



Waipapa
Taumata Rau
**University
of Auckland**

Responsible AI for Research



Monday, 06 July 2026



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



Responsible AI for Research

- How AI works
- Responsible AI
- Methodological considerations



Dr Erin Wood (she/her)

Research Services Adviser

Te Kahu Tuaira | Student and Scholarly Services
Waipapa Taumata Rau | University of Auckland



Dahlia Han (she/her)

Research Services Adviser

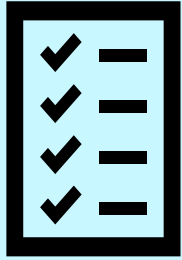
Te Kahu Tuaira | Student and Scholarly Services
Waipapa Taumata Rau | University of Auckland



Pātai

What are you responsible for as a researcher?

How does AI impact this?



Policy & Ethics

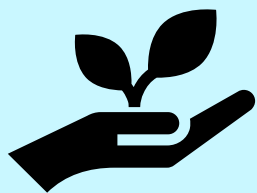
Am I allowed to use AI this way under institutional rules and regulations?



Accuracy & Verification

Is this information correct, reliable and properly checked?

Responsible research with AI



Environment

Is my use of AI purposeful and justified, given its environmental impact?



Skill Development

Am I using AI in a way that helps me learn and develop as a researcher?

Part 1:

How AI works



AI use in research

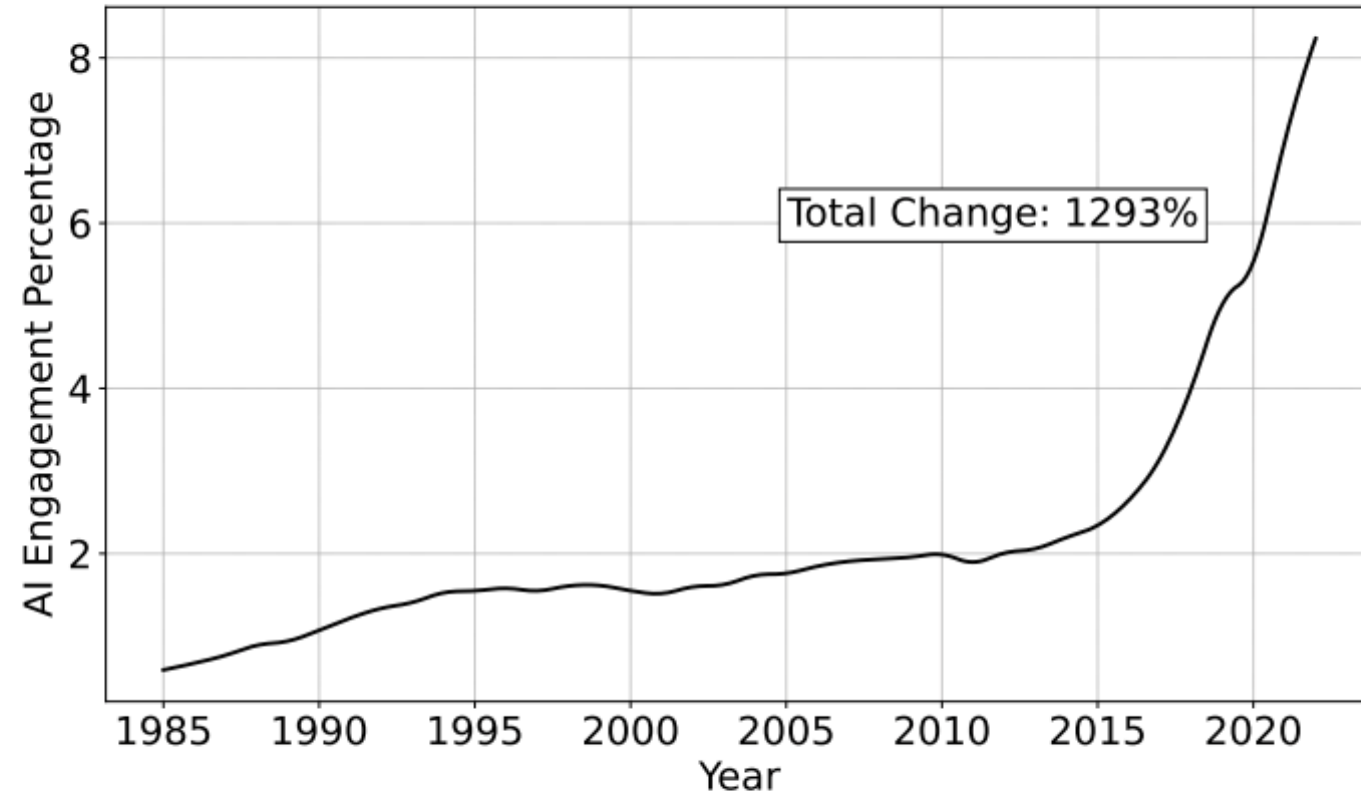


Figure 1: Change in AI engagement across all fields from 1985 - 2022

<https://doi.org/10.48550/arXiv.2405.15828>

AI use in research



Source: [ExplanAltions 2025: The evolution of AI in research – results by discipline](#)

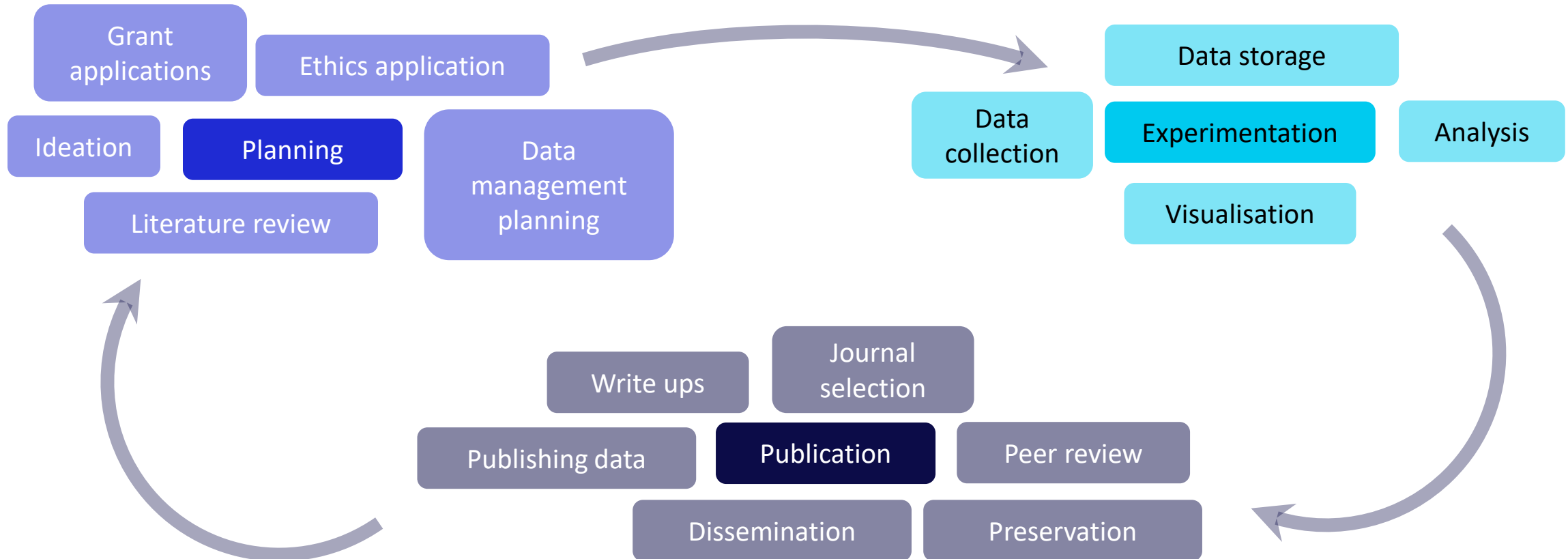


Pātai

What AI tools do you use?

What do you use them for?

Where might you use or encounter AI in the research cycle



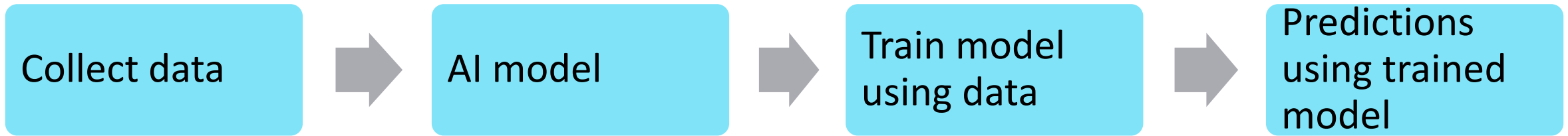
Defining artificial intelligence

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.

