

CENTRUM
SPOLOČENSKÝCH
A PSYCHOLOGICKÝCH VIED



LARGE LANGUAGE MODELS IN (NOT ONLY) PSYCHOLOGY TRAINING AND RESEARCH: A BRIEF INTRODUCTION

Matúš Adamkovič

LARGE LANGUAGE MODELS IN (NOT ONLY) PSYCHOLOGY TRAINING AND RESEARCH: A BRIEF INTRODUCTION

Matúš Adamkovič

2025

Centre of Social and Psychological Sciences, Slovak Academy of Sciences

Author

Matúš Adamkovič, PhD.
Slovak Academy of Sciences (Slovakia)
Charles University (Czechia)
University of Jyväskylä (Finland)

Reviewers

doc. Mgr. Pavol Kačmár, PhD.
Mgr. Jakub Rajčáni, PhD.

Graphical design

Natália Partilová

Funding

This work was supported by the Slovak Research and Development Agency (APVV-23-0336) and by the project PRIMUS/24/SSH/017.

Word count

18 575 words (126 561 characters)

Publisher

© Centre of Social and Psychological Sciences, Slovak Academy of Sciences
The publication was approved by © Editorial Committee, Centre of Social and Psychological Sciences, Slovak Academy of Sciences

**License**

CC BY-NC-ND 4.0

DOI

<https://doi.org/10.31577/2025.9788082980144>

ISBN

978-80-8298-014-4

Proofreading

The textbook was proofread and its writing style refined by an LLM.

Table of Contents

Introduction	4
Ethical Considerations	11
1 Large Language Models – An Overview	15
1.1 <i>Understanding Large Language Models</i>	15
1.1.1 What are Large Language Models?	15
1.1.2 How do Large Language Models Work?	15
1.1.3 Tokens and Contextual Windows	16
1.1.4 User Interface vs API Usage of LLMs.....	16
1.2 <i>Prompting: The Key to Leveraging LLMs</i>	17
1.2.1 What is Prompting?	17
1.2.2 How Prompting Works.....	17
1.2.3 Types of Prompts	18
1.2.4 Examples of Good and Bad Prompts by Category.....	18
1.2.5 Examples of Prompts in Psychology Research.....	20
1.2.6 Considerations and Limitations of Prompting in LLMs	21
1.3 <i>LLM Fine-tuning and Output Control</i>	25
1.3.1 Fine-tuning LLMs.....	25
1.3.2 Leveraging Follow-up Prompts	27
1.3.3 Determinism versus Creativity in LLM Outputs: Temperature Settings.....	28
1.3.4 Other Methods for Controlling LLM Outputs.....	32
2 Large Language Models in Psychology Training	37
2.1 <i>Adaptive Learning</i>	37
2.2 <i>AI-Generated Case Studies and Scenarios</i>	38
2.2.1 Generating Diverse Scenarios	38
2.3 <i>Programming Skills</i>	40
2.3.1 Programming Tools and Languages for Psychology Students.....	40
2.3.2 Programming in Psychology Curricula.....	40
2.3.3 Programming Learning Examples from R	41
3 Large Language Models in Psychology Research	46
3.1 <i>Research Design Improvement</i>	46
3.1.1 Generating Scenarios.....	46
3.1.2 Tweaking Designs	48
3.1.3 Reviewing Your Research.....	49
3.2 <i>Data Extraction</i>	50
3.2.1 Extracting Specific Information.....	50
3.2.2 Summarizing Texts.....	51
3.3 <i>Writing Refinement and Proofreading</i>	52
Conclusion	55
References	58

This world doesn't need a hero, it needs a professional.

— Geralt of Rivia

Introduction

Convergence of generative AI and Behavioral Research: Bridging the Gap

The way we analyze information, make decisions, and push the boundaries of knowledge is undergoing a fundamental shift. At the center of this transformation is artificial intelligence (AI), a technology that has moved far beyond simple automation to become a powerful tool for uncovering patterns, generating insights, and solving complex problems. What began as an effort to streamline tedious, repetitive tasks has evolved into a system that enhances human cognition, accelerates discovery, and optimizes resource allocation. AI is not merely a substitute for human effort—on the contrary, it expands our ability to process vast amounts of data, recognize hidden relationships, and tackle challenges that once seemed difficult to overcome. Its applications span across disciplines, from healthcare and finance to education and behavioral science, where it is reshaping methodologies, fostering innovation, and redefining what is possible in modern science.

In the rapidly advancing field of technology, the integration of generative artificial intelligence (AI) and psychology has emerged as a promising area for exploration. This development is not just a technological advancement but a major shift that has the potential to transform how we conduct and understand training and research in the behavioral sciences.

The utilization of gen AI tools is steadily rewriting the rulebook of traditional learning and research methodologies. Imagine a world where learning about complex problems is interactive and tailored to one's specific needs, where vast reservoirs of data can be navigated with ease, and where patterns and insights that once took months to uncover are revealed in moments. The implications of such technology in psychology are profound.

The introduction of advanced generative AI tools, such as large language models (LLMs), marks a significant milestone. LLMs are a class of AI systems designed to understand and generate human-like text based on the vast amounts of data they are trained on. These models, such as GPT, Gemini, Bard, Claude, are built on transformer architectures

that allow them to process and produce text with high coherence and relevance (see Solanki & Khublani, 2024). By leveraging deep learning techniques, LLMs can perform a variety of tasks, from answering questions and summarizing documents to generating creative content and engaging in conversation. Their ability to grasp context, interpret nuances, produce detailed responses, and understand different languages makes them powerful tools in both academic and commercial applications.

Certainly, skepticism is well-placed. It would be naïve to expect gen AI to instantly solve the complex problems and questions of behavioral sciences, especially given the nuanced, intangible, and often ephemeral nature of the phenomena. However, generative AI can provide highly useful tools for enhancing the effectiveness of the research process. For example, LLMs can extract information from large volumes of data for meta-scientific purposes, assist in refining experimental designs, increase reproducibility (e.g., through generating or debugging analytical code), and help reduce human biases or genuine errors (e.g., Guo et al., 2024; Han et al., 2023; Konet et al., 2024; Yenduri et al., 2023). Nonetheless, LLMs also present a relatively easy means to misuse or even abuse AI power, making ethical considerations and careful use essential. LLMs are a significant addition to the long tradition of human-created tools: they can be enormously beneficial, but we must learn to use them properly. The use of LLMs, therefore, is not only about discovery and academic pursuit but also about responsibility and foresight (Jobin et al., 2019). This comprehensive approach ensures that while we push the boundaries of what is possible, we remain anchored to the principles of ethical conduct and societal benefit.

This textbook aims to be a comprehensive guide, illuminating the possibilities of using LLMs in psychology training and research. It is intended for readers, primarily university students, who are looking to keep pace with these rapid advancements, as well as those eager to actively engage with, adapt to, and harness the potential of LLMs in their careers or even daily lives. We will delve into the intricate mechanics and theoretical underpinnings of these AI models, exploring their multifaceted applications in psychology—ranging from adaptive learning and simulation-based training, to the augmentation of research capabilities and the offering of new insights into longstanding questions.

Why Focus on LLMs in Psychology?

The realm of psychology, with its intricate and diverse exploration of human behavior, cognition, and emotion, presents an ideal landscape for the integration of advanced gen AI technologies like LLMs. These sophisticated tools, with their capacity to process, generate, and analyze language, are not just supplementary resources; they are transformative agents that can significantly enhance the scope, efficiency, and depth of

psychological research, education, and clinical practice. A sneak peek at some of the applications is provided below, with these possibilities explored in greater detail in the corresponding chapters.

Advancing Psychology Education and Training

LLMs can serve as interactive partners in the learning process (e.g., Holmes et al., 2019). Here are some of the main applications:

Improve Learning

LLMs can be programmed to adapt teaching methods and materials to individual students' learning styles and needs, making psychology education more accessible and effective. They can provide personalized learning experiences by adjusting the difficulty of material based on student performance, ensuring that each student progresses at an optimal pace.

Interactive Learning Environments

LLMs can enhance psychology training by simulating various scenarios that can be interactively analyzed and discussed. For instance, in professional development, LLMs can create realistic workplace situations, allowing trainees to practice and refine their conflict resolution, leadership, and decision-making skills. In educational settings, LLMs can simulate classroom dynamics, helping teachers develop and test new instructional strategies. These simulations generate synthetic data, providing a risk-free environment for trainees to learn and hone their skills through detailed analysis and feedback, ultimately leading to more effective training outcomes.

Automated Feedback and Assessment

LLMs can be used to provide instant, personalized feedback on students' assignments or assessments, enhancing the learning process and freeing up time for educators to focus on more complex teaching tasks. This can include evaluating written assignments, generating constructive feedback, and even suggesting areas for further study based on student performance on various tasks.

Enhancing Research Methodologies

LLMs can be used for numerous tasks, significantly advancing traditional research techniques in psychology (e.g., Bubeck et al., 2023; Polak & Morgan, 2024). Few of the possibilities are depicted below.

Design Improvement

LLMs can significantly improve multiple stages of the research process. For instance, they can assist with generating experimental materials, modifying existing designs, and introducing novel elements that researchers might not have considered. Beyond design, LLMs can serve as valuable consultants, helping to evaluate the robustness of proposed methodologies and suggesting refinements or alternative approaches. Additionally, they can help identify potential weaknesses in research protocols, offering insights into areas that may benefit from further development or adjustment, leading to more rigorous and innovative studies.

Data Extraction

With the ever-growing volume of scientific literature, LLMs can be highly effective in streamlining processes related to data extraction and summarization. They can condense complex texts, highlight key points, and respond to specific prompts from the reader. LLMs are also highly useful for systematic reviews, as they can efficiently extract specific data, enhancing the feasibility and scalability of meta-studies.

Data Interpretation

Trained on vast amounts of data, LLMs are well-versed in a broad spectrum of methodologies and analyses, both quantitative and qualitative. In addition to helping with analytical code, these models can assist in interpreting the results of complex statistical analyses, offering additional explanations and insights that help clarify findings for researchers. Beyond numerical data, LLMs are also highly effective in text analysis, making them particularly useful for examining open-ended responses, identifying recurring themes, or even performing sentiment analysis.

Conclusion

In conclusion, the application of LLMs in psychology promises not only to enhance the efficiency and effectiveness of research, education, and clinical practice but also to revolutionize our understanding of the human mind and behavior. By harnessing the capabilities of these gen AI models, psychologists can gain deeper, more complex insights into various phenomena, while also improving learning experiences and clinical practices. This synergy between psychology and AI could open the door to a new era of innovation and improved mental health outcomes.

Human-in-the-Loop

A reader might ask—can we really trust AI that much? Where is the human in all of this? Indeed, the concept of human-in-the-loop involves maintaining active human oversight and intervention at various stages of an LLM's operation, ranging from initial prompt design to the final evaluation of generated content. Rather than allowing the model to function fully autonomously and accepting its output at face value, users should continuously guide and review the results to ensure alignment with task goals, ethical considerations, and domain-specific standards. In psychology research and training, this oversight is particularly crucial due to the sensitive and nuanced nature of the subject matter, where an unchecked LLM might produce misleading or ethically problematic content. A human-in-the-loop framework integrates expert supervision throughout the entire process, primarily focusing on prompt review and iterative refinement, output monitoring, and adherence to ethical standards while mitigating biases. These aspects, which we will touch upon further throughout the textbook, are essential for ensuring that AI remains a responsible and effective tool in AI-assisted research and training.

Aims and Scope

This book aims to cover several important goals that connect traditional psychological research and training methods with the new possibilities offered by generative AI technologies. By focusing on these goals, the book seeks to show how these advanced tools can be effectively and ethically used in psychology. In particular, we will:

1. Explore the Integration of Generative AI in Psychology

This book aims to explore the integration of LLMs in the field of psychology. It will provide a comprehensive overview of how these advanced AI tools can be utilized to enhance psychology, offering insights into both established methods and innovative approaches.

2. Present Practical Applications and Case Studies

The book will present a variety of practical applications and case studies demonstrating the effective use of generative AI in psychology training and research. The hands-on examples of integrating LLMs into these processes will provide readers with concrete illustrations of AI's transformative impact on the field.

3. Address Methodological Considerations

An important aim of the book is to address the ethical and methodological considerations associated with the use of generative AI in psychology. This includes discussing the

challenges of bias, privacy, and data integrity, as well as offering guidelines for the responsible and ethical implementation of AI tools in research.

