

AI in economic research

A guide for students and instructors

Version: 17 July 2024

Marc Burri
Daniel Kaufmann
Nima Ostovan

Disclaimer: The views expressed in this report are those of the authors and do not necessarily represent those of the University of Neuchâtel. Specifically, the guidelines in this report do not represent binding directives of the University of Neuchâtel.

Abstract: This report documents the use and misuse of generative artificial intelligence in academic economic research and provides guidelines for university students and instructors. It primarily addresses students and instructors in a master's program in economics; however, the use cases and guidelines may be useful in other fields and academic research in general.

Recommended citation: Burri, M., D. Kaufmann, and N. Ostovan, 2024. "AI in economic research – A guide for students and instructors," IRENE Policy Reports 24-03, IRENE Institute of Economic Research, University of Neuchâtel.

Corresponding author:

Daniel Kaufmann
Université de Neuchâtel
Institut de recherches économiques
Rue A.-L. Breguet 2
CH-2000 Neuchâtel
Switzerland
✉ daniel.kaufmann@unine.ch

Team:

Marc Burri
✉ marc.burri@unine.ch
Nima Ostovan
✉ nima.ostovan@unine.ch

Acknowledgments: We thank Sarah Cotelli Kureth, Paul Cotofrei, Marika Fenley, Bruno Lanz, and Tim Reinicke for helpful discussions and comments, as well as the communication offices of the École Polytechnique Fédérale de Lausanne and the Université de Genève for directing us to their guidelines. Also, we gratefully acknowledge financial support from the University of Neuchâtel for the project titled *ChatGPT for Academic Writing*.

Neuchâtel, July 2024

Contents

1	Introduction	1
2	AI assistants	4
2.1	ChatBots	4
2.2	Reviewing literature	6
2.3	Coding	7
2.4	Writing	8
2.5	Presenting	9
3	Use cases	11
3.1	Brainstorming and feedback	12
3.1.1	Generating ideas	12
3.1.2	Generating hypotheses	14
3.1.3	Obtaining feedback	15
3.2	Background research	17
3.2.1	Finding literature	17
3.2.2	Checking literature	20
3.2.3	Formatting citations	21
3.2.4	Summarizing findings	22
3.2.5	Translating	23
3.2.6	Simplifying language	24
3.2.7	Explaining concepts	25
3.3	Coding	25
3.3.1	Writing	26
3.3.2	Explaining	27
3.3.3	Translating	28
3.3.4	Debugging	29
3.3.5	Commenting	31
3.4	Data analysis	32
3.4.1	Downloading data	32
3.4.2	Creating figures	33
3.4.3	Extracting data from text	34
3.4.4	Reformatting data	36
3.4.5	Classifying text	36
3.4.6	Extracting sentiment	37
3.4.7	Simulating human subjects	38
3.5	Math	39
3.5.1	Setting up models	39
3.5.2	Deriving equations	40
3.5.3	Explaining equations	41
3.6	Writing	43

3.6.1	Writing text	43
3.6.2	Editing text	44
3.6.3	Generating titles and headlines	45
3.6.4	Generating graphs	46
3.6.5	Generating equations from pictures	47
3.6.6	Detecting AI-generated text	48
3.7	Presentation	50
3.7.1	Generating bullet points	50
4	Guidelines	53
4.1	Existing rules and guidelines	53
4.1.1	Rules on scientific integrity at the University of Neuchâtel	53
4.1.2	Existing guidelines on AI	54
4.2	Recommendations for guidelines	58
4.2.1	For students	58
4.2.2	For instructors	61
A	Declaration on the use of AI	66
	References	67

Figures

3.1	Literature review	19
3.2	AI detection	49

Tables

2.1	AI assistants	5
3.1	Summary and rating of use cases	13
4.1	Guidelines and rules at Swiss universities	55

Imagine if we had a ChatBot in 1633, trained on all existing scientific texts, including the works of Plato and Socrates. If people would use this model to evaluate Galileo's revolutionary idea, that the Earth orbits the sun, they would dismiss it. The model would more likely answer that the sun orbits the Earth because this is the view that is more frequently expressed in existing texts. Because the model lacks the ability to see things in new ways, it merely echoes the prevalent beliefs of its time. And if many people go outside to verify what the model says, they would conclude: 'Nope, not moving over here; not moving over here.' and we could all watch the sun go around.

— Felin (2024) paraphrased with ChatGPT, rewritten by the authors.^a

A useful model for interacting with LLMs [Large Language Models] is to treat them like an intern who is highly motivated, eager to help, and smart in specific domains, but who has just walked into the job, lacks the context of what you are doing, and is prone to certain types of errors.

— Korinek (2023)

1 Introduction

Generative Artificial Intelligence (AI) is a rapidly evolving field. Every day, we learn about the new capabilities of AI assistants. However, because these tools are trained on specific data sources, many of which are far from academic, their output is often vague, opaque, misleading, or outright wrong. Indeed, the quotation and interview excerpt at the beginning of this page shows that AI assistants can be very useful, but we should also use them with care and understand their limitations.

This report sheds light on the use and misuse of AI assistants in academic economic research.¹ Although excellent work by Korinek (2023) already exists on this topic, we differ for the following reason: We aim to provide guidance for master students in economics, in particular for writing their theses, and their instructors.

We start by presenting a set of AI assistants that may be particularly useful for economic research. Although the specific assistants may be quickly outdated because of the rapid

^aOriginal text paraphrased by ChatGPT-4, 16 April 2024, OpenAI, chat.openai.com.

¹The use of AI raises issues related to intellectual property and data protection laws (see e.g. [The Economist, 2024](#)). Although we mention this problem several times, our legal expertise does not suffice to assess whether specific uses indeed violate legal provisions.

technological process in this area, we briefly review tools that help with reviewing literature, coding, writing texts, and preparing presentations.

Then, we evaluate the usefulness of various AI assistants for these tasks. We do this by presenting use cases and illustrating the strengths and weaknesses of the most important AI assistants. While the focus stays on ChatGPT, the currently most commonly used AI assistant, we discuss new and traditional alternative tools that can be used for the following tasks: Brainstorming and feedback, background research, coding, data analysis, math, writing, and presenting. We show that AI assistants can significantly speed up some tasks in economic research. In particular, AI assistants make researchers much more productive in solving repetitive and well-structured problems. Examples of this are coding or correcting grammar. However, these assistants are no substitute for understanding the underlying concepts because the researcher is still solely responsible for the correctness and relevance of the AI's output. In addition, there is a danger that using AI violates rules of scientific integrity or data confidentiality. At most universities, AI output is not regarded as a scientific source. In addition, uploading data or documents to an AI assistant for analysis may violate privacy or intellectual property rights.

Finally, we propose a set of guidelines that should help students and instructors assess whether the use of AI is permitted or not. Although instructors and students should discuss case-by-case whether a certain use is permitted, we hope to provide some basic principles that will apply more permanently despite the rapid technological progress:

1. **Permission:** AI assistants are allowed in academic work subject to scientific integrity rules, legal provisions, or explicit constraints imposed by instructors or the university.
2. **Originality:** Students must submit original work and are responsible for its correctness and relevance.
3. **Transparency:** The use of and content created by AI assistants has to be declared.
4. **Quality:** Students must be able to judge the quality of AI output. Therefore, AI assistants are no substitute for developing academic skills; they make skilled academics more productive.

These principles are based on our assessment of existing guidelines at Swiss universities and our teaching experience and do not necessarily represent the official view or binding directives of the University of Neuchâtel. We will make these principles more concrete when proposing specific guidelines in [Chapter 4](#).

In what follows, [Chapter 2](#) presents some helpful AI assistants beyond the most prominent

ChatBots. Chapter 3 provides use cases in academic research and assesses their usefulness. Finally, Chapter 4 summarizes existing guidelines at Swiss universities. In addition, we formulate guidelines for students and instructors that we find most useful.

2 AI assistants

Generative AI underlies a wide variety of recent tools facilitating academic work. The most widely known tools are so-called ChatBots, providing content (text, images, documents, etc.) in response to prompts entered by the user (e.g., ChatGPT).¹ However, a wide variety of specialized apps exist for specific tasks (e.g., language translation with DeepL). Moreover, generative AI will almost certainly become more and more integrated into existing software (e.g. grammar checks in MS Word). We summarize all these tools under the label AI assistants (see Table 2.1 for a non-representative selection of AI assistants). For most of these assistants, the basic version with limited functionality is free of charge. Full versions are often paid services.

Which tool is suitable for a specific task? Most researchers turn to a ChatBot by default, mostly ChatGPT. However, specialized AI assistants may perform a specific task better or more efficiently. Context is essential; AI assistants are often trained on a wide variety of data, scientific and non-scientific. Tuning the assistant to an academic setting is, therefore, key. On the one hand, this can be achieved by providing context in a ChatBot prompt before asking the actual question or by asking follow-up questions. On the other hand, this can be achieved using specialized AI assistants already tuned to an academic setting. This section aims to briefly describe AI assistants with a focus on academic research in economics. An illustration of their current capabilities will be shown in the next chapter on various use cases.

2.1 ChatBots

The most prominent ChatBots were developed (or acquired) by big-tech companies (e.g. chat.openai.com, copilot.microsoft.com, gemini.google.com). A ChatBot is a Large Language Model (LLM) where a user can enter text in a prompt, and the model returns an answer. This answer, mostly text, is a prediction based on the input provided by the user and the

¹In addition, Universities started to develop specialized tools for teaching and learning to ensure equal access to all students (see, e.g., [ETH Zürich, 2024c](#)). Such tools can assist students in the learning process, for example, by generating exam questions. The focus of this report is on tools that help with writing a research paper.

Tab. 2.1: AI assistants

Tool	Description
ChatGPT (GPT-4, GPT-3.5)	Conversational AI model. chat.openai.com
ChatGPT Plugin: AskYourPDF	Plugin for ChatGPT that allows users to submit PDF files to generate a review for the submitted document. www.askyourpdf.com
ChatGPT Plugin: Consensus	Plugin for ChatGPT that provides access to a database of over 150 million research articles. consensus.app
ChatPDF	Platform where users can upload PDF documents and receive responses to their queries from the model. chatpdf.com
Gemini (formerly Bard)	Conversational AI model. gemini.google.com
Copilot (GPT-4, formerly Bing)	Conversational AI model. copilot.microsoft.com , www.bing.com/chat
Perplexity	Conversational AI model. www.perplexity.ai
Grammarly	Writing assistant that checks grammar, spelling, punctuation, and style. app.grammarly.com
Writefull	Writing assistant that checks grammar, spelling, punctuation, and style. www.writefull.com
Hemingway	Writing assistant to make text more readable. hemingwayapp.com
DeepLWrite	Writing assistant that checks grammar, spelling, punctuation, and style. www.deepl.com/write
DeepLTranslate	Translation tool. www.deepl.com/translator
GitHub Copilot	Code completion tool. github.com/features/copilot
Mathpix	Tool that can convert images of equations into MS Word or $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ code. mathpix.com
Compilatio	Tool to detect plagiarism that can also detect AI-generated text. www.compilatio.net
GPTZero	Tool that can detect AI-generated text. gptzero.me
Connectedpapers	Service that identifies papers related to a specific reference paper. www.connectedpapers.com
Litmaps	Service identifies papers related to a specific reference paper. www.litmaps.com
Elicit	Platform enables users to pose questions and obtain responses from a language model with access to over 150 million research articles. elicit.com

information set used to estimate the model's parameters. Many specialized functionalities are now seamlessly integrated into these ChatBots (for example, image and document creation and access to web resources). Besides the big-tech companies, some new players aim to challenge the incumbents in this dynamic field (e.g. www.anthropic.com/claude, www.perplexity.ai, www.mistral.ai).

While these ChatBots are the Swiss army knife of AI assistants (they can do almost everything), researchers must provide context in the prompt to fine-tune them to an academic setting (choose the right knife). Without context, the ChatBots will usually provide vague, unscientific answers that are inadequate for research.

An alternative to providing context in the prompt is to use specialized pre-tuned tools from other providers. These can take the form of plugins, that is, pre-trained and fine-tuned versions of ChatGPT-4.² Yet another alternative is to use specialized AI assistants for specific tasks. In what follows, we will discuss various tools for specific academic activities.

2.2 Reviewing literature

An essential step in a research project is to get an overview of the existing literature. Recent AI assistants promise to make this process much more efficient. One option is simply to ask a ChatBot for relevant literature. However, ChatBots may provide inaccurate results, which would be attributed to the researcher. Therefore, there are specialized tools for this task that may provide more accurate results.

Elicit.com has two main features. First, it finds relevant papers based on a research question. Second, it summarizes the most important papers and key concepts related to that research question. The site, therefore, allows the researcher to obtain a quick overview of the most relevant literature related to his/her research question.

Connectedpapers.com allows the researcher to upload a research paper based on which the site suggests related literature. This lowers the probability of missing a relevant paper in a specific field. In addition, the site has the functionality to create bibliographies.

Litmaps.com suggests related papers based on a collection of input papers. The functionality allows the researcher to construct maps of related and relevant literature.

²The number of plugins is growing rapidly. Therefore, we could not cover a relevant amount of plugins in the test. Plugins are available only in the paid version of ChatGPT.

[Consensus.app](#) provides relevant papers related to a research question. An important point about this website is that it gives the researcher feedback on whether a research question has already been answered. The website also provides a summary based on all highly related papers it finds.

An open question for all these new tools is whether the underlying corpus includes the most relevant papers in economics and related fields. Clearly, the results will only be as good as the training data set. Moreover, it may be difficult for inexperienced researchers to assess the quality of the output from the AI assistant. Therefore, searching literature using traditional approaches remains important (e.g., Google Scholar, IDEAS/RePEc, Scopus, internet search, library, and discussions with experienced researchers).

2.3 Coding

Theoretical or empirical analyses in economics usually require writing code in a programming language (e.g. Matlab, Python, or R). AI assistants can help find errors in existing codes or propose code snippets that solve specific problems. ChatBots like chatGPT have been shown to produce impressive results. For example, a researcher can post code snippets with an error and ask the ChatBot to correct it. However, one can also ask ChatBot to provide the first version of the code based on a problem description. Finally, it is possible to ask a ChatBot to translate a code snippet from one programming language to another.

More recently, a specialized coding assistant based on an LLM has been published by GitHub (github.com/features/copilot). This coding assistant can be integrated into various programming platforms (e.g. Visual Studio Code or RStudio) and provides real-time proposals for written code in various programming languages. The main advantage compared to a ChatBot is the real-time feedback from the coding assistant.

