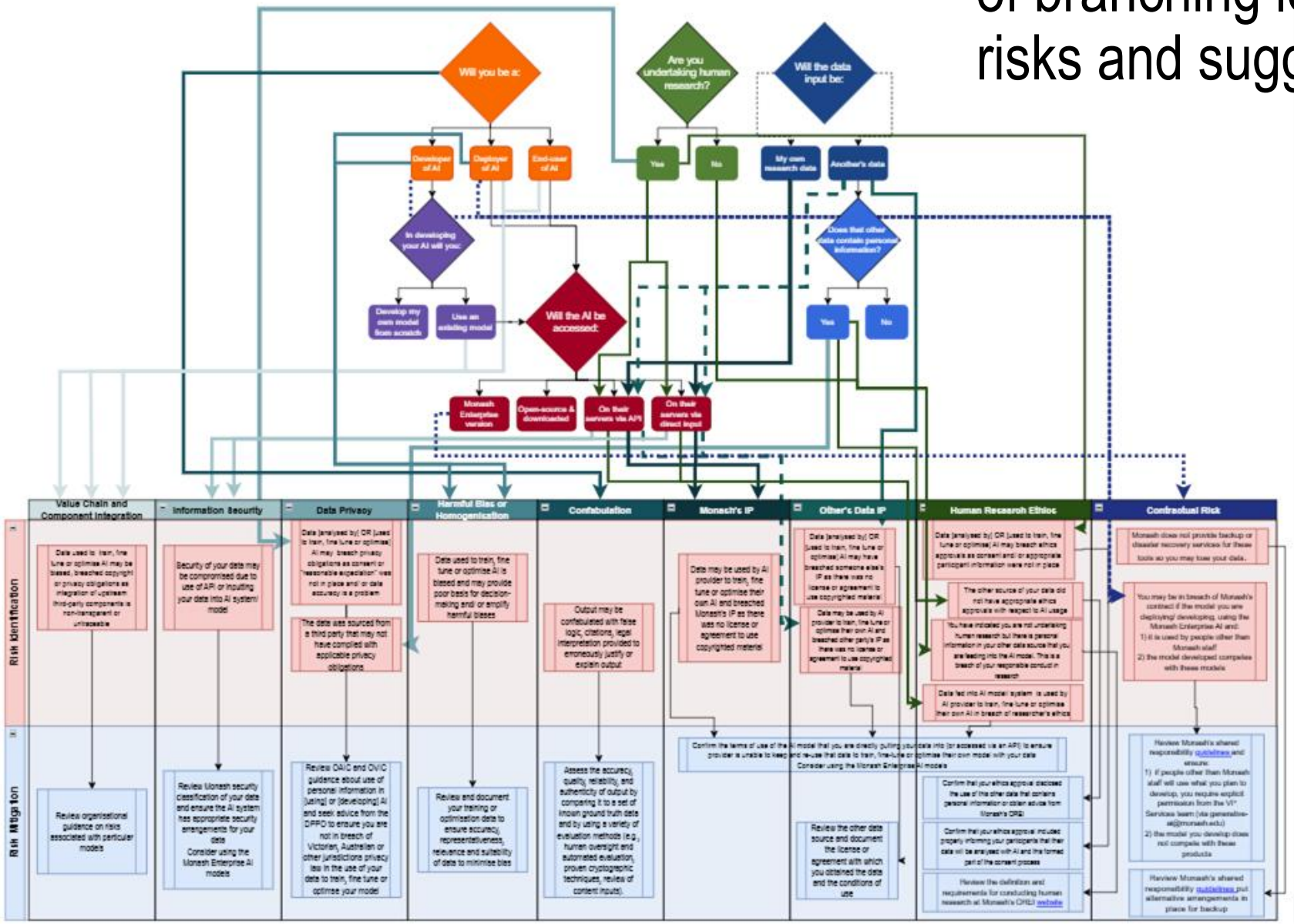
Are you undertaking human research?

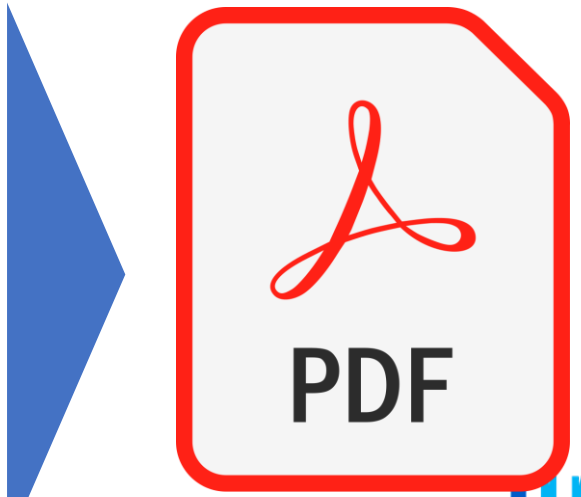
Will the data input be your own data or another's?