4. Foster Innovation and Future Directions

This book aims to inspire innovation by exploring future directions for the use of generative AI in psychology training and research. It discusses emerging trends, potential new applications, and the evolving role of AI in psychology. By fostering a forward-thinking mindset, the book encourages students and researchers to explore possibilities and push the boundaries of integrating psychology with LLMs.

Structure of the Book

The book begins by laying a foundation in the basics of large language models (LLMs), covering their architecture, training processes, and the revolutionary advancements they represent. Following this, we delve into diverse applications of LLMs in behavioral research, showcasing case studies and real-world examples that highlight how these models can enhance research practices as well as learning methods. For each topic, the book provides examples and practical tips for efficient integration of gen AI tools into various stages of training and research. We will also engage in a critical discussion of the ethical implications of using AI in psychological research, addressing complex issues such as biases, privacy, data security, and the future of AI-human interaction. This comprehensive guide aims to equip both students and researchers with the knowledge, skills, and ethical awareness needed to leverage gen AI's transformative potential in the field of psychology.

Who This Book Is For

The book is primarily aimed at psychology students, but also researchers and practitioners interested in implementing novel methods in their work and education. This book might also provide some food for thought for other enthusiasts, including, but not limiting to meta-researchers and data scientists. As we embark on this journey together, the hope is that this textbook not only informs and educates but also inspires the readers to innovate and push the boundaries of what is possible in the realm of behavioral research through the transformative power of gen AI.

Author's Positionality

I have seen many people trying to integrate LLMs into their daily lives, work, or studies. Sometimes they succeeded, but often they did not. Based on my anecdotal experiences, most people struggle with LLMs and become frustrated with their outputs, primarily because they lack a clear understanding of how these models operate and how to use

them effectively. In my team, we began incorporating LLMs into our research in various ways, primarily to make the process more efficient (and fun). We have experienced many failures but learned a lot along the way. These observations and the learning process are what inspired me to write this book. I would like to disclose my strong support for integrating LLMs into psychology research and training. The immense capabilities of these tools have the potential to revolutionize the field. Much like the introduction of calculators and computers, there is initial skepticism surrounding technological advancements, yet they have become indispensable in both our daily lives and professional (academic) activities. It is neither functional nor practical to ignore or neglect their presence; instead, we should actively explore how to benefit from their capabilities while remaining mindful of their limitations. It is crucial to approach the implementation of LLMs with both care and transparency. While they are not miraculous or error-free, they can greatly benefit the field when used properly. As a matter of fact, LLMs—GPT-4o and Claude 3.5 Sonnet, in particular—have greatly helped me in writing this textbook. Nonetheless, by the time this textbook is published, these models are likely to be outdated. Despite the vast benefits, we must remain vigilant about ethical considerations and ensure that these tools are used responsibly. By doing so, we can harness the power of LLMs to advance psychology while maintaining the integrity and trust essential to the discipline.

Ethical Considerations

I have long considered how and where to include a discussion about the ethical considerations of generative AI usage. After completing the beta version of the textbook, I decided to place this discussion before diving into the specific topics. However, ethical concerns related to the use of LLMs are also addressed in various sections throughout the book.

The integration of LLMs into psychological training, research, and practice presents numerous opportunities for advancing the field. However, it also introduces a range of ethical considerations that users must carefully navigate. It is crucial to recognize that the ethical and legal obligations surrounding LLM usage are in a state of rapid evolution. As these technologies continue to advance and their applications expand, regulatory frameworks and ethical guidelines are continuously being developed, updated, and implemented. LLMs users must remain vigilant, staying informed about the latest ethical standards and legal requirements governing LLM use in their field, as what is considered acceptable practice today may change significantly in the near future. While this book will occasionally address ethical and legal considerations, it is not intended to serve as a definitive authority on these matters. Nevertheless, I open with this section examining the critical ethical implications of integrating LLMs into psychological research and practice. This foundational discussion aims to provide readers with a framework for considering the broader impacts and responsibilities associated with LLM usage. I summarize WAME's recommendations on using generative AI in scholarly communication, and then proceed with broader, common-sense considerations. However, more guidelines and perspectives on the responsible and ethical usage of LLMs (e.g., Cacciamani et al., 2023; Knoechel et al., 2024) are being published these days and might be of interest to you.

Recommendations on Generative AI in Academic Publications

Following the rapid rise in popularity of generative AI tools, organizations have sought to catch up by establishing guidelines for their responsible use. Here, I refer to two such examples: first, the World Association of Medical Editors (WAME) recommendations (Zielinski et al., 2023) on the use of generative AI in scholarly publishing; and second, Harvard's guidelines (HUIT, 2025) for the use of generative AI. WAME outlined five main principles: (1) LLMs cannot be listed as authors of academic work; (2) authors should transparently disclose how LLMs were used, including providing the prompts used for novel text generation and analytical work; (3) authors must take full responsibility for the content generated by an LLM in their publication; (4) the use of LLMs in the peer review process should also be acknowledged; and (5) editors should have appropriate tools to detect AI-generated content. Similarly, HUIT advises to (1) protect confidential data; (2)

review content (emphasizing the author's accountability); (3) adhere to local policies, (4) be cautious of phishing; and (5) use approved AI tools.

Before employing LLMs in academic communication, I highly recommend consulting local guidelines and standards (e.g., those of universities, journals, or academic societies) on LLM usage.

Privacy and Confidentiality

LLMs process vast amounts of text data, which in psychological contexts may include sensitive personal information. Students, researchers, and practitioners have to ensure:

- Individual privacy is protected.
- Intellectual property rights are protected.
- Clients and research participants are informed about the use of LLMs in data analysis or interaction.

Consider: Develop strict protocols for data handling when using LLMs, and regularly audit these processes to ensure compliance with ethical standards and legally binding data protection regulations.

Bias and Accountability

LLMs can perpetuate or amplify existing biases present in their training data:

- Racial, gender, or cultural biases may influence LLM outputs in ways that could skew research results or affect (clinical) judgments.
- Over-reliance on LLMs could lead to the marginalization of underrepresented groups whose experiences may not be well-represented in the training data.
- Given the potential for errors in numerical tasks, it is crucial to double-check the results produced by LLMs to ensure accuracy and reliability.
- LLMs occasionally make mistakes and produce non-existent content (hallucinate). Users are advised to verify the credibility of the content produced by LLMs.

Consider: Maintain clear lines of human oversight and responsibility when using LLMs. Document and be prepared to explain the role of LLMs in research methodologies or training tools. Implement bias-checking procedures in research methodologies, and critically examine LLM outputs for potential biases. Strive for diverse representation in any data used to fine-tune or prompt LLMs. Overall, psychologists need to develop competence in understanding LLMs and using them ethically

Deception and Authenticity

LLMs can generate human-like text, raising concerns about:

- The potential for generating made-up primary research, reviews, data or case studies that seem authentic.
- Blurring the lines between human-generated and AI-generated content.

Consider: Disclose the use of LLMs and maintain a commitment to authenticity in all aspects of your work.

Impact on Human Relationships

In psychological training and clinical settings, the use of LLMs could affect the therapeutic relationship:

- Over-reliance on AI-generated insights might diminish the human aspect of psychological work.
- Participants and clients might develop unrealistic expectations about the capabilities of AI in mental health care.

Consider: The role of LLMs is supportive. They should not replace human expertise and empathy in psychological practice.

Ecological and Social Impact

The development and use of LLMs have broader implications:

- Environmental concerns due to the high energy consumption of training and running large AI models.
- Potential exacerbation of digital divides, as access to advanced LLMs may be limited to well-resourced individuals and institutions.

Consider: Consider the broader impacts of LLMs and advocate for responsible and sustainable generative AI usage.

Long-term Consequences

The long-term effects of integrating LLMs into psychological research and practice are not yet fully understood:

- How might reliance on LLMs change the nature of psychology research and practice over time?

- What are the potential (unforeseen) consequences of integrating this technology?

Consider: Engaging in ongoing dialogue and research on the long-term implications of LLM use in psychology training and research is crucial for the meaningful progression of the field.

Conclusion

As LLMs become increasingly integrated into psychological research and practice, it is incumbent upon students and professionals in the field to approach their use with careful ethical consideration. By proactively addressing these ethical challenges, psychologists can harness the potential of LLMs while upholding important ethical principles such as beneficence, integrity, and respect for people's rights and dignity, which are essential to responsible psychological practice. Ethical guidelines for LLM use in psychology should be regularly reviewed and updated to keep pace with technological advancements. Psychologists must remain vigilant, critically evaluating the role of these powerful tools in their work and always prioritizing the well-being and rights of the individuals and communities they serve.

Questions before we start...

1. What is generative AI?
2. How should psychologists approach the integration of LLMs in their field?
3. What are the potential ethical concerns associated with the use of LLMs in psychological research and practice.
4. How might the use of LLMs in psychological research impact the reproducibility of studies? Consider both potential improvements and challenges.
5. What are potential benefits and potential risks of using LLMs in psychology practice?
6. What are novel ways that LLMs could be applied in psychological research or practice that were not explicitly mentioned in the introduction?

Chapter 1

Large Language Models – An Overview

Large Language Models (LLMs) have emerged as powerful tools in the field of artificial intelligence, with significant potential applications in psychological research. These models, trained on vast amounts of textual data, have demonstrated remarkable capabilities in understanding and generating human-like text. This chapter explores the relevance of LLMs to psychology, their potential applications in research methodologies, and the ethical considerations surrounding their use.

1.1 Understanding Large Language Models

1.1.1 What are Large Language Models?

LLMs are advanced artificial intelligence systems designed to process and generate human language. They are built using deep learning techniques, particularly transformer architectures, which allow them to capture complex patterns and relationships in language (e.g., Bharathi Mohan et al., 2024). LLMs are trained on diverse and extensive text corpora, enabling them to perform a wide range of language tasks without task-specific training.

1.1.2 How do Large Language Models Work?

LLMs operate on the principle of predicting the next word in a sequence based on the context provided by the preceding words. At their core, LLMs are complex neural networks trained on vast corpora of text data, learning patterns and relationships between words and concepts. During training, the model adjusts its internal parameters (which can number in the billions) to minimize the difference between its predictions and the actual next words in the training data. This process allows the model to capture intricate linguistic patterns, semantic relationships, and even some aspects of world knowledge. When given a prompt or a partial sentence, the model uses its learned parameters to generate probability distributions over possible next words, allowing it to complete sentences or generate coherent text. Importantly, the scale of these models—both in terms of the amount of training data and the number of parameters—allows them to capture nuanced linguistic phenomena and generalize to a wide range of tasks without task-specific training.

1.1.3 Tokens and Contextual Windows

Tokens are the fundamental units of text that language models process. They can be words, parts of words, or individual characters, depending on the tokenization method used (Yang, 2024). For example, the phrase "I love AI!" might be tokenized as ["I", "love", "AI", "!"]. There are two types of tokens based on their role in the model's processing: input tokens and output tokens. Input tokens are those provided to the model, representing the user's prompts, questions, or context. Output tokens are those generated by the model in response to the input. Both input and output tokens consume space in the contextual window.

The contextual window, also known as the context length or context window, refers to the maximum number of tokens an LLM can process or generate in a single operation. This window determines how much text the model can "see" and consider at once when analyzing or generating content. For instance, as of May 2024, GPT-4o has a context length of 128k tokens with output token limit capped to 2,048 (OpenAI, 2024).

The relationship between tokens and contextual windows is crucial for several reasons. The number of tokens in your input text must not exceed the model's context window. If it does, the text needs to be split into smaller chunks. The model can only generate responses up to the remaining space in the context window after considering the input. Larger windows allow the model to maintain coherence over longer texts, as it can reference information from further back in the conversation or document. Larger context windows also require more computational power and memory to process. As LLM technology advances, we're seeing a trend towards larger context windows, allowing for more sophisticated analysis and generation of longer, more complex texts. However, there's often a trade-off between context length and other aspects of model performance or efficiency.

1.1.4 User Interface vs API Usage of LLMs

LLMs can be accessed and utilized through two primary interfaces: User Interfaces (UIs) and Application Programming Interfaces (APIs).

UIs provide a graphical, often web-based platform where users can interact with the LLM directly, inputting prompts and receiving responses in a user-friendly format. This approach is intuitive, requires minimal technical expertise, and is ideal for exploratory research, brainstorming, or occasional use. Examples include ChatGPT's web interface or Claude's chat window.

