UNIVERSITY
OF WOLLONGONG
AUSTRALIA

HDR HELPFUL HINTS

From the Dean of Graduate Research, Simon Moss.

Uses of AI in research - data analysis

SUMMARY

Generative AI tools can now significantly facilitate the analysis of data. For example, AI tools can

- offer advice on the best methods or techniques to utilise—and how to conduct these techniques effectively,
- generate the code that is necessary to apply these methods in R, Python, or other software,
- interpret the output these techniques generate,
- actually conduct the analyses—without the need to use other software, such as R or SPSS.

CAUTIONS

Generative AI has transformed the practices that researchers apply to analyse data. Whereas researchers might have dedicated weeks, or even months, to analyse their data, they can now complete the same analysis within days or even hours. Nevertheless, before you become too excited about these possibilities, you need to be sensitive to some complications. The following case studies illustrate these complications.

Use of social media
A PhD candidate, Sal, has collated social media posts from Facebook and other platforms about a particular topic: vaping. The candidate now wants to upload these data into Chat GPT and prompt

Chat GPT to uncover themes or insights from these data. The candidate believes that only Chat GPT can extract unbiased themes.

Problem 1: Inappropriate use of data

- When users of social media submit posts, they may not consider the possibility their data will be used in generative AI tools—especially if the social media group is private. To illustrate the problem, they may be concerned that employees of Open AI could access and read their posts. They may also be concerned their comments could shape future versions of these generative AI tools and perhaps in a way they had not intended.
- Partly to address this concern, social media platforms stipulate conditions that limit how the data can be used. HDR candidates would thus need to check they comply with these conditions—sometimes a challenging task.
- To address these matters, at the very least, HDR candidates should inform the human ethics committee of this use of generative AI and seek approval. For more information, read about "Using social media data in research" from this webpage.

Problem 2: Unsuitable research practices

- Generative AI tools are not unbiased. The biases that pervade society also shape the responses of these tools. In addition, to prevent offensive or inappropriate responses, these companies deliberately adjust their models to bias answers—biases that may diminish the degree to which the ensuing themes represent the underlying data.
- Unlike many research projects in which researchers convey their background to readers and thus acknowledge their biases, generative AI cannot delineate or regulate these biases in the same way.
- Similarly, many research paradigms, such as descriptive phenomenology, clarify how researchers can identify and perhaps temper their biases. In contrast, research paradigms have not clarified how to identify and temper the biases in many generative AI tools.

Reliance on generative AI

A Master of Philosophy candidate, Ram, used an AI tool, called Julius, to analyse quantitative data. Ram has accrued limited knowledge of statistics. However, to analyse the data, Ram merely uploaded an Excel spreadsheet, containing the data, into Julius and then entered the following prompt:

I am a researcher exploring whether relaxation exercises and diabetes—as well as the characteristics of individuals that affect this association. Can you clean the data, present relevant descriptive statistics, conduct statistical analyses to explore my research question, and generate useful graphs?

The tool then conducted the relevant statistical tests and reported the output. Ram then inserted a paraphrased variant of this output to his thesis.

Problem 1: Limited skills

- Several years later, to secure a job, Ram needed to demonstrate his skill in data analysis.
- In an interview, Ram indicated that he could use generative AI tools to analyse data.
- Ram was rejected. The employer recognised that anyone could use generative AI tools to analyse data. But the employer wanted someone who could identify when generative AI tools may generate misleading results—and thus sought an employee who understood statistics comprehensively.

Problem 2: Sensitive data

- The data that Ram collected contained sensitive information about individuals.
- Ram failed to check whether the data he uploaded was secure or could be released elsewhere.

ROLES OF AI IN QUANTITATIVE DATA ANALYSIS: GUIDANCE

Many generative AI tools, including Chat GPT, can advise you on how to analyse data. This advice might complement the guidance you receive from textbooks, YouTube videos, supervisors, and NIASMA: a data analysis service at UOW. The following table illustrates some prompts you can enter to guide your data analysis.

PRACTICE	EXAMPLES
First, ask the tool to identify a suitable technique to analyse your data—after describing your participants, manipulations, measures, and research questions.	• I asked 100 participants to specify the number of hours they engage in meditation and their age. Half of these participants had been diagnosed with diabetes, and half these participants had not been diagnosed. • I want to determine whether age affects the relationship between meditation and diabetes.

	• Which statistical technique should I utilise?
To learn about this technique, ask the tool to present an example.	• Can you provide an example to help me understand this technique?
Ask the tool to identify the software or platforms you could use to conduct this technique.	• You suggested that I conduct an ANCOVA. What software or tools can I use to conduct an ANCOVA?
Clarify some challenges you might need to consider when completing this technique.	• You suggested that I conduct an ANCOVA. What are some mistakes that researchers sometimes commit when they conduct an ANCOVA?
Continue to ask more questions about which methods can be applied to overcome these challenges or prevent these mistakes.	• You indicated that I should control for confounding variables. How should I achieve this goal?
If relevant, you can ask Generative AI tools to generate computer code to execute common operations, such as a statistical test. Chat GPT can generate helpful code, partly because computer codes together with annotations are often stored on Guthub or similar platforms.	• I asked 100 participants to specify the number of hours they engage in meditation and their age. Half of these participants had been diagnosed with diabetes, and half these participants had not been diagnosed. • What python or R code should I use to complete an ANCOVA?
After you conduct these analyses, prompt these tools to interpret these data.	• The ANCOVA generated a table. The first row indicates that exercise (1,43) = .543. The heading of this table is …. How do I interpret these results?