Will the AI be accessed via Enterprise Agreemt, API/ direct entry, on Uni Servers



This allowed us to create a set of branching logic to identify risks and suggested mitigations





MONASH University

AAA

Monash AI Use & Research Data Risk Tool

HUMAN RESEARCH ETHICS RISK

Risk Identification	Using AI to analyse participant data breaches ethics approvals as consent and/ or appropriate participant information was not in place
Risk Mitigation	Confirm that your ethics approval included properly informing your participants that their data will be analysed with AI and this formed part of the consent process

UPSTREAM/ VALUE CHAIN RISK

Risk Identification	Data used to train, fine tune or optimise AI model you are using may have breached ethics, may be biased, breached copyright or privacy obligations as there is non-transparent or untraceable integration of upstream third-party components in the model
Risk Mitigation	Review organisational guidance on ethic's risks associated with particular models

PRIVACY RISK

Risk Identification	Data you are analysing with AI breaches privacy obligations as consent or "reasonable expectation" was not in place and/ or data accuracy is a problem
Risk Mitigation	Review Office of Australian Information Commissioner and Office of Victorian Information Commissioner guidance about use of personal information in using AI and seek advice from the DPPO to ensure you are not in breach of Victorian, Australian or other jurisdictions privacy law

CONFABULATION RISK

Risk Identification	Output is confabulated with false logic, citations, legal interpretation provided to erroneously justify or explain output
Risk Mitigation	Assess the accuracy, quality, reliability, and authenticity of output by comparing it to a set of known ground truth data and by using a variety of evaluation methods (e.g., human oversight and automated evaluation, proven cryptographic techniques, review of content inputs).

Do you plan to use artificial intelligence (AI) with your research data?

This is includes all types of Generative AI, General-purpose AI, Agentic AI, AI models or systems.

*This does **NOT** include using AI to formulate your funding applications, protocols or manuscripts.*

How do you intend to use Artificial Intelligence (AI):

*Will you just use an AI model/ service offered by a third party without further training, fine tuning or optimisation of that AI model **(end-user)**?*

*Will you use an AI model/ system that you will optimise/ fine with your research data (**deployer**)?*

*Will you design, build, train, adopt, or combine AI models and applications (**developer**)?*

Yes

No

reset

My own research data

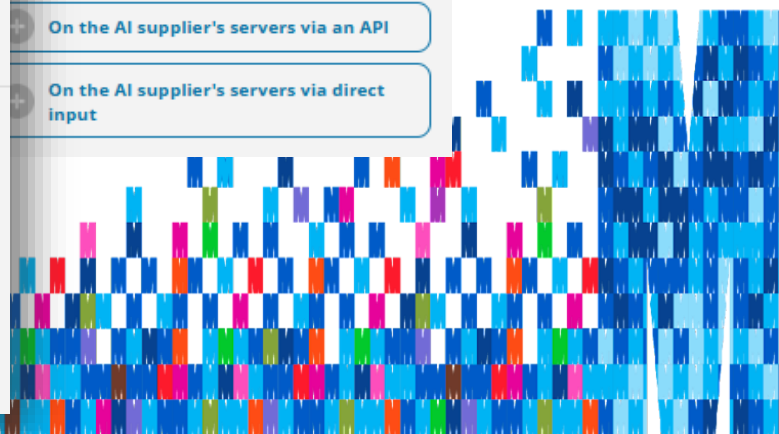
Another's data

Monash Enterprise version of AI

Open source and downloaded onto Monash servers

On the AI supplier's servers via an API

On the AI supplier's servers via direct input



MONASH University

AAA

Monash AI Use & Research Data Risk Tool

Welcome to this tool to determine the risks and appropriate mitigation strategies in using AI with your research data.

This tool is designed to ask a few key questions and provide you guidance to ensure you meet all your research data obligations.

At the end of the questions you will see a set of risks and mitigations that you can follow. You can choose to have this emailed as a pdf to you at the end of the tool or you can download it yourself.

If you have any questions about this tool or the suggestions it produces, please contact us at researchdatagovernance@monash.edu

Do you plan to use artificial intelligence (AI) with your research data?

This is includes all types of Generative AI, General-purpose AI, Agentic AI, AI models or systems.

