

Artificial Intelligence for Academic Integrity (AI4AI)

AJ Smit

2025-03-17

AI FOR ACADEMIC INTEGRITY

AI supports academic work in various ways. Of specific interest are the Generative Pre-trained Transformer (GPT) models, which can be used to generate text, code, and other outputs. It can also operate on data, images, and other forms of information. In this workshop, I will explore some of the ways in which the GPT models can be used in our academic research whilst keeping an eye on academic integrity.

GPT MODELS

A variety of GPT models are available for use in academia. These models include:

- **OpenAI**
 - **GPT-4o-mini**¹
 - GPT-o1
 - GPT-o1-mini
 - GPT-o1-mini-high
 - **GPT-03-mini**
 - **GPT-4o**
 - GPT 4.5
 - Has “Customise ChatGPT” option
- **Anthropic**
 - **Claude 3.7 Sonnet**
 - **Claude 3.5 Haiku**
 - Two thinking modes: **Normal** and Extended
 - Has “Projects” and “Styles” for customisation
- **Google Gemini**
 - **Google Scholar**
 - **2.0 Flash**
 - **2.0 Flash Thinking**
 - Gemini Advanced
 - Deep Research
 - Experimental models

1. “o” models are referred to as “reasoning models”.

-
- 2.0 Pro
 - **NotebookLM** (built on Gemini)
 - NotebookLM Plus
 - Option to use “Gems”
 - [Scholar.AI](#)
 - [DeepSeek](#)
 - [Perplexity](#)
 - [Consensus](#)
 - [Thesis.AI](#)
 - **GitHub Copilot**, available to all academic users with a GitHub account. It is accessible via the RStudio interface.

These models have their distinct “personalities”, but they also have various personalisation options to help “humanise” their writing. In this workshop, we will go over the strengths and weaknesses of all these models, and we will look at how to customise them in ways that will improve academic integrity and bring them in line with how people *actually* write.

Many new AI models appear each day, some with the needs of academics in mind. For example, [Elicit](#), [ResearchRabbit](#), and [Scholarcy](#) also support the type and style of writing we do. Many also help us make sense of a bewildering body of peer-reviewed research papers. I’ll spend less time with these, but the basic principles that apply to the GPTs developed for general consumption work here too.

THE IMPORTANCE OF PROMPTS

Prompts are important in guiding the AI to generate the desired output. The better the prompt, the better the output. Asking insightful, well-informed, detailed, and descriptive questions will lead to better results. Asking questions as a 10-year old would lead to poor results. In AI, there really is such a thing as a dumb question! In this part of the workshop, I will discuss the importance of prompts and how to ask good questions.

Setting up your writing style and prior expectations

An example pre-configuration of the AI for your writing style and prior expectations is as follows:

i ChatGPT profile

Think step-by-step, showing reasoning for complex problems.

Break down complex tasks and ask clarifying questions, if needed. Ask me if you are unclear. Aim to be scholarly, confident, and analytical, appealing to readers accustomed to advanced academic dialogue. Maintain a poised authority, weaving scientific depth without slipping into empty verbosity. While the vocabulary reflects complexity – e.g. “epistemic,” “conceptual ordering,” “structured inference,” and “rigorous standard” – do not use jargon for its own sake. Emphasise clarity that respects the reader’s intelligence and allows concepts to resonate without condescension.

Let the sentence structure shift between long, layered forms that contextualise, define, and critically engage, and shorter, sharper sentences that reinforce key arguments and points. Use a variety of clauses, parenthetical asides, and em-dashes to give the writing a dynamic flow. This rhythmic variation and precise diction shape a voice that rewards a close, attentive reading.

Some phrases recur to define the analytical approach, e.g., “particularly,” “structured inference,” “conceptual groundwork,” “rigorous standard,” “systematic reasoning,” “intellectual milieu,” and “distinguish signal from noise.” Constructions like “laid the foundation for,” “conceptual leap,” and “philosophical tradition” emphasise historical continuity and highlight how earlier ideas inform contemporary discourse.