On the other hand, APIs offer programmatic access to LLMs, allowing researchers to integrate these models directly into their research workflows, scripts, or applications. API

usage provides greater flexibility, automation capabilities, and the ability to process large volumes of data. It is particularly useful when consistent, repeatable interactions with the LLM are required (e.g., data extraction for meta-analysis). However, API usage typically requires programming skills and a deeper understanding of the LLM's parameters and functionalities. While both methods have their merits in psychological research, this textbook will primarily focus on UI LLM usage. This choice is made for two main reasons. First, to ensure accessibility for a broader audience and allow for immediate practical application without requiring programming skills, UI-based interactions with LLMs can effectively demonstrate core concepts, promote hands-on learning, and encourage creative exploration of LLM capabilities in psychology research and training contexts. Second, if a reader is using LLMs via API, it's safe to assume they're already familiar with everything discussed here—this book might just be a refresher for them.

1.2 Prompting: The Key to Leveraging LLMs

1.2.1 What is Prompting?

Prompting is the process of providing input or instructions to an LLM to elicit a desired output or behavior. It is the primary method of interaction with LLMs and can be thought of as a way of "programming" the model to perform specific tasks or generate particular types of content. The prompt serves as the initial context that guides the model's text generation process.

Please note that effective prompting is a skill that requires practice and refinement. Crafting a good prompt takes time—one often needs to experiment, iterate, and adjust inputs to achieve optimal results. Do not get frustrated—learning to communicate effectively with an LLM is a process that improves with experience. Patience and resilience are thus key.¹

1.2.2 How Prompting Works

Prompting works by leveraging the LLM's ability to predict the most likely continuation of a given text. By carefully crafting the input, researchers can guide the model towards producing outputs that are relevant to their specific needs. The effectiveness of prompting relies on several factors:

- **Clarity and specificity:** The more clear and specific the prompt, the more likely the model is to produce the desired output.
- **Context setting:** Providing relevant context in the prompt helps the model understand the desired frame of reference.

- **Task description:** Explicitly stating the task or desired output format can improve the model's performance.
- **Examples:** Including examples of the desired output (few-shot learning) can significantly enhance the model's ability to generate appropriate responses.

1.2.3 Types of Prompts

Several types of prompts exist, each suited to slightly different tasks. For the purposes of this textbook, I have selected four that I consider the most important for this context:

- **Zero-shot prompts:** These provide instructions without examples, relying on the model's general knowledge (Kojima et al., 2022).
- **Few-shot prompts:** These include one or more examples of the desired input-output pair, helping the model understand the task better (Brown et al., 2020).
- **Chain-of-Thought prompts:** These encourage the model to break down complex tasks into steps, often improving performance on reasoning tasks (Wei et al., 2022).
- **Role-Based prompts:** These instruct the model to assume a specific role or perspective when generating responses (Kong et al., 2023).

1.2.4 Examples of Good and Bad Prompts by Category

Several examples of both good and bad prompts are illustrated below.

1.2.4.1 Zero-shot Prompts

Bad Prompt: "Explain depression."

Why it is bad: This prompt is too vague and does not specify the desired depth, focus, or audience for the explanation.

Good Prompt: "Provide a concise explanation of major depressive disorder suitable for undergraduate psychology students. Include:

1. Key diagnostic criteria
2. Prevalence rates
3. Two main theories about its causes
4. Common treatment approaches Limit your response to approximately 200 words."

Why it is good: This prompt specifies the topic, target audience, desired content, and response length, guiding the model to produce a focused and appropriate output.

1.2.4.2 Few-shot Prompts

Bad Prompt: "Classify these statements as showing high or low self-esteem:

1. 'I'm proud of my achievements.'
2. 'I'm not good at anything.'

Now classify: 'I sometimes make mistakes but I learn from them.'"

Why it's bad: While this prompt does provide examples, it does not explicitly state the task or explain the classification criteria.

Good Prompt: "Classify the following statements as indicating either high self-esteem or low self-esteem. Base your classification on whether the statement reflects a generally positive or negative self-view.

Examples:

1. 'I'm capable of achieving my goals.' - High self-esteem
2. 'I always mess things up.' - Low self-esteem

Now classify the following statements:

1. 'I'm not as smart as everyone else.'
2. 'I can handle challenges that come my way.'
3. 'Nobody likes me.'

For each classification, provide a brief explanation of your reasoning."

Why it is good: This prompt clearly states the task, provides diverse examples with classifications, explains the classification criteria, and asks for explanations, which helps ensure the model understands the task correctly.

1.2.4.3 Chain-of-Thought Prompts

Bad Prompt: "Diagnose this case: A patient reports feeling sad and tired all the time."

Why it is bad: This prompt does not encourage a step-by-step reasoning process and provides very limited information.

Good Prompt: "You are a clinical psychologist conducting an initial assessment. A new patient reports feeling sad and tired all the time. Walk through your diagnostic reasoning process step-by-step. At each step, explain your thought process and what additional information you would seek. Consider:

1. What are the possible diagnoses based on the initial complaint?
2. What key diagnostic criteria would you need to check for each possibility?
3. What differential diagnoses should be considered?
4. What additional assessment tools or questions would you use?
5. How would you rule out medical causes?

Conclude with a preliminary diagnostic impression and explain what further steps you would take to confirm or refine this impression."

Why it is good: This prompt encourages a detailed, step-by-step reasoning process, mimicking the complex decision-making involved in clinical diagnosis. It guides the model to consider multiple factors and explain its reasoning at each stage.

1.2.4.4 Role-Based Prompts

Bad Prompt: "Act like a therapist and help with anxiety."

Why it is bad: This prompt is too vague and does not provide enough context about the client or the specific therapeutic approach to use.

Good Prompt: "Assume the role of a cognitive-behavioral therapist conducting a session with a client who has generalized anxiety disorder. The client is a 28-year-old software engineer who reports constant worry about work performance, social interactions, and health. In your response:

1. Start with a brief introduction and check-in with the client.
2. Ask 2-3 questions to gather more information about their anxiety symptoms and triggers.
3. Explain one CBT technique (e.g., cognitive restructuring or progressive muscle relaxation) that could help the client.
4. Guide the client through a brief exercise using this technique.
5. Assign a homework task related to the technique you taught.

Maintain a compassionate and professional tone throughout the simulated interaction."

Why it is good: This prompt provides a clear role (CBT therapist), client background, and a structured outline for the interaction. It guides the model to demonstrate key aspects of CBT while maintaining an appropriate therapeutic tone.

1.2.5 Examples of Prompts in Psychology Research

Here are some examples of how prompting could be used in psychology research contexts:

Thematic Analysis of Interview Transcripts:

Prompt: "Analyze the following interview transcript and identify the main themes related to social anxiety. List each theme and provide a brief explanation with supporting quotes from the transcript: [Insert anonymized transcript]"

Generating Experimental Stimuli:

Prompt: "Create 5 short vignettes (2-3 sentences each) describing everyday situations that might provoke social anxiety in adolescents. Ensure the scenarios are appropriate for a 14-16 year old age group and vary in intensity."

Literature Review Assistance:

Prompt: "Summarize the key findings from the following extended abstract on cognitive behavioral therapy for depression. Then, suggest two potential gaps in the research based on this study: [Insert abstract]"

Coding Open-ended Survey Responses:

Prompt: "The following are responses to the question 'How has the COVID-19 pandemic affected your mental health?' Code each response into one of the following categories: Negative Impact, Positive Impact, Mixed Impact, or No Impact. Provide a brief justification for each coding decision.

1. [Response 1]

2. [Response 2]

3. [Response 3]"

Generating Hypotheses:

Prompt: "Based on the following summary of research on social media use and adolescent well-being, generate three plausible hypotheses for future studies. For each hypothesis, provide a brief rationale based on existing findings: [Insert research summary]"

1.2.6 Considerations and Limitations of Prompting in LLMs

While prompting is a powerful tool for leveraging LLMs in psychological research and practice, it comes with several important considerations and limitations. Researchers must be aware of these to ensure the responsible and effective use of LLMs in their work.

1.2.6.1 Prompt Sensitivity

Prompt sensitivity refers to the phenomenon where small changes in prompt wording or structure can lead to significant differences in the LLM's output.

Implications:

- Inconsistent results: Slight variations in prompts might produce substantially different responses, potentially affecting research outcomes.
- Reproducibility challenges: reproducing findings becomes more complex when results are highly dependent on exact prompt wording.

Examples in psychology:

- Asking "Describe depression" versus "List the clinical indicators of major depressive disorder" might yield different levels of detail or focus.

Best practices:

- Standardize prompts: Develop and document standardized prompts for common research and training tasks.
- Test variations: Experiment with different phrasings to understand the range of possible outputs.
- History tracking: Keep track of specific wording of your prompts and changes over time.

1.2.6.2 Bias Amplification

Bias amplification occurs when poorly designed prompts inadvertently trigger or exacerbate biases present in the LLM's training data.

Implications:

- Skewed results: Biased prompts can lead to outputs that reinforce stereotypes or overlook diverse perspectives. Amplifying biases can have serious consequences, especially in clinical or educational settings.

Examples in psychology:

- A prompt asking about "typical family dynamics" might produce responses biased towards specific cultural norms.

Mitigation strategies:

- Use inclusive language: Craft prompts that are neutral and avoid assumptions about gender, culture, or other demographic factors.
- Explicit instructions: Include directives for the LLM to consider diverse perspectives or avoid stereotypes.
- Bias checking: Implement systematic review processes to identify and correct biased outputs.
- Collaborate: Work with experts in ethics and bias to review prompts and methodologies.

1.2.6.3 Over-reliance on LLM Outputs

Over-reliance refers to the tendency to accept LLM outputs as authoritative or definitively correct without appropriate scrutiny.

Implications:

- Misinterpretation of results: Treating speculative or generated content as factual can lead to erroneous conclusions.
- Neglect of empirical validation: Over-trusting LLM outputs might reduce emphasis on traditional research methods and peer review.

Examples in psychology:

- Accepting an LLM's interpretation of a novel therapeutic technique without empirical validation.
- Using LLM-generated case studies as definitive examples without acknowledging their hypothetical nature.

Safeguards:

- Critical evaluation: Always cross-reference LLM outputs with established literature and empirical data.
- Clear documentation: Explicitly state when content is LLM-generated and its limitations.
- Education: Train researchers and students in the appropriate use and interpretation of LLM outputs.
- Complementary use: Utilize LLMs as assistive tools rather than primary sources of information.

1.2.6.4 Ethical Considerations

Ethical considerations in prompting involve the responsible design and use of prompts, especially when dealing with sensitive topics or vulnerable populations.

Implications:

- Potential harm: Inappropriately designed prompts could lead to harmful or misleading information, particularly in clinical contexts.
- Privacy concerns: Prompts might inadvertently encourage the model to generate content that compromises individual privacy. Always consider what data could be used to feed a model.
- Informed consent: Participants should be aware when AI tools are being used in research or therapeutic contexts.

Examples in psychology:

- Queries that might lead to the generation of personally identifiable information in research contexts.

Best practices:

- Ethics review: Include LLM use and prompt design in institutional review board (IRB) processes.
- Sensitivity training: Educate researchers on crafting prompts that are sensitive to diverse populations and experiences.
- Clear guidelines: Develop and adhere to ethical guidelines specific to LLM use in psychological research and practice.
- Transparency: Be open with participants about the use of AI tools and their limitations.

Consider: If you are unfamiliar with certain functions or steps, consider breaking down the process into smaller, manageable parts, prompting for each one individually. Additionally, it is beneficial to ask the LLM for explanations of the code it generates, which can enhance your understanding of the underlying concepts and methods. Be attentive to common errors that can prevent your code from running correctly, such as typos in variable names or mismatched parentheses. Even small mistakes can lead to frustrating issues, so it's important to carefully review your prompt and code for accuracy.

1.2.6.5 Additional Considerations

Breaking the prompt: If a prompt does not produce the expected results, consider breaking it down into smaller, more manageable parts and prompting for each one individually.

Ask for reasoning: If an LLM produces an unexpected response or if the output's reliability feels uncertain, it can help to prompt the LLM to explain its reasoning (some models do this by default).

Main goal reminder: Engaging in a conversation with an LLM can be a lot of fun and a wonderful learning experience, sometimes leading to unexpected directions. If the flow of the conversation drifts too far, remind both yourself and the model of the main goal by asking it to summarize the initial or main purpose of the discussion.

Contextual limitations: LLMs may not always understand the full context of psychological theories or recent research developments.

Cultural sensitivity: Prompts and interpretations need to be culturally informed and sensitive, especially in cross-cultural psychology studies.

Temporal limitations: LLMs' knowledge cutoff dates may limit their awareness of recent developments in psychology.

Lack of personal experience: LLMs cannot provide personal insights or experiences, which are often valuable in psychological research.