*This does **NOT** include using AI to formulate your funding applications, protocols or manuscripts.*

Yes

No

reset

Are you undertaking human research?

i.e. is your research with or about people or their data or tissue?

* must provide value

Yes

No

reset

INTELLECTUAL PROPERTY RISK

Risk Identification	Data analysed by AI breached **someone else's IP** as there was no license or agreement to use copyrighted material
Risk Mitigation	Review the other data source and document the license or agreement with which you obtained the data and the conditions of use

How do you intend to use Artificial Intelligence (AI):

*Will you just use an AI model/ service offered by a third party without further training, fine tuning or optimisation of that AI model **(end-user)**?*

*Will you use an AI model/ system that you will optimise/ fine tune with your research data (**deployer**)?*

*Will you design, build, train, adopt, or combine AI models and applications (**developer**)?*

Developer

reset

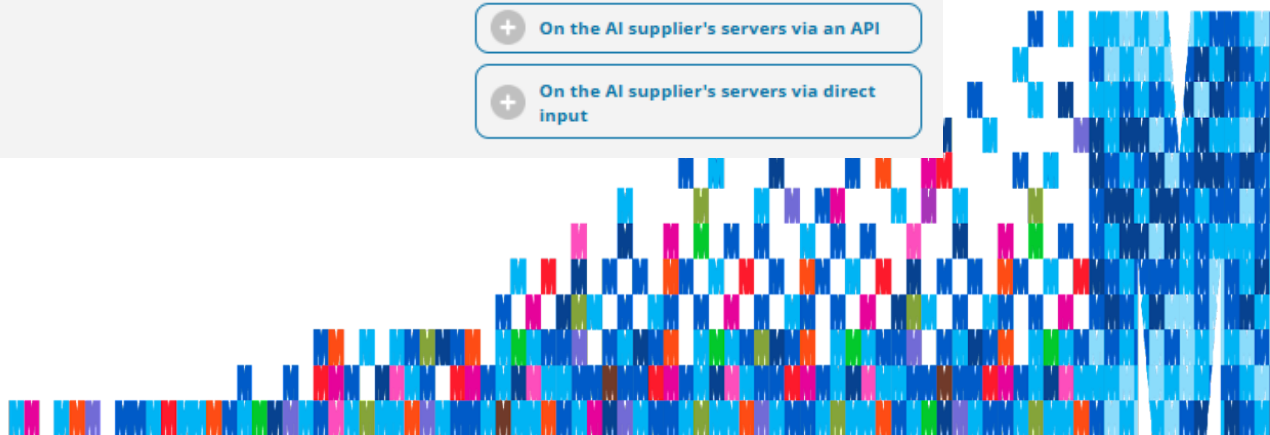
Tick all that apply

Monash Enterprise version of AI

Open source and downloaded onto Monash servers

On the AI supplier's servers via an API

On the AI supplier's servers via direct input



MONASH University

AAA

Monash AI Use & Research Data Risk Tool

Welcome to this tool to determine the risks and appropriate mitigation strategies in using AI with your research data.

This tool is designed to ask a few key questions and provide you guidance to ensure you meet all your research data obligations.

At the end of the questions you will see a set of risks and mitigations that you can follow. You can choose to have this emailed as a pdf to you at the end of the tool or you can download it yourself.

If you have any questions about this tool or the suggestions it produces, please contact us at researchdatagovernance@monash.edu

Do you plan to use artificial intelligence (AI) with your research data?

This is includes all types of Generative AI, General-purpose AI, Agentic AI, AI models or systems.

*This does **NOT** include using AI to formulate your funding applications, protocols or manuscripts.*

Yes

No

reset

How do you intend to use Artificial Intelligence (AI):

*Will you just use an AI model/ service offered by a third party without further training, fine tuning or optimisation of that AI model **(end-user)**?*

*Will you use an AI model/ system that you will optimise/ fine tune with your research data (**deployer**)?*

*Will you design, build, train, adopt, or combine AI models and applications (**developer**)?*

End-user

Deployer

Developer

reset

Are you undertaking human research?

i.e. is your research with or about people or their data or tissue?

