Waipapa Taumata Rau
University of Auckland

Responsible AI for Research

We will get started soon!
Meanwhile, introduce yourself and your research topic in the chat.

14 April 2026

Ngā Ratonga Manaaki Rangahau | Research Services
Te Kahu Tauira | Student & Scholarly Services

Quick poll
& Session roadmap

1. How AI works
2. Responsible AI
3. AI shortfalls

Pātai

What does responsible use of AI mean to you as a researcher?

Part 1:
How AI works

AI use in research

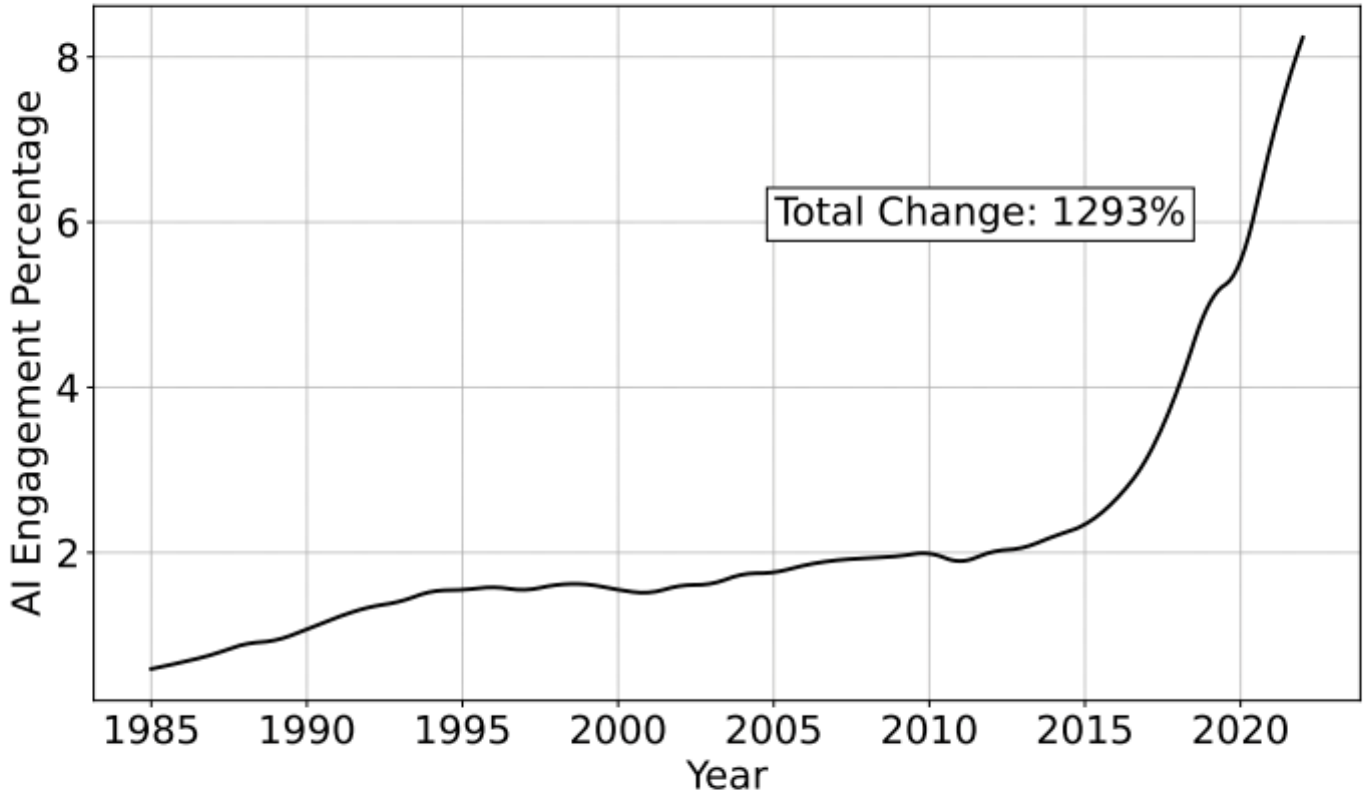
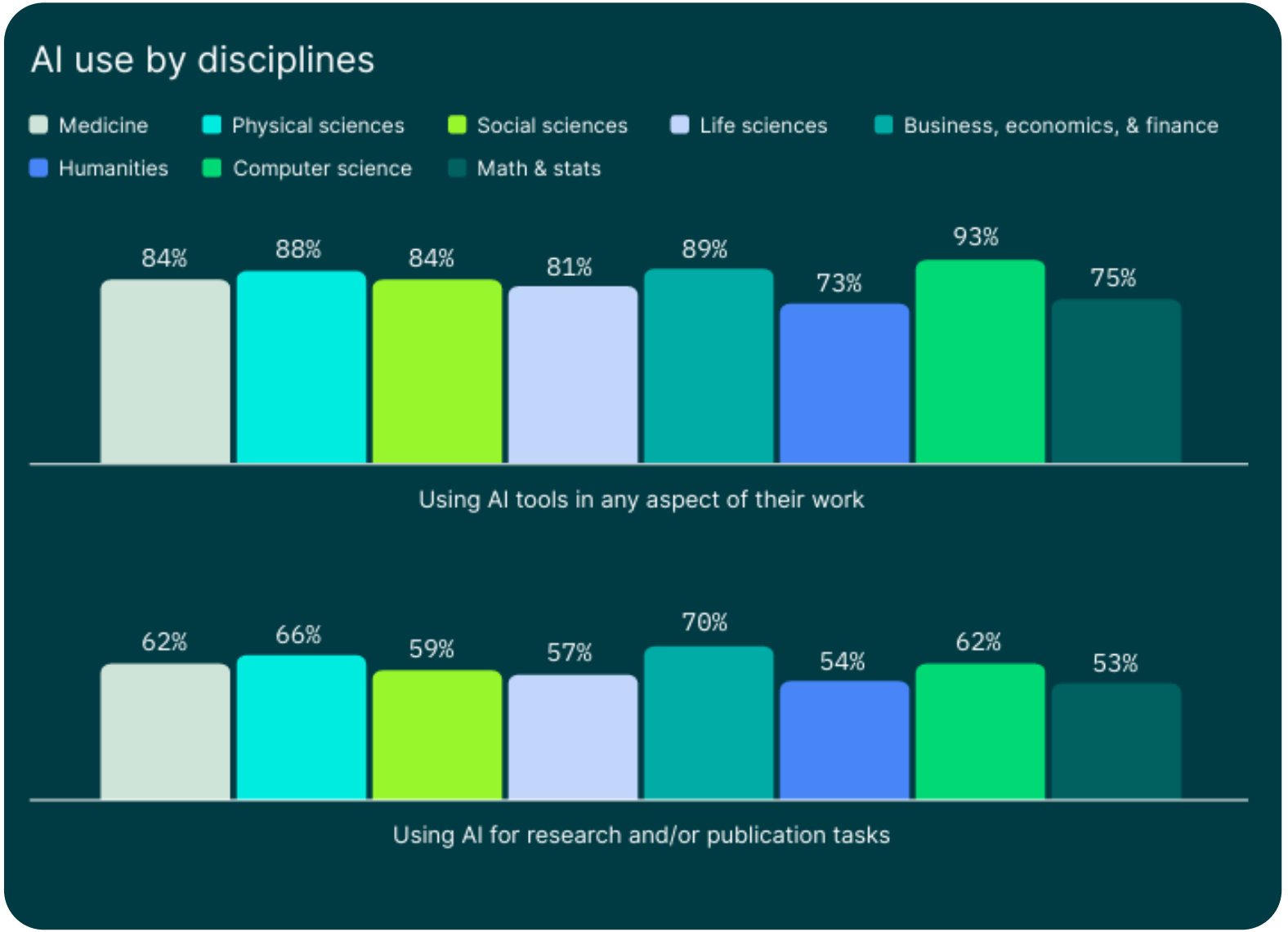