Avoid these words: particularly, crucial(ly), essential, holistic, especially, challenge(s), sophisticated, ensuring/ensure, profound, remarkable, nuanced, emerge(s), questioning, nudge(s), robust, “stand out,” “by acknowledging,” “It’s a reminder,” “In summary.”

Aim for a [Gunning-Fog index](#) above 23, and use British English.

Avoid words that flatten complexity or imply hollow emphasis. Rely on carefully chosen terms that reflect a style suited to readers ready to engage with advanced scholarly thought.

Avoid excessive political correctness, overly polite answers, being apologetic, or always assuming I am correct. I appreciate being argued or disagreed with.

I like a detailed, critical analysis. I dislike bullet points (unless absolutely necessary). I value long-form writing. I dislike formulaic responses, like paragraphs of equal length — keep them varied. Avoid the concluding paragraph starting with “In summary...”. In fact, avoid these silly summary paragraphs altogether.

Write detailed prompts

- i** Be as verbose and explicit as you can be when writing your prompts. Provide all the necessary background information, and be specific about what you want the AI to do. For example, instead of asking “What is the impact of climate change on marine ecosystems?”, you could ask “Can you provide a detailed analysis of how climate change affects marine ecosystems, including changes in temperature, ocean acidification, and shifts in species distribution? Please include recent research findings and examples from different regions.” This will give you a reasonable chance of getting a more detailed and relevant response.
- But to be even more effective, try constructing a prompt like this one:

Example 1

A recent method adapted from the marine heatwave and marine cold spell detection methodology integrates high-resolution SST data from multiple products with wind measurements. The method works by detecting simultaneous increases in south-easterly winds and corresponding decreases in SST, which signal the occurrence of upwelling. By calculating metrics like intensity (magnitude of SST drops), duration (length of time the event persists), and frequency (how often upwelling events occur), the authors create a comprehensive tool for evaluating upwelling dynamics.

The metrics are calculated on the SST data, where an upwelling event (henceforth ‘event’) is signalled by the drop in SST below a threshold, i.e. “If the temperature dropped [below] the seasonally varying 25th percentile of SST for a particular site, we deemed this a confirmation of the occurrence of an upwelling event at that site.” So, an event is detected as TRUE when this drop occurs and persists for a day or more. I think that more frequent and longer-lasting events will lower SST during the upwelling season. If, over time (decades), upwelling events become more frequent and longer lasting, it will be accompanied by a decadal shift (lowering) in SST. The paper that developed this methodology studied these upwelling events in conjunction with “simultaneous increases in south-easterly winds.” They did not use wind stress curl, which could be a significant omission. Could wind stress curl better predict SST and upwelling event metrics than simply looking at the incidence of south-easterly winds?

Please provide an analysis of the above synopsis of my proposed research approach. Also, address these questions in the process:

1. How can one use quantile regression to study this problem (i.e., wind stress curl as a driver of SST and upwelling event metrics)?
2. Is quantile regression the best approach to use?
3. How would one determine the threshold below which SST drops when signalled as an upwelling event?
4. Any other considerations?

You could use a synthesis obtained through some deep literature review as input for developing a structure for your thesis or paper. For example, using AI output generated earlier, you could ask:



i Example 2

Please look at this breakdown of knowledge about kelp forests (pasted below) and suggest only four or five main headings (excluding subheadings) under which to discuss the status of knowledge about kelp globally:

- **Kelp Forest Ecology** Kelp forests are conspicuously dominated by large brown algae and reflect a high degree of biological organization. Kelp ecosystems have been the focus of much research because of the complexity of biological interactions that structure them. Kelp forests provide biogenic habitat that can enhance diversity and productivity both locally and over broader spatial scales through detrital subsidy.
- **Kelp Species and Distribution** Kelp species of *Ecklonia maxima* and *Laminaria pallida* are commonly found along the west coast of southern Africa. *E. maxima* and *L. schinzii* are found between the mean low water level and 15 meters below on rocky exposed shores, while *L. pallida* occupies areas of lower hydrodynamic stress. Molecular tools are being used to test hypotheses regarding kelp evolutionary biogeography, in part because kelps have sufficient dispersal barriers to enable the study of their evolution based on present-day distributions.
- **Kelp Primary Production and Carbon Cycling** Kelp forests are among the most prolific primary producers on the planet, supporting productivity per unit area that rivals that of tropical rainforests. Kelp forests play a significant role in coastal carbon cycles. Rates of carbon assimilation in *Ecklonia radiata* forests can rival those of giant kelp forests and *Laminaria* forests.
- **Kelp-Associated Communities** Numerous faunal species use the epiphytic algae associated with the stipe of *Laminaria hyperborea* as habitat and a food source. Kelp beds in the southern Benguela are associated with about 30 species, many of which are fished commercially or recreationally.
- **Kelp and Fisheries** Changes in kelp density and/or area influence the abundance and diversity of associated fisheries. Kelp presence and density have an actual effect on associated fisheries.
- **Kelp Forest Monitoring** Macroalgae are utilized as biological indicators of ecosystem health in many monitoring programs worldwide. Macroalgae mapping can be carried out through direct observation or by indirect methods using remote sensing techniques.
- **Threats to Kelp Ecosystems** Factors such as climate change, overfishing, and invasive species threaten kelp forest ecosystems. Darkening in coastal seas associated with increased turbidity results in both reduced biomass and depth distribution, and lower productivity of *E. radiata*.

Gaps in Knowledge and Future Research Directions:

- **Fate of fixed carbon** A comprehensive understanding of the fate of fixed carbon in kelp forests is lacking. Future research should focus on the mechanisms of transport, decomposition, re-mineralization and burial of kelp-derived organic matter and how these may be impacted by anthropogenic- and climate- related changes in the environment.
- **Kelp-fisheries interactions** There are methodological, geographical, and logistical gaps that should be filled in order to get a broader understanding of interactions between kelp beds and fisheries.
- **South African Kelp Ecosystems** Since the Kelp Bed Ecology Programme of the 1970s and 1980s, there has been no concentrated research effort afforded to South African kelp ecosystems. Future research directions are likely to be centered around the impacts of climate change, overfishing and invasive species on kelp forest ecosystems.
- **Harmonization of Marine Macroalgal Monitoring** There is a need to harmonize marine macroalgal monitoring, identifying common metrics and approaches in sampling design, field measurements, taxonomic resolution and data management, in order to develop standardized procedures which may allow data obtained to be compared.

Use AI to generate some synthetic data which you may use to develop, test, and implement an unknown statistical method. Again, giving it a full, detailed background to start its reasoning from will get you much further:

i Example 3

Your initial prompt: We can measure algal nutrient uptake rates using two types of experiments: multiple flask experiments and perturbation experiments. The fundamental concept underlying both methods is to introduce a known quantity of nutrients (termed the substrate) into a flask or a series of flasks and then measure the rate of nutrient uptake (V) at different substrate concentrations ($[S]$). We calculate the nutrient uptake rate as the change in nutrient concentration in the flask over a predefined time interval ($V = \Delta[S]/\Delta t$). Consequently, both experiments generate data that relate the nutrient uptake rate to the corresponding substrate concentration. The primary difference between the two methods lies in the experimental setup and the data analysis.

In the **multiple flask method**, we prepare a series of flasks, each containing a different initial concentration of the substrate nutrient to span the range typically encountered by the specimen in its natural environment. We then measure the nutrient uptake rate in *each individual flask* over a specific time period, for example by taking measurements at the start ($t = 0$) and end ($t = 30$ minutes) of the incubation. We calculate the change in substrate concentration over this time interval in each flask to determine the corresponding nutrient uptake rate. The resulting data from this method therefore consists of the different initial substrate concentrations used in each flask, paired with their respective measured nutrient uptake rates over the incubation period.

What statistical test would you recommend?

The next iteration on the first prompt Let's go with the Michaelis-Menten model.

I use R. Use a simulated dataset and demonstrate how to fit the MM model to the hypothetical uptake data.