However, these assistants are also prone to making mistakes. For example, they may invent functions or commands that do not exist in a specific programming language. Consequently, the researcher will have to debug the code provided by the ChatBot again to make it work. This problem can be particularly severe for programming languages without a large online community (recall that the models are mainly trained based on data from the internet).

The upshot is that AI assistants are currently no substitute for coding skills. Researchers must be able to write and understand code themselves to assess the correctness of AI-written code. While AI assistants can increase the productivity of a researcher who knows how to code, a

researcher without coding abilities cannot produce correct and reliable code even with an AI assistant.

2.4 Writing

ChatBots can be used to write and correct written text in different languages. However, the results are often inappropriate for economic research. Without any context, the writing style of ChatBots is hardly scientific because it represents the average writing style of the training data (mainly the internet).³

ChatGPT has the functionality to offer fine-tuned GPTs. For academic research, Michael Haenlin provides an academic writing editor with better results than the standard chatGPT prompt (see chat.openai.com/g/g-W7TtM1Vne-academic-editor). Researchers can paste their written text and receive a refined result with better academic wording and fewer language errors.

essayailab.com checks grammar, writing style, and plagiarism. It also has the possibility to directly write a text online, where the tool makes real-time suggestions.

app.grammarly.com is an app that corrects grammar and spelling and improves writing style (paid version). A browser integration can be used to correct documents written with online tools (e.g., Overleaf) and an integration for MS Office.

www.deepl.com is mostly known for translations. However, there is also a writing optimizer for English (www.deepl.com/write). It features a browser integration. However, the tool may soon require a subscription.

hemingwayapp.com is an app that makes English text more readable. It aims at short and clear sentences and paragraphs. Usually, this is more consistent with a scientific writing style. There is a desktop app, and the paid version allows the researcher to upload files and let the tool rewrite text. One advantage of the tool is that it gives feedback on the reasons why a text is hard to read. This will result in better text and improve the researcher's writing skills.

www.writefull.com is a specialized AI assistant for academic writing. It corrects mistakes and offers suggestions for improvement. The main advantage is that it is trained on scientific literature, and therefore, its style is more academic than that of a standard ChatBot. It integrates various text editors (e.g. Overleaf and MS Word).

³In addition, letting a ChatBot write entire sections may violate the rules of scientific integrity (see Chapter 4).

Translation tools may also come in handy when writing or reading academic texts. For example, a paper may be written in another language and requires translation. Or a non-native English speaker may prefer to write a first draft in their native language and then use the translation as a first version of the text. The standard tools for this task are [deepl.com](https://www.deepl.com), [reverso.net](https://www.reverso.net), [translate.google.com](https://www.translate.google.com), sider.ai, www.modernmt.com, www.machinetranslation.com. Note, however, that translating text may also result in a change of meaning or loss of information.

The problem with most of these AI assistants is that an inexperienced researcher cannot check the quality of the suggestions. The assistants are useful for correcting obvious errors (typos, punctuation, grammar). However, they perform less well in terms of style and structure. If the suggested writing style is not academic and the structure is unclear, it may worsen the text's quality. Researchers should, therefore, improve their writing style. A good starting point is the writing tips by [Nikolov \(2023\)](#), which also contain references to standard works related to good writing in economics or academic writing courses at the University.⁴

2.5 Presenting

ChatBots can be used to create presentations (content and graphics). More and more of these capabilities are integrated in the most prominent ChatBots or even in traditional software. For example, PowerPoint includes a Copilot add-in, which can create a presentation based on a source file. However, a wide variety of specialized platforms are also available for this task.

gamma.app is a tool to prepare presentations based on content provided by the researcher. It also allows the researcher to choose from various themes.

www.popai.pro is based on chatGPT. One functionality is to construct a PowerPoint presentation based on an existing document. This can serve as a first pass of a presentation based on a written academic text.

For L^AT_EX beamer slides, there are fine-tuned GPTs that allow us to convert a LaTeX document into beamer slides with references (e.g. chat.openai.com/g/g-ynWaaMhkf-latex-beamer-assistant). The AI assistant will automatically propose slide titles and structure the presentation.

There are a variety of tools that can be used to create images and visual aids. However, disputes about intellectual property rights are ongoing. In fact, legal provisions may have been violated when training such models. Therefore, output from such models may violate intellectual property rights, and some universities forbid using such output (see Chapter 4).

⁴See, e.g., www.unine.ch/cdl/en/home.html.

openai.com/dall-e-3 is integrated into ChatGPT and trained to generate images from textual descriptions. Some examples of its capabilities include creating plausible images of fictional creatures, modifying existing images in specific ways based on textual input, generating images of objects in specific styles or settings, and constructing images based on abstract prompts. app.leonardo.ai similarly creates images, but in better quality and fitting the desired artistic style (fantasy, classic, painting, etc.).

What makes a good scientific presentation depends on the field of research. A complete discussion is beyond the scope of this report. In economics, simple presentations without dynamic elements are usually preferred. The focus should remain on the content and clarity of exposition. Again, context is essential. Researchers should create presentations in a format and style that meets scientific standards. In addition, creating presentations based on an existing paper may violate scientific integrity rules and infringe intellectual property laws (see Chapter 4).

3 Use cases

This section shows how AI assistants can support economic research and discusses their limitations. We categorize these examples into seven activities: brainstorming and feedback, preliminary research, writing, coding, data analysis, math, and presentations. For each activity, we provide use cases and assess the usefulness of AI. For each use case, We evaluate the ability of the AI assistant to save time and offer recommendations on which tools to use and which to avoid. We aim to offer a balanced view of their usefulness, acknowledging their current strengths and weaknesses.

We relied heavily upon the work by [Korinek \(2023\)](#), with some extensions and modifications. Many examples and further resources are also available on the website by [Korinek and Taliaferro \(2024\)](#). The purpose of this section is to evaluate the tools on our own, take into account recent improvements in AI tools since [Korinek \(2023\)](#), and then incorporate this experience into guidelines in the next section. Some prompts are indeed identical to [Korinek \(2023\)](#), but we modified the applications to form our own opinion on the usefulness of specific tasks.

We evaluate tools for a wide variety of tasks. The tools discussed in the previous Chapter are the basis of the analysis. However, we focus on the most common ones, particularly the paid version ChatGPT-4, which also allows uploading and analyzing PDFs and other documents.

We also evaluated other ChatBots and obtained similar results. Unless otherwise stated, the performance of ChatGPT and other ChatBots is, therefore, similar.¹

Table 3.1 summarizes the results, which are broadly in line with [Korinek \(2023\)](#). ChatBots based on large language models (LLMs; GPT-4 in particular) are helpful for various tasks. However, the advantage of other tools such as Grammarly, Writefull, or GitHub Copilot is that they are already integrated into traditional software. This makes them more convenient to use than ChatBots. Moreover, some of them are already tuned to an academic setting. For all

¹In what follows, we use the terms LLM and ChatBot interchangeably. Note that LLMs underlie most AI assistants, whereas a ChatBot is primarily a way to interact with an LLM.

these examples, note that context is essential. If the answer is unsatisfactory, it may help explain to the AI assistant what purpose the answer is needed for (e.g., for an academic paper or a master's thesis).

In what follows, we show the use cases with the input we provided (prompt) and the output generated. Several comments are in order. ChatBots usually provide different answers when asked the same question twice. They introduce some randomness into the answers. Therefore, the output we provide in the report is not precisely reproducible.² We show use cases based on the most effective models, predominantly the paid version of ChatGPT (GPT-4). Finally, in some instances, we have condensed or omitted output for the sake of brevity.

Note that AI assistants may provide vague or incorrect information in all these use cases. All output should, therefore, be carefully evaluated before being used in a research paper (see Chapter 4).

3.1 Brainstorming and feedback

Research projects begin with a brainstorming phase. This stage involves generating ideas and hypotheses. Feedback on preliminary work is also important. Researchers can ask AI assistants if no colleague or supervisor is available to talk.

3.1.1 Generating ideas

AI assistants can be used to generate ideas for a research project by providing a starting point for further exploration. Prompt 1 asks ChatGPT (GPT-4) to generate five research ideas for a Master's thesis in empirical macroeconomics.

Prompt 1: I would like to write a Master thesis in the field of empirical macroeconomics. Specifically, I'd like to dig deeper into VARs (vector autoregressions). Brainstorm 5 research ideas with data for Switzerland (one line of explanation). ChatGPT (GPT-4)

- 1. Examining the Impact of Swiss Monetary Policy on Inflation and Unemployment:** Utilize VAR models to explore how changes in the Swiss National Bank's policies influence key macroeconomic variables.
- 2. Assessing the Effects of Global Economic Shocks on the Swiss Economy:** Study how international financial crises or major global economic events impact Switzerland's

²LLMs incorporate a tuning parameter known as temperature, denoted as τ , which regulates the level of randomness in their generated output. A temperature setting of 0 indicates deterministic output. However, the models under consideration utilize a temperature setting of $\tau > 0$, with no provision for user adjustment. Consequently, inherent randomness exists in the output, rendering it irreproducible. The temperature can be adjusted with the Application Programming Interface (API) to access GPT-3 and GPT-4 offered by OpenAI.

Tab. 3.1: Summary and rating of use cases

Category	Task	Recommended tools	Useful
3.1 Brainstorming & Feedback	Idea generation	LLMs	●
	Generating hypotheses	LLMs	◐
	Obtaining feedback	LLMs, ChatGPT Plugin: AskYourPDF	◐
3.2 Background research	Find literature	LLMs, Connectedpapers	◐
	Check literature	LLMs with internet	◐
	Format citations	LLMs	◐
	Summarize findings	LLMs	◐
	Translation	LLMs, DeepL	●
	Simplifying language	LLMs	◐
	Explanation of concepts	LLMs	●
3.3 Coding	Writing	LLMs, GitHub Copilot	●
	Explaining	LLMs	●
	Translating	LLMs	●
	Debugging	LLMs	●
	Commenting	LLMs, GitHub Copilot	●
3.4 Data analysis	Downloading data	LLMs	◐
	Creating figures	ChatGPT (GPT-4)	●
	Extracting data from text	LLMs	◐
	Reformatting data	LLMs	◐
	Classifying and scoring text	LLMs	◐
	Extracting sentiment	LLMs	◐
	Simulating human subjects	LLMs	◐
3.5 Math	Setting up models	LLMs	◐
	Deriving equations	LLMs	◐
	Explaining equations	LLMs	◐
3.6 Writing	Writing text	LLMs	◐
	Editing text	LLMs, Grammarly, Writefull, Hemingway	●
	Generating titles and headlines	LLMs	●
	Generating graphs	ChatGPT (GPT-4)	●
	Generating equations from pictures	ChatGPT (GPT-4), Mathpix	●
	Detecting AI-generated text	GPTZero, Compilatio	○
3.7 Presentation	Generating bullet points	LLMs	◐

Notes: LLMs include ChatGPT (GPT-3.5 and GPT-4), Bard (now Gemini), and Bing (now Copilot). ChatGPT (GPT-4) and Bing (now Copilot) are LLMs with internet access. The third column is our subjective rating of the tools' usefulness. ●: Very useful and saves a lot of time, ◐: Useful, might save some time ○: Not useful.

economic indicators through a VAR analysis.

3. **Analyzing the Relationship Between Swiss Franc Exchange Rates and Trade Balances:** Investigate how fluctuations in the Swiss Franc affect the country's trade balance, using VAR to understand these dynamic interactions.

4. **Exploring the Interplay Between Swiss Tourism and National GDP:** Use VAR models to assess the impact of tourism, a significant sector in Switzerland, on the country's overall economic growth.

5. **Studying the Influence of European Union Policies on the Swiss Financial Sector:** Analyze how EU financial regulations and policies affect Switzerland's banking and finance industries, despite not being an EU member, using VAR methodologies.

The ideas are relevant to empirical macroeconomics and specific to Switzerland. However, the ChatBot does not discuss relevance and feasibility. As such, it provides a good starting point for further exploration. It is then helpful to brainstorm the most promising topics further. Focusing on the fourth point, we can ask follow-up questions like, "Where can I find tourism data for Switzerland?". Note that finding data requires a ChatBot with internet access, which is the case for the paid version of ChatGPT. However, using an internet search engine will usually yield similar results.

3.1.2 Generating hypotheses

AI assistants can provide arguments or counterarguments to a statement. This can be helpful to formulate hypotheses as well. Prompt 2 asks ChatGPT (GPT-4) to generate five arguments supporting a statement. The arguments are relevant and plausible.

Prompt 2: Provide five important arguments for the following (one line of explanation): More than 50% of people rate their behavior as more environmentally friendly than that of their peers. ChatGPT (GPT-4)

1. **Self-Perception Bias:** People tend to view themselves in a positive light, leading them to overestimate their own environmentally friendly actions compared to others.
2. **Social Desirability Bias:** Due to the societal praise for eco-friendly behavior, individuals might claim to be more environmentally conscious to align with these favorable perceptions.
3. **Illusory Superiority Effect:** Common in human psychology, this effect leads people to believe they are better than the average, including in environmental practices.
4. **Information Asymmetry:** Individuals are more aware of their own efforts in being eco-friendly and less aware of the efforts made by others, skewing their comparative assessments.
5. **Varying Standards and Definitions:** Different personal standards for what constitutes 'environmentally friendly' can lead people to believe their actions are superior to those of their peers.

Prompt 3 asks to generate five plausible counterarguments to a statement. Again, the arguments brought forward seem plausible.

Prompt 3: Provide five important counterarguments for the following statement (one line of explanation): Immigration leads to higher unemployment. ChatGPT (GPT-4)

1. **Complementary Job Skills:** Immigrants often bring diverse skill sets that complement the local workforce, filling gaps in the job market rather than displacing native workers.
2. **Job Creation through Entrepreneurship:** Immigrants are statistically more likely to start businesses, which creates new jobs and stimulates economic growth.
3. **Increased Demand for Goods and Services:** A larger population due to immigration boosts demand in the economy, leading to job creation in various sectors.
4. **Labor Market Flexibility:** Immigrants often take jobs that native workers are less willing to do, providing essential labor in certain sectors and contributing to overall economic efficiency.
5. **Positive Impact on Aging Workforce:** In countries with aging populations, immigrants can mitigate workforce shrinkage and the associated economic challenges, thus supporting employment stability.

Such prompts can be very helpful in an initial research phase for an inexperienced researcher without an overview of a field. They serve as a starting point. In any case, the arguments should be vetted for correctness, and the original source should be obtained and cited.