Source: "[Guidelines for the best-practice use of generative artificial intelligence in research in Aotearoa New Zealand \(Online\)](#)" published by Royal Society Te Apārangi (retrieved 12 June 2026).

AI chatbots (GenAI)



Why did the chicken cross the _____?

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

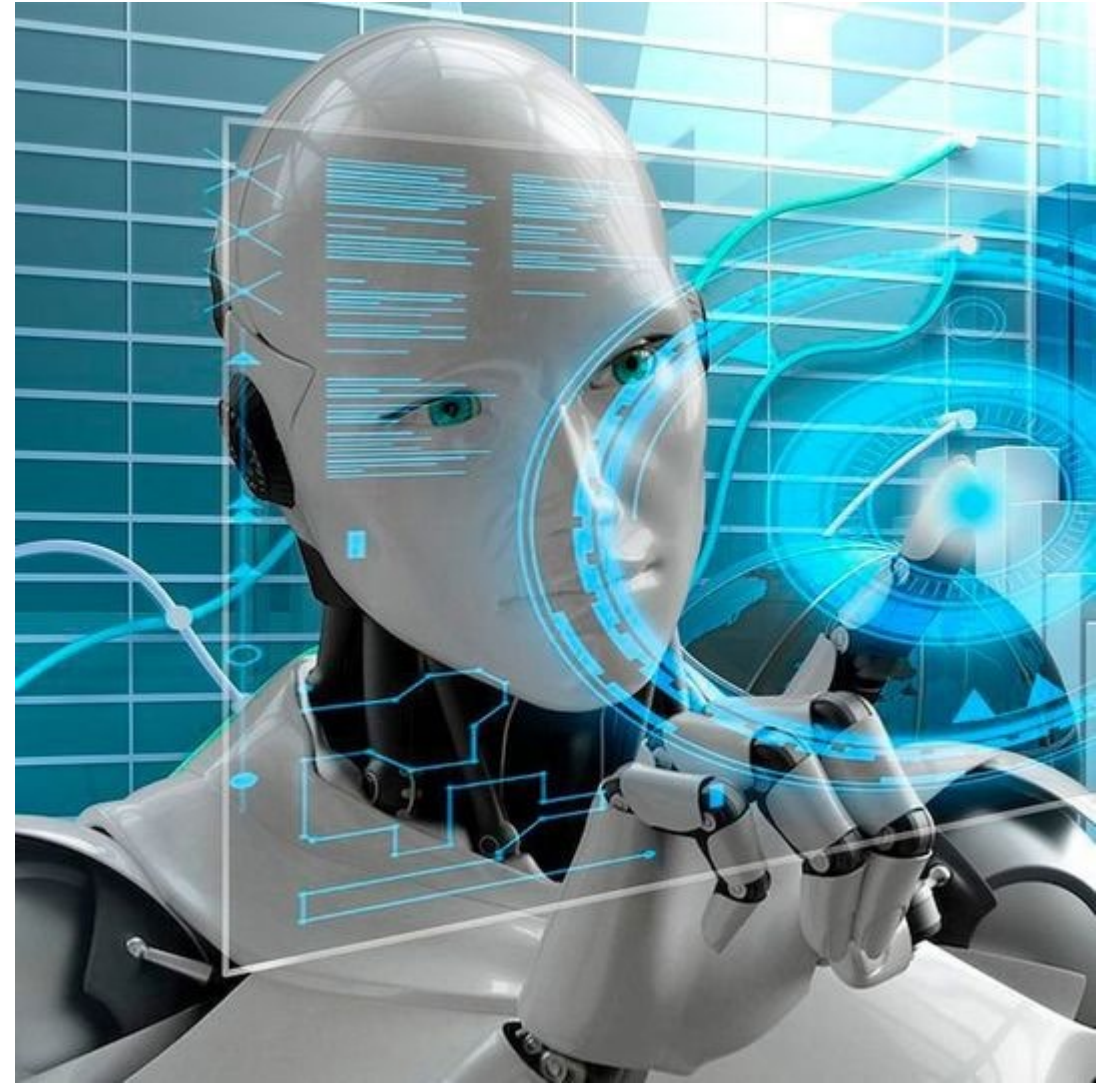
Based on data collected

It makes predictions – but it is not always accurate

Demo

Platform: Gemini

Prompt: Create three sentences 10 words long with the word "Auckland" as the seventh word



Agentic AI

“Agentic AI” is defined as 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.”**

Any questions?