By understanding these considerations and limitations, psychology students and researchers can better harness the capabilities of LLMs while maintaining scientific rigor, ethical standards, and the integrity of their work. Effective prompting requires a balanced approach that leverages the strengths of LLMs while actively mitigating their limitations.

1.3 LLM Fine-tuning and Output Control

LLMs can be adapted for specific tasks through a process called fine-tuning (see Solanki & Khublani, 2024). This section explores the concept of fine-tuning and how researchers can control LLM outputs, with a focus on temperature settings.

1.3.1 Fine-tuning LLMs

Fine-tuning is the process of further training a pre-trained LLM on a specific dataset or for a particular task. This allows users to tailor the model's capabilities to their specific needs.

Key aspects of fine-tuning:

- **Dataset preparation:** Curate a high-quality dataset relevant to the target task or domain in psychology.
- **Training process:** The model learns to adjust its weights based on the new data, enhancing its performance on domain-specific tasks.
- **Evaluation:** Assess the fine-tuned model's performance against baseline models and specific metrics relevant to psychological research.

Benefits for psychology research:

- **Improved domain specificity:** Enhance the model's understanding of psychological concepts and terminology.
- **Task adaptation:** Optimize the model for specific tasks like therapy dialogue simulation or psychological assessment.
- **Reduced hallucination and sloppiness:** Fine-tuning can help minimize the generation of false or irrelevant information in psychology-related queries.

1.3.1.1 Minimizing Hallucinations and Sloppiness

Sometimes, users may notice an LLM underperforming or generating responses that look plausible but are not true. This issue often arises from two related phenomena: hallucination and sloppiness. **Hallucination** refers to an LLM's tendency to generate plausible-sounding but inaccurate or entirely fabricated content (Huang et al., 2025). In this case, fictional references, unsubstantiated claims, or erroneous citations. This phenomenon arises because LLMs rely on probabilistic text generation rather than direct fact-checking, leading to outputs that may sound authoritative while lacking factual grounding. **Sloppiness** describes an LLM's propensity to produce imprecise, vague, or poorly reasoned responses—often due to insufficient context, ambiguous prompts, or limitations in the training data. Such output may contain half-formed arguments, truncated explanations, or inaccuracies that undermine clarity and reliability, especially in scholarly work. The accuracy and quality of LLM outputs are, however, improving steadily. One of the simplest ways to reduce hallucinations or sloppiness is to use newer versions of these models, which often draw on larger training datasets and refined architectures. Another effective strategy is to provide the model with clear, context-rich prompts and then follow up with additional queries to clarify ambiguous or incomplete answers. If you notice a potential inaccuracy or fabrication, challenge or correct the model and request a more detailed explanation. You should routinely verify the model's responses against reputable sources, and be especially cautious if the initial output seemed suspicious.

1.3.2 Leveraging Follow-up Prompts

Follow-up prompts are additional instructions or questions given to an LLM after its initial response. They are crucial for refining outputs, gathering more specific information, or steering the conversation in a desired direction. In psychology, effective use of follow-up prompts can significantly enhance the quality and relevance of LLM-generated content.

Key benefits of follow-up prompts:

- **Clarification:** Obtain more precise or detailed information on specific aspects of the initial response.
- **Exploration:** Delve deeper into particular topics or ideas mentioned in the initial output.
- **Correction:** Address any misunderstandings or inaccuracies in the model's initial response.
- **Expansion:** Encourage the model to consider additional perspectives or factors.
- **Focusing:** Direct the model's attention to specific areas of interest that may have been overlooked.

Strategies for effective follow-up prompts:

- **Be specific:** Craft prompts that target exact information or aspects you want to explore further.
- **Use open-ended questions:** Encourage more comprehensive and thoughtful responses.
- **Provide context:** Remind the model of relevant information from previous interactions if necessary.
- **Challenge assumptions:** Ask the model to consider alternative viewpoints or explanations.
- **Request examples or elaboration:** Ask for concrete instances or more detailed explanations of concepts.

Examples of follow-up prompts in psychology:

- *Initial prompt:* "Explain the concept of cognitive dissonance."
Follow-up: "How might cognitive dissonance manifest in decision-making processes?"
- *Initial prompt:* "Describe the symptoms of major depressive disorder."
Follow-up: "Can you differentiate these symptoms between adolescents and older adults?"

- *Initial prompt:* "Outline the key principles of cognitive-behavioral therapy (CBT)."
Follow-up: "Provide a hypothetical case study demonstrating the application of these CBT principles in treating anxiety disorders."
- *Initial prompt:* "Explain the bystander effect in social psychology."
Follow-up: "How might social media and online environments influence or alter the traditional understanding of the bystander effect?"
- *Initial prompt:* "Describe Piaget's stages of cognitive development."
Follow-up: "How have modern research findings supported or challenged Piaget's theory? Provide specific examples."

Best practices for using follow-up prompts:

- **Sequential refinement:** Use a series of follow-up prompts to progressively refine and focus the LLM's outputs.
- **Cross-verification:** Use follow-up prompts to check the consistency of the model's responses across different phrasings or contexts.
- **Hypothetical scenarios:** Employ follow-up prompts to apply concepts to specific, hypothetical situations relevant to psychological practice or research.
- **Critical analysis:** Use prompts that encourage the model to critically evaluate its own responses or common assumptions in the field.
- **Interdisciplinary connections:** Craft follow-up prompts that explore links between psychological concepts and other fields of study.

Follow-up prompts can significantly influence the depth and quality of LLM-generated content, potentially affecting outcomes by guiding the model towards more relevant and detailed responses. To ensure reproducibility and transparency, researchers should document their prompting pipelines. The strategic use of follow-up prompts not only helps generate more nuanced and context-specific insights, but also enhances the overall reliability and richness of the output. This practice is crucial for maintaining scientific rigor and advancing the understanding of complex psychological phenomena through the effective use of LLMs.

1.3.3 Determinism versus Creativity in LLM Outputs: Temperature Settings

Temperature is a key hyperparameter that controls the randomness of the model's outputs. Understanding and manipulating temperature allows users to balance between deterministic and creative responses (Döderlein et al., 2022).

1.3.3.1 Temperature scale

Temperature settings in LLMs typically range from 0 to 1, or higher in some implementations. Lower temperatures, closer to 0, result in more deterministic outputs,

ensuring consistency and predictability in responses. Conversely, higher temperatures, closer to 1 or above, lead to more diverse and creative outputs, allowing for greater variability and novelty in the generated content.

Please note that different LLMs have different possibilities for temperature adjustment in their UI. If unsure, consult the documentation of your preferred model.

1.3.3.2 Low temperature (Deterministic mode)

In low temperature, LLMs generate more consistent and predictable responses, with higher precision and a focus on the most likely next token. This setting tends to produce more factual and conservative outputs, ensuring reliability and accuracy in the generated content.

In psychology, these might include generating standardized responses for psychological assessments, producing consistent explanations of psychological theories that are faithful to originals, and simulating structured clinical interviews. These applications benefit from the mode's reliability and accuracy, ensuring uniformity and clarity in the generated content.

1.3.3.3 High temperature (Creative mode)

In high temperature, LLMs generate more diverse and varied responses, with an increased likelihood of novel or unexpected outputs. This setting can lead to more creative text generation, allowing for greater flexibility and originality in the responses.

In psychology, these might include brainstorming research hypotheses, generating diverse therapy scenarios for training purposes, and exploring potential interpretations in qualitative data analysis. These applications benefit from the mode's ability to produce flexible and original responses, fostering creativity and innovation in psychological research and practice

Best practices for temperature manipulation:

- **Experiment with settings:** Test different temperatures to find the optimal balance for each task.
- **Document choices:** Record temperature settings used in research protocols and publications.
- **Consider multiple settings:** Use various temperatures to compare outputs when appropriate.
- **Awareness of variability:** Understand that the same prompt at different temperatures can yield significantly different results.

Implications of temperature settings in LLMs are significant. The choice of temperature can substantially impact the model's outcomes. Users should well-consider and balance the need for creativity against the requirement for reproducibility and consistency. Different research questions or learning tasks may benefit from different temperature settings or from comparing outputs across multiple settings.

A few examples of different temperature settings and their corresponding outcomes are provided below:

Prompt: "What are the symptoms of major depressive disorder?"

Response (Temperature == 0.1):

"Major depressive disorder is characterized by a persistent feeling of sadness or a lack of interest in outside stimuli. Symptoms include significant weight change, sleep disturbances, fatigue, feelings of worthlessness or excessive guilt, difficulty concentrating, and recurrent thoughts of death or suicide. These symptoms must be present for at least two weeks and represent a change from previous functioning."

Response (Temperature == 1):

"Major depressive disorder manifests in various ways, often characterized by a pervasive sense of sadness and a loss of interest in activities that were once enjoyable. People may experience significant changes in weight, either gaining or losing a lot. Sleep disturbances are common, ranging from insomnia to excessive sleeping. There's a constant feeling of fatigue and a lack of energy, making it hard to get through the day. Emotional symptoms include feelings of worthlessness, excessive guilt, and a persistent sense of hopelessness. Concentration and decision-making can become difficult, and there may be recurrent thoughts of death or suicide. These symptoms need to be present for at least two weeks and must represent a change from previous functioning to be diagnosed as major depressive disorder."

Response (Temperature == 10):

"Major depressive disorder can show up in so many ways, it's like a roller coaster of feelings. One minute you might feel like there's a heavy cloud over you, making everything dull and lifeless, and the next you could be experiencing extreme changes in weight, like suddenly eating everything in sight or having no appetite at all. Sleep becomes a big issue too—either you can't sleep at all and find yourself staring at the ceiling all night, or you might sleep way too much and still feel exhausted. There's this overwhelming sense of fatigue that doesn't go away, no matter how much you rest. Emotionally, you might feel worthless or guilty about things that don't even make sense, and concentrating on even

simple tasks can feel impossible. The thoughts can get really dark too, with frequent thoughts about death or even planning suicide. It's this mix of physical, emotional, and cognitive symptoms that can last for weeks and make everyday life feel like a huge challenge."

Prompt: "Please suggest a twist to the Stanford Prison Experiment. Temperature == 0.1. Make it one paragraph long."

Response (Temperature == 0.1):

"A potential twist on the Stanford Prison Experiment could involve reversing the power dynamics by assigning the "prisoners" hidden authority over certain aspects of the guards' lives outside the experiment. For instance, the prisoners could be given the ability to impact the guards' academic standings or future job prospects through covert evaluations. This reversal could explore how perceived powerlessness in one context interacts with hidden influence in another, potentially revealing complex patterns of behavior and psychological stress when individuals navigate multiple, conflicting power structures simultaneously. The twist might shed light on how people behave when they hold power in some areas of life while appearing subordinate in others, and how this affects group dynamics and individual decision-making in high-pressure situations."

Response (Temperature == 1):

"Imagine a radical twist on the Stanford Prison Experiment where the participants are unknowingly placed in a simulated post-apocalyptic bunker environment. They're told that a catastrophic event has occurred outside, and they must work together to survive and potentially rebuild society. The experiment would introduce artificial scarcity of resources, simulated external threats, and periodic "updates" on the outside world. Unbeknownst to the participants, some are secretly assigned roles as "information brokers" with access to contradictory news about the state of the world. This setup could explore how humans organize themselves in extreme circumstances, how power structures emerge when traditional societal rules break down, and how the control of information influences group dynamics and individual behavior. The twist might reveal insights into human nature under existential stress and the malleability of perceived reality in isolated groups."

Response (Temperature == 10):

"Behold, the Stanford Prison Experiment: Interstellar Cheese Edition! Picture this: participants are whisked away to a top-secret moon base constructed entirely of various cheeses. Guards wield fondue forks as batons, while prisoners don spacesuits made of stretchy mozzarella. The twist? Every 12 hours, the entire base rotates, causing the cheese to melt and reform, constantly altering the power dynamics. Participants must

navigate an ever-changing labyrinth of Gouda corridors and Camembert cells, all while battling the maddening aroma of aged cheddar. The experiment aims to study how humans adapt to absurdist power structures and lactose-induced hallucinations. Will they crumble like feta or band together in dairy solidarity? Only time and stomach acid will tell in this utterly cheesy exploration of human behavior under fromage-based duress.”

1.3.4 Other Methods for Controlling LLM Outputs

While temperature settings are a crucial aspect of controlling LLM outputs, there are several other methods that psychology researchers can employ to fine-tune and guide LLM responses. Understanding these techniques can significantly enhance the utility and reliability of LLMs in psychological research and applications.