3.1.3 Obtaining feedback

AI assistants can also be used to provide feedback, which can be particularly useful when seeking feedback on an early draft. We evaluated three promising ways to get feedback on research papers.³ The first approach uses ChatGPT (GPT-4), which allows uploading files. Second, we can use the AskYourPDF plugin for ChatGPT. Finally, ChatPDF is a web-based platform where the researcher can upload PDF documents and receive responses to their queries from the model.⁴ We ask the models to provide feedback on Burri (2023), who forecasts Swiss GDP growth using daily lead texts from newspaper articles. Note that a referee is an academic expert who is tasked with reviewing a paper sent to an academic journal. Therefore, we ask ChatGPT to provide a report from such a referee. Prompt 4 shows the result.

³A fourth possibility would be to use Claude AI (see e.g. Korinek, 2023). However, at the time of writing this report, Claude AI was not available in Switzerland.

⁴Only ChatPDF is available for free.

Prompt 4: Please write a referee report for the following working paper: <https://www.econstor.eu/bitstream/10419/278605/1/1855951029.pdf>

ChatGPT (GPT-4), Plugin: AskYourPDF

Summary

The working paper by Marc Burri, titled "Do daily lead texts help nowcasting GDP growth?", investigates the utility of daily news lead texts in nowcasting Swiss GDP growth. The author collects titles and lead texts from three Swiss newspapers, creating text-based indicators for various economic concepts. A composite indicator, derived from these text-based indicators, demonstrates a high correlation with low-frequency macroeconomic data and survey-based indicators. The paper's primary contribution is the demonstration that a composite indicator, calculated from daily news texts, outperforms a monthly Swiss business cycle indicator in a pseudo-out-of-sample nowcasting exercise for Swiss GDP growth, particularly during economic distress periods.

Evaluation

Originality and Contribution: The paper's approach to utilizing daily news texts for nowcasting GDP growth is innovative, especially in the context of Swiss economic data. The development of the "Short Economic News Indicator" (SENI) and its application in nowcasting exercises adds valuable insights into the literature on mixed-frequency data analysis and economic forecasting. The focus on the effectiveness of text-based indicators during economic distress periods is particularly timely and relevant.

Policy Implications and Practical Relevance: The practical implications of this research are significant for policymakers and economic analysts. The ability to nowcast GDP growth accurately and timely, especially during periods of economic uncertainty, can enhance decision-making processes. The paper successfully argues for the inclusion of high-frequency text data in economic forecasting models, highlighting its cost-effectiveness and timeliness compared to traditional survey data.

Suggestions for Improvement:

Data Sources and Bias: Expand on the selection process for the newspapers and address potential biases in news coverage.

Robustness Checks: Conduct additional robustness checks to ensure the findings are not specific to the particular economic events covered in the study period.

Comparative Analysis: Include a comparative analysis with other high-frequency indicators to contextualize the SENI's performance further.

Technical Details: Provide more technical details on the factor model and the criteria for selecting specific economic concepts for sentiment indicators.

Conclusion

Overall, the paper makes a valuable contribution to the literature on economic forecasting and nowcasting. The innovative use of daily news texts for nowcasting GDP growth, particularly in the Swiss context, offers new insights and methodologies for economic analysis. With some enhancements, especially in addressing potential biases and expanding on methodological details, this paper could significantly impact the field.

The report is well structured and provides an evaluation of the paper. It highlights the

paper's originality, contribution, policy implications, and practical relevance. It also offers suggestions for improvement, such as expanding on the selection process for the newspapers and conducting additional robustness checks.

However, the feedback is relatively superficial and vague. In addition, the suggestions are based on the "opinion" of the AI assistant and do not necessarily reflect the views of an experienced researcher in the field. Therefore, it is helpful to get quick initial feedback but does not substitute for discussions with other researchers or academic supervisors, who can tell in detail what and where something needs to be improved.

Other useful Prompts to get feedback include (see also [Korinek, 2023](#)):

- I am an economist writing my master's thesis on the effects of monetary policy in Switzerland. What are the main challenges in researching this topic, and how can I address them best?
- Please provide ways how to improve the following text/paper.
- What are the main strengths and weaknesses of this paper?
- What are the main novel ideas in the paper that are not sufficiently emphasized?
- Can you identify any instances of bias in this paper?

3.2 Background research

Once a topic is found, researchers must grasp the existing literature. Several AI assistants help find, summarize, and understand the literature.

3.2.1 Finding literature

Prompt 6 asks ChatGPT (GPT-4) to provide a list of five influential papers on the topic of empirical macroeconomics, explicitly focusing on vector autoregressions (VARs).

Prompt 5: I would like to write a Master thesis in the field of empirical macroeconomics. Specifically, I'd like to dig deeper into VARs. Provide 5 top influential papers (Only author, year, title and journal). ChatGPT (GPT-4)

1. **Sims, Christopher A. (1980)**, "Macroeconomics and Reality", *Econometrica*, Vol. 48, No. 1.
2. **Bernanke, Ben S. (1986)**, "Alternative Explanations of the Money-Income Correlation", *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25.

3. **Blanchard, Olivier J., and Quah, Danny (1989)**, "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 79, No. 4.
4. **Christiano, Lawrence J., Eichenbaum, Martin, and Evans, Charles L. (1999)**, "Monetary Policy Shocks: What Have We Learned and to What End?", *Handbook of Macroeconomics*, Vol. 1A.
5. **Clarida, Richard, Gali, Jordi, and Gertler, Mark (2000)**, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", *Quarterly Journal of Economics*, Vol. 115, No. 1.

We obtain a list of influential papers that are relevant to the topic, including the correct authors, journals, and publication years. Therefore, ChatGPT performs relatively well. In our experiments, other ChatBots without internet access often failed to provide relevant literature and sometimes invented papers that do not exist. Therefore, the output should always be checked for relevance and correctness.

To illustrate additional limitations, we ask ChatGPT to provide ten papers on vector autoregressions published in the *American Economic Review* (AER). Suggestion 3 was published in the *Journal of Economic Perspectives*, and suggestion 10 does not exist (there is a similarly titled paper in the journal *Econometric Theory*). The example shows that the output from AI assistants regarding academic literature should always be treated with care.

Prompt 6: Can you list 10 papers published in the AER that use vector autoregressions? ChatGPT (GPT-4)

Here are 10 papers published in the *American Economic Review* that use vector autoregressions (VAR):

...

3. **Vector Autoregressions**, by James H. Stock and Mark W. Watson, Fall 2001.

...

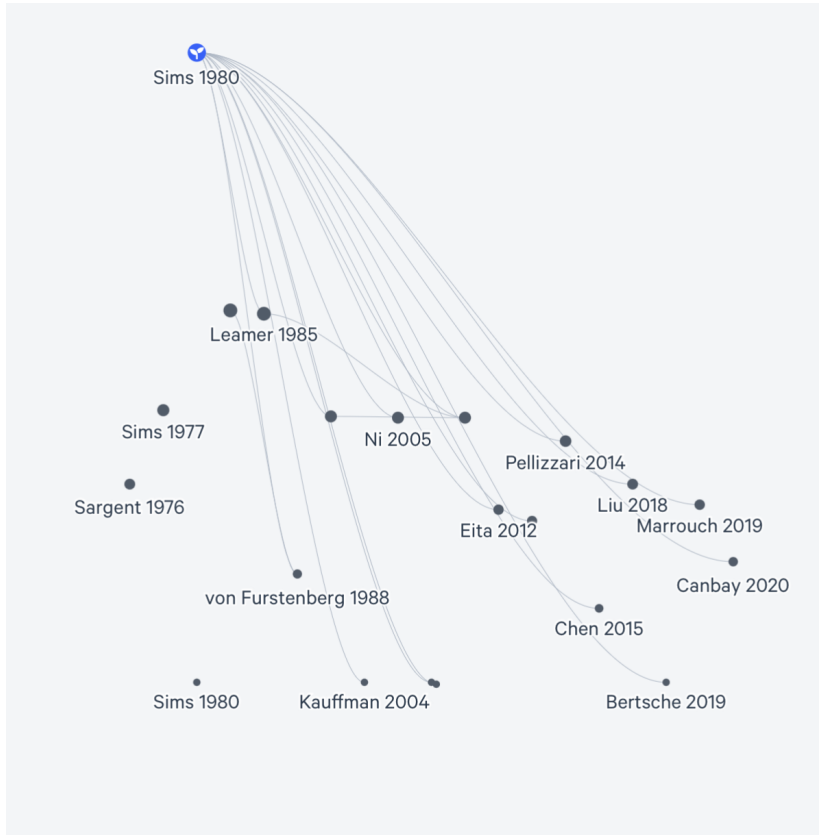
10. **Advances in Using Vector Autoregressions to Estimate Structural Parameters**, December 2021.

Do specialized tools perform better? Figure 3.1 presents the results from two platforms using [Sims \(1980\)](#) as the reference paper. These tools then provide related literature and offer a graphical representation. We see that the two platforms identify a different set of papers. Indeed, for a researcher familiar with the field, it is clear that Litmaps does not identify the most relevant papers. Connectedpapers seem to perform better. In addition, it provides abstracts to get a quicker grasp of the main results of the papers.

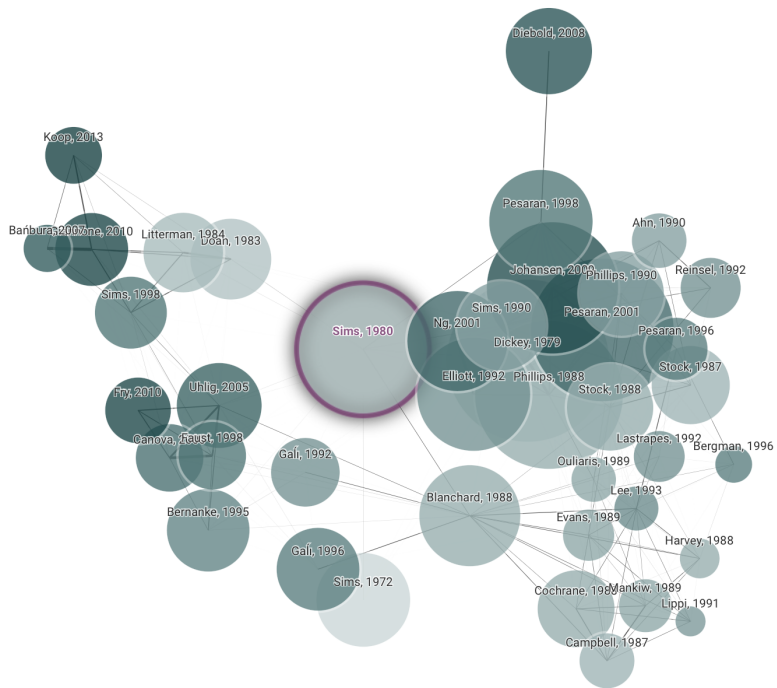
We see that identifying the relevant literature using AI assistants is challenging. They merely provide a starting point but do not substitute for traditional approaches to literature review

Fig. 3.1: Literature review

a) Litmaps



b) Connectedpapers



Notes: The seed paper provided on both platforms is [Sims \(1980\)](#)

(e.g. traditional archives like Scopus, IDEAS/RePEc, JSTOR, or Google Scholar, as well as references in relevant papers and survey articles; finally, read abstracts and introductions, go to the library and discuss with experienced researchers).

3.2.2 Checking literature

We have seen that literature provided by AI assistants is often incorrect or does not exist. However, AI assistants can also be used to check these references. Prompt 7 asks Bing (GPT-4, now copilot) to validate a list of publications, with the first four being genuine and the fifth fabricated.

Prompt 7: Please verify whether the following publications are real and correct:

1. **Sims, Christopher A. (1980)**, "Macroeconomics and Reality", *Econometrica*, Vol. 48, No. 1.
2. **Bernanke, Ben S. (1986)**, "Alternative Explanations of the Money-Income Correlation", *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25.
3. **Blanchard, Olivier J., and Quah, Danny (1989)**, "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 79, No. 4.
4. **Christiano, Lawrence J., Eichenbaum, Martin, and Evans, Charles L. (1999)**, "Monetary Policy Shocks: What Have We Learned and to What End?", *Handbook of Macroeconomics*, Vol. 1A.
5. **Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. (2005)**. Why Doesn't the News Impact Asset Prices as Much as Theory Predicts? *Journal of Finance*, 60(4), 1801-1837.

Bing (GPT-4)

Here are the verifications for the publications you mentioned:

1. **Sims, Charles A. (1980)**, "Macroeconomics and Reality", *Econometrica*, Vol. 48, No. 1. This publication is real and correctly cited.
2. **Bernanke, Ben S. (1986)**, "Alternative Explanations of the Money-Income Correlation", *Carnegie-Rochester Conference Series on Public Policy*, Vol. 25. This publication is real and correctly cited.
3. **Blanchard, Olivier J., and Quah, Danny (1989)**, "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 79, No. 4. This publication is real and correctly cited.
4. **Christiano, Lawrence J., Eichenbaum, Martin, and Evans, Charles L. (1999)**, "Monetary Policy Shocks: What Have We Learned and to What End?", *Handbook of Macroeconomics*, Vol. 1A. This publication is real and correctly cited.
5. **Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. (2005)**. Why Doesn't the News Impact Asset Prices as Much as Theory Predicts? *Journal of Finance*, 60(4), 1801-1837. I'm sorry, but I couldn't find any information about this publication. It's possible that the citation is incorrect or the publication is not widely available online. You may want to check the details again or try to find it in a specific database or library.

The model accurately identifies the fabricated reference. Note that this result depends on

Bing's (now Copilot) internet access compared to other ChatBots. ChatBots without internet access did not identify the erroneous reference in our experience. Clearly, checking references remains an important task. Alternatively, researchers can use [Concensus.app](#) plugin for ChatGPT, which provides access to a database containing over 150 million research papers to validate references. Even without internet access, the model noted that it is not sure whether the fifth publication exists indeed. However, the plugin may not include working papers or work-in-progress, which could be a limitation for those citing the most current research developments.

To conclude, AI assistants with internet access can help verify references. However, we still recommend a traditional approach, using the repositories mentioned in the previous section to double-check the results. In any case, citing literature without knowing its content is not recommended.

3.2.3 Formatting citations

AI assistants can format citations. Prompt 8 shows how ChatGPT (GPT-4) can be used to bring the paper by [Sims \(1980\)](#) into BibTeX format (the format used in \LaTeX documents).

Prompt 8: Please bring the following paper into bibtex format:

Sims, Christopher A. (1980), "Macroeconomics and Reality", *Econometrica*, Vol. 48, No. 1.