Learn more: [Generative AI vs. other AI types](#)



Part 2:

Responsible AI





Pātai

**What research data are you
working with?**

Terms of use example



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

[Privacy policy](#)

Based on the Privacy Act 2020



Responsible AI in research

Use is guided by existing policy

Privacy

- [Privacy policy](#)
- [Privacy Impact Assessment \(PIA\)](#)

Ethics

- [Ethical Guidelines](#)

Security

- [IT Security Policy](#)
- [IT Acceptable Use Policy](#)

Integrity

- [Research Integrity Policy](#)
- [Authorship and Publishing Guidelines](#)
- [Doctoral guidelines](#)

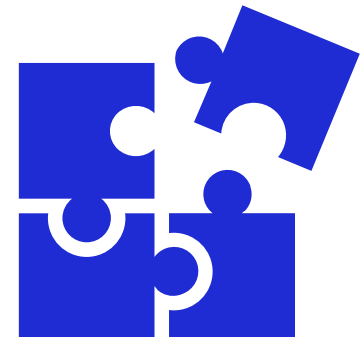
Data management

- [Research Data Management Policy](#)
- [Data classification standard](#)
- [GenAI Usage Standard](#)

IP


- [Intellectual Property Created by Staff and Students Policy](#)

Plus national and publisher policies



Ethical Guidelines

- Key messages for researchers from 1.12
- Faculty-base ethics advisers
- [Research ethics team](#)



The screenshot shows the 'Ethical Guidelines' page for '1.12 Artificial Intelligence in Research'. The page header includes 'EG Ethical Guidelines' and navigation links for 'Home', 'Part 1 - Ethical principles, policies and standards', and 'Part 2 - Ethical considerations for specific groups and contexts'. The main content area features a yellow background with the title '1.12 Artificial Intelligence in Research' and a section titled 'Key messages for researchers'. Below this, it states 'Researchers must:' followed by a list of five bullet points: 'Name the AI Systems you are intending to use in their ethics application', 'Use institutionally approved tools or seek approval from IT security before using other tools', 'Comply with data, privacy and security requirements when processing Human Participant data', 'Be transparent with participants about how AI will be used in the research', and 'Monitor risks – including bias, harm, and the potential for re-identification'. A final bullet point reads 'Engage with Māori, Pacific or Indigenous advisors when relevant to data, methods, or outputs.'

Learn more: [Ethical Guidelines - 1.12 Artificial Intelligence in Research](#)

Learn more: [Human ethics – research involving human participants – University of Auckland](#)

1. Understand your research data

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, and analyse data, etc

Learn more: [Data classification standard](#)

2. How to handle that data

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.”

Learn more: [Research data management policy guidance](#)

Research Data Support Services: researchdata@auckland.ac.nz

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](#) and [Claude Code](#) are 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 our data protection processes and needs to manage research software sustainably. To protect staff, students, research participants, and the reputation of the University, all digital applications, services, and systems will undergo a security assessment by Digital Services to ensure compliance with these policies:

- [IT Security Policy](#)
- [Privacy Policy](#)
- [Research Data Management Policy](#) including the [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 contacting your [Digital Services Business Relationship Manager](#) or the [Centre for eResearch](#) to check. Previous or current use of a tool does not mean it is currently University approved

Learn more: [AI at the University \(staff intranet\)](#) - [Get started with AI](#)

Learn more: [Research AI – Generative AI tools, approvals process](#)



Generative AI usage standard

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

Learn more: [Generative AI usage standard](#) (staff intranet)

Generative Artificial Intelligence Usage Standard

On this page:

- [Application](#)
- [Purpose](#)
- [Standards](#)
- [Definitions](#)
- [Key documents](#)
- [Document management and control](#)

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.