Figure 1: Change in AI engagement across all fields from 1985 - 2022
https://doi.org/10.48550/arXiv.2405.15828

AI use in research



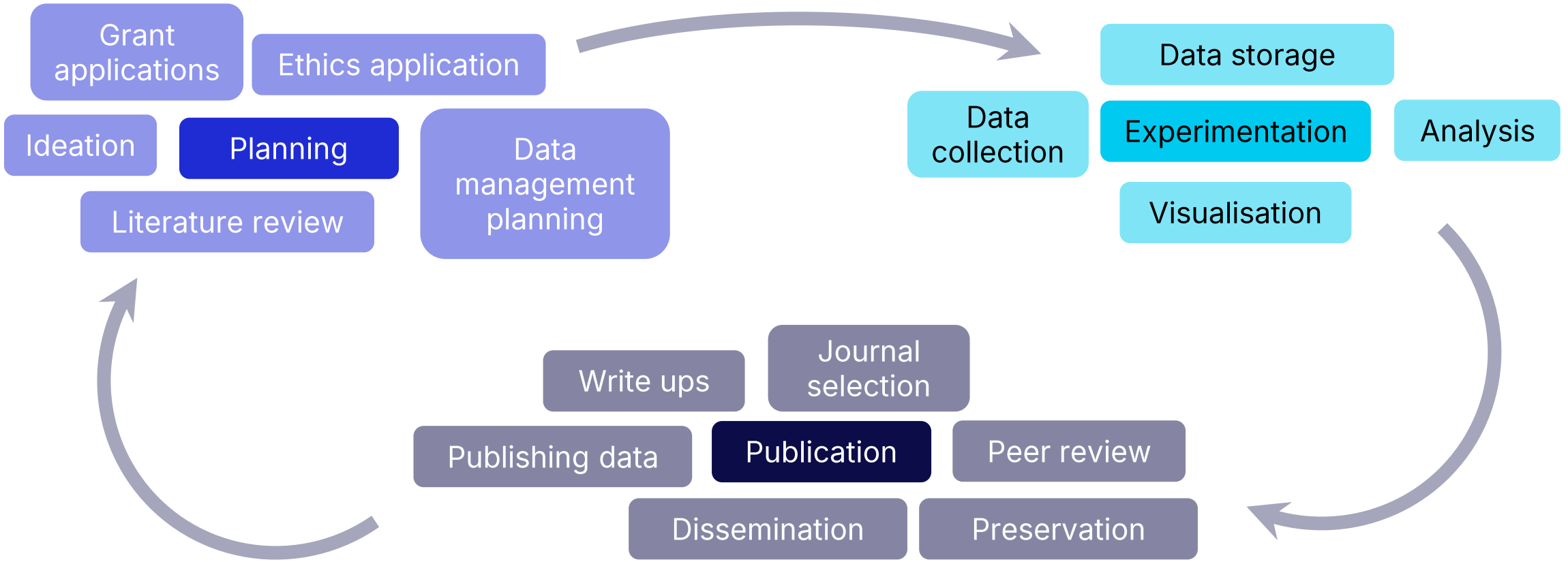
Source: [ExplanAItions 2025: The evolution of AI in research – results by discipline]