ChatGPT (GPT-4)

```
@article{sims1980macroeconomics,  
title={Macroeconomics and Reality},  
author={Sims, Christopher A.},  
journal={Econometrica},  
volume={48},  
number=1,  
year={1980}  
}
```

We see that this works well. As \LaTeX is a programming language for document creation, LLMs understand the syntax well and produce accurate results. We also experimented with providing an MS Word document with references and letting the ChatBot reformat them. The results were sensible, but often not without error. Several tools are available to organize the literature and create a reference list, such as Zotero, Mendeley, and EndNote. Most of these offer integration with word processors, such as MS Word or \LaTeX , allowing for easy citation in different styles and bibliography creation. In addition, Google Scholar is still a very useful and often faster tool for obtaining references for academic papers in various formats.

Other useful Prompts include:

- Please bring the following paper into APA format.
- Please bring the following paper into Chicago format.
- Please bring the following paper into MLA format.
- Please bring the following paper into Harvard format.

3.2.4 Summarizing findings

Sometimes, research papers are written in lengthy and complex language, making it difficult to understand the main findings. AI assistants can summarize and simplify texts. Prompt 9 shows how ChatGPT (GPT-4) can be used to summarize the main findings of the paper by Sims (1980).

Prompt 9: Please summarize the main findings of the attached paper. [Attachment: SimsMacroReality1980.pdf] ChatGPT (GPT-4)

The paper "Macroeconomics and Reality" by Christopher A. Sims critically evaluates existing macroeconomic econometric strategies, highlighting their limitations and proposing an alternative approach. Sims argues that conventional macroeconomic models often rely on over-identifying restrictions based on dubious a priori assumptions, which might not hold true in reality. He suggests that these models are unlikely to accurately capture the true dynamics of macroeconomic variables. Instead, Sims advocates for a more flexible, less restrictive modeling approach that does not depend heavily on theoretical preconceptions. He illustrates this with an example of econometric work that accounts for the critique of standard practices, proposing a shift towards models that better reflect the data without imposing unnecessary theoretical constraints. This approach aims to enhance the empirical relevance and reliability of macroeconomic econometric analysis.

The model summarizes the paper, highlighting the main findings and contributions. The answers of ChatPDF and the AskYourPDF plugin for ChatGPT are similar. The other AI assistants are not able to read PDF files; however, they can summarize the main findings of a paper if the text is provided in the prompt (given it does not exceed the context window). Other specialized tools exist to summarize the findings of a paper. These are usually connected to literature review tools such as [Elicit.com](https://elicit.com).

We recommend this function only in an early phase of the literature review to get a quick grasp of the literature. As an alternative, the classical approach is to read the abstract and introduction of a paper. Of course, if the paper is used in the research paper, for example, in a

literature review, the researcher must read it and know its content. There is no guarantee that the ChatBot will get it right. In addition, beware that uploading PDFs to a ChatBot may violate the rules of scientific integrity or legal provisions (see Chapter 4).

3.2.5 Translating

AI can translate written text into different languages. Prompt 10 shows how ChatGPT (GPT-4) can translate a complex text from English to French.

Prompt 10: Translate the following text to French:

The monetary policy paradigmatic discourse, inextricably interwoven within the macroeconomic milieu, endeavors to optimize the intricate equilibrium between inflationary pressures and economic growth trajectories. Central banks, vested with the solemn fiduciary responsibility to steward the monetary aggregates, deploy a panoply of instruments, such as the federal funds rate, open market operations, and reserve requirements, in a concerted effort to calibrate the money supply dynamics. The overarching objective of these deliberations, couched in the rhetoric of price stability, is to mitigate the proclivity for undue price-level fluctuations while preserving the underpinning integrity of financial markets, thereby fostering an environment conducive to sustainable economic prosperity.

ChatGPT (GPT-4)

Le discours paradigmatique de la politique monétaire, inextricablement entrelacé au sein du milieu macroéconomique, s'efforce d'optimiser l'équilibre complexe entre les pressions inflationnistes et les trajectoires de croissance économique. Les banques centrales, investies de la solennelle responsabilité fiduciaire de gérer les agrégats monétaires, déploient une panoplie d'instruments, tels que le taux des fonds fédéraux, les opérations d'open market et les exigences de réserves, dans un effort concerté pour calibrer la dynamique de l'offre de monnaie. L'objectif principal de ces délibérations, exprimé dans la rhétorique de la stabilité des prix, est de mitiger la propension aux fluctuations indésirables du niveau des prix tout en préservant l'intégrité sous-jacente des marchés financiers, favorisant ainsi un environnement propice à la prospérité économique durable.

The model provides a translation that enables a non-English speaker to understand the paper's major points. The translations of DeepL and the other AI assistants are similar. However, DeepL only allows the free translation of a limited amount of text. With ChatBots, the amount of text that can be translated is only limited by the context window.⁵ To translate large amounts of text, the GPT API is a good alternative.

We do not aim to assess the quality of the translation. However, a common problem in text generated with ChatGPT is that it uses complicated words and long sentences. If this is

⁵A context window refers to the span of text that the model considers when generating responses. It encompasses the words preceding and following the current word being processed. The size of the context window determines how much information the model can utilize to generate its response.

the case, asking the ChatBot to use less complicated words and sentences may improve the readability. Another caveat for translations is that the ChatBot may change the meaning of the text or miss nuances. However, this is also the case for any human who reads a text in a foreign language.

3.2.6 Simplifying language

ChatBots can simplify complex language. For example, the text we have translated in Prompt 10 is unnecessarily complex and difficult to understand. Prompt 11 shows how ChatGPT (GPT-4) can simplify such a complex text.

Prompt 11: Please simplify the following text:

The monetary policy paradigmatic discourse, inextricably interwoven within the macroeconomic milieu, endeavors to optimize the intricate equilibrium between inflationary pressures and economic growth trajectories. Central banks, vested with the solemn fiduciary responsibility to steward the monetary aggregates, deploy a panoply of instruments, such as the federal funds rate, open market operations, and reserve requirements, in a concerted effort to calibrate the money supply dynamics. The overarching objective of these deliberations, couched in the rhetoric of price stability, is to mitigate the proclivity for undue price-level fluctuations while preserving the underpinning integrity of financial markets, thereby fostering an environment conducive to sustainable economic prosperity.

ChatGPT (GPT-4)

The text talks about how the conversation around monetary policy, which is deeply connected to the overall economy, aims to balance controlling inflation and supporting economic growth. Central banks have the important job of managing the country's money supply using various tools like interest rates, buying and selling government securities, and setting bank reserve requirements. Their main goal is to keep prices stable, avoiding too much fluctuation in prices, which helps maintain a healthy financial market and supports ongoing economic growth.

AI assistants can help a researcher to understand complex papers.⁶ However, by doing so, it is also possible that ChatBots change the meaning of the text.

Other useful Prompts include:

- Explain the following text to a five-year-old.
- Explain the following text to a high school student.
- Explain the following text to a non-economist.

⁶In addition, unnecessarily complex writing is inconsistent with a good academic writing style in economics. ChatBots could help improve the writing style of a research paper. We will discuss these applications in a later section.

3.2.7 Explaining concepts

AI assistants can explain complex concepts. Prompt 12 shows how ChatGPT (GPT-4) explains the concept of fixed effects.

Prompt 12: Please explain in one paragraph what fixed effects are. ChatGPT (GPT-4)

Fixed effects are a statistical technique used in the analysis of panel data, where the same subjects are observed across multiple time periods. This approach is designed to control for unobserved heterogeneity when this heterogeneity is constant over time and unique to the individual or entity being studied. By including individual-specific variables (fixed effects) that capture these unobserved characteristics, the model aims to isolate the effect of independent variables on the dependent variable by removing the influence of time-invariant characteristics. This is particularly useful in econometrics and social sciences for analyzing the effects of variables over time while accounting for inherent differences across subjects that could bias the results.

The model provides a detailed explanation. If parts of the explanation are still unclear, follow-up questions such as “What do you mean by unobserved heterogeneity?” or “Can you give an example of fixed effects?” can be asked.

3.3 Coding

LLMs can generate code in multiple programming languages. They can also explain, translate, debug, clean, and document existing code. The advantage of LLMs is that the internet comprises large repositories of code used for training. Moreover, LLMs find the syntax of computer code relatively easy to grasp and learn (Korinek, 2023).

In what follows, we will apply ChatBots based on LLMs for coding, particularly with the statistical language R. However, these AI assistants likely perform well for all programming languages with large online communities (e.g. C++ or Python). We will focus on ChatGPT; another helpful tool is GitHub Copilot, which is a plugin for coding environments (e.g. Visual Studio Code or RStudio). Copilot is useful because it suggests code completions and can generate entire functions as the researcher writes the code. A back-and-forth between a ChatBot and the coding environment can be avoided.⁷

One crucial caveat applies. As we discussed in Chapter 2, there is currently no way to check whether the code is correct without some prior knowledge of programming languages. Therefore, these tools are not a substitute for coding skills but rather increase the productivity

⁷GitHub Copilot is free for students, but it is a paid service for others.

of researchers with knowledge in this area.

3.3.1 Writing

Prompt 13 shows how ChatGPT (GPT-4) writes a function performing a linear regression in R.

Prompt 13: Create a function in R for a linear regression with the following properties:

- * Do not use any existing functions from R
- * data input is a data frame or a tibble
- * model specification is done with a formula, where `.(dot)` represents all but the dependent variable
- * assume all variables are numeric
- * return coefficients, p-values as well as R^2 .

ChatGPT (GPT-4)

```
custom_lm <- function(formula, data) {
  # Step 1: Parse formula
  dependent_var <- all.vars(formula)[1]
  independent_vars <- all.vars(formula)[-1]
  if (independent_vars == ".") {
    independent_vars <- setdiff(names(data), dependent_var)
  }

  # Step 2: Construct design matrix X and response vector y
  X <- as.matrix(cbind(1, data[independent_vars])) # Add intercept
  y <- as.matrix(data[[dependent_var]])

  # Step 3: Compute coefficients
  beta <- solve(t(X) %*% X) %*% t(X) %*% y

  # Calculate residuals
  residuals <- y - X %*% beta

  # Calculate R-squared
  rss <- sum(residuals^2)
  tss <- sum((y - mean(y))^2)
  r_squared <- 1 - rss/tss

  # Step 4: Compute and p-values
  sigma_squared <- rss / (nrow(X) - ncol(X))
  var_beta <- sigma_squared * solve(t(X) %*% X)
  se_beta <- sqrt(diag(var_beta))
  t_values <- beta / se_beta
  p_values <- 2 * pt(abs(t_values), df=nrow(X) - ncol(X),
    lower.tail = FALSE)

  # Output
  return(list(coefficients = beta, p_values = p_values,
    r_squared = r_squared))
}
```

```
}  
# Example usage  
# result <- custom_lm(y ~ ., data)
```

The model generates a well-structured function that fulfills the specified criteria. In this case, however, the task was relatively easy because R is open-source software, and its function performing a linear regression is available on the Internet. Therefore, it was very likely used to train the model. The results may be less impressive when coding functions that do not yet exist.

In any case, it is recommended to cross-check the correctness of the code (e.g. using simulated data, manual checks, comparing it to existing functions in other languages, and showing it to experienced researchers).

3.3.2 Explaining

Prompt 14 shows how ChatGPT (GPT-4) can explain what the linear regression function from Prompt 13 is doing. We ask the model to keep its answer short; otherwise, it will explain in detail what every line is doing.

Prompt 14: Explain in 4-5 sentences what the following R function is doing:

`custom_lm()` (Function from Prompt 13)

ChatGPT (GPT-4)

The `custom_lm` function in R is designed to manually perform linear regression analysis, closely mimicking the functionality of builtin R functions like `lm`. Initially, it interprets a formula to distinguish between the dependent variable and the independent variables, adjusting for scenarios where all other dataset variables are to be used as predictors. The function then constructs a design matrix with an intercept term and a response vector from the provided data. It computes the regression coefficients using the normal equation, a method involving matrix operations. Following this, it calculates residuals, the R^2 statistic to gauge model fit, and performs further statistical calculations to determine the standard errors, *t* values, and *p* values for the regression coefficients. These computations allow for the evaluation of the significance of each predictor in the model. The function ultimately aggregates these results and returns them, offering insights into the model's parameters and its explanatory power regarding the dependent variable.

The model explains the function. If there are specific questions, it is always possible to ask follow-up questions like “Can you explain how *p*-values are calculated?”.

3.3.3 Translating

LLMs can translate code from one language to another. This is particularly useful when a function is unavailable in the programming language the researcher is most familiar with. Prompt 15 shows how ChatGPT (GPT-4) can translate the linear regression function generated in Prompt 13 from R to Matlab. T

Prompt 15: Please translate the following function to matlab:
custom_lm() (Function from Prompt 13)

ChatGPT (GPT-4)

```
function result = custom_lm(formula, data)
    % Step 1: Parse formula
    formulaParts = strsplit(formula, '~');
    dependent_var = strtrim(formulaParts{1});
    independent_vars_str = strtrim(formulaParts{2});
    if strcmp(independent_vars_str, '.')
        independent_vars = setdiff(fieldnames(data), ...
            dependent_var);
    else
        independent_vars = strsplit(independent_vars_str, '+');
    end

    % Step 2: Construct design matrix X and response vector y
    X = ones(size(data.(independent_vars{1}), 1),
        length(independent_vars) + 1);
    for i = 1:length(independent_vars)
        X(:, i+1) = data.(strtrim(independent_vars{i}));
    end
    y = data.(dependent_var);

    % Step 3: Compute coefficients
    beta = (X'*X)\(X'*y);
    % Calculate residuals
    residuals = y - X * beta;
    % Calculate R-squared
    rss = sum(residuals.^2);
    tss = sum((y - mean(y)).^2);
    r_squared = 1 - rss / tss;

    % Step 4: Compute p-values for coefficients
    n = size(X, 1);
    p = size(X, 2);
    sigma_squared = rss / (n - p);
    var_beta = sigma_squared * inv(X' * X);
```

```

    se_beta = sqrt(diag(var_beta));
    t_values = beta ./ se_beta;
    p_values = 2 * (1 - tcdf(abs(t_values), n - p));
    % Output
    result.coefficients = beta;
    result.p_values = p_values;
    result.r_squared = r_squared;
end

```

The model translates all relevant parts of the function to Matlab. A caveat of translating code to a proprietary programming language (e.g., Matlab or Stata) is that certain functions might be available only in paid packages and, therefore, not immediately available to the researcher. In addition, ChatGPT likely translates the code line by line. This may result in inefficient code if the target language has its own data structures and functions that perform a specific task more efficiently than in the source language. Of course, it is also essential to check the translated code for correctness (see previous section).