1.3.4.1. Top-k Sampling

Top-k sampling restricts the model to choose from only the k most likely next tokens at each step of generation.

Use in psychology:

- Useful for maintaining focus on the most relevant psychological concepts or terminology.
- Can help prevent the introduction of irrelevant or less common terms in clinical descriptions.

Example:

Prompt: "List the primary symptoms of major depressive disorder." With k=5, the model might output: "The primary symptoms of major depressive disorder include:

1. Depressed mood
2. Loss of interest or pleasure
3. Significant weight changes
4. Sleep disturbances
5. Fatigue"

This output sticks to the most common and well-established symptoms, avoiding less frequent or controversial indicators.

1.3.4.2. Top-p (Nucleus) Sampling

Also known as nucleus sampling, Top-p (Nucleus) Sampling method dynamically chooses from the smallest possible set of words whose cumulative probability exceeds the probability p .

Use in psychology:

- Provides a balance between diversity and relevance in responses.
- Useful for generating varied but pertinent examples in psychological case studies.

Example:

Prompt: "Describe a possible manifestation of social anxiety in a workplace setting." With $p=0.9$, the model might generate: "A 28-year-old marketing executive experiences intense fear when asked to present in team meetings. They might stutter, blush, or even call in sick to avoid such situations. Their heart races, palms sweat, and they struggle to make eye contact with colleagues."

This output provides a diverse yet relevant description, including physical, emotional, and behavioral aspects of social anxiety.

1.3.4.3 Length Control

Specify the desired length of the output, either in tokens or words.

Use in psychology:

- Useful for generating responses of consistent length for standardized assessments.
- Can help in creating concise summaries of psychological theories or research findings.

Example:

Prompt: "Summarize Piaget's theory of cognitive development in about 50 words."

Output: "Piaget's theory posits four stages of cognitive development: sensorimotor, preoperational, concrete operational, and formal operational. Children progress through these stages, developing increasingly complex thinking. Each stage builds on the previous, with children actively constructing knowledge through interaction with their environment. The theory emphasizes qualitative changes in thinking processes."

This example demonstrates how length control can be used to create concise, standardized explanations of psychological theories.

1.3.4.4 Guided Generation (Prompt Engineering)

Use carefully crafted prompts to guide the model's output in specific directions.

Use in psychology:

- Crucial for eliciting responses tailored to specific psychological frameworks or theories.
- Can help in generating outputs that adhere to ethical guidelines in clinical psychology.

Example:

Prompt: "Describe the symptoms of depression using the DSM-5 criteria, focusing on how they might present in adolescents. Include both emotional and behavioral manifestations, and consider how these symptoms might impact academic performance."

This prompt guides the model to provide a specific, targeted response that adheres to DSM-5 criteria while focusing on a particular demographic and context.

1.3.4.5 Few-shot Learning

Providing the model with a few examples of the desired input-output pattern within the prompt.

Use in psychology:

- Useful for teaching the model to follow specific formats or styles relevant to psychological reporting.
- Can help in generating responses that mimic particular therapeutic approaches.

Example:

Prompt: "Generate a response in the style of Cognitive Behavioral Therapy (CBT) to a client's negative thought. Follow this pattern:

Client: 'I'm going to fail this job interview.' Therapist: 'That sounds like a challenging thought. Let's examine the evidence. What makes you think you'll fail? Have you had successful interviews in the past?'

Client: 'Everyone at the party will think I'm boring.' Therapist: 'I understand you're worried about how others perceive you. Can we look at this thought more closely? What evidence do you have that everyone will find you boring?'

Now, respond to this client statement:

Client: 'I'll never be good enough at my job.'"

This prompt uses few-shot learning to guide the model in generating a response that follows the CBT approach of challenging negative thoughts and seeking evidence.

Ethical considerations:

- Be mindful of potential biases introduced through the framing of prompts.
- Ensure that chains of prompts do not lead the model towards predetermined conclusions, potentially skewing the outcomes.
- When using LLMs in sensitive settings, carefully consider the impact of prompts on the information obtained and presented.

Summary

This chapter has explored the emerging role of LLMs in psychological research and practice. It covered the fundamental concepts of LLMs, their operational mechanisms, and the critical skill of prompting for effective utilization. Key topics included the basics of LLMs, the importance of prompting, methods for controlling LLM outputs, and potential applications in psychology. LLMs offer exciting possibilities for enhancing research methodologies, assisting in learning and clinical practice, and providing novel ways to explore psychological concepts. However, their use comes with important caveats, including potential bias amplification and the need for empirical validation. As LLMs continue to evolve, it will be crucial for psychologists to stay informed about their capabilities and limitations, develop best practices for their use, and contribute to the ongoing dialogue about the ethical implementation of AI in mental health and behavioral sciences.

Questions

1. What are LLMs, and what are their primary components?
2. What is prompting in the context of LLMs, and why is it important?
3. What are the different types of prompts, and can you provide examples of both good and bad prompts?
4. What are the main ethical considerations in using LLMs for psychological research and practice?

5. Can you explain the concept of fine-tuning LLMs and its benefits for psychology research?
6. What are temperature settings in LLM outputs, and how do they affect the results?
7. Describe two potential applications of LLMs in clinical psychology. What are the benefits and risks of each?

Exercises

1. Using an LLM, develop a quiz with multiple-choice and short-answer questions covering the basics of LLMs, including their architecture and functioning.
2. Write a set of prompts for an LLM to generate different types of content (e.g., a summary, a detailed explanation, and a creative story) on a psychological concept. Evaluate the quality of the outputs.
3. Use an LLM to simulate a counseling session, focusing on a specific psychological issue. Reflect on the session's effectiveness and areas for improvement.
4. Experiment with different temperature settings in an LLM to generate responses to a psychological query. Analyze how changing the temperature affects the creativity and accuracy of the responses. Present your findings in a report.
5. Use an LLM to generate content on a psychological topic. Analyze the outputs for potential biases related to gender, culture, or other factors. Propose strategies to mitigate these biases.
6. Consider how LLMs might be used to bridge psychology with another field of your choice (e.g., education, marketing, public health). Propose a research study that leverages LLMs to explore an interdisciplinary question. Outline the methodology, potential benefits, and challenges of using LLMs in this context.
7. Ask an LLM to generate a prompt for a purpose of your choice. Analyze the suggested prompt (its type, structure, etc.) and compare it to the prompting strategies outlined above.

Chapter 2

Large Language Models in Psychology Training

LLMs can be smoothly integrated into psychology training and learning processes, offering innovative ways to enhance educational experiences and outcomes (e.g., Holmes et al., 2019). As we explore in this chapter, LLMs are not just add-ons to existing methodologies, but powerful tools for creating more dynamic, adaptive, and effective learning environments. From personalizing learning pathways to generating complex, realistic case studies, they might help psychology teachers to generate materials, as well as students to acquire knowledge and develop critical skills. This is nicely demonstrated by various learning platforms that have incorporated LLM-powered chatbots to enhance the learning process. The present chapter will depict several ways LLMs can be integrated into psychology education. Specifically, we will touch upon (1) how LLMs facilitate adaptive learning experiences tailored to individual students' needs and learning styles; (2) the use of AI-generated case studies and scenarios to provide diverse and challenging training materials; (3) the potential for LLMs to assist in the development of critical thinking and diagnostic skills through interactive problem-solving exercises; and (4) the integration of LLMs into supervision and mentorship programs for psychology trainees.

2.1 Adaptive Learning

LLMs offer unprecedented opportunities to create personalized learning experiences, for instance, by tailoring content delivery and difficulty levels to match individual needs.

Imagine a student that excels in other fields but struggles with statistical concepts. An LLM might help explain a concept in the required depth, offering a personalized learning experience. Importantly, LLM explains the concept with unending patience. For instance, take the concept of linear regression.

Prompt: "I'm a psychology student. Please explain the concept of linear regression to me."

Follow-up 1: "I'm very bad at math. Please explain it to me without using formulas."

Follow-up 2: "I'm still unsure about the concepts of standardized and unstandardized regression coefficients. Could you please explain them to me in simple terms with real-life examples?"

Follow-up 3: "Please plot a regression analysis model with $b = 0.2$."

Follow-up 4: “Here is a regression analysis table from a paper I just read. Please interpret the results. [Table inserted]”

After interacting with an LLM, a student can test their knowledge through quizzes generated with the help of the LLM, with instant feedback from the model. Here are example prompts to create such quizzes:

Prompt: “Please create a quiz with five multiple-choice questions and five short-answer questions to test my understanding of the basic concepts of linear regression. Please provide feedback after I answer the quiz.”

Consider: To enhance your learning experience, tailor the interaction to your level of understanding and background knowledge, using follow-up prompts for clarifications. Promote active engagement by asking questions and completing exercises that encourage critical thinking about the material. Request a variety of examples and analogies to explain concepts, and start with simpler ideas, gradually introducing more complex ones to ensure mastery of each step before progressing further. Some LLMs also offer the ability to create a specialized workspace that can be tailored to a specific task. For example, you can upload a specialized textbook you want to study, allowing the model to primarily reference and work with that material.

2.2 AI-Generated Case Studies and Scenarios

LLMs can generate diverse, realistic case studies and scenarios, providing students with a rich array of training materials. If done correctly, they might be crucial for developing comprehensive counseling and clinical skills. It is important to remember that LLMs can be used as a part of training. Psychological practice in the real world requires supervision, peer consultation, and ongoing professional development to ensure accurate diagnosis and effective interventions.

2.2.1 Generating Diverse Scenarios

An LLM could be used to generate a series of fictional client profiles and therapy simulations, each presenting unique challenges. Integrated with complex feedback from the model, they can be used for training in diagnostics, therapeutic skills, or problem-solving.

2.2.1.1 Diagnostics

Prompt: “Think of a person with mental health issues and generate a page-long case study where the symptoms are not entirely obvious. Please avoid stating the diagnosis/diagnoses in the description. My task will be to come up with the

diagnosis/diagnoses based on the case study. After I provide my diagnosis and reasoning, I would like you to give me feedback on my response and the correct answer.”

Consider: In the real world, the inclusion of subtle and complex symptomatology that requires critical thinking and thorough analysis is crucial, as mental health cases are often messy and not straightforward. This necessitates a balance between clarity and ambiguity in requesting symptom descriptions from the model to adequately challenge the trainee’s diagnostic skills. Furthermore, incorporating cultural, socioeconomic, and personal history aspects is essential to create comprehensive and realistic client profiles.

2.2.1.2 Clinical Scenarios

Prompt: “I’m a psychology student training for clinical practice. Could you generate an example scenario of a teenage client struggling with generalized anxiety and school refusal? Please provide a paragraph-long description of the case. Then, we can proceed with me being the therapist and you being the client.”

[Simulated interview is over]

Follow-up prompt: “Okay, we’re done with the roleplay. Could you please provide me with feedback on my interactions?”

Consider: When creating clinical scenarios, ensure the clarity and depth of the case description, providing enough detail to create a meaningful and engaging roleplay. After the roleplay, do not forget to ask LLM for detailed feedback that includes specific suggestions for improvement.

2.2.1.3 Other Scenarios

A similar structure can be applied to a wide range of psychological scenarios, facilitating effective training processes across different settings. These simulations allow students to encounter a broad range of scenarios than might otherwise be unavailable through traditional approaches.

Prompt: “I’m a psychology student training my HR skills. Could you generate an example scenario of two employees experiencing a conflict in a marketing department? Please provide a paragraph-long description of the case. Then, we can proceed with me being the mediator and you playing the roles of both employees.”

[Simulated mediation is over]

Follow-up prompt: “Alright, we’ve completed the roleplay. Could you please provide me with feedback on my mediation techniques?”

Consider: When creating different scenarios, ensure the description includes enough context, including the nature of the problem. After the roleplay, remember to request comprehensive feedback from the LLM that includes specific observations on your mediation style, communication techniques, and suggestions for improving your skills.

2.3 Programming Skills

The integration of programming skills into psychology education is increasingly essential in an era where data analysis, artificial intelligence, and technology play significant roles in research and practice. Programming can empower psychology students to analyze data (ideally in a reproducible way), modify and run computer-based experiments, or develop computational models of psychological processes.