Pātai

What AI tools do you use?

What do you use them for?

Where might you use or encounter AI in the research cycle



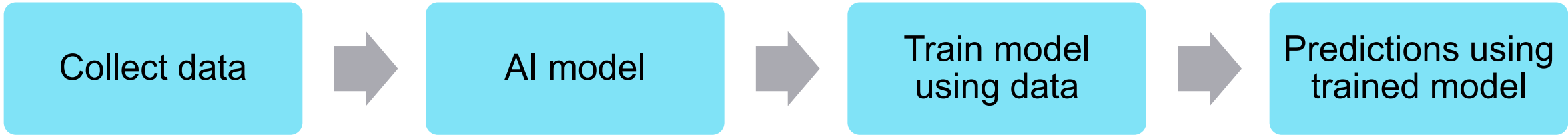
Defining artificial intelligence

Artificial intelligence (AI) refers to machines programmed to mimic human intelligence. AI is used to analyse data to recognise patterns and make predictions based on those patterns.

Generative AI (GenAI) is a subset of AI. In response to human instructions or 'prompts', GenAI can create new content, such as text, images, video, audio, and code.
It does this through being trained on very large collections of existing content – scraped from the internet and other sources.

This definition is used in the University's proposed **'Guidelines for the Use of Generative AI in Doctoral Research'**, which are based on the **"Guidelines for the best-practice use of generative artificial intelligence in research in Aotearoa New Zealand (Online)"** published by Royal Society Te Apārangi (retrieved 26 Feb 2026).

AI chatbots (GenAI)



Why did the chicken cross the _____?

Road	90%
Street	6%
Field	3%
Mafia	1%

Based on data collected

Agentic AI

Artificial intelligence systems that can act autonomously, reason toward specific goals, and operate independently with minimal human oversight.