ROLES OF AI IN QUANTITATIVE DATA ANALYSIS: TOOLS THAT CONDUCT ANALYSES

Several tools will not only teach you to conduct quantitative data analysis but will also conduct the analyses. These tools include Deepnote, Grapha.ai, Julius AI, TalktoData, and Vizley. Chat GPT also occasionally release plugins to facilitate data analysis. To utilise these tools, you merely need to

- locate the buttons that enable you to upload your data—often as an Excel file or csv file,
- enter something like the following sequence of prompts.

PROMPTS	OUTPUT
Can you please tell me about the data—such as the number of cases for each variable and the prevalence of missing data?	• The tool will present information about the variables, such as missing data or duplicate records. • The tool may even interpret the variables. • For example, if a variable name is BMI, the tool will define this variable as body mass index.
Can you please clean the data and identify outliers	• The tool may identify and remove outliers. • The tool may perform other operations, such as replace missing values and so forth. • For each operation, the tool will generate the Python code—and you can then include this code in your appendices.
Can you please conduct exploratory data analysis	• The tool may conduct a range of analyses, depending on the data, such as correlation analyses or factor analyses.
Can you help me answer my research question. My research question is what characteristics of individuals affect the association between meditation and diabetes. Which statistics should I conduct? Can you please conduct these techniques?	• The tool will conduct the right analyses and generate the output.

Can you use simple language to interpret this output?	• The tool may generate a narrative that you can paraphrase and include in your chapter or paper.
Can you please generate some graphs to represent the findings	• The tool will generate a range of graphs. • If you do not like the graphs, you can ask "Can you improve the appearance of these graphs?" • You can also specify the journal requirements and ask the tool to comply with these requirements.
Can you create a decision tree as well?	• That is, the tool can perform some machine learning algorithms as well.

So, which of these AI tools could you utilise to analyse your data. Here are some attributes you should consider.

ATTRIBUTES	DETAILS
Cost	• The cost of these tools varies but is usually reasonable. Typically, you can trial the tools at no cost. • Some tools might cost only US $20 a month—and you might only need to use the tool for one month. • Some tools might cost more—such as US $300—but you can retain the tool indefinitely.
Security	• Some tools do not use the data to train future models and, therefore, the data is secure. • Some tools, such as Julius, also impose strict data retention policies—such as delete the data after one hour of inactivity.
Useability	• Most of these tools are simple to use. • To use Deepnote, you need to click a purple button to open the relevant box. • To use Julius, click "Start a chat" to enter a prompt. • Nevertheless, you should utilise trial the tool, usually at no cost, to decide which alternative is simplest.
Specialisation	• Some tools specialise in specific activities. • Grapha.ai, for example, is especially useful if you want to generate useful graphs.

You need to consider two complications of these tools. First, if you depend on these tools, however, you may not develop the requisite quantitative skills to secure relevant jobs. So, you should use these tools not only to conduct the analyses but to learn about data analysis as well. Second, the data may not always be secure. Therefore, whenever using these tools, you should either

- analyse data that are not private—such as deidentified data or public datasets,
- check the tool is entirely secure.

Tools are secure if

- they indicate they do not use the data to train their models,
- they do not retain data for more than a limited time,
- they have introduced policies to manage data responsibly.

In practice, to confirm that an AI tool is secure, rather than merely read the policies, you could

- contact the human or animal ethics committee, depending on your project,
- ascertain whether your supervisor or other colleagues in your school have utilised this tool and established the tool is secure..

ROLES OF AI IN QUALITATIVE DATA ANALYSIS: GUIDANCE

Generative AI can also be useful to guide or to conduct analysis of qualitative data, such as interview transcripts. First, you can use generative AI to recommend suitable practices, as illustrated in the following table.

PRACTICE	EXAMPLES
First, ask the tool to identify a suitable method or technique to analyse your data—after describing your research.	• I have interviewed 20 participants about problems they experienced during internships. • What techniques could I use to identify the main problems—and to understand the relationship between these problems?
Perhaps clarify which variant of this technique is most suitable	• You recommended thematic analysis. Have researchers developed multiple version of thematic analysis?

	• If so, can you describe each variant—and recommend which variant may be suited to my circumstances.
Then, seek advice about these methods or techniques	• You recommended reflexive thematic analysis. Can you summarise the key features of this technique? • What steps do I need to complete?
Clarify the challenges of this technique?	• What are common mistakes that researchers commit when they conduct reflexive thematic analysis? • How can I avoid these mistakes?

Second, you can ask an AI tool to identify themes from some qualitative data. However,

- you would need to acknowledge the use of AI and the biases of these AI tools,
- you would need to recognise that AI tools have not developed your expertise about the topic and thus might not uncover suitable themes,
- you would need to check the tool is secure and the data would not be released.