* must provide value

Yes

No

reset

Will the data input into the system be:

Tick all that apply

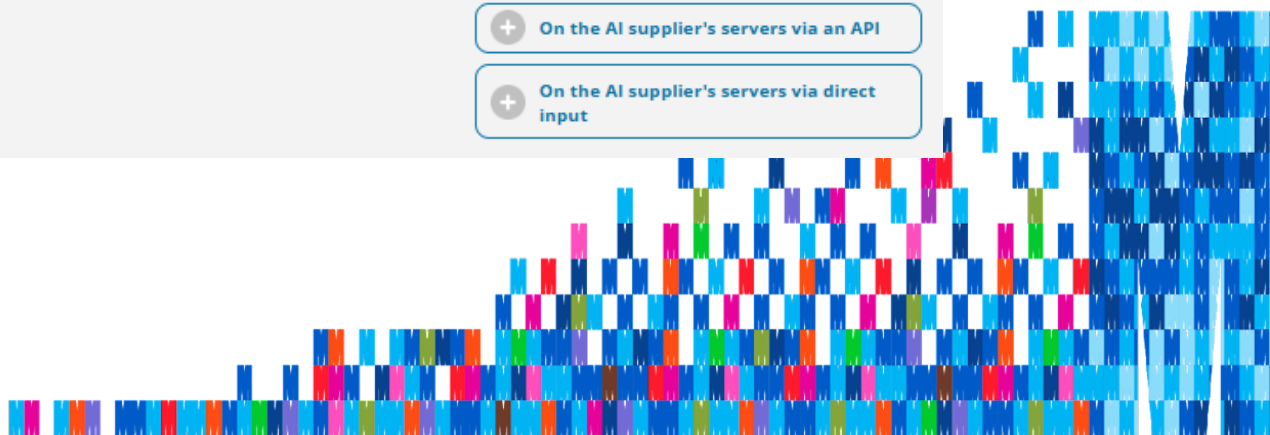
- ✓ My own research data
- ✓ Another's data

Will the AI model/ system be accessed via:

Tick all that apply

- ✓ Monash Enterprise version of AI
- + Open source and downloaded onto Monash servers
- + On the AI supplier's servers via an API
- + On the AI supplier's servers via direct input

CONTRACTUAL RISK



MONASH University

AAA

Monash AI Use & Research Data Risk Tool

Welcome to this tool to determine the risks and appropriate mitigation strategies in using AI with your research data.

This tool is designed to ask a few key questions and provide you guidance to ensure you meet all your research data obligations.

At the end of the ques… you will see a set of risks and mitigations that you're to follow. You can choose to have this emailed as a pdf to yo…

If you have any questi… s at researchdatagovernan…

Do you plan to use artificial intelligence (AI) with your research data?

This is includes all types of Generative AI, General-purpose AI, Agentic AI, AI models or systems.

*This does **NOT** include using AI to formulate your funding applications, protocols or manuscripts.*

Yes

No

reset

How do you intend to use Artificia…

*Will you just use an AI model/ service o… without further training, fine tuning or … model (**end-user**)?*

*Will you use an AI model/ system that … with your research data (**deployer**)?*

*Will you design, build, train, adapt, or … applications (**developer**)?*

…rch?

…ople or their data or …

…be:

My own research data

Another's data

Monash Enterprise version of AI

Open source and do… Monash servers

On the AI supplier's servers via an API

On the AI supplier's servers via direct input

INTELLECTUAL PROPERTY RISK

Risk Identification	Data is used by … Monash's and … copyrighted ma…
Risk Mitigation	Confirm the ter… accessed via an… fine-tune or op… Consider using…

HUMAN RESEARCH ETHICS RIS…

Risk Identification	Data fed into A… their own AI in breach…
Risk Mitigation	Use only AI services where further use of data by AI provider is clearly known Consider using the Monash Enterprise AI models

HUMAN RESEARCH ETHICS RISK

Risk Ident…

The other …

Risk Mitiga…

Confirm th… advice from …

HUMAN RESE…

Risk Ident…

Data fed in… researcher's …

Risk Mitiga…

Use only A… Consider usi…

UPSTREAM/ …

Risk Ident…

Data used … breached co… third-party c…

Risk Mitiga…

Review org…

INFORMATIO…

Risk Ident…

Security of …

Risk Mitiga…

Review Mo… arrangement… Consider usi…

PRIVACY RISK

Risk Identification

Data you are analysing with AI breaches privacy obligations as consent or "reasonable expectation" was not in place and/ or data accuracy is a problem

Risk Mitigation

Review Office of Australian Information Commissioner and Office of Victorian Information Commissioner guidance about use of personal information in using AI and seek advice from the DPPO to ensure you are not in breach of Victorian, Australian or other jurisdictions privacy law

PRIVACY RISK

Risk Identification

The data was sourced from a third party that may not have complied with applicable privacy obligations

Risk Mitigation

Review Office of Australian Information Commissioner and Office of Victorian Information Commissioner guidance about use of personal information in using AI and seek advice from the DPPO to ensure you are not in breach of Victorian, Australian or other jurisdictions privacy law

CONFABULATION RISK

Risk Identification

Output is confabulated with false logic, citations, legal interpretation provided to erroneously justify or explain output

Risk Mitigation

Assess the accuracy, quality, reliability, and authenticity of output by comparing it to a set of known ground truth data and by using a variety of evaluation methods (e.g., human oversight and automated evaluation, proven cryptographic techniques, review of content inputs).