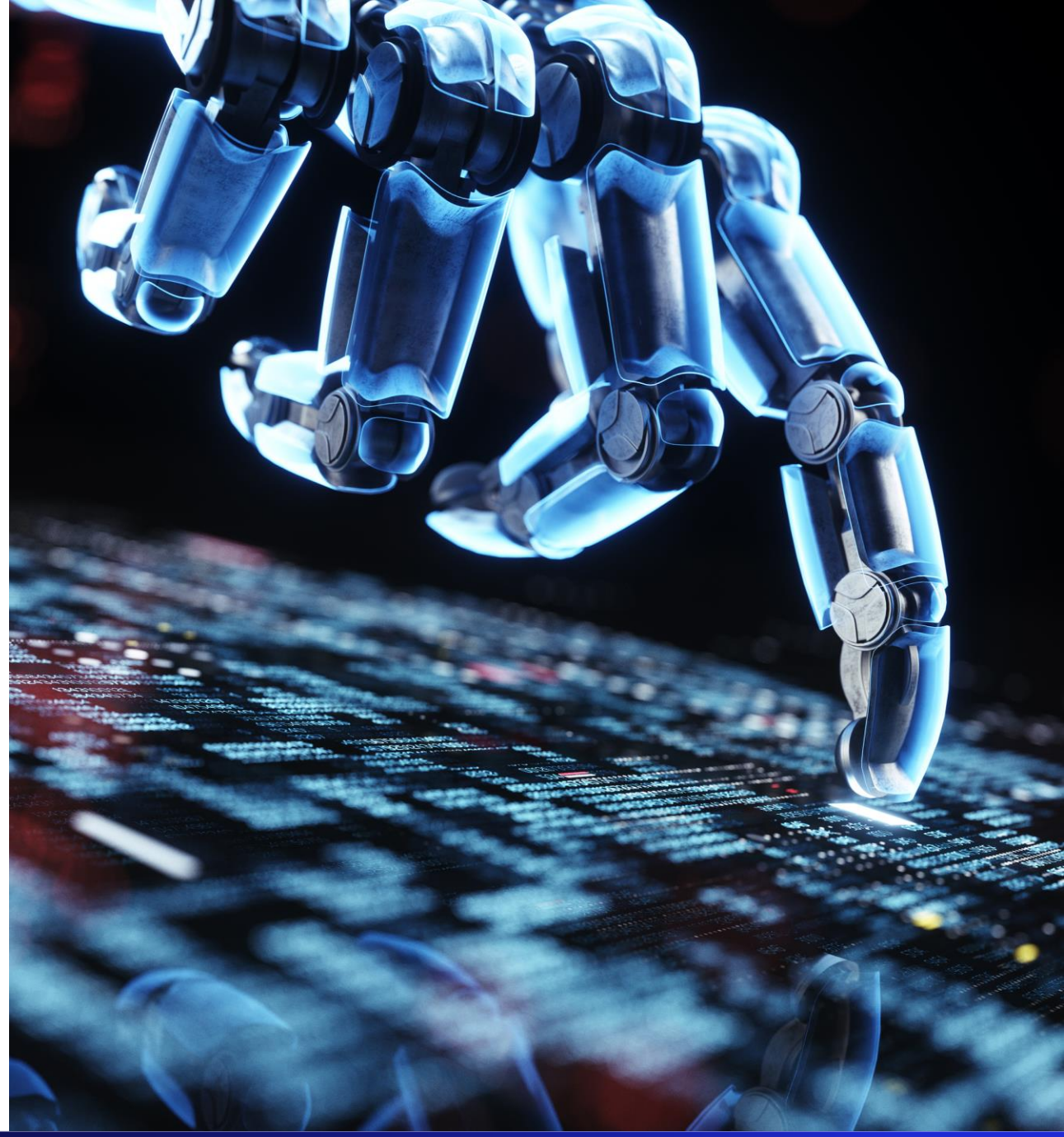
Definition from Research Data Alliance Global Community Priorities for Agentic AI in Research: Community Consultation Results

Learn more:
IBM's What is agentic AI?

Part 1 Summary

- **AI use is increasing rapidly in research**, across disciplines and throughout the research cycle.
- **Different types of AI work in different ways**, from predictive tools to generative and agentic systems.
- Generative AI tools **predict - not "know."**
- **Use AI responsibly** by choosing approved tools and matching them to the data you are working with.

Any questions?



Part 2:
Responsible AI

Pātai

What research data are you working with?

Terms of use example

Otter.ai

2. HOW WE USE YOUR PERSONAL INFORMATION

We use your Personal Information to:

- **Set up your account.** We use your registration information, device information and information received from third parties (such as your username, email address) in order to set up an account for you to use our Services. We do so in accordance with our contractual and precontractual obligations to you in order to provide you with an account to use the Services.
- **Provide you with the Services**. We use your audio recordings, usage information and platform information in order to provide you with the Services. In addition, we use your communication information to facilitate support (e.g. retrieval of a forgotten password). We do so in accordance with our contractual obligations to you in order to provide you with the Services.
- **Improve and monitor the Services**. We use information we automatically collect or generate about you when you use the Services, as well as information about your device such as device manufacturer, model and operating system, and the amount of free space on your device, to analyze the use of and improve our Services. We train our proprietary artificial intelligence technology on de-identified audio recordings. We also train our technology on transcriptions to provide more accurate services, which may contain Personal Information. We obtain explicit permission (e.g. when you rate the transcript quality and check the box to give Otter.ai and its third-party service provider(s) permission to access the conversation for training and product improvement purposes) for manual review of specific audio recordings to further refine our model training data.

https://otter.ai/privacy-policy

Privacy principles

Based on the Privacy Act 2020:

[Privacy policy]

Responsible AI

- **No AI Policy - use guided by existing policies**
- Privacy Policy
- Research Data Management
  - Data classification
- IT Security Policy
- GenAI usage standard
- Research Integrity Policy
  - Authorship and Publication Guidelines
  - **COMING SOON** Guidelines on the Use of Generative AI in Doctoral Research

Policies impacting use of AI in research

An overview of institutional policies, broader guidance, and resources to inform how researchers can use generative AI responsibly.

acting use of AI in research

Waipapa Taumata Rau, University of Auckland and research funders, data providers and publishers all have requirements and expectations for how Artificial Intelligence (AI) is used in research. Researchers are responsible for ensuring they are familiar with relevant policies and reflect them in the decisions they make around AI in research.