AI in doctoral research guidelines

Maintain the originality and integrity of your work and align with institutional policies and scholarly standards.

- [Guidelines](#)
- [Staff intranet pages](#) More information about guideline implementation.
- **AI essentials:** [AI for postgraduate research](#) Support materials for students such as recordkeeping document examples and templates as well as case studies of responsible use, in alignment with the guidelines.



September 1st
2026



AI in doctoral research guidelines

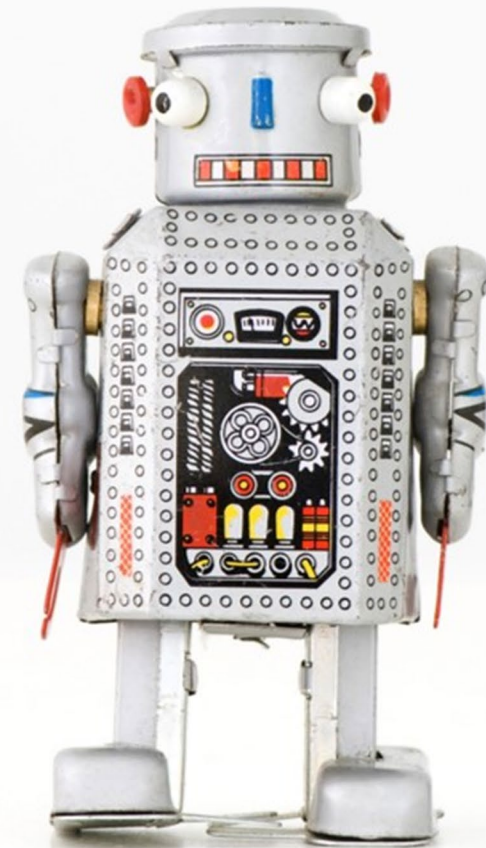
Maintain the originality and integrity of your work and align with institutional policies and scholarly standards.

- **Supervisor discussions:** AI use to be approved and decisions recorded. Record of approval used should be submitted in confirmation proposal and in final submission. A candidate-supervisor discussion guide is available on AI essentials
- **Policy adherence:** Use must comply with policy and good research practice.
- **Ethics:** If wanting to share research data covered by ethics approval, apply ahead of time and get participant consent.
- **Recordkeeping:** Detail in a [Data Management Plan](#) and create a usage log. Details of the log to be determined with your supervisor, log examples are available on AI essentials.

Part 2 Summary

- AI use is guided by different good research practice policies
 - Ethics, privacy, security, academic & research integrity
- Data classification determines what tool you should use
 - New tools need approval
- Doctoral guidelines
 - Approval, documentation, declaration

Any questions?



Part 3:

Methodological considerations



Essential parts

Data

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

Computing power

- [Access, technical ability?](#)
(Is it worth it?)

Tools, algorithms, models

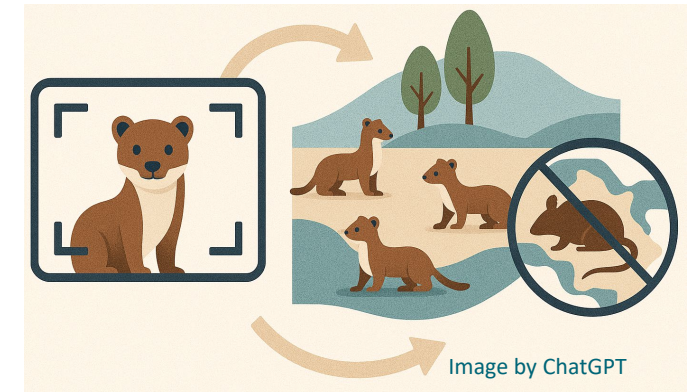
- Peers using it?
(Consult peers and literature)

DREAM BIG

What questions would change your field?
(If we could know X, then we can do Y)



[Clearshores project](#)



[Stoa Re-Identification with Vision Language Models](#)

Environmental considerations

Computing power

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

Energy consumption

- ~3% of NZ's total electricity use¹
- Comparable to Auckland's electric rail network or the output of a full coal power plant²