INTELLECTUAL PROPERTY RISK

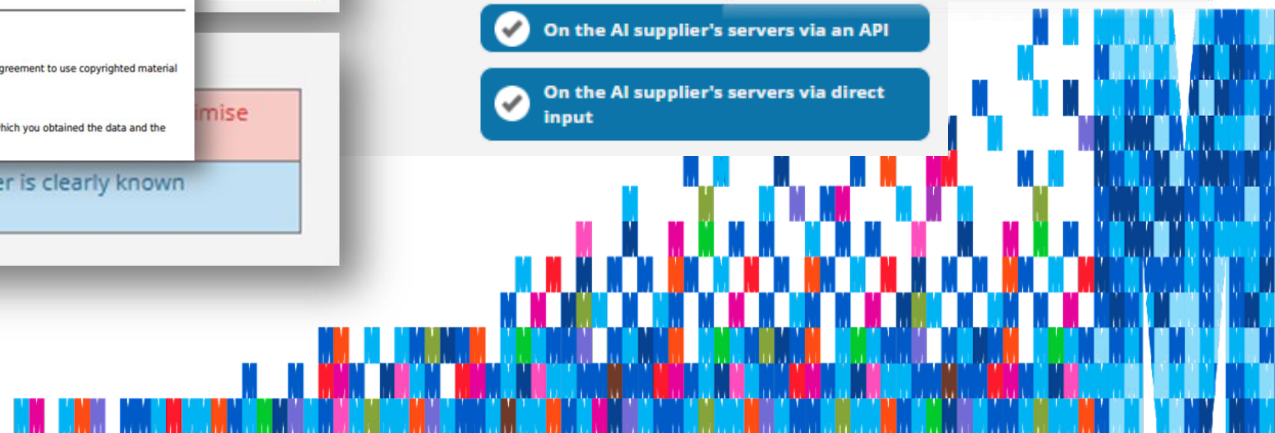
Risk Identification

Data analysed by AI breached someone else's IP as there was no license or agreement to use copyrighted material

Risk Mitigation

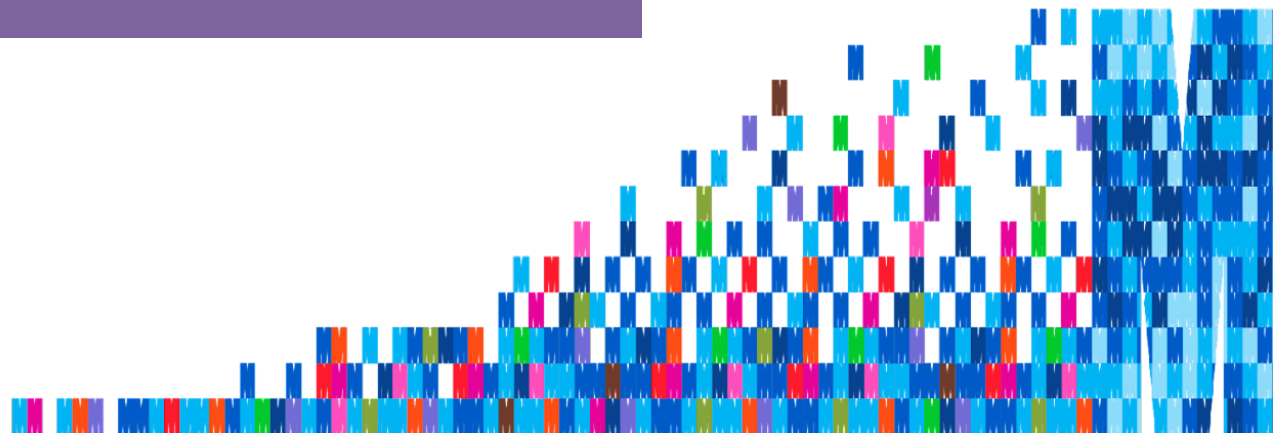
Review the other data source and document the license or agreement with which you obtained the data and the conditions of use

INFORMATION SECURITY RISK



Lessons Learnt

- Cutting through the noise
- Finding the right people to work
- Integration
- The unexpected



MONASH University

Thank you

Acknowledgement: Komathy Padmanabhan