University policies and guidelines

- The Privacy Policy aligns with the Privacy Act 2020 and provides rules for protecting personal information. When collecting or using personal information with generative AI tools, you must first complete a Privacy Impact Assessment.
- The Ethical Guidelines coming into effect in February 2026 contain comprehensive guidance on Artificial Intelligence in Research (section 1.12).
- The IT Security Policy and IT Acceptable Use Policy outline responsibilities for the use of IT resources, including the requirement to use University-approved generative AI tools or to seek approval for something else.
- The Research Integrity Policy outlines responsibilities around conducting research in accordance with the highest standards of research integrity, and is supported by:
  - Authorship and Publishing Guidelines (clauses 19-21)
  - Doctoral policies and guidelines, including for editing and proofreading theses, and additional guidance is in development
- The Research Data Management Policy outlines responsibilities around how data is collected, generated, stored, and analysed, and is supported by:
  - Research Data Management Policy guidance
  - Research data classification standard to inform choice of generative AI tool
  - Data Management Plans to record use of AI tools
- The GenAI usage standard lays out steps and considerations for use of generative artificial intelligence tools, including when generating, summarising, processing or analysing research data.
- The Intellectual Property Created by Staff and Students Policy provides a framework for managing rights and responsibilities relating to intellectual property.

Other policies and guidelines

- The Royal Society Te Apārangi has published Guidelines for the best-practice use of generative artificial intelligence in research in

Research Data Management

Data classification

Four levels

- Public
- Internal
- Sensitive
- Restricted

All University data types:
administrative, teaching & learning, research

Security orientated

Aligns with government

Data classification informs selection of systems and tools used to collect, process, analyse etc. data

Research Data Management

Policy

Use University-approved systems and tools for research data capture and storage

"Ensuring that digital forms of research data are stored on an appropriate University-managed research storage service or other trusted storage service approved by the University Chief Information Security Officer."

Research Data Management Policy Guidance

University approved AI tools

Enterprise-wide tools

For secure access, sign in to these tools using your University account. This is your abcd123@aucklanduni.ac.nz address.

- Microsoft 365 Copilot Chat is approved for public, internal and sensitive data.
- Google Gemini is approved for public, internal and sensitive data.
- Google AI Studio is approved for public, internal and sensitive data.
- NotebookLM is approved for public and internal data.

Free or user-pays tools

Before using these tools, we recommend going to settings and turning off permissions for Large Language Model (LLM) training.

- Claude Pro is approved for public and internal data.
- ChatGPT is approved for public data.
- Consensus.ai is approved for public data.
- Perplexity is approved for use with public and internal data.

University approval of applications, services and systems (tools)

Why is this required?

The University is strengthening its data protection processes and needs to manage research software sustainably. As a result, all digital applications, services, and systems must be University-approved. This is to protect staff, students, research participants, and the reputation of the University. This means the tool will undergo a security assessment by Digital Services to ensure compliance with the following policies:

- IT Security Policy
- Privacy Policy
- Research Data Management Policy including the Research data classification standard
- Generative AI tools are also subject to the GenAI Usage Standard.

Completion of ethics application and IT procurement processes rely on tools being confirmed as University-approved.

Identifying tools you intend to use

1. Consider research software and tool requirements for your project and document this in your Data Management Plan.
2. Identify if the tool is University-approved by searching the ResearchHub or contacting your Digital Services Business Relationship Manager or the Centre for eResearch to check. Note, previous or current use of a tool does not mean it is currently University-approved.

Research AI | Generative AI tools

University approved systems

Approval process

Seeking University-approval

1. Plan for the University-approval process to take **3–5 weeks**.
2. Start the process by contacting your Digital Services Business Relationship Manager or the Centre for eResearch. Be prepared to state what tool(s) are needed, the intended purpose, including data classification, and how currently approved tools do not meet your needs.
3. Work with Digital Services staff to compile information including a completed Privacy Impact Assessment (PIA) and security information. This will be assessed by the Digital Services Security team.
4. Digital Services' approval will be for a period of time, aligned with the University's data classification, and may include additional conditions of use (e.g. limited to a specific research project or prescribed settings).

For further advice on the approval process, contact your Digital Services Business Relationship Manager or the Centre for eResearch.

Research software & computing / AI in research / GenAI tools

Generative AI usage standard

1. Select Data classification
3. Undertake Privacy Impact Assessment
6. Understand AI limitations and biases

Intranet / ... / Generative AI usage standard

Generative Artificial Intelligence Usage Standard

On this page:

Application

The standard applies to the use of Generative Artificial Intelligence (GenAI) tools by all members of the University community in connection with any process or activity undertaken at the University of Auckland. This includes all tools, services, and applications which leverage or integrate GenAI capabilities.

Purpose

This standard serves to ensure safe, ethical, and legal use of GenAI tools and services.