3.3.4 Debugging

Even experienced researchers encounter many error messages when writing code. LLMs can help identify the causes of such errors. We intentionally provided a data format to the linear regression function discussed earlier that generates an error (converting one of the independent variables into a factor/categorical variable). The function is not able to handle this scenario because it does not automatically convert factor variables into dummy variables. When asked what the error is (see Prompt 16), ChatGPT (GPT-4) proposes various explanations and solutions. One of the suggestions – ensure numeric data types – is indeed correct, as the error was due to the data not being numeric.

Prompt 16: I get the following error when executing the function `custom_lm`, can you help debugging:

Error in h(simpleError(msg, call)) : error in evaluating the argument 'a' in selecting a method for function 'solve': requires numeric/complex matrix/vector arguments
`custom_lm()` (Function from Prompt 13) ChatGPT (GPT-4)

The error you're encountering, "requires numeric/complex matrix/vector arguments," suggests that the issue arises from the `solve` function call. This problem typically occurs when the argument passed to `solve` is not a numeric or complex matrix or vector, which is required for the operation. In the context of your function, this could be happening at two points where `solve` is used: when calculating the coefficients (`beta`) and when calculating the variance of `beta` (`var_beta`).

To debug this error, consider the following steps:

1. **Ensure Numeric Data:** Make sure that the data within the `data` argument, particularly for the independent and dependent variables, are of numeric type. If any of these are not numeric (for example, they could be factors, characters, or logical), the matrix operations will fail.
2. **Check for Singular Matrix:** The error might occur if the matrix $t(X) \%*\% X$ is singular or nearly singular, meaning it does not have an inverse. This situation can arise if there is multicollinearity in your data (i.e., some of the independent variables are linearly dependent on others) or if there are too few observations relative to the number of variables. To check for this, you could add a condition to verify if $\det(t(X) \%*\% X)$ is close to zero (which indicates a singular matrix) before attempting to solve it.

Prompt 17 shows how we can provide an erroneous code to ChatGPT, and we will get a corrected version. The erroneous parts are highlighted in red. We omitted the explanations for the sake of brevity.

Prompt 17: Can you help debug the following code (programming language R):

```
library(tsbox)
CPI = load_series('CPIAUSL')
INFyoy = ts_pcy(INF)
Model : ARMA(INFyoy, p = 1, q = 0)
Forecast = predict(Model)
graph(Forecast)
?
```

ChatGPT (GPT-4)

```
library(tsbox)
# Load the 'forecast' library for ARIMA modeling and forecasting
library(forecast)

# Assuming CPIAUSL is a time series in a CSV file
CPI_data <- read.csv("path_to_CPIAUSL.csv")
# Adjust frequency as needed
CPI <- ts(CPI_data$column_name, frequency = 12)

# Calculating year-over-year percentage change
INFyoy <- ts_pcy(CPI)

# Fitting an ARMA model, here using ARIMA with p = 1, d = 0
# (as ARMA is a special case of ARIMA), q = 0
model <- arima(INFyoy, order = c(1, 0, 0))

# Forecasting
forecast <- forecast(model)

# Plotting the forecast
plot(forecast)
```

The model realizes that various functions do not exist and that the code contains logical errors. It suggests downloading the data manually and importing it via a CSV file (see further examples for import via the internet below). Although the code will not automatically produce the correct result (we have to download the correct data first), it is a first step in the right direction. In addition, the explanations we omitted help us understand the mistakes made in the code.

3.3.5 Commenting

Comments help other people, and after a while, the original author understand the code. Prompt 18 shows how ChatGPT (GPT-4) adds comments to the linear regression function from Prompt 13. The model adds sensible comments that explain the purpose of each line in more detail than before.

Prompt 18: Please add comments to the following R function:

custom_lm() (Function from Prompt 13 without comments)

ChatGPT (GPT-4)

```
# Define a custom linear model function
custom_lm <- function(formula, data) {
  # Extract the dependent variable from the formula
  dependent_var <- all.vars(formula)[1]
  # Extract the independent variables from the formula
  independent_vars <- all.vars(formula)[-1]
  # If independent variables are specified as '.',
  # use all other variables in data as independent variables
  if (independent_vars == ".") {
    independent_vars <- setdiff(names(data), dependent_var)
  }
  # Construct matrix X with intercept and independent variables
  X <- as.matrix(cbind(1, data[independent_vars]))
  # Construct the response vector y from the dependent variable
  y <- as.matrix(data[[dependent_var]])

  # Compute the coefficients (beta) using the Normal Equation
  beta <- solve(t(X) %*% X) %*% t(X) %*% y
  # Residuals = the difference between observed and predicted values
  residuals <- y - X %*% beta
  # Residual sum of squares (rss) and total sum of squares (tss)
  rss <- sum(residuals^2)
  tss <- sum((y - mean(y))^2)

  # Calculate R-squared
  r_squared <- 1 - rss/tss
  # Calculate the estimated variance of the error term (sigma squared)
  sigma_squared <- rss / (nrow(X) - ncol(X))
  # Calculate the variance of the beta coefficients
```

```

var_beta <- sigma_squared * solve(t(X) %*% X)
# Calculate the standard errors of the beta coefficients
se_beta <- sqrt(diag(var_beta))
# Calculate the t-values for the beta coefficients
t_values <- beta / se_beta
# Calculate the p-values for the beta coefficients
p_values <- 2 * pt(abs(t_values), df=nrow(X) - ncol(X),
lower.tail = FALSE)

# Return a list containing the output
return(list(coefficients = beta, p_values = p_values,
r_squared = r_squared))
}

```

3.4 Data analysis

LLMs can assist in various aspects of data analysis, including data cleaning, visualization, and statistical analysis. They can also help generate data, extract data from text, and reformat data. The following sections will demonstrate how LLMs can support these tasks.

However, uploading proprietary, confidential, or personal data to an LLM for analysis is usually prohibited. This will likely violate legal provisions. It is better to ask the LLM how to analyze data, for example, based on simulated data with a similar structure, and then use its tips and codes on actual data with a statistical program on a local computer.

3.4.1 Downloading data

LLMs often know where to find relevant data and can help download them. Prompt 19 shows how ChatGPT (GPT-4) generates R code to download data for Apple stock prices and US real GDP. The advantage of having an R script to download the data instead of manually downloading Excel files is that the data can be updated automatically. A disadvantage is that if the internet data source is deleted, your analysis will not work anymore.

Prompt 19: Provide a R script to download data for the Apple stock prices as well as US real GDP. ChatGPT (GPT-4)

```

# Load the libraries
library(quantmod)
library(tidyquant)
library(fredr)

```

```

# Set FRED API key (get your free API key from the FRED website)
fredr_set_key("YOUR_API_KEY_HERE")

# Function to get stock prices using quantmod
get_stock_prices <- function(symbol, start_date, end_date) {
  getSymbols(symbol, src = "yahoo", from = start_date,
             to = end_date, auto.assign = TRUE)
  return(get(symbol))
}

# Function to get GDP data using fredr
get_gdp_data <- function(series_id) {
  fredr(series_id)
}

# Define date range for stock data
start_date <- as.Date("2020-01-01")
end_date <- as.Date("2024-01-01")

# Download Apple stock prices
apple_stock_prices <- get_stock_prices("AAPL", start_date, end_date)

# Download US Real GDP data (Series ID for Real GDP is "GDPC1")
us_real_gdp <- get_gdp_data("GDPC1")

```

The model provides code that downloads the required data using common R packages.⁸ Be aware, however, that LLMs tend to provide an answer even if the accuracy is doubtful. Therefore, they sometimes invent data sources that do not exist or replace them with simulated data. Checking whether the series the LLM recommends is the correct data is, therefore, key.

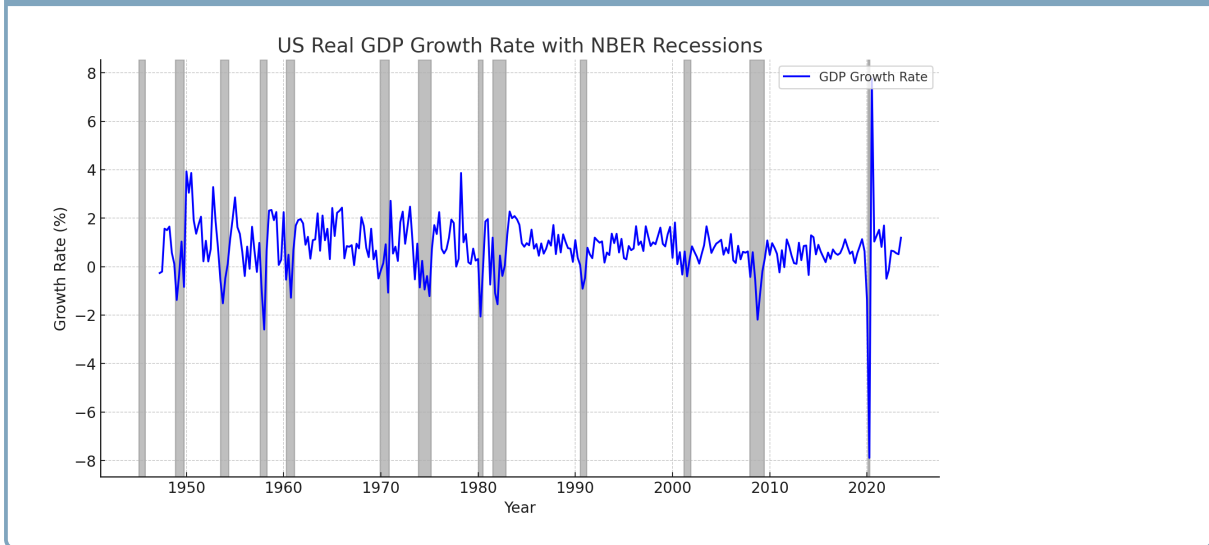
3.4.2 Creating figures

ChatGPT (GPT-4) is the only model that can execute code (in Python). It is the only one that can directly generate figures. However, the other models help generate the code, which also creates figures in other programming languages. Prompt 20 shows how ChatGPT (GPT-4) generates a graph with the US real GDP growth rate and adds NBER recessions as gray-shaded areas.

⁸Note that you need to register on the website of the St. Louis FRED to get an API key to be able to use the fredr package.

Prompt 20: The attached excel file contains real GDP for the US. Please create a nice graph with the growth rate and add NBER recessions as gray-shaded areas.
Attached file: US_Real_GDP.xlsx

ChatGPT (GPT-4)



Because the environment where the code is executed is not connected to the internet, the data has to be attached with an Excel file downloaded from the FRED website. Importantly, as for any data analysis, do not upload proprietary, confidential, or personal data to an LLM to generate graphs.

3.4.3 Extracting data from text

As LLMs are models trained on text, they are also useful for textual analysis. Prompt 21 shows how ChatGPT (GPT-4) extracts countries from a text containing the results of soccer world cups since 1998 and creates a \LaTeX table with the countries in the first four places.

Prompt 21: The following text contains place 1 - 4 from soccer world cups since 1998, please extract countries on these places and create a latex table: 16 1998 France France 3–0 Stade de France, Saint-Denis Brazil Croatia 2–1 Parc des Princes, Paris Netherlands 32 17 2002 South Korea Japan Brazil 2–0 International Stadium, Yokohama Germany Turkey 3–2 Daegu Stadium, Daegu South Korea 32 18 2006 Germany Italy 1–1 (a.e.t.) (5–3 p) Olympiastadion, Berlin France Germany 3–1 Gottlieb-Daimler-Stadion, Stuttgart Portugal 32 19 2010 South Africa Spain 1–0 (a.e.t.) Soccer City, Johannesburg Netherlands Germany 3–2 Nelson Mandela Bay Stadium, Port Elizabeth Uruguay 32 20 2014 Brazil Germany 1–0 (a.e.t.) Maracanã, Rio de Janeiro Argentina Netherlands 3–0 Estádio Nacional, Brasília Brazil 32 21 2018 Russia France 4–2 Luzhniki Stadium, Moscow Croatia Belgium 2–0 Krestovsky Stadium, Saint Petersburg England 32 22 2022 Qatar Argentina 3–3 (a.e.t.) (4–2 p) Lusail Stadium, Lusail France Croatia 2–1 Khalifa International Stadium, Al Rayyan Morocco 32

ChatGPT (GPT-4)

```

\begin{tabular}{|l|l|l|l|l|}
\hline
\textbf{Year} & \textbf{1st Place} & \textbf{2nd Place} & \textbf{3rd Place} & \textbf{4th Place} \\ \hline
1998 & France & Brazil & Croatia & Netherlands \\ \hline
2002 & Brazil & Germany & Turkey & South Korea \\ \hline
2006 & Italy & France & Germany & Portugal \\ \hline
2010 & Spain & Netherlands & Germany & Uruguay \\ \hline
2014 & Germany & Argentina & Netherlands & Brazil \\ \hline
2018 & France & Croatia & Belgium & England \\ \hline
2022 & Argentina & France & Croatia & Morocco \\ \hline
\end{tabular}

```

Compiled table

Year	1st Place	2nd Place	3rd Place	4th Place
1998	France	Brazil	Croatia	Netherlands
2002	Brazil	Germany	Turkey	South Korea
2006	Italy	France	Germany	Portugal
2010	Spain	Netherlands	Germany	Uruguay
2014	Germany	Argentina	Netherlands	Brazil
2018	France	Croatia	Belgium	England
2022	Argentina	France	Croatia	Morocco

The structured data contains the correct information, even though the table is copy-pasted from Wikipedia and hardly readable. An important caveat is that, for large data sets, it will almost be impossible to verify whether the LLM generated the correct data. The approach can be risky, and the results depend on the quality of the LLM. Therefore, we currently recommend using such analysis only when the output can be manually checked.

3.4.4 Reformatting data

Prompt 22 shows how ChatGPT (GPT-4) converts the \LaTeX table from Prompt 21 to a tibble in R (we could do the same for a data frame, of course).

Prompt 22: Create a tibble in R for the following latex table:

Year	1st Place	2nd Place	3rd Place	4th Place
1998	France	Brazil	Croatia	Netherlands
2002	Brazil	Germany	Turkey	South Korea
2006	Italy	France	Germany	Portugal
2010	Spain	Netherlands	Germany	Uruguay
2014	Germany	Argentina	Netherlands	Brazil
2018	France	Croatia	Belgium	England
2022	Argentina	France	Croatia	Morocco

ChatGPT (GPT-4)

```
world_cup_standings <- tibble(  
  Year = c(1998, 2002, 2006, 2010, 2014, 2018, 2022),  
  '1st Place' = c("France", "Brazil", "Italy",  
    "Spain", "Germany", "France", "Argentina"),  
  '2nd Place' = c("Brazil", "Germany", "France",  
    "Netherlands", "Argentina", "Croatia", "France"),  
  '3rd Place' = c("Croatia", "Turkey", "Germany",  
    "Germany", "Netherlands", "Belgium", "Croatia"),  
  '4th Place' = c("Netherlands", "SouthKorea", "Portugal",  
    "Uruguay", "Brazil", "England", "Morocco")  
)
```

Again, the result is correct, and the code-snipped can be used directly in R for further analysis. But the same caveat applies as before. It will be difficult to verify the correctness of larger data sets.