Water use

- Large volumes used for data centre cooling



The planned Datagrid data centre in Southland, [Why this AI data centre might be New Zealand's most important climate project](#), The Spinoff

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

- Predicts patterns

Bias comes from data

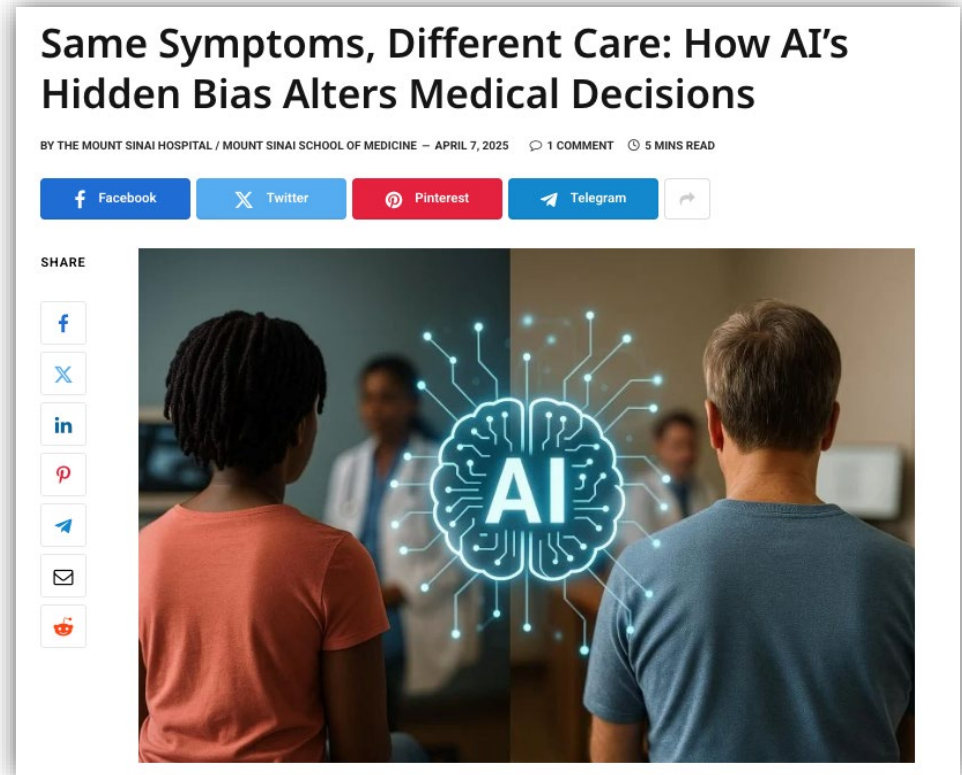
- Reflects training data
- Can include gaps and biases

Not neutral or objective

- May reinforce existing assumptions

Te Rēo

- Meaning can be lost in translation



[Same symptoms, different care](#)

[Learn more: AI essentials – Bias in AI](#)

Limitations of AI: Hallucinations

- **Citation:** Fabricated paper titles, DOIs, and authors.
- **Data:** Fake specific numbers, sample sizes, and statistics.
- **Conceptual:** Inventing non-existent theoretical frameworks, scientific principles or laws.
- **Temporal:** Mixing up timelines, publication dates, and eras.



Source: <https://www.inra.ai/blog/ai-hallucinations>

Citation hallucination

Unreliable citations

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

Confident but false outputs

- Errors often appear plausible ("hallucinations")

Learn more: [Generative Artificial Intelligence in Doctoral Research Guidelines – University of Auckland](#)

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 Linardon¹](#); [Hannah K Jarman¹](#); [Zoe McClure¹](#); [Cleo Anderson¹](#); [Claudia Liu¹](#); [Mariel Messer¹](#)

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.