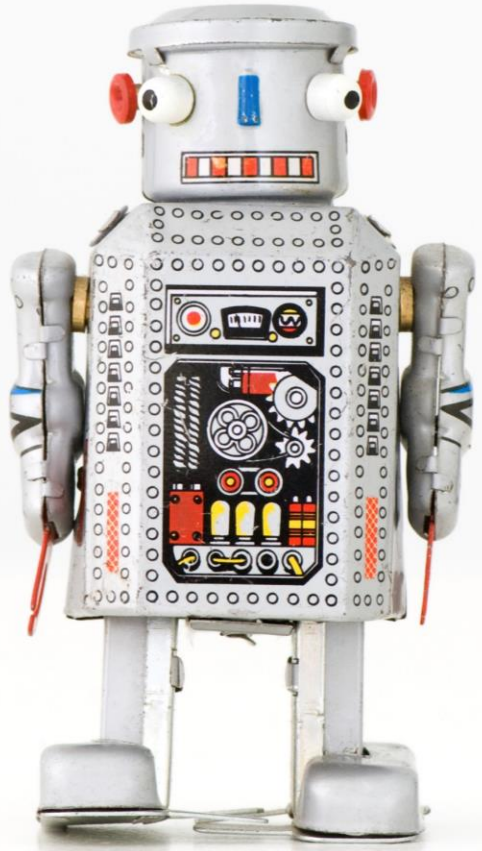
Steps for students & their supervisors

COMING SOON from the School of Graduate Studies: Guidelines on the Use of Generative Artificial Intelligence in Doctoral Research

- **Continued discussion together**: Talk with your supervisor about AI use.
  Expect continuous developments in this space too – e.g. Upcoming guidelines
- **Review applicable policies and guidelines**
- **Reflect:** Tool specific benefits and limitations for use case?
- **Identify tools**: Is it approved for the intended use?
- **Ethics:** If wanting to share research data covered by ethics approval, apply ahead of time. Detail if you are consuming or developing AI, why the tool is appropriate and how you will manage risks.
- **Document:** Detail in a Data Management Plan and create usage log.

Part 2 Summary

- Use of AI is guided by **existing University policies**
- **Data classification determines tool selection** – use only tools approved for your data classification
- **Want to use a new tool?** It must go through the approval process before use
- AI use must align with **ethical, privacy, and security** requirements in your research project
- **For doctoral candidates:** there are guidelines coming, start with discussing AI use with your supervisor



Any questions?

Part 3:
AI shortfalls

Methodological considerations

Data

- Do you have the data for it? (Size, representation)

Computing power

- Access, technical ability?

(Is it worth it?)

Algorithm (tools)

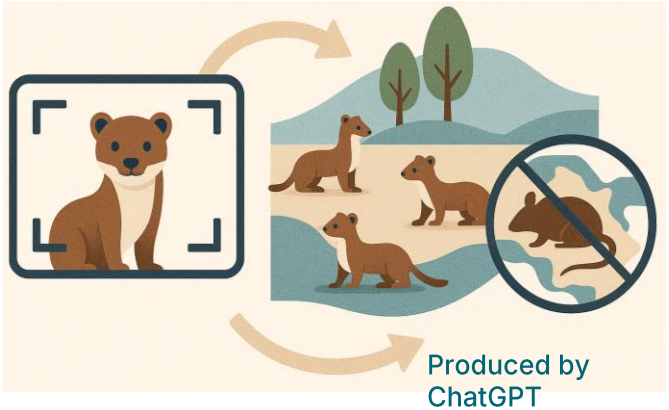
- Peers using it?

(Consult lab, literature)

DREAM BIG

What questions would change your field?

(If we could know X, then we can do Y)



Produced by ChatGPT

Unseen AI impacts

Computing power

- Data centres running constant queries
- Model training and day-to-day usage

Energy consumption

- ~3% of NZ's total electricity use[1]
- Comparable to Auckland's electric rail network or the output of a full coal power plant[2]

Water use

- Large volumes used for data centre cooling
- Often invisible but significant, especially during peak demand

Data Centre in UK, Unsplash

1 How much electrical power will be required due to New Zealanders adopting the use of generative AI?
2 Auckland Transport's Energy Story