3.4.5 Classifying text

Textual classification is a technique that is increasingly used in economics. For example, it can be applied to construct business cycle indicators from newspaper texts. Prompt 23 shows how ChatGPT (GPT-4) extracts economic concepts discussed in a text.

Prompt 23: The following text contains a description of the US economy in 1894 and 1895. Extract economic variables described. Do not provide an explanation:

1894: "Deep depression. Severe stagnation; widespread unemployment; bituminous coal strike, second quarter; railroad strike, June; many failures; further decline in commodity prices; reduction in foreign trade. Money very easy; declining railroad stock prices, rising industrials; higher bond prices. Average wheat crop with low price; corn failure, high price; large cotton crop and very low price. Floods in Middle West and Pacific states; Coxey's armies of tramps, spring; lower tariff into operation, August."
1895: "Depression; revival. Dullness yields to gradual improvement, late spring; boom in iron and steel industry; continued numerous failures; commodity prices . rise; revival of imports. Money market easy to December; temporary boom advances industrial stocks to peak, June, and railroads, September; further advance in bond prices. Excellent corn, good wheat, and short cotton crops; prices low, except cotton. Diplomatic difficulties with Great Britain over Venezuela, December."
ChatGPT (GPT-4)

Economic Variables in 1894:

1. Economic depression
2. Severe stagnation
3. Widespread unemployment
4. Bituminous coal strike
5. Railroad strike
6. Business failures
7. Decline in commodity prices
8. Reduction in foreign trade
9. Money supply (very easy)
10. Declining railroad stock prices
11. Rising industrial stock prices
12. Higher bond prices
13. Wheat crop yield and price
14. Corn crop failure and price
15. Cotton crop yield and price
16. Natural disasters (floods)
17. Lower tariff implementation

Economic Variables in 1895:

1. Economic depression and revival
2. Improvement in economic conditions
3. Boom in iron and steel industry
4. Business failures
5. Rise in commodity prices
6. Revival of imports
7. Money market conditions (easy to December)
8. Industrial and railroad stock prices
9. Bond prices
10. Corn crop yield and price
11. Wheat crop yield and price
12. Cotton crop yield and price
13. Diplomatic tensions with Great Britain over Venezuela

The concepts seem sensible. Doing this manually would take much longer, especially for more extended periods. Again, the quality of the output is hard to assess without manually checking the results. In addition, the results were slightly different when we used other LLMs. But, of course, this may also be the case for a human reading these texts. One major issue is that the results are not reproducible and may violate scientific integrity rules if the analysis constitutes an integral part of the thesis (see Chapter 4).

3.4.6 Extracting sentiment

Textual analysis can also be used to extract a sentiment from text. Prompt 24 shows how ChatGPT (GPT-4) assigns a sentiment score (an assessment of the positivity/negativity of the text) on a scale from 1 to 5 based on the results from Prompt 23.

Prompt 24: Continued from Prompt 23. Please rate each of these extracted concepts from 1 (worst) to 5 (best) and calculate an average for each year. ChatGPT (GPT-4)

1894 Economic Variables:

1. Economic depression: 1 2. Severe stagnation: 1 3. Widespread unemployment: 1
4. Bituminous coal strike: 2 5. Railroad strike: 2 6. Business failures: 1 7. Decline in commodity prices: 2 8. Reduction in foreign trade: 2 9. Money supply (very easy): 3 10. Declining railroad stock prices: 2 11. Rising industrial stock prices: 4 12. Higher bond prices: 4 13. Wheat crop yield and price: 3 14. Corn crop failure and price: 2 15. Cotton crop yield and price: 3 16. Natural disasters (floods): 1 17. Lower tariff implementation: 3

1895 Economic Variables:

1. Economic depression and revival: 3 2. Improvement in economic conditions: 4 3. Boom in iron and steel industry: 5 4. Business failures: 2 5. Rise in commodity prices: 4 6. Revival of imports: 4 7. Money market conditions (easy to December): 3 8. Industrial and railroad stock prices: 4 9. Bond prices: 4 10. Corn crop yield and price: 4 11. Wheat crop yield and price: 3 12. Cotton crop yield and price: 3 13. Diplomatic tensions with Great Britain over Venezuela: 2

The average ratings for the economic variables of each year are as follows:
For the year 1894, the average rating is approximately 2.18. For the year 1895, the average rating is approximately 3.46. These averages suggest a more positive economic outlook in 1895 compared to 1894.

We calculated the average sentiment score for each year based on the ratings provided for each economic variable. As one would expect from reading the texts, the sentiment scores for 1894 are lower than those for 1895, indicating a worse economic situation in 1894. An important caveat is that it is difficult to verify whether the results make sense and are accurate for larger data sets, and the results are not reproducible.

3.4.7 Simulating human subjects

LLMs learn from vast datasets of human-generated text and inevitably include opinions, biases, and stereotypes. LLMs can, therefore, be used to simulate the opinion of human subjects or predict human behavior. Prompt 25, adapted from Korinek (2023), shows how ChatGPT (GPT-4) predicts voter groups' approval of a policy proposal. Such an approach is sometimes used to pre-test online surveys. But one could also consider using such an approach to predict human behavior in marketing.

Prompt 25: For each of the following voters, provide a one-word rating to indicate whether they are more likely to approve or disapprove of a policy proposal to loosen Swiss immigration restrictions (do not provide any explanation):

1. A 76 year old male retired road worker living in Appenzell, who is Caucasian.
2. A 68 year old male retired high school teacher living in Lausanne, who is Caucasian.
3. A 50 year old female artist from Bern, who is Caucasian.
4. A 29 year old female manager living in Geneva, who is Asian.

ChatGPT (GPT-4)

1. Disapprove
2. Approve
3. Approve
4. Approve

The results may or may not make sense based on the biases and stereotypes of the reader. Several issues arise when using LLMs to predict human behavior. There is a risk that they might not only reflect but also amplify existing stereotypes because a supposedly external and independent source confirms our prior views. This can reinforce harmful prejudices and misconceptions about certain groups of people. In addition, AI is trained on a non-representative body of data. This suggests that, depending on the application, using AI-generated predictions about human behavior can be ethically and methodologically problematic or simply wrong.

3.5 Math

LLMs can do math, from elementary arithmetic to more advanced topics in calculus, algebra, and differential equations. They will not always get it right, however. Therefore, every equation has to be verified. In addition, note that struggling to solve a mathematical problem is part of the learning process. The goal is not to learn the solution by heart but to understand how to solve it. We are unlikely to forget how to solve a problem once we solve it ourselves. Therefore, it is generally recommended to use AI assistants to provide tips along the way or to check and give feedback on a solution. Of course, this applies to the other academic tasks equally well.

3.5.1 Setting up models

The following example (Prompt 26, adapted from Korinek (2023)), shows how ChatGPT (GPT-4) can be used to set up the optimization problem of a consumer with Cobb-Douglas preferences over two goods. The model explains the optimization problem, including the utility function, budget constraint, and the Lagrangian for optimization.

Prompt 26: Set up the optimization problem of a consumer with Cobb-Douglas preferences over two goods in LaTeX code. ChatGPT (GPT-4)

To set up an optimization problem for a consumer with Cobb-Douglas preferences over two goods, we define the utility function, budget constraint, and the Lagrangian for optimization. Let's denote the two goods as X and Y , their respective prices as P_X and P_Y , and the consumer's income as I . The Cobb-Douglas utility function is typically written as $U(X, Y) = X^\alpha Y^{1-\alpha}$, where $0 < \alpha < 1$ is a parameter representing the preference weight.

The consumer's problem is to maximize utility subject to the budget constraint. The budget constraint is that the total amount spent on goods X and Y cannot exceed the consumer's income, i.e., $P_X X + P_Y Y \leq I$.

The optimization problem can be set up as follows in LaTeX code:

$$\begin{aligned} \max_{X, Y} U(X, Y) &= X^\alpha Y^{1-\alpha} \\ \text{s.t. } P_X X + P_Y Y &= I \end{aligned} \quad (3.1)$$

Here, $\max_{X, Y}$ indicates that we are maximizing with respect to X and Y , $U(X, Y) = X^\alpha Y^{1-\alpha}$ is the Cobb-Douglas utility function, and $P_X X + P_Y Y = I$ is the budget constraint.

Note: For ease of readability, we compile the latex output of the model.

3.5.2 Deriving equations

In a follow-up to Prompt 26, we ask ChatGPT (GPT-4) to solve the maximization problem.

Prompt 27: Continued from Prompt 26. Solve the maximization problem.

ChatGPT (GPT-4)

To solve the optimization problem for a consumer with Cobb-Douglas preferences, we will use the method of Lagrange multipliers. The problem is:

Maximize $U(X, Y) = X^\alpha Y^{1-\alpha}$

Subject to $P_X X + P_Y Y = I$

Where: - X, Y are the quantities of the two goods. - P_X, P_Y are the prices of the goods. - I is the income. - α is a parameter ($0 < \alpha < 1$).

We set up the Lagrangian, \mathcal{L} , as follows:

$$\mathcal{L} = X^\alpha Y^{1-\alpha} + \lambda(I - P_X X - P_Y Y)$$

Where λ is the Lagrange multiplier.

To find the maximum, we take the first-order partial derivatives of \mathcal{L} with respect to X , Y , and λ , and set them equal to zero:

$$\frac{\partial \mathcal{L}}{\partial X} = \alpha X^{\alpha-1} Y^{1-\alpha} - \lambda P_X = 0$$

$$\frac{\partial \mathcal{L}}{\partial Y} = (1 - \alpha)X^\alpha Y^{-\alpha} - \lambda P_Y = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = I - P_X X - P_Y Y = 0$$

Now, we solve these equations simultaneously. From the first two equations, we can derive the ratio of X to Y that the consumer will choose:

$$\frac{\alpha/P_X}{1 - \alpha/P_Y} = \frac{X}{Y}$$

$$X = \frac{\alpha Y}{P_X} \left(\frac{P_Y}{1 - \alpha} \right)$$

Substituting this into the budget constraint:

$$P_X \left(\frac{\alpha Y}{P_X} \left(\frac{P_Y}{1 - \alpha} \right) \right) + P_Y Y = I$$

$$\alpha Y \left(\frac{P_Y}{1 - \alpha} \right) + P_Y Y = I$$

$$Y \left(\alpha \frac{P_Y}{1 - \alpha} + P_Y \right) = I$$

$$Y = \frac{I}{\left(\alpha \frac{P_Y}{1 - \alpha} + P_Y \right)}$$

And solving for X :

$$X = \frac{\alpha I}{P_X(1 - \alpha)}$$

These values of X and Y maximize the consumer's utility given the budget constraint and the prices of the goods.

The model uses the Lagrange multipliers method to solve the optimization problem. It correctly takes the first-order partial derivatives and solves for the correct X and Y . Note that the solution for Y could be further simplified by factoring out P_Y from the denominator. Of course, this is a textbook example, and the solution is well-known. More complex problems might not be solved as easily.

3.5.3 Explaining equations

Prompt 28 asks ChatGPT (GPT-4) to explain the meaning of the optimization function associated with elastic net regularization, a technique frequently used in statistics.

Prompt 28: Please explain what the following equation is representing:

$$\text{Minimize: } \frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$$

ChatGPT (GPT-4)

The equation you've provided represents an optimization problem commonly used in machine learning and statistics, particularly in the context of regularized linear regression. Let's break it down:

1. Objective to Minimize

- The goal is to minimize the function that is specified. In this case, the function consists of two main parts.

2. The First Part - Mean Squared Error: $\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2$

- This part represents the mean squared error (MSE) between predicted values and actual values.

- y_i is the actual value for the i -th observation.

- x_i is a vector of the features (or independent variables) for the i -th observation.

- β is a vector of coefficients (parameters) to be estimated.

- $x_i^T \beta$ represents the predicted value for the i -th observation (the dot product of the feature vector and the coefficients).

- The MSE is averaged over all n observations.

3. The Second Part - Regularization Term: $\lambda \sum_{j=1}^p \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$

- This part is the regularization term, which is used to prevent overfitting by penalizing the magnitude of the coefficients.

- λ is a regularization parameter that controls the strength of the regularization.

- α is a parameter that balances the type of regularization: L1 (Lasso) and L2 (Ridge).

- $|\beta_j|$ is the L1 norm (Lasso regularization), which encourages sparsity in the coefficients.

- β_j^2 is the L2 norm (Ridge regularization), which encourages smaller, more diffuse coefficients.

- The sum is taken over all p features.

Overall, this equation is a formulation of Elastic Net regularization, which combines both L1 and L2 regularization. The parameters α and λ allow for flexible control over the regularization process, balancing between Ridge and Lasso regularization. This is particularly useful in scenarios where some features are correlated or when you want to include feature selection as part of the regression process.

The model correctly recognizes and explains the equation. This includes an analysis of the various terms of the function, such as the mean squared error and the regularization term. An LLM will perform better for equations well described on internet resources. For other equations, the explanations may be more vague or incorrect.

3.6 Writing

Many tools, such as Grammarly, Writefull, and Hemingway, help improve written text's spelling, grammar, and style. In what follows, we examine how LLMs can be used for these tasks.

An important caveat is that an inexperienced writer cannot judge whether the AI output's quality meets academic standards. Therefore, improving writing skills remains indispensable (see Chapter 2 for references).

3.6.1 Writing text

Prompt 29, adapted from Korinek (2023), shows how ChatGPT (GPT-4) writes a paragraph in the role of an academic economist by transforming bullet points into text.

Prompt 29: Write a paragraph in the style of an academic economist that integrates the following points. Include a topic sentence.