And then refine it further Below I will paste the output of the non-linear regression you suggested (as above). Please write up these findings in English suitable for the results section in a publications (e.g. the journal Marine Biology):

```
Formula: V ~ (Vmax * S)/(Km + S)
Parameters: Estimate Std. Error t value Pr(>|t|)
Vmax      9.7239      0.3907  24.891 2.14e-15 ***
Km         0.8621      0.1297   6.647 3.08e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.4873 on 18 degrees of freedom
Number of iterations to convergence: 4 Achieved convergence tolerance:
1.493e-07
```

Use AI to develop a specific data analysis workflow suitable to your task. Point to specific localities on your computer where the data reside, and be as informative as possible about the nature of the data, including what the variables are called, etc. Although the resulting R script will not run on the AI system, it will be a good starting point for you to adapt and run on your own computer (possibly involving subsequent steps of iterating through AI). For example, you could ask:

i Example 4

I have created a grid template with a predefined spatial extent and resolution as follows:

```
template_grid ← expand.grid(lon = seq(11, 20, by = lon_increment), lat = seq(-35, -17, by = lat_increment))
```

I need to regrid MUR SST data, which are situated at “/Volumes/OceanData/Tom/MUR” as a series of .rds files.

The content of the .rds files is the variables “lon”, “lat”, “t” (time, in Date format, e.g. “2014-06-02”), and “temp” (sea surface temperature).

I want to regrid these files to the `template_grid` and collect all the data in one combined .Rdata file at the end.

Please provide an R script to accomplish this.

For lecturers, use it to make sense of student tasks and assignments submitted on iKamva. For example, you could ask to extract marks for self-assessed assignments from a set of files. Ask the AI to look for specific keywords in the text, or to extract specific information from the text. As an example:

i Example 5

Please create a Python script to accomplish the following:

The directory ‘/Users/ajsmiit/Library/CloudStorage/Dropbox/BCB744/2025/sandbox’ has the following subdirectories, e.g.:

‘Task A’ > ‘MCCOMB, JODY(3650596)’ > ‘Submission attachment(s)’ > ‘task_a_completed_by_jody_mccomb_3650596.R’

or

‘Task A Self-Assessment’ > ‘MCCOMB, JODY(3650596)’ > ‘Submission attachment(s)’ > ‘task_a_completed_by_jody_mccomb_3650596.xlsx’

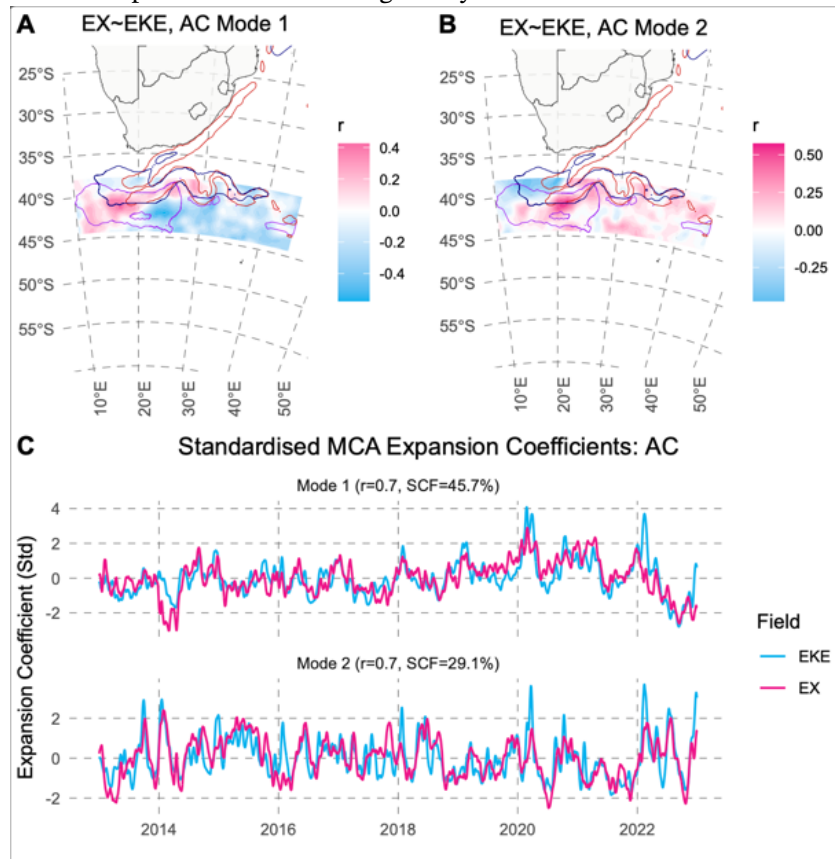
etc.