Limitations of AI: Bias

AI does not think or reason

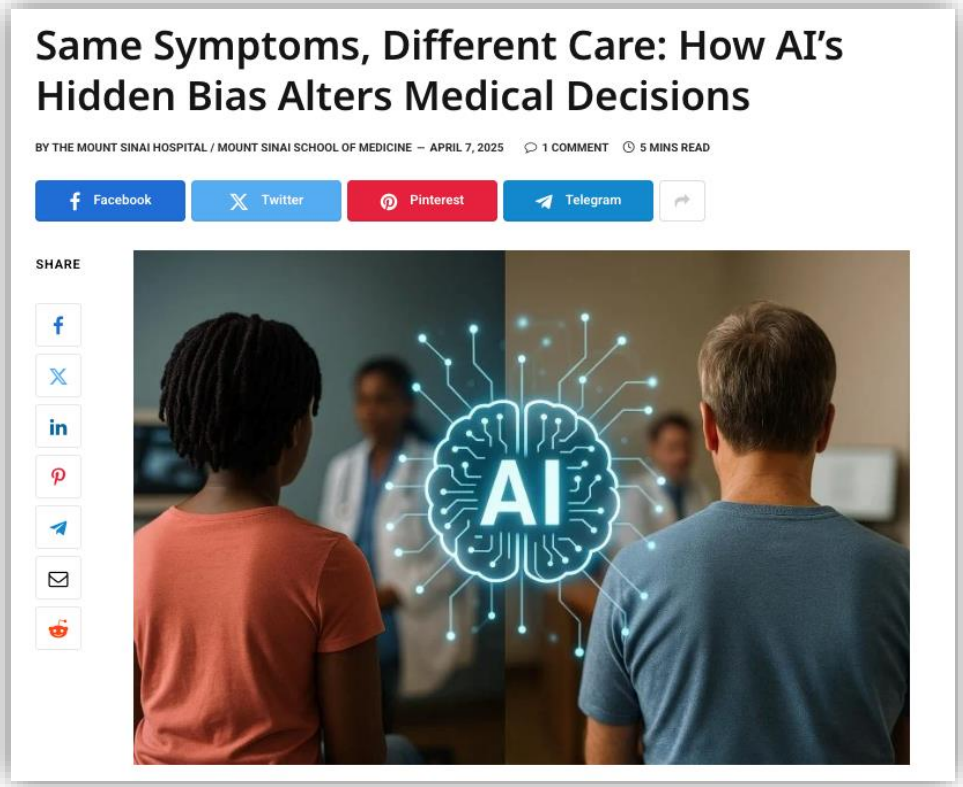
- It only predicts patterns

Bias comes from data

- AI reflects the data it was trained on
- This includes historical biases, gaps and under-representation

Not neutral or objective

- Outputs may reinforce dominant perspectives or existing assumptions



[Same symptoms, different care]

Limitations of AI: Hallucinations

Unreliable citations

- Over half of AI-generated citations fail
- 20% entirely fabricated
- 45% of real papers had major DOI errors

Confident but false outputs

- Errors often appear plausible ("hallucinations")

Published on 12.Nov.2025 in Vol 12 (2025)

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/80371, first published 09.Jul.2025.

Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication in Mental Health Research Using Large Language Models: Experimental Study

Jake Linardon[1]; Hannah K Jarman[1]; Zoe McClure[1]; Cleo Anderson[1]; Claudia Liu[1]; Mariel Messer[1]

Article | Authors | Cited by (7) | Tweetations (25) | Metrics

- Abstract
- Introduction
- Methods
- Results
- Discussion
- References
- Abbreviations
- Copyright

Abstract

Background:
Mental health researchers are increasingly using large language models (LLMs) to improve efficiency, yet these tools can generate fabricated but plausible-sounding content (*hallucinations*). A notable form of hallucination involves fabricated bibliographic citations that cannot be traced to real publications. Although previous studies have explored citation fabrication across disciplines, it remains unclear whether citation accuracy in LLM output systematically varies across topics within the same field that differ in public visibility, scientific maturity, and specialization.