- The macroeconomic consequences of climate change are significant [add 2 examples].
- There are several ways to mitigate these effects [add 2 examples] ChatGPT (GPT-4)

Climate change poses profound macroeconomic challenges, manifesting in both direct and indirect repercussions. For instance, the increased frequency of extreme weather events such as hurricanes and droughts disrupts agricultural productivity and infrastructure stability, leading to substantial economic losses. Similarly, the gradual rise in global temperatures alters ecosystems, impacting industries reliant on natural resources and forcing labor adjustments in sectors like agriculture and fishing. To mitigate these effects, one approach involves investing in sustainable infrastructure, which not only reduces carbon emissions but also enhances resilience against climate-related disasters. Additionally, transitioning towards a green economy, with a focus on renewable energy and eco-friendly practices, can stimulate job creation in new sectors while curbing the environmental degradation responsible for many climate-induced economic challenges. These strategies not only address the immediate impacts of climate change but also pave the way for a sustainable and resilient economic future.

However, the writing style is not very academic. The verbs used are often quite complicated (e.g., “manifesting”, “impacting”, “enhance”), and more common alternatives exist. The text could be easily rewritten to make it shorter, more direct, and easier to understand. This is confirmed by hemingwayapp.com; the tool gives the paragraph a poor readability rating (5 of 6 sentences are too long with too complicated words).

Therefore, generated text should be carefully reviewed and edited to ensure it aligns with the intended style and content of the research paper. This is particularly important as LLMs may

inadvertently introduce errors or inaccuracies in the text. Therefore, the AI assistant's output should be used as a starting point and refined as necessary to ensure it meets academic standards.

In addition, ChatGPT adds two examples that were not provided in the prompt. This may violate the rules of scientific integrity because the sources of these examples are unknown. Therefore, the result should be taken as a starting point, an idea generator, and then the researcher should find literature that is consistent with the examples mentioned.

3.6.2 Editing text

LLMs can detect and correct grammatical or spelling errors and improve style. Prompt 30, adapted from Korinek (2023), shows how ChatGPT (GPT-4) improves the quality of the original text and provides feedback on the errors.

Prompt 30: Correct (and explain the corrections) the following text for mistakes, style, clarity and simplicity:

Climate change cannot be stopped. It has dramatic macoeconomic effects on the macroeconomy. For example a higher number of extreme wheather as e.g. hurricanes makes agriculture less productive, leading to economic losses. Or the rise in temperature has a negative impact on industries that are dependent on natural resources. A possinle solution to counteract these negative effects are pricing of carbon emissions or investments in green energy.

ChatGPT (GPT-4)

Correction: "Climate change is unstoppable. It has significant effects on the economy. For example, increased extreme weather events like hurricanes reduce agricultural productivity, causing economic losses. Similarly, rising temperatures adversely affect industries reliant on natural resources. Potential solutions to mitigate these effects include implementing a carbon emissions pricing system and investing in green energy."

Explanation of Corrections:

1. **Redundancy Removal:** The original phrase "dramatic macoeconomic effects on the macroeconomy" was redundant. It's simplified to "significant effects on the economy."
2. **Spelling Correction:** Corrected "wheather" to "weather" and "possinle" to "potential."
3. **Style and Clarity:** The original text used complex structures and some informal or less precise terms. I simplified these for clarity and formality. For example, "as e.g." was changed to "like" for simplicity.
4. **Sentence Structure:** Improved sentence structure for readability. The original "Or the rise in temperature..." began abruptly and lacked coherence with the previous sentence. It's rephrased to flow logically from the previous point.
5. **Subject-Verb Agreement:** Corrected the grammatical error in "A possible solution... are pricing." Since the subject "solution" is singular, the verb should be "is" instead of "are."

6. Simplification and Conciseness: General simplification was done throughout to make the text more concise and straightforward while retaining the original meaning. The overall aim was to enhance readability, ensure grammatical accuracy, and maintain a formal and clear style.

The text becomes shorter and easier to understand. The explanations help us to understand the rationale behind each change and avoid similar mistakes in the future.

Tools such as Hemingway, Grammarly, and Writefull are also helpful for editing text. Their advantage is that they are already integrated into word processors and that they are often fine-tuned for particular settings. We expect that LLMs like Copilot will soon be integrated for such tasks in traditional word processors as well.

An important caveat is that a ChatBot may change the meaning of the text. In the end, the researcher is responsible for correctness and relevance, as well as for a writing style that meets academic standards. Therefore, AI assistants are no substitute for improving writing skills. It is better to use tools that provide feedback on why a text is complicated. Such tools are discussed in Chapter 2. Of course, data protection and privacy issues may arise with such specialized tools as well because it is challenging to verify what happens with the submitted data.

3.6.3 Generating titles and headlines

Prompt 31 shows how ChatGPT (GPT-4) can be used to generate five short and concise titles for a paragraph. The model provides a list of titles and captures the essence of the paragraph.

Prompt 31: Generate 5 short and concise titles for the following paragraph:

Climate change poses profound macroeconomic challenges, manifesting in both direct and indirect repercussions. For instance, the increased frequency of extreme weather events such as hurricanes and droughts disrupts agricultural productivity and infrastructure stability, leading to substantial economic losses. Similarly, the gradual rise in global temperatures alters ecosystems, impacting industries reliant on natural resources and forcing labor adjustments in sectors like agriculture and fishing. To mitigate these effects, one approach involves investing in sustainable infrastructure, which not only reduces carbon emissions but also enhances resilience against climate-related disasters. Additionally, transitioning towards a green economy, with a focus on renewable energy and eco-friendly practices, can stimulate job creation in new sectors while curbing the environmental degradation responsible for many climate-induced economic challenges. These strategies not only address the immediate impacts of climate change but also pave the way for a sustainable and resilient economic future.

ChatGPT (GPT-4)

1. "Economic Impacts of Climate Change"
2. "Climate Risks to Global Economy"
3. "Green Economy as a Climate Solution"
4. "Building Climate-Resilient Economies"
5. "Climate Change: Economic Strategies and Solutions"

One caveat for using a ChatBot to generate section titles is that a researcher should think about and sometimes struggle with a paper's structure. The structure should support the argument the researcher wants to make. Struggling to set up a structure with good titles often implies that the researcher is not fully sure what argument to make in the first place. If this is the case, the ChatBot will not suddenly provide you with a good structure. Instead, writing is a key step in the research process.

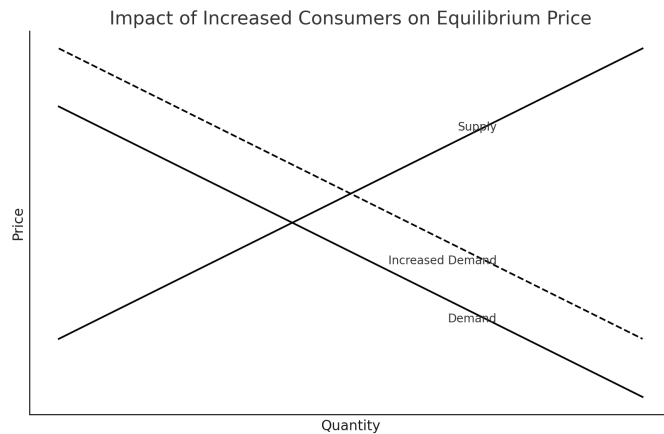
3.6.4 Generating graphs

Currently, only ChatGPT (GPT-4) can run code (in Python) and therefore generate reproducible graphs. Prompt 32 illustrates how ChatGPT (GPT-4) can depict the effects of an increased number of consumers on the equilibrium price within a competitive market. The graph shows how a rise in demand shifts the demand curve rightward, leading to higher prices and quantities in the new equilibrium. To modify the graph (e.g. color or axis titles) follow-up prompts prompt, such as "Please indicate the old and new prices and quantities on the axes." can be used.

Prompt 32: Generate a graph with the impact of an increase of the number consumers on the equilibrium price.

- * Remove any numbers on the axes
- * Use black solid lines for all curves
- * Instead of a legend, label the curves directly with demand and supply

ChatGPT (GPT-4)



An alternative is to use ChatBots to recommend code that generates graphs in another programming language (e.g. Matlab or R) and then run the code on a local computer. This is usually the preferred method because it allows us to modify and reproduce the graph later on.

3.6.5 Generating equations from pictures

Prompt 33 shows how ChatGPT (GPT-4) can be used to generate an equation in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ code from a picture.

Prompt 33: Please generate latex code with the equations on the attached image.

$$\begin{aligned}\frac{P - MC}{P} &= -\frac{dP}{dQ} \frac{Q}{P} \\ &= -1 / \left(\frac{dQ}{dP} \frac{P}{Q} \right) \\ &= -\frac{1}{E_d},\end{aligned}$$

ChatGPT (GPT-4)

The LaTeX code for the equations in the image you provided is as follows:

$$\frac{P - MC}{P} = \frac{dP}{dQ} \frac{Q}{P} \quad (3.2)$$

$$= -\frac{1}{\left(\frac{dQ}{dP} \frac{P}{Q} \right)} \quad (3.3)$$

$$= \frac{1}{E_d}, \quad (3.4)$$

Author's note: For ease of readability we compile the latex output of the model.

The result is almost correct, but the model missed the minus sign in the first and third equations.⁹ The error in the equation highlights that output from AI assistants may be erroneous. Therefore, it should be mainly used to speed up writing equations in MS Word or L^AT_EX but the researcher has to redo the calculations to verify that the formulas are correct. Another tool for this is Mathpix, a web-based tool that correctly recognizes the equation.

3.6.6 Detecting AI-generated text

Some AI assistants promise to detect AI-generated text, such as GPTZero and Compilatio, a software for detecting plagiarism. Figure 3.2 shows a screenshot of these platforms. The text provided on both platforms is the answer from Prompt 29.

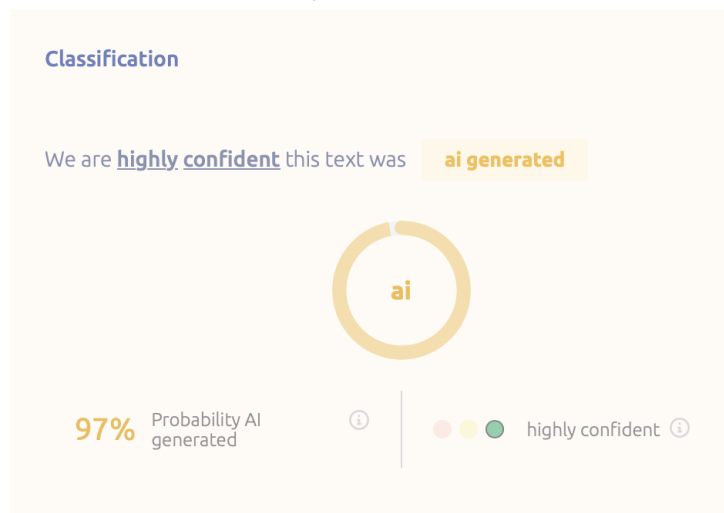
In this example, both platforms correctly identified the text as AI-generated. However, it is widely acknowledged that detecting AI-generated text is not reliable. The problem is that there is still a high rate of false positives (Fowler, 2023).¹⁰ Moreover, texts from non-native speakers are often

⁹It should read $-\frac{dP}{dQ} \frac{Q}{P}$ and $-\frac{1}{E_d}$.

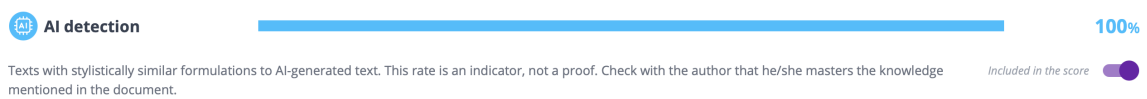
¹⁰We checked this report using a prominent AI detection tool. Interestingly, the answers to the prompts were sometimes, but quite rarely, identified as AI-generated. By contrast, many text passages we have written ourselves were also identified as AI-generated. This confirms that AI-detection tools are still relatively unreliable.

Fig. 3.2: AI detection

a) GPTZero



b) Compilatio



Notes: The text provided on both platforms is the answer from Prompt 29.

falsely identified as AI-generated (Myers, 2023). In addition, it is difficult to distinguish whether AI has been used for correcting grammar and spelling (allowed under some circumstances) or whether the content stems from the AI assistant itself (forbidden under most circumstances, see Chapter 4). Therefore, giving too much weight to the results of these tools is not recommended.

3.7 Presentation

LLMs can help create presentations. This includes generating bullet points or creating presentations in \LaTeX using the beamer package. More and more tools exist that also directly create presentations in other formats (e.g. PowerPoint). In addition, newer versions of MS PowerPoint include a Copilot plugin that provides support directly in the software. We discuss these tools in Chapter 2.

3.7.1 Generating bullet points

Prompt 34 shows how ChatGPT (GPT-4) generates bullet points for a beamer presentation summarizing the model's answer to Prompt 28.

Prompt 34: Please create a beamer presentation summarizing the following:
(Answer to Prompt 28) ChatGPT (GPT-4)

```
\documentclass{beamer}
\usetheme{Madrid}
\usecolortheme{dolphin}
\title{Understanding Regularized Linear Regression}
\subtitle{A Deep Dive into Elastic Net Regularization}
\author{Your Name}
\institute{Your Institution}
\date{\today}

\begin{document}
\begin{frame}
  \titlepage
\end{frame}

\begin{frame}{Objective to Minimize}
  \begin{itemize}
    \item The goal is to minimize a specific function that consists of two main components.
    \item This optimization problem is crucial in machine learning and statistics, particularly in regularized linear regression.
  \end{itemize}
\end{frame}
```

```
\end{frame}
```

```
\begin{frame}{Mean Squared Error}
```

```
\begin{itemize}
```

```
\item  $\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2$ 
```

```
\item Represents the mean squared error (MSE) between predicted and actual values.
```

```
\item  $(y_i)$ : actual value,  $(x_i)$ : feature vector,
```

```
 $(\beta)$ : coefficients.
```

```
\item MSE is averaged over all observations  $(n)$ .
```

```
\end{itemize}
```

```
\end{frame}
```

```
\begin{frame}{Regularization Term}
```

```
\begin{itemize}
```

```
\item  $\lambda \sum_{j=1}^p$ 
```

```
 $\left( \alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right)$ 
```

```
\item Regularization prevents overfitting by penalizing large coefficients.
```

```
\item  $(\lambda)$ : regularization strength,  $(\alpha)$ : balance between L1 (Lasso) and L2 (Ridge).
```

```
\item Encourages sparsity (L1) and smaller, more diffuse coefficients (L2).
```

```
\end{itemize}
```

```
\end{frame}
```

```
\begin{frame}{Elastic Net Regularization}
```

```
\begin{itemize}
```

```
\item Combines L1 (Lasso) and L2 (Ridge) regularization.
```

```
\item Flexible control over regularization process through  $(\alpha)$  and  $(\lambda)$ .
```

```
\item Useful in scenarios with correlated features or when feature selection is part of the regression process.
```

```
\end{itemize}
```

```
\end{frame}
```

```
\begin{frame}{Conclusion}
```

```
\begin{itemize}
```

```
\item The discussed equation is a cornerstone of Elastic Net regularization in regularized linear regression.
```

```
\item Proper selection of  $(\alpha)$  and  $(\lambda)$  allows for a balanced regularization process, aiding in feature selection and preventing overfitting.
```

```
\item Understanding and applying these concepts is crucial for effective model development in machine learning and statistics.
```

```
\end{itemize}  
\end{frame}  
  
\end{document}
```

The answer comprises \LaTeX code for a beamer presentation that includes slides on relevant points of the elastic net regularization. With ChatGPT (GPT-4), it is also possible to attach a paper in PDF format and ask the model to create a first draft of a presentation. Note, however, that depending on the context, this may violate the rules of scientific integrity (see Chapter 4). For example, if an assignment is to read and present a paper, it would not be allowed to let an AI assistant create the presentation based on the uploaded paper.

4 Guidelines

In what follows, we first discuss existing rules and guidelines at Swiss universities and then recommend some specific guidelines for students and instructors in a master's program in economics. We keep these guidelines as general as possible to make them useful in the coming years despite the rapid progress of AI assistants.

4.1 Existing rules and guidelines

4.1.1 Rules on scientific integrity at the University of Neuchâtel

We believe that several issues regarding the use of AI assistants are covered by existing rules on scientific integrity ([Université de Neuchâtel, 2014](#)).

Article 3 al. 1 suggests that *behavior consistent with scientific integrity implies compliance not only with legal norms but also with the duties of care and standards or rules generally recognized within the scientific and academic community (rules of good scientific practice), as well as with the requirements specific to each field of science.*¹

Then, it lists examples that may violate the rules of scientific integrity. For example, it may be violated if a member of the University of Neuchâtel

- *passes off the work or ideas of others as one's own ("plagiarism").*² Article 4 al. 3 a)
- *omits essential contributions by other authors on the same subject (omission of quotations).*³ Article 4 al. 3 d)
- *erroneously or incompletely cites existing or supposed works by third parties.*⁴ Article 4 al. 3 e)

¹Text translated by DeepL, 16 April 2024, DeepL, www.deepl.com

²Text translated by DeepL, 16 April 2024, DeepL, www.deepl.com

³Text translated by DeepL, 16 April 2024, DeepL, www.deepl.com

⁴Text translated by DeepL, 16 April 2024, DeepL, www.deepl.com

This suggests that the scientific community's existing work must be cited. Therefore, directly using AI-generated output, which may be based on the scientific work published on the internet, may violate the rules on scientific integrity. Although different universities do not agree on whether using AI output directly is plagiarism in the formal sense (as AI is not a person and, therefore, not an author), it is clear that AI has been trained on existing work of the scientific community. Therefore, not citing these original sources violates scientific integrity. When using scientific ideas, models, and results reported by an AI assistant, it is necessary to cite the original source rather than the AI assistant. Importantly, this does not apply to use cases such as correcting grammar or typos. Clearly, citing every document used to train an AI assistant to help with these tasks is impossible.

The rules of scientific integrity also require the researcher to verify the correctness and relevance of the information. It is well known that AI assistants make errors because they are based on probabilistic models. These models predict text based on data available to the model and its parameters. Of course, predictions can be wrong. Therefore, such models sometimes “hallucinate”, that is, make statements that are actually nowhere written in the existing data. The output of AI assistants is, therefore, generally not regarded as a scientific source. Instead, the AI output has to be verified for correctness and relevance, and the original source has to be cited.

4.1.2 Existing guidelines on AI

Table 4.1 provides a summary of existing guidelines at various Swiss universities.⁵ Currently, AI is not prohibited at any Swiss university.⁶ Often, the instructor or the university retains the right to prohibit the use of AI for certain courses or certain activities. Instructors are often asked to clarify whether AI can be used and for what purposes. If AI is used, it usually has to be declared (where and how). Swiss universities seem relatively open to and actively discuss the use of AI.⁷ They also emphasize that it is important that students learn how to use AI in scientific research appropriately.

⁵Similar principles and guidelines have been developed at foreign universities (see e.g. [Garber et al., 2023](#), [UC Berkeley, 2023](#)). See also guidelines by the [SNSF \(2024\)](#). However, we focus on Swiss universities for the sake of brevity, as the main points are similar.

⁶This is confirmed by another review of the guidelines by [Solis \(2023\)](#).

⁷This is based on a non-representative review of various well-known universities in the US and our assessment of the guidelines published by these institutions, as well as [Solis \(2023\)](#).

Tab. 4.1: Guidelines and rules at Swiss universities

Source	Summary
Université de Neuchâtel (2024)	<p>22/04/2024: A ban is regarded as problematic because AI is difficult to detect; because a ban would be to the detriment of students who need to learn how to deal with AI for their careers; because AI will become indispensable tools. Instructors should integrate AI assistants and discuss advantages and limitations. Presenting the output of an AI assistant as one's own work is against the rules of scientific integrity. Instructors can adjust the evaluation mode to consider AI (e.g. different mediums, oral evaluation, in-class presentations or assignments, asking open questions that require critical thinking). Also, they can integrate AI in the assignments (e.g. brainstorming, AI for feedback on grammar mistakes, evaluating the research process rather than the final output with intermediate evaluations)</p>
Université de Fribourg (2024)	<p>22/04/2024: Offers courses for appropriate use of AI. No tool is available to reliably detect output generated by AI. They rely on instructors to detect inappropriate use of AI, which requires adapting evaluation modes. They also refer to the guidelines on using AI in education and research published by the UNSECO (Miao and Holmes, 2023)</p>
Université de Genève (2024)	<p>22/04/2024: Realize that AI tools are already used at the university, but their use must be accompanied by training. Development of such courses is underway. Faculties and institutes are responsible for guidelines. Students should be instructed concerning the use of AI. Assessment modes should account for the existence of AI. Full transparency on the use of AI is required (citation rules). Recommendations include training students, establishing guidelines, recalling the notion of academic integrity and objectives of university education, reiterating the importance of writing skills, recall the importance of the author and their responsibility in academic work. Some guidelines about image generation (but generally allowed). They have a secretariat at the university level that is responsible for issues related to AI</p>
EPF Lausanne (2023) and e-mail sent to students	<p>09/10/2023: Students should follow instructions by teachers. Declare the use of AI. Beware that you are responsible for your work, and it has to respect scientific integrity rules. Do <i>not</i> use AI to learn new things, search for information, or generate content that you cannot check for correctness (the risk that information is false)? You may use AI to generate ideas if you can check the correctness of the output for help with format rather than content. There are various risks: Plausible nonsense, environmental damage, privacy, and bias.</p>

Continued on next page

Table 4.1: Guidelines and rules at Swiss universities – *Continued from previous page*

Source	Summary
Universität (2024)	Basel 16/04/2024: No general ban planned. Students should learn how to handle these tools sensibly and responsibly. The use of AI must be cited. ⁸ They recommend saving a full transcript of all prompts and answers outside of the tool and keeping a log of which tools are used and which sections. AI outputs are not scientific sources; the author's responsibility is to ensure relevance and correctness. Papers and exams must be solely the work of the student. Instructors decide case-by-case whether to allow AI. The university embraces AI tools by offering training and providing licensing. But also assess tools from the point of view of data protection. Consider AI in the declaration of independent authorship and have technical solutions to prevent access during evaluations. Also, guidelines on how to cite AI assistants should be provided. Established a working group on the topic.
Universität de Lausanne (2024)	16/04/2024: Permitted under the condition of data protection and confidentiality restrictions. Largely highlight the limits of these tools for summarizing literature and identifying gaps in the literature, but also reworking existing texts. Using AI is not plagiarism. However, the work submitted must be the student's independent work, and failing to cite AI appropriately violates scientific integrity rules. Also, it is problematic if the author's original work based on which the AI is trained is not cited. Also, how and when AI was used to generate certain results has to be reported. Parts that have been produced by AI are consequently not attributed to the student in the evaluation. Producing images for manuscripts and presentations is prohibited because the legal questions are unresolved. AI is difficult to detect. Students are explicitly allowed to use AI for brainstorming research ideas, improving codes, creating learning tools (like multiple-choice questions), ask for improvements in their written work. Teachers cannot ask students to use AI if they do not want to.

Continued on next page

⁸They refer to the following guidelines: style.mla.org/citing-generative-ai/

Table 4.1: Guidelines and rules at Swiss universities – *Continued from previous page*

Source	Summary
Universität Luzern (2023)	19/10/2023: Copy-paste AI output in a thesis is prohibited. This includes creating parts of the thesis (such as abstract, introduction, etc.). Data should not be uploaded to AI assistants for analytical purposes because correctness cannot be verified. Data must be collected independently or by third parties for the same reason. Exceptions to these rules have to be discussed with the supervisor. AI is allowed to edit, translate, and enhance language. Furthermore, idea generation and assisting in programming are allowed. However, students have to document how they used AI. Students are always responsible for the correctness of the content. The use of AI without disclosure violates the rules of academic integrity. It has to be declared in an appendix of the thesis (whether it has been used, how, and how the correctness has been verified; an example is provided). Recommend traditional tools for literature search.
ETH Zürich (2024b) and ETH Zürich (2024a)	19/01/2024: No official regulation, only guidelines. Allowed unless explicitly prohibited. Instructors decide whether to permit AI and modify the evaluation mode if necessary. It is not permitted to assign grades using AI automatically. Copy-pasting from chatGPT in any work is unacceptable (students must submit their own work). Even when citing AI as a source, these sources will be less convincing than well-published journal articles. The responsibility for the correctness of the output lies with the user. It is not possible to reliably detect the use of AI. But there are hints (e.g. wrong references and the like). Instructors can also demand a presentation of the work for a secondary check. Give worse grades for false statements. If instructors have a strong suspicion of inappropriate use of AI, they can start a disciplinary process. The Declaration of Independence was changed to cover AI. They offer a course on using chatGPT in scientific writing.
Universität Zürich (2023)	26/9/2023: Instructors teach appropriate use of AI and highlight their limitations. The use can be prohibited by instructors; they ensure that the evaluation conditions remain fair. Instructors and students transparently label the use of AI. The responsibility for the produced output lies with the corresponding user. Violation of scientific integrity that emanates from the use of AI will be punished. Faculties are responsible for establishing appropriate guidelines.

Continued on next page

Table 4.1: Guidelines and rules at Swiss universities – *Continued from previous page*

Source	Summary
Universität Bern (2023)	16/5/2023: Permitted, unless prohibited in single cases by the instructor. Instructors should consider AI assistants in the evaluation (evaluation mode, tasks to be completed). Instructors are responsible for verifying the use of AI. Written work has to be independent work by students. Students are responsible for the work's correctness, truthfulness, and relevance. AI outputs are not scientific sources but tools that facilitate research. The use of AI has to be appropriately indicated. The Declaration of Honor also covers the use of AI (there are separate versions for works where AI is permitted/prohibited). An FAQ on the use of AI has been established
Università della Svizzera italiana (2023)	28/6/2023: Permitted, unless explicitly prohibited for a specific activity (e.g. exam, assignment, etc.). Their use must always be correctly acknowledged (I used it / how I used it). Instructors should proactively discuss the use of generative AI tools with students and explicitly state what works and what does not for the various activities (e.g. for my exam, the use of generative AI is not allowed/is allowed). Trans-faculty working group established

Notes: The date either corresponds to the publication date of the guidelines or the date of access to the web resource. Information on the Universities of Applied Sciences can also be found in [Solis \(2023\)](#).

4.2 Recommendations for guidelines

This section provides some general guidelines for students and instructors. These guidelines do not represent the official view of the University of Neuchâtel but are solely our opinion based on a review of the existing guidelines at other universities, their relevance to the field of economics, and our own teaching experience.

4.2.1 For students

In the introduction, we mentioned the basic principles we regard as relevant: Permission, originality, transparency, and quality. In what follows, we add specific guidelines for each principle.

Permission

AI assistants are allowed if their use does not violate the rules of scientific integrity, legal provisions, or specific rules of the instructor or the university We believe that AI

assistants will become indispensable tools in the workplace. Therefore, we think the university should be a place where students learn how to use these tools and what their limitations are.

AI assistants are allowed to facilitate repetitive and secondary tasks The faster repetitive tasks are dealt with, the more time is left to think about the research question and how to answer it. Such repetitive tasks include spell-checking, improving grammar, debugging code, formatting the references list, brainstorming ideas, etc. This list may differ for different courses (e.g. debugging code using AI may be forbidden in a programming course).

AI assistants are allowed as a search tool Like the internet, AI assistants give access to vast amounts of information that can provide ideas, literature, and examples for motivating a research question. Therefore, such assistants are allowed as search tools (like search engines that give access to information on the web). Of course, any output has to be vetted for correctness and relevance in all cases, and the original source has to be cited.

AI assistants are allowed to get feedback AI assistants will give students feedback on how writing or coding skills can be improved. Students should compare what they wrote and what the AI assistant suggests to improve their writing and coding. Feedback is an essential part of the learning process. Although this feedback may not always be correct and is certainly not a perfect substitute for discussions with colleagues and instructors, it is an additional means for pondering over one's work and improving one's skills in the process.

If it is uncertain whether AI assistants are allowed, contact the instructor Instructors may forbid using AI assistants entirely or for specific tasks. If the AI assistant does the task that has been assigned or relevant parts of it, this is usually a sign that its use is forbidden.

Originality

AI assistants are not allowed to generate the work itself A research paper must be the student's original work. Therefore, creating an integral part of the work using AI assistants is forbidden. Examples include writing the literature review, writing the introduction, collecting the data to be analyzed, performing a descriptive analysis of the data, or writing a code in a programming course. This list is not exhaustive and may differ depending on the specific course.

What is original work?

The concept of original work is something difficult to grasp. Fenley (2023) provides a good summary of criteria that have to be satisfied for original work.

Original work implies that you incorporate, integrate, and extend existing work in an academic field in a novel way.

Some examples of original work include (but are not limited to):

- Create a coherent argument for or against a theory based on information collected from various sources
- Apply existing theories to a new situation or context
- Collect new data to test an existing theory
- Analyze existing data with a new methodology
- Review an existing body of literature to identify gaps or patterns that have not yet been identified

Everything you do not mark with a quotation or reference is considered your own writing and idea. It is key for the reader to be able to judge what is new in your work.

Another aspect of originality is your contribution. Clearly, if everything you write is in quotation marks, which is formally correct, your contribution to the academic literature is zero