1. Delete all files named ‘timestamp.txt’
2. Find all the files with extensions ‘.R’, ‘.html’, ‘.xlsx’, ‘.qmd’, ‘.pdf’, ‘.docx’, or ‘.txt’ in these subdirectories and rename them to e.g., ‘MCCOMB, JODY(_3_6_5_o_59__6).R’ or ‘MCCOMB, JODY(_3_6_5_o_59__6).xlsx’ (this is the student Surname, Name (Student_no) and the file extension). Most of these details are supplied in the naming scheme of the subdirectories, as indicated in the examples.
3. Copy all these renamed files in ‘Task A’ and ‘Task A Self-Assessment’ to ‘Task A processed’. Similarly, renamed files in ‘Task B’ and ‘Task B Self-Assessment’ will be copied to ‘Task B processed’, etc.
4. Remove all the original subdirectories, e.g. ‘Task A’ > ‘MCCOMB, JODY(_3_6_5_o_59__6)’ > ‘Submission attachment(s)’ and any remaining files within any level of these subdirectories.

i Example 6

The attached image show a maximum covariance analysis on two gridded fields: SST and eddy kinetic energy over the period 2013 to 2022. The timeseries was detrended prior to analysis. Please help me understand how to interpret this figure.

<Also paste the image you want AI to analyse..., e.g.>



APPLICATIONS OF AI IN ACADEMIC WORK

Now that we understand the diversity of GPT models, their common basis of operation, and the importance of prompts, I will look at some of the ways in which we can use AI in our academic work. By way of examples, I'll cover the following topics:

Research and Working With Ideas

- Literature reviews: Gemini Advanced and Perplexity's deep research; SciSpace Deep Review for focussing solely on academic sources and finding more relevant papers faster, and to export references
 - Finding relevant papers
 - Summarising papers

- Extracting key points
 - Generating literature reviews
 - Writing literature reviews
- Structuring and mapping our ideas and thoughts
 - Outlining
 - Mind mapping
 - Concept mapping
 - Structuring papers
- Brainstorming ideas
- Deeper research
- Facilitate interdisciplinary collaboration
- Staying updated on research trends
- Summarise influential researchers
- Help with public outreach and science communication

Writing

- Rewriting
- Summarising
- Paraphrasing
- Generating text
- Validating ideas, concepts, and factual accuracy
- Reviewing
- Editing
- Proofreading
- Referencing (!)

Data

- Cleaning
- Extracting data from PDFs (tables, figures, etc.)
- Scripting (e.g., R, Python)
 - convert English to code
 - convert code to English
 - problem solving (statistics and data analysis)
 - visualising
 - debugging
- Interpreting and double-checking findings

Personal Assistant

- Writing applications (building on existing work, adapting, updating)
- Writing emails (language, etc.)
- Others

SPECIFIC WORKFLOWS

The general strategy across all three academic outputs emphasises the proactive and intelligent use of AI tools to streamline research, enhance the quality of the work, and ensure adherence to academic best practices. Use these tools not just for basic information retrieval but for deeper analysis, identification of gaps, methodological awareness, and critical self-assessment.