2.3.1 Programming Tools and Languages for Psychology Students

Several programming languages and tools are particularly valuable for psychology students, each offering unique strengths that align with different aspects of psychological research. R and Python are among the most common ones. R is highly regarded for its capabilities in statistical analysis. It provides robust tools for a wide range of statistical modeling tasks, from basic data summaries to complex analyses. Additionally, R is known for its strength in creating detailed and informative visualizations, making it a great choice for presenting research data in a clear and compelling way. Python is one of the most popular programming languages in psychology due to its versatility and ease of learning. It is well-suited for data analysis, making it a powerful tool for handling large datasets and performing complex computations. It is also widely used for machine learning and natural language processing, which are increasingly important in psychological research, enabling innovative approaches to analyzing data and modeling cognitive processes. For experimental design and implementation, tools like PsychoPy and E-Prime provide platforms for creating and running psychological experiments. PsychoPy, in particular, is an open-source Python library that allows researchers to design complex experiments without extensive programming knowledge, making it accessible to students who are new to coding.

2.3.2 Programming in psychology curricula

Let's be honest—it wasn't long ago that point-and-click software, especially SPSS, dominated the landscape of statistical analysis in psychology worldwide. However, teaching research methods and statistical analyses using programming languages is becoming increasingly popular (Masuadi et al., 2021). Many students and researchers, though, lack a sufficient background in programming. Teaching programming to psychology students can thus be challenging due to the diverse backgrounds and varying

levels of prior experience with coding. Furthermore, some students may feel intimidated by programming, perceiving it as difficult or unrelated to their primary interests in psychology. Despite the availability of excellent introductory textbooks, video tutorials, and freely accessible code, gaining proficiency in programming for statistical analysis remains a challenging task. Learning coding skills becomes significantly more accessible when LLMs are integrated into the process. Beyond their interactive nature, LLMs can debug code, provide detailed explanations of each step, and suggest more efficient coding solutions. This support makes the learning experience smoother and more effective. Below, I present a few examples of learning programming using R. These examples can easily be adapted to other software or tasks.

2.3.3 Programming Learning Examples for R

We will go through four main routes of learning programming using LLMs. An LLM will be asked to (1) teach how to use R, (2) produce a code given a specific input, (3) debug a code, and (4) explain a code.

2.3.3.1 Teach Me Some R

Prompt: “I’m a psychology student completely new to R programming. Could you please teach me the very basics, such as how to create vectors, perform simple data manipulation, calculate basic statistics like the mean and standard deviation, and save my work? Please guide me through each step with clear explanations.”

[Code walkthrough is complete]

Follow-up prompt: “I’ve followed the steps. Could you provide feedback on my code and suggest any improvements?”

Consider: When crafting prompts for learning R programming, it’s important to be clear and specific about the tasks you want to learn. Start with the most basic concepts, ensuring that your prompts focus on one step at a time, such as creating vectors or performing a simple operation. This approach helps in building a strong foundation. Additionally, be explicit in requesting detailed explanations, which will aid in your understanding of each step. After completing the tasks, ask for feedback in your prompts, specifically requesting suggestions for improving your coding practices and understanding. This will help you to not only learn the basics but also to refine your skills as you progress.

2.3.3.2 Code Generation

Forgot the function? Or just not sure how to proceed? No worries, LLMs can help. Just be precise with your instructions.

Prompt: “I have a dataset called “dat” that contains two variables: “ageGroup” and “workingMemoryScore.” Please subset the data to include only individuals with an ageGroup less than 3. Then, compare their working memory performance using a Welch’s t-test and visualize the results with box plots. Could you generate the R code for this?”

Consider: When crafting prompts to generate R code using an LLM, it’s important to be clear and specific in your requests. Clearly outline each task, such as subsetting data, performing a t-test, or creating visualizations, to ensure the LLM understands the full scope of what you need. If you are unfamiliar with certain functions or steps, consider parts. Additionally, it is beneficial to ask the LLM for annotations and explanations of the code it generates. Be attentive to common errors that can prevent your code from running correctly, such as typos in variable names or mismatched parentheses. Even small mistakes can lead to frustrating issues, so it’s important to carefully review your prompt and code for accuracy.

2.3.3.3. Code Debugging

Ran into an error? LLMs can assist you in identifying and fixing issues in your code. Whether it’s a syntax error, a logical mistake, or a simple typo, the LLM can guide you through the debugging process.

Prompt: “I’m getting an NA when trying to calculate the mean of a variable that contains missing data (NA values). Here’s the code I’m using:

```
reaction_times <- c(251, 300, 322, 571, NA, 378, 292, 312, 446, NA, 339, 492)
mean(reaction_times)
```

Can you help me figure out what’s wrong and how to fix it?”

Consider: When using an LLM to debug code, it is crucial to provide as much relevant information as possible about the error. Copy-paste the specific error message, the portion of the code where the error occurs, and any relevant context that might help the LLM understand the problem. Clearly describe the expected outcome versus the actual outcome, and ask the LLM to explain the issue in simple terms. This approach not only helps in resolving the current error but also enhances your understanding of common

debugging techniques. Additionally, after fixing the issue, consider running the code incrementally to ensure each part works as expected before moving on to the next step.

2.3.3.4. Code Explanation

LLMs can help break down complex code into simpler terms, making it easier for you to grasp the underlying concepts and logic.

Prompt: “I have a piece of R code that I don’t fully understand, and it’s poorly annotated. Could you please explain what each part of this code does, especially how the datasets are merged and how the dplyr functions work? Here’s the code: [Code inserted].”

Consider: When seeking an explanation for complex code, it’s crucial to provide all relevant details, especially when the code involves multiple operations like merging datasets, creating new variables with conditional logic, and performing filtering. Be clear about which parts are most confusing to you.

Ethical Considerations

- While using AI-generated case studies can provide diverse scenarios, it's crucial to ensure that these do not inadvertently reflect real individuals' behaviors. Keep in mind that the real world is very complex and people’s motivations, emotions, cognitions, and experiences are not always straightforward, rational, and predictable.
- LLMs may perpetuate existing biases in their generated content. It's important to regularly audit and correct for biases in the scenarios and case studies to ensure diverse and fair representation.
- Relying too much on AI can lead to the neglect of critical thinking skills and the acquisition of real-world experience. Strive to strike a balance between utilizing AI-assisted learning and engaging in traditional methods, such as independent problem-solving, hands-on experiences, and in-depth discussions.

Summary

This chapter explores the integration of LLMs into psychology education, showcasing their potential to revolutionize training methods. It demonstrates how LLMs can create personalized, adaptive learning experiences, tailoring content delivery to individual student needs and learning styles. The chapter highlights LLMs' capacity to generate diverse, realistic case studies and scenarios, crucial for developing comprehensive counseling and clinical skills. Key applications include concept explanation, document summarization, and interactive simulations for diagnostics and therapeutic practice. The text emphasizes the importance of balancing AI-assisted learning with traditional

methods, using LLMs as supplements rather than replacements for real-world experience. It provides practical examples of prompts and considerations for effectively utilizing LLMs in various psychological training contexts, from explaining complex statistical concepts to simulating client interactions. By offering a broader range of scenarios than traditional approaches, LLMs present unprecedented opportunities to enhance the depth and breadth of psychology education, preparing students more comprehensively for the complexities of professional practice.

Questions

1. What are the potential benefits and limitations of using LLMs for generating case studies in psychology training?
2. In what ways can LLMs assist psychology students in developing critical thinking and diagnostic skills?
3. What is the role of programming skills in modern psychology education and how LLMs can support students in acquiring these skills?
4. What ethical considerations should be taken into account when using LLMs for psychology education and training?
5. How can LLMs help in explaining complex psychological concepts to students with varying levels of understanding?
6. What challenges might arise from integrating LLMs into psychology curricula, and how can they be addressed?
7. What roles can LLMs play in fostering critical thinking skills in psychology students?
8. In what ways can LLMs be used to support the supervision of psychology trainees?
9. What are the potential risks of relying too heavily on LLMs in psychology education and training?

Exercises

1. Using an LLM, simulate an adaptive learning session for a psychological concept. Start with a basic explanation and gradually introduce more complex ideas based on your own responses. Reflect on the effectiveness of the adaptive learning process and suggest improvements.
2. Prompt an LLM to generate a fictional case study involving a client with ambiguous symptoms. Diagnose the case based on the information provided, then request feedback from the LLM on your diagnosis. Reflect on how this exercise helps develop diagnostic skills.
3. Use an LLM to set up a series of programming challenges related to psychological research (e.g., data cleaning, hypothesis testing, visualizations). After completing

the tasks, ask the LLM for feedback on your code and suggestions for improvement.

4. Generate content on a psychological topic using an LLM and analyze the output for potential biases (e.g., gender, cultural, socioeconomic). Propose prompting strategies to mitigate these biases and discuss the importance of bias detection.
5. Conduct a role-playing exercise where the LLM simulates a client with a specific psychological disorder. After the session, evaluate the effectiveness of the role-play and discuss how it could be improved for better educational value.
6. Work through a psychological concept or statistical method with the guidance of an LLM. After each step, ask for detailed feedback on your understanding and performance. Summarize the feedback and discuss how it helped improve your learning.
7. Use an LLM to simulate a research collaboration where you and the LLM design a study together. The LLM will assist in hypothesis generation, methodology design, and data interpretation. Reflect on the collaborative process and the insights gained.

Chapter 3:

Large Language Models in Psychology Research

Traditionally, decisions made throughout the research process have been considered the sole responsibility of the researcher, who exercises full control over each step. However, technology has already been integrated into many aspects of current research—from randomization in experimental designs and keyword highlighting for systematic review screenings to the execution of complex statistical analyses. LLMs extend this assistance even further, offering new opportunities for innovation in psychology research. They can be applied in various stages of the research process that were once thought to require exclusively human input. For example, they can generate scenarios for vignette designs, allowing for a broader, yet systematic exploration of hypothetical situations. They can creatively tweak standard experimental tasks, introducing nuanced variations that may lead to novel insights (e.g., Baek et al., 2024; Si et al., 2024). Additionally, LLMs can automate the extraction of information from large volumes of text, streamlining the literature review process and saving valuable time and resources. Moreover, these models can significantly enhance the clarity and readability of research papers by refining writing style and providing thorough proofreading. These LLM capabilities will be explored and demonstrated in this chapter.

3.1 Research design improvement

Occasionally, research designs require adjustments that are time-consuming or could greatly benefit from external assistance. With the right prompts, LLMs can efficiently generate tasks that would otherwise be tedious and prone to errors. Moreover, the inherent creativity of these models can be leveraged to introduce novel elements and unique twists, enhancing and refining the research design.

3.1.1 Generating Scenarios

Imagine you are conducting an experimental study for your final thesis. The topic focuses on people’s perceptions of biobanks and biobanking. In some countries, biobanking is a relatively new concept, leading to a lack of information and subsequent natural skepticism and fear toward biobanks. Your aim is to explore the most effective way to present information about biobanks to increase trust and willingness to participate in biobanking. To achieve this, you plan to test the effect of different framing techniques when presenting foundational information about biobanks on people’s trust and willingness to participate in biobanking. You identified four main variables (one can easily come up with even more factors—e.g., information about benefits, information about data privacy, information about

type of data collected, etc.) that could potentially affect the message you convey: 1. Purpose of biobanks (minimal information vs. detailed information); 2. Information about biobank ownership (no mention vs. private company vs. governmental); 3. External partners (no mention vs. partnerships mentioned). 4. Incentives (no mention vs. incentives mentioned).

Altogether, this yields $2 \times 3 \times 2 \times 2 = 24$ different scenarios presentable to research participants. LLMs can speed up the process of scenario generation, especially when a large number of scenarios is involved, with a simple prompt.

Prompt: “I’m conducting research on people’s willingness to participate in biobanking. The study will use a vignette design, in which I will vary four main variables:

1. Purpose of biobanks (minimal information vs. detailed information)
2. Information about biobank ownership (no mention vs. private company vs. governmental)
3. External partners (no mention vs. partnerships mentioned)
4. Incentives (no mention vs. incentives mentioned).

Could you please generate these vignettes accordingly?”

It might be that you already have your vignettes generated. However, you may want to examine whether the participants’ responses are robust to different wordings. Therefore, one might wish to duplicate the existing vignettes with different phrasings and then randomly administer one from each pair. Again, LLMs can come in handy here.

Prompt: “I have developed 12 stories to explore the effect of friendship on children’s morality. In each story, I vary three key variables: 1. The relationship with the person (good friend vs. acquaintance); 2. The type of intervention during a confrontation (confront the person vs. support the person despite knowing they are wrong); and 3. The subsequent feelings of the character (feeling better vs. feeling the same vs. feeling worse).

[stories and how the variables vary here]

To improve the robustness of the design, please rephrase each story while staying faithful to the variations of the variables.”