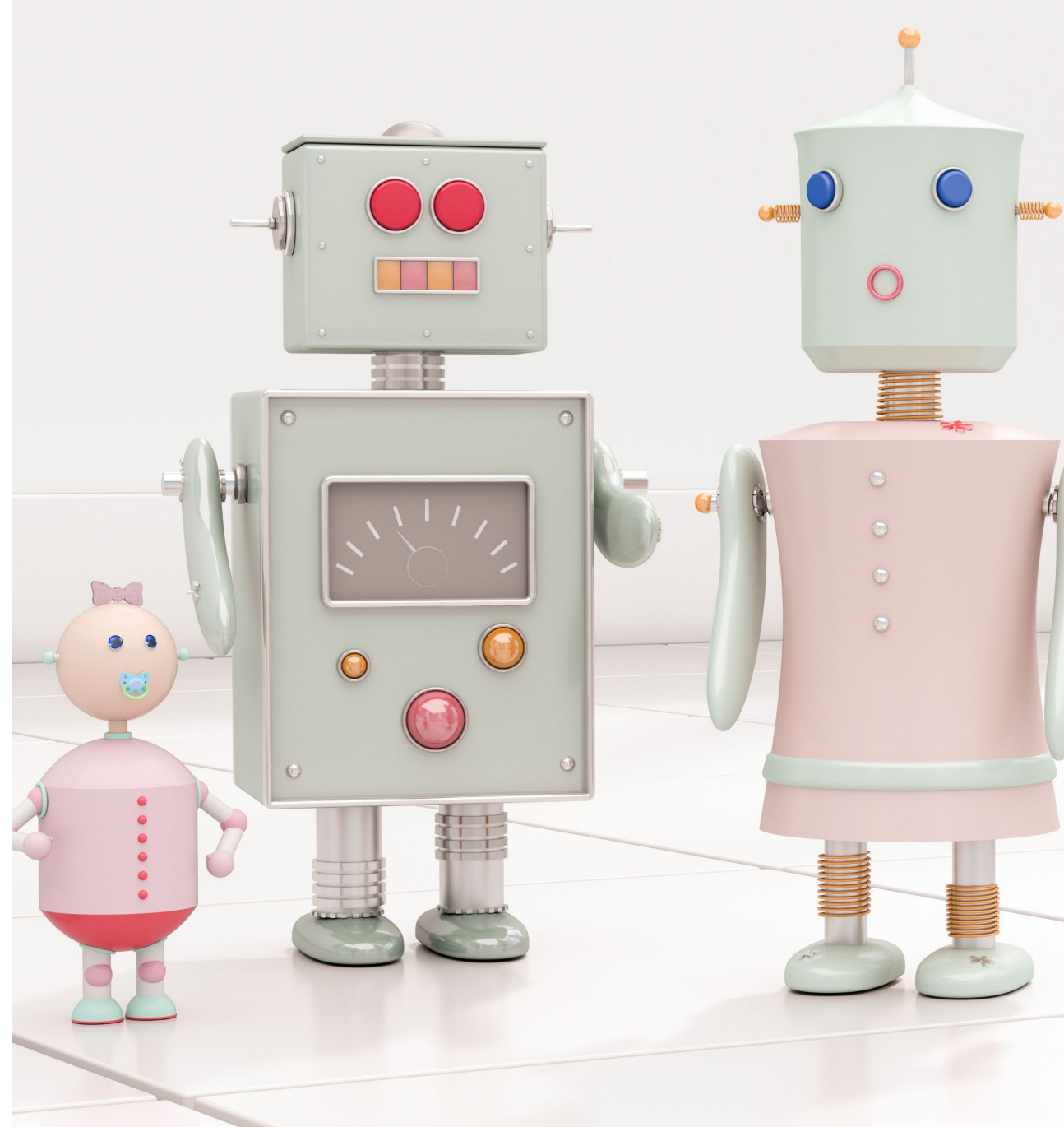
Objective: This study aims to examine the frequency and nature of citation fabrication and bibliographic errors in GPT-4o (Omni) outputs when generating literature reviews on mental health topics that varied in public familiarity and scientific maturity. We also tested whether prompt specificity (general 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 has limitations and risks**
- Outputs may be **inaccurate, biased or misleading**
- AI has **environmental impacts** (energy and water used in data centres)
- **Uncritical use can cause harm**
- **Careful validation and responsible judgment**

Any questions?





Any
questions?

AI tools for literature reviews

ResBaz Day 5

Friday, July 3rd, 10:00AM

[Register](#) or scan



More University of Auckland resources:

Contact: [AskUs](#)

- [Research AI](#)
- [Responsible AI in research workshop](#)
- [AI in literature review work flows](#)
- [AI Essentials](#)
- [AI for literature reviews workshop](#)