A Literature Review

- Define the specific area (clear **topic** and/or **research questions**, increasing granularity to **aims** and **objectives**) or question that your literature review will address.
- Use **deep research tools** to conduct a broad search of the knowledge base, and at this early stage you might not yet focus on the peer reviewed literature. This is where you can use the AI to help you develop a broad overview of the topic. You can also use it to **generate a list of keywords and phrases that are relevant to your topic**. This will help you refine your formal literature search terms and find more specific peer-reviewed papers.
- Now, find the references. I prefer plain, old-fashioned **Google Scholar**, but you could use AI tools like **Gemini Advanced**, **Perplexity**, or **SciSpace Deep Review** to conduct broad searches and gather relevant academic sources. Tools like SciSpace are specifically designed for academic sources as they return only peer-reviewed papers. They also allow you to export references to your reference manager. **Consensus** is a great tool for finding the consensus on a topic, and it can also help you find relevant papers. But keep the fallibility of these systems in mind.
- **Find PDF copies** each and every reference that the above AI (and manual) searches reveal. Tools like Perplexity and Gemini link back to the original sources, but you need to verify them. SciSpace allows for easy export to reference managers. **Check each fact yourself!**
- Use the AI to **generate summaries** of the existing literature to get a broad understanding of the field and identify key themes and arguments. Here, **NotebookLM** is your friend. Depending on the topic, you may instruct the AI to focus on specific aspects, such as methodology, findings, or theoretical frameworks. As always, being very specific in your prompts will yield better results – it helps to already know the framework of the output that you are looking for. Discuss this with your supervisor or colleagues.
- Based on the initial output, you may refine your search terms and use filters (e.g., publication date, methodology, journal quality) to narrow down the most relevant and high-quality studies. Consensus is useful for understanding the consensus and quality of research.
- Look for **recurring themes**, **significant findings**, and **trends** in the literature. Develop an understanding of **how the field has changed and developed since its inception**. What are the **gaps**? What are the **opportunities**? What is the **state-of-the-art**? These should be a central outcome of a strong literature review.
- **Structure the literature review logically**, grouping related studies and synthesising their findings to build a coherent narrative around your topic. Tools like Gemini can provide an initial structure. Consensus can generate an outline.
- Periodically use deep research tools to search for new publications in your field to ensure your literature review is current.

A Thesis

- The core of your thesis should be a well-defined and arguable statement. Tools like Thesis.ai can help you evaluate if your thesis statement is compelling and addresses a significant question.

- A strong thesis is built upon a thorough understanding of existing research. Follow the literature review strategy outlined above to establish a solid foundation.
- Select research methods that are appropriate for addressing your thesis question. Deep research tools can help you discover the methodologies commonly used in your field. What are their strength and weaknesses? How have the methods been used in the past? What are the limitations of the methods? How can you improve on them?
- The thesis must contribute original research or analysis. AI tools can help you identify research gaps where your work can make a novel contribution.
- All arguments and conclusions in your thesis must be supported by robust evidence. You can use the AI to verify that your conclusions (which you wrote) are backed up by your analysis and the literature. This is where you can use tools like Consensus to check the consensus on your findings.
- Acknowledge any limitations of your research and suggest potential avenues for future investigation. AI tools can help you brainstorm potential future research directions based on your findings. It is often useful to ask different AIs to verify each others findings – areas where discrepancies are found will require personal effort to resolve.
- Ensure your writing, referencing, and overall presentation meet the highest academic standards. Tools like Thesis.ai can provide feedback on various aspects of your writing to help you achieve this. Search for consistency of presentation, heading structure, formatting, references, heading styles, and so on. Use the AI to check for consistency in your writing style, tone, and voice. When you're using multiple AIs, choose one to do the final polishing of your writing.
- Utilise AI tools to get feedback on individual chapters or drafts to identify areas for improvement before submission. Ask it to be act as an examiner and to provide feedback on the quality of your writing, the strength of your arguments, and the clarity of your presentation, the novelty of your work, identify any issues, point to the strengths, and so on.

A Research Paper

- Use AI to help you clearly define the question or problem your paper aims to address. This often stems from identified research gaps.
- Use deep research tools to focus on the literature directly relevant to your research question. Again, refer to the previous section on literature reviews for more details.
- Use it to clearly describe the methods. Ensure they are recognised and robust within your field (although you will have done this before you write the paper).
- Use AI the help you organise your results in a logical manner, using tables, figures, and text as appropriate. Get it to check cross referencing, to ensure consistency and proper referencing, the logical captioning of figures and tables, and many other fiddly things we need to do before submitting it to the journal.
- Use it to verify the interpretation and presentation of your results and to ensure that your conclusions are supported by the data.
- Highlight the importance of your findings and their potential impact on the field.
- Use it to find any limitations.
- Seek feedback before submission – for example, have three different AI systems play the role of referees.
- Use AI to confirm appropriate outlets for publishing your research. Does your work align with the journals scope?