Consider: When using an LLM to generate scenarios or vignettes, it is essential to start by clearly outlining the variables you want to manipulate, the context of the study, and the desired outcome of each scenario. For example, if you are testing different ways to frame information about biobanks, specify the exact framing variations, such as the level of

detail or the nature of the ownership. Include examples or templates if possible, as this helps the LLM understand the structure and tone you're aiming for. If you need to test the robustness to different phrasings, request the LLM to reframe scenarios while maintaining the core meaning. Review the new scenarios to ensure they align with your research goals and check for consistency in style and language.

3.1.2 Tweaking Designs

Replication and generalizability studies are crucial for a credible understanding of human behavior. Replications can range from direct (closely replicating the original study) to conceptual (testing the original claims using slightly different methods), each offering valuable insights. Here, we will focus on the spectrum closer to conceptual replications. Consider a well-established experiment, such as the trolley dilemma—a task commonly used to explore moral decision-making. While numerous studies have utilized this task, you might aim to enhance it with a unique twist. You can seek primary inspiration from LLMs to generate its new, creative variations.

Prompt: “Can you suggest a couple of twists to the trolley dilemma task? Temperature == 2”

[LLM's reply]

Follow-up prompt: “Could you please elaborate on the moral integrity score assigned to each person in the task? What additional insights does it provide, and how can I potentially integrate it into the interpretation of the results?”

LLMs can be valuable for brainstorming. Your task is then to select the most reasonable and suitable suggestions, and refine them based on your needs and expertise. For instance, consider using LLMs to generate creative variations of the digit span test.

Prompt: “Can you suggest 5 creative twists to the digit span task? Temperature == 5.”

In this particular case, LLM provided this list:

1. Digit Origami: After seeing the digits, participants are given paper and must fold it into shapes that represent each digit in the sequence.
2. Gravity Reversed: Participants must “recall” by writing the digit upside down.
3. Digital Ecosystem: Each digit corresponds to a specific element in an imaginary ecosystem (e.g., 1 = river, 2 = mountain, 3 = forest). Participants must recall the digits by describing how these elements interact and evolve within the ecosystem in the order of the sequence.

4. Emotion-Fueled Recall: Each digit is linked to a specific emotion (e.g., 1 = happiness, 2 = anger). Participants must recall the sequence by describing a scenario in which they experience these emotions in the order of the digits.
5. Digit Countdown: Participants are presented with the digit sequence in reverse order (from last to first). They must recall the digits in the original forward sequence, challenging their ability to mentally reverse the order.

If the generated list does not fully meet your needs or provide the desired level of inspiration, you can easily follow up with the LLM to request additional options. This iterative process allows you to refine and expand upon the initial suggestions, tailoring the results more closely to your specific requirements.

Consider: When using an LLM to brainstorm creative twists for experimental tasks or scenarios, it's important to experiment with temperature settings to achieve meaningful and diverse outputs. Consider defining the elements of the task or scenario you want to modify. If you're exploring conceptual replications, consider asking the LLM to elaborate on the implications of the suggested modifications. After receiving the LLM's suggestions, carefully review them to ensure they align with your research goals. This is where your expertise comes into play. Make sure the creative twists preserve the integrity of the original experiment while providing new insights.

3.1.3 Reviewing Your Research

The process of peer review—external feedback from an expert in the field—is vital for research quality and credibility. When working on research, and before sharing it with colleagues or supervisors, it can be useful to get a quick, independent check on aspects such as the research rationale, study design, analysis plan, and overall clarity of writing. Recent research (Liang et al., 2024) analyzing human-generated and AI-generated peer reviews of Nature portfolio journals has shown that LLMs perform about as well as human experts and often provide feedback that is even more helpful. This can be particularly valuable in the early stages of research. You can request critical yet constructive feedback as follows.

Prompt: “Imagine you are a reviewer for a high-quality scientific journal. I am attaching a study protocol that includes the theoretical background, study goals, methodology, and analytic plan. It does not yet contain results or discussion, as this is a pre-study protocol. Your task is to provide a scientific peer review of the protocol. Please be critical yet constructive in suggesting ways to improve it.

[study protocol attached]”

Consider: While LLMs can provide valuable feedback, it is important to assess their suggestions critically, as they may lack the depth of field expertise. Therefore, LLM feedback is best used in the early stages of research preparation or as a complement rather than a replacement for expert evaluation. Do not forget to ask follow-up questions to explore the suggestions further and request clarifications whenever needed.

3.2 Data Extraction

The volume of published literature has been growing exponentially, often making it impractical for researchers to thoroughly review all available texts. When a researcher needs to extract specific information—such as data for a meta-analysis—or condense complex (or messy) texts, LLMs can be particularly useful. Although data extraction methods and possibilities by LLMs are still developing and the available evidence shows mixed results, it is a promising avenue for substantial improvement in the near future. Below, we explore several simple examples of how LLMs can automate data extraction and conserve valuable time.

3.2.1 Extracting Specific Information

Sometimes, especially for meta-analytic purposes, one needs to extract specific pieces of information from a text. For example, one might be interested in whether a study was preregistered, what the sample size was for the focal analysis, or which measures were used to assess a specific construct. Of course, this can all be done manually, but such manual coding of information is tedious and error-prone, especially when numerous variables from hundreds of sources need to be extracted. Despite the potential to increase the scalability of such research, work still needs to be done to standardize these extraction processes—teams from all around the world, including ours, are working on this. Here, I provide only basic examples of how this can be done. In fact, large-scale data extraction requires constructing advanced prompt engineering pipelines and the use of APIs, which is beyond the scope of this textbook. So, let's move on to the examples.

Imagine that you are interested in finding out whether a study was preregistered or not. Most of the time, but definitely not always, information about preregistration is explicitly stated in a separate section. Sometimes, however, it is not so obvious. Additionally, there are multiple platforms and ways to preregister a study.

Prompt: “Please review this scientific paper and determine whether the study was preregistered. This is typically indicated by phrases such as “the study was preregistered at...” or “the registration of the study can be found at...”. Common platforms for study preregistration include OSF, AsPredicted.org, ClinicalTrials.gov, etc.. Provide your response in a Yes/No format. Temperature == 0.2.

[study attached]”

Now imagine that you have a study with several measures of a construct, say, socioeconomic situation. You need to extract all of them. You can prompt the LLM like this.

Prompt: “Please review the attached scientific paper and identify all the measures related to socioeconomic status (e.g., scales, items, author-constructed items). Socioeconomic status is typically assessed using both subjective and objective indicators. Examples of these measures include income, household income, economic position, Cantril’s ladder, ability to make ends meet, subjective assessment of wealth, etc. List all identified measures and separate them with “;”. Temperature == 0.5.

[study attached]”

Consider: Before applying a prompt to multiple papers, first test the output for accuracy to ensure the correctness of the data produced. It is important to have realistic expectations about what can be accurately extracted by a model, as the complexity of scientific literature can make it challenging to gather the required information without extensive extrapolation and field-specific expertise. While LLMs can be valuable in this process, they are not infallible. When seeking explicit information, remember to set the temperature settings to support deterministic answers. Sometimes, LLMs might underperform due to text formatting issues (e.g., two-column layouts). If this occurs, try reformatting the document to a single-column format before inputting it into the model. Additionally, always consider the legal and ethical guidelines when using LLMs with scientific papers or other copyrighted materials.

3.2.2 Summarizing Texts

Reading an entire study is often beneficial, but practical constraints can sometimes prevent a reader from doing so. Abstracts, extended abstracts, summaries, and conclusions serve as great tools for conveying the main information (based on the study authors’ perception) in a condensed format. However, these sections might not always be sufficiently informative, and a reader (i.e., researcher, practitioner) may need to focus on specific aspects, such as the methods used or the practical implications of the results. When time for reading is limited, it can be useful to prompt an LLM to generate summaries tailored to one’s specific needs. Additionally, one can further interact with the model to request both specific and general follow-ups.

Prompt: “Here is a scientific paper [paper attached]. Please provide a summary of the statistical methods used, and explain the key features of network analysis to help me better understand the results.”

Follow-up prompt: “What is the difference between the measures of strength and expected influence in this context? What are the practical implications of these differences?”

Similar principles can be applied to reading the discussion, which might often be lengthy.

Prompt: “Here is a scientific paper [paper attached]. Please summarize the discussion into several key points, with an emphasis on the practical implications of the findings.”

Follow-up prompt: “Can you help interpret the [finding] in the context of self-determination theory?”

Consider: Although LLMs can significantly enhance understanding of a topic by summarizing complex documents, it is important to verify the accuracy of the extracted information. In optimal scenarios, summaries should be used as a supplement to one’s reading, not as a replacement. As always, carefully craft your prompts to target the specific information you are interested in. Keep in mind that, particularly in highly specialized fields, LLMs may miss subtle nuances or oversimplify the content. Again, be sure to adhere to legal and ethical guidelines when using LLMs with scientific papers or other copyrighted materials.

3.3 Writing Refinement and Proofreading

Not a native speaker, or just looking to improve writing clarity and flow? Using LLMs to enhance your text is incredibly easy. In fact, this entire textbook has been reviewed for language by an LLM. While these language refinement services are most effective with English, they can also be (surprisingly) useful for other languages. The process is simple—one just has to specify whether they want proofreading alone or if additional enhancements are required. Typically, when asked for proofreading, an LLM not only corrects grammatical errors but also substitutes words and adjusts the text to flow more naturally.

Prompt: “Please proofread and improve: [insert text]”

Sometimes, you may have specific terminology that you want to preserve. An LLM might not always be fully familiar with highly specialized terms. If this occurs, simply include these phrases in your prompt to ensure they are maintained.

Prompt: “Please proofread and enhance the following text. Ensure that the terms “verisimilitude,” “meta-synthesis,” and “problematic gaming” remain unchanged.

[text inserted]”

One can easily request further revisions by asking the model to rewrite a specific part of the text or to make any additional adjustments specified in the follow up prompt.

Consider: It is possible that a part of the “improved” text may not actually be improved at all. Always verify that the refined version accurately conveys the original idea you intended. Refining text is an iterative process, so keep in mind that an LLM is a tool to help you express your ideas—not the other way around.

Summary

This chapter highlights the transformative potential of LLMs across various stages of the research process. Traditionally, researchers have maintained control over the entirety of research design and execution. However, LLMs offer a novel extension of technological assistance, enabling more efficient and creative approaches to research. The chapter demonstrates how LLMs can generate experimental scenarios, tweak established research designs, and even aid in conceptual replications, thereby enhancing the originality and robustness of studies. Additionally, LLMs can automate data extraction from large volumes of text, significantly streamlining literature reviews and meta-analyses. The text also explores the role of LLMs in improving the clarity and readability of research papers through language refinement and proofreading. While LLMs present exciting new possibilities, human expertise is crucial to ensure that tools provided by the models are used effectively and responsibly.

Questions

1. How do LLMs extend the traditional role of researchers in the design and execution of psychology studies?
2. How can LLMs be used to create conceptual replications of established studies?
3. What role might LLMs play in the automation of data extraction for systematic reviews, particularly meta-analyses?
4. What are the ethical considerations when using LLMs to summarize and extract information from scientific literature?
5. What are the risks of over-reliance on LLMs for language refinement in research papers?
6. How can LLMs contribute to the development of novel research methodologies in psychology?

7. How can LLMs help researchers stay updated with the rapidly growing body of scientific literature?
8. How can LLMs be utilized to enhance the reproducibility of psychological studies?

Exercises

1. Use an LLM to generate a series of vignettes for a psychology experiment on decision-making. Ensure that each vignette varies systematically according to predefined variables. Analyze the generated vignettes for consistency and relevance.
2. Use an LLM to suggest modifications to a classic psychology experiment for a conceptual replication with a twist. Implement these suggestions and discuss how they may influence the study's outcomes.
3. Prompt an LLM to review some of your research protocols.
4. Apply an LLM to extract key information from a study such as information about participants, measures used, performed statistical analyses, and whether the study was preregistered. Review the accuracy of the extracted information.
5. Use an LLM to summarize the methodology section of a research paper. Compare the summary to the original text and try to identify any missing details or inaccuracies.
6. Input a detailed methodology section into an LLM and ask it to suggest improvements for clarity and completeness. Discuss the practicality of these suggestions in the context of the research.
7. Input a draft of a research paper into an LLM and ask it to proofread and improve the text. Reflect on the changes made and whether they enhanced the clarity and readability of the section.
8. Engage with an LLM in a brainstorming session to generate new research ideas in psychology. Evaluate the feasibility and originality of the ideas generated.
9. Provide an LLM with a dense theoretical psychology paper and ask it to summarize the key arguments and conclusions. Review the summary for completeness and accuracy.