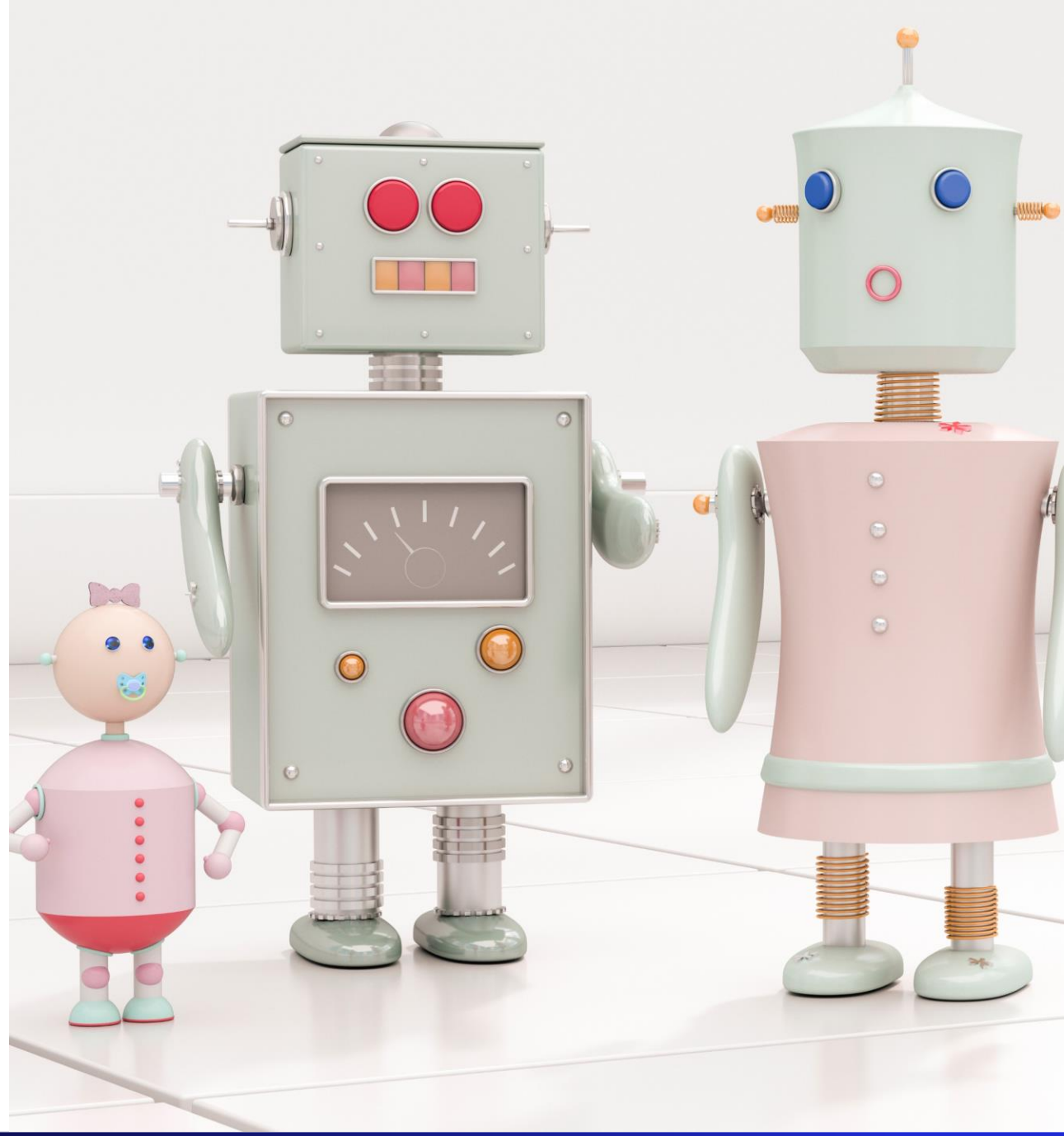
Objective:
This study aims to examine the frequency and nature of citation fabrication and bibliographic errors in GPT-4o (Omni) outputs when generating literature reviews on mental health topics that varied in public familiarity and scientific maturity. We also tested whether prompt specificity (general vs specialized) influenced fabrication or accuracy rates.

Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication in Mental Health Research Using Large Language Models: Experimental Study

Part 3 Summary

- **AI methods have limitations and risks**, including non-representative outcomes
- Outputs may be **inaccurate, fabricated, biased or misleading**
- AI use has **environmental impacts**, including high energy and data-centre demand
- **Uncritical use can cause harm**, especially when informing decisions
- **Careful validation and responsible judgment** are essential for ethical research

Any questions?



Pātai

What is one take away you have from this session?

More resources

AI Essentials | UoA

Build your AI literacy for academic success

AI at the University

We will adopt and embrace Artificial Intelligence confidently and ethically in ways that maximise value and benefit for our people, our institution and our world.

Quick poll
Last questions?

Research AI

Services, tools and training opportunities supporting AI-powered research.

Service

GPUs for research computing and AI

Advice for accessing GPUs for research computing and AI applications.

Event

AI for literature reviews workshop

Online workshop discussing benefits and risks of using AI in literature reviews for researchers, particularly postgraduate and doctoral students.

Event

Transcription using AI workshop

Online workshop introducing AI tools for transcribing research data and important considerations for their use.

Event

Introduction to machine learning workshop

Online workshop providing an introduction to machine learning using python designed for postgraduate research students and researchers with no prior machine learning experience.

Event

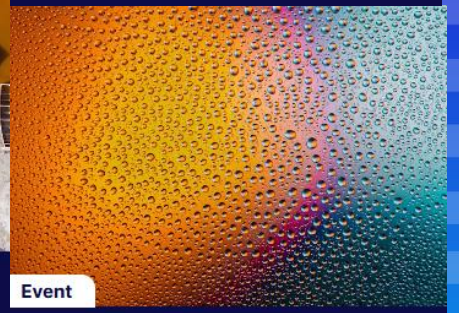
Responsible AI in research for supervisors workshop

Online workshop introducing the benefits and considerations of using AI tools for conducting research designed for supervisors.

Event

Introduction to deep learning workshop

Online workshop providing an introduction to deep learning using python designed for postgraduate research students and researchers to build on their existing machine learning knowledge.



Event

Introduction to AI-assisted workflows for qualitative analyses workshop

An online workshop demonstrating the benefits of using Large Language Models (LLMs) to label and analyse qualitative research data.