Conclusion

The integration of LLMs into the field of psychology research has the potential to fundamentally transform the way we approach learning, training, and the conduct and communication of research. The provided examples, questions, and exercises clearly demonstrate that these advanced models enable creative and efficient collaboration between human intelligence and machine capabilities. However, as with any powerful technology, the successful integration of LLMs into psychology training and research requires careful consideration, ethical oversight, and a thorough understanding of both their potential and their limitations.

As illustrated in this textbook, LLMs bring a wealth of possibilities to psychology education and research. They offer opportunities to enhance both the breadth and depth of learning. These models can serve as interactive learning partners, capable of providing instant feedback, generating practice scenarios, and even simulating complex psychological cases. This interactivity allows students to engage with the material in a more dynamic and personalized way than traditional methods typically allow. For example, students can use LLMs to generate diverse case studies that cover a wide range of psychological disorders, each with varying degrees of complexity and symptoms. These case studies can be used for diagnostic practice, allowing students to refine their clinical reasoning skills. Furthermore, LLMs can provide immediate feedback, suggest alternative interpretations, and prompt deeper exploration of the material. This kind of personalized, iterative learning is invaluable in developing the critical thinking and problem-solving skills necessary for effective psychological practice. Moreover, LLMs can assist in the development of soft skills, which are crucial for any practicing psychologist. By simulating client-therapist interactions, LLMs can serve as a valuable tool to help students practice and improve their counseling techniques in a safe and controlled environment. This type of training allows students to explore different approaches and receive feedback on their performance, ultimately leading to greater effectiveness in real-world settings.

Researchers can greatly benefit from LLMs by transcending the traditional boundaries of research design, coding, and data extraction. These models enable a more innovative approach to the research process by automating routine and labor-intensive tasks, such as generating scenarios for experimental designs, refining research methodologies, and systematically extracting data from vast amounts of literature. In particular, the automation of data extraction by LLMs, although currently in its early stages, represents a significant advancement in the scalability and efficiency of research. As the volume of published literature continues to grow exponentially, the manual extraction of data for systematic reviews and meta-analyses becomes increasingly challenging and time-consuming. LLMs can streamline this process by rapidly identifying and extracting specific pieces of information from large datasets, allowing researchers to focus on synthesizing

and analyzing the data. This not only saves time but also reduces the potential for human error in data coding and extraction, thereby enhancing the reliability of research findings. By offloading such tasks to LLMs, researchers can dedicate more time and mental energy to complex, high-level tasks that require human expertise, such as formulating hypotheses, interpreting data, and developing new theoretical frameworks. This shift not only accelerates the research process but also enhances its creativity and rigor, allowing researchers to explore a broader and more diverse range of research questions.

Despite the impressive capabilities of generative AI, it is crucial to emphasize that LLMs are most effective when used in conjunction with human expertise. The concept of “human-in-the-loop” is paramount: LLMs should be seen as collaborative partners that augment, rather than replace, the researcher’s role. Human oversight is essential at every stage of the research process—from initial design and hypothesis generation to the final analysis and interpretation of results. Users of LLMs must critically evaluate the outputs generated by these models, ensuring that they are meaningful and align with ethical standards. Although LLMs are sophisticated, they lack the ability to fully grasp the nuanced ethical considerations that often arise in psychological training and research. Operating based on patterns in data, LLMs may inadvertently suggest information or interpretations that could be misleading or ethically problematic. Therefore, both students and researchers must remain vigilant, applying their knowledge and judgment to ensure that the final research product is both scientifically sound and ethically responsible.

To sum up, here are four key best practices and common pitfalls to avoid, which will help you fully harness the potential of LLMs:

- **Clearly define objectives:** Before employing an LLM, users should have a clear understanding of what they aim to achieve. This includes specifying the input, context, and desired outcomes. A well-defined objective ensures that the prompts given to the LLM are precise and relevant, leading to outputs that align with the intended goals.
- **Iterative refinement:** The use of LLMs should be an iterative process. Students and researchers should typically not settle for the first output generated by the model. Instead, they should continuously refine their prompts, experiment with different settings (such as temperature for creativity), and critically evaluate the results.
- **Ethical considerations:** Adherence to legal and ethical standards is critical. Users must be aware of potential biases that LLMs might produce, especially since these models are trained on large datasets that may reflect existing societal biases. For example, cultural, social, and linguistic subtleties may be overlooked by a model that processes text based on statistical patterns. Furthermore, when using LLMs,

students and researchers must be vigilant about data privacy and intellectual property issues.

- **Over-reliance on LLMs and outputs validation:** Users should avoid the temptation to rely too heavily on LLMs and should instead view them as one tool among many in their arsenal. Critical thinking and human judgment should always take precedence. Outputs generated by LLMs should be cross-verified with established literature and empirical data. Students and researchers should not take the model's suggestions at face value but should critically evaluate them. This step is essential for ensuring the accuracy and reliability of the outputs.

Looking ahead, the role of LLMs in psychology training and research is poised to expand even further. As these models continue to evolve, they will likely become more sophisticated, offering even greater support for students and researchers alike. To navigate this evolving landscape effectively, it is essential to stay informed about advancements in AI and continuously develop critical thinking skills. Nonetheless, the fundamental principle that human expertise must guide the use of LLMs will likely remain unchanged. The future incorporation of LLMs into psychology lies in a dynamic collaboration between human expertise and experience and machine capabilities.

In conclusion, LLMs represent a powerful advancement in the research toolkit, offering unprecedented opportunities for innovation and efficiency. When used responsibly and in conjunction with human oversight, they can significantly enhance the quality and scope of psychological training and research. By adhering to best practices, remaining mindful of potential pitfalls, and prioritizing the human-in-the-loop approach, users can fully harness the potential of LLMs, driving the field of psychology toward new horizons of discovery and understanding.

References

- Baek, J., Jauhar, S. K., Cucerzan, S., & Hwang, S. J. (2024). ResearchAgent: Iterative research idea generation over scientific literature with large language models. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2404.07738>
- Bharathi Mohan, G., Prasanna Kumar, R., Vishal Krishh, P., Keerthinathan, A., Lavanya, G., Meghana, M. K. U., Sulthana, S., & Doss, S. (2024). An analysis of large language models: their impact and potential applications. *Knowledge and Information Systems*, 66(9), 5047–5070. <https://doi.org/10.1007/s10115-024-02120-8>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & ... Amodi, D. (2020). Language Models are Few-Shot Learners. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2005.14165>
- Bubeck, S., Chandrasekaran, V., et al. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2303.12712>
- Cacciamani, G. E., Collins, G. S., & Gill, I. S. (2023). ChatGPT: standard reporting guidelines for responsible use. *Nature*, 618(7964), 238–238. <https://doi.org/10.1038/d41586-023-01853-w>
- Döderlein, J.-B., Acher, M., Khelladi, D. E., & Combemale, B. (2022). Piloting Copilot and Codex: Hot temperature, cold prompts, or black magic? *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2210.14699>
- Guo, E., Gupta, M., Deng, J., Park, Y.-J., Paget, M., & Naugler, C. (2024). Automated paper screening for clinical reviews using Large language models: Data analysis study. *Journal of Medical Internet Research*, 26, e48996. <https://doi.org/10.2196/48996>
- Han, T., Adams, L. C., Bressemer, K. K., Busch, F., Nebelung, S., & Truhn, D. (2024). Comparative analysis of multimodal large language model performance on clinical vignette questions. *JAMA: The Journal of the American Medical Association*, 331(15), 1320. <https://doi.org/10.1001/jama.2023.27861>
- Harvard University Information Technology. (2025). Generative AI guidelines. <https://www.huit.harvard.edu/ai/guidelines>
- Holmes, W; Bialik, M; Fadel, C; (2019) Artificial Intelligence in Education Promises and Implications for Teaching and Learning. (1st ed.). Center for Curriculum Redesign: MA, USA
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*. <https://doi.org/10.1145/3703155>

- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Knoechel, T.-D., Konrad Schweizer, Acar, O. A., Akil, A. M., Al-Hoorie, A. H., Buehler, F., Elsherif, M. M., Giannini, A., Heyselaar, E., Hosseini, M., Ilangoan, V., Kovacs, M., Lin, Z., Liu, M., Peeters, A., van Ravenzwaaij, D., Vranka, M. A., Yamada, Y., Yang, Y.-F., & Aczel, B. (2024). Principles for responsible AI usage in research. *PsyArXiv*. <https://doi.org/10.31234/osf.io/g3m5f>
- Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2205.11916>
- Konet, A., Thomas, I., Gartlehner, G., Kahwati, L., Hilscher, R., Kugley, S., Crotty, K., Viswanathan, M., & Chew, R. (2024). Performance of two large language models for data extraction in evidence synthesis. *Research Synthesis Methods*. <https://doi.org/10.1002/jrsm.1732>
- Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., Zhou, X., Wang, E., & Dong, X. (2023). Better zero-shot reasoning with role-play prompting. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2308.07702>
- Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Vodrahalli, K., He, S., Smith, D. S., Yin, Y., McFarland, D. A., & Zou, J. (2024). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*, 1(8). <https://doi.org/10.1056/aioa2400196>
- Masuadi, E., Mohamud, M., Almutairi, M., Alsunaidi, A., Alswayed, A. K., & Aldhafeeri, O. F. (2021). Trends in the usage of statistical software and their associated study designs in health sciences research: A bibliometric analysis. *Cureus*. <https://doi.org/10.7759/cureus.12639>
- OpenAI (2024). OpenAI platform. Available at: <https://platform.openai.com/docs/models>
- Polak, M. P., & Morgan, D. (2024). Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nature Communications*, 15(1). <https://doi.org/10.1038/s41467-024-45914-8>
- Si, C., Yang, D., & Hashimoto, T. (2024). Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2409.04109>
- Solanki, S. R., & Khublani, D. K. (2024). Large language models. In *Generative Artificial Intelligence* (pp. 173–228). Apress.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *arXiv: Computer Science*. <https://doi.org/10.48550/arXiv.2201.11903>

- Yang, J., (2024). Rethinking Tokenization: Crafting Better Tokenizers for Large Language Models. arXiv: Computer Science. <https://doi.org/10.48550/arXiv.2403.00417>
- Yenduri, G., M, Ramalingam, G, Chemmalar Selvi, Y, Supriya, Srivastava, G., Maddikunta, P. K. R., G, Deepti Raj, Jhaveri, R. H., B, Prabadevi, Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). Generative Pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. arXiv: Computer Science. <https://doi.org/10.48550/arXiv.2305.10435>
- Zielinski, C., Winker, M. A., Aggarwal, R., Ferris, L. E., Heinemann, M., Lapeña, J. F., Pai, S. A., Ing, E., Citrome, L., Alam, M., Voight, M., & Habibzadeh, F. (2023, May 31). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. World Association of Medical Editors.

Register

A

Adaptive Learning – 37
API (Application Programming Interface)
– 16-17
(Automated) Feedback – 6, 38-41, 49-50

B

Bias in LLMs – 12, 21-23

C

Case Studies – 38-40
Chain-of-Thought Prompts – 18-19
Contextual Window – 16
Cultural Sensitivity in AI – 25

D

Data Extraction – 7, 50
Data Interpretation – 37-38
Data Privacy in LLM Use – 12
Determinism (Temperature Settings) –
29
Design Tweaking – 48-49

E

Ethical Considerations in LLMs – 11

F

Few-shot Learning – 18-19
Fine-tuning – 25-26

G

Guided Generation – 34

H

Hallucinations – 26
Human-in-the-Loop – 8, 56

I

Interactive Learning Environments – 6

L

Large Language Models (LLMs) – 15
Length Control – 33-34

N

Nucleus Sampling (Top-p Sampling) – 33

O

Over-reliance on LLM Outputs – 12, 23

P

Programming – 40-41
Prompting – 17-18

R

Replication – 48-49
Reproducibility – 22
Role-based Prompts – 20

S


Sloppiness – 26

T

Temperature settings – 28-32
Top-k Sampling – 32

U

User Interface (UI) – 16-17

The background features a series of overlapping, wavy, grey lines that create a sense of movement and depth. These lines are set against a light grey background with larger, semi-transparent, wavy shapes that further enhance the abstract, organic feel of the design.

I believe this publication can enrich the perspective on the use of LLMs among psychology students and researchers, and offer practical guidance—whether the reader is a beginner or more advanced in the topic.

— Pavol Kačmár

This textbook brings novel information in the emerging field and can be very beneficial to students, teachers, or researchers who aim to use or are currently using LLMs. The manuscript is written in simple, accessible language, it is clearly structured and provides many useful examples.

— Jakub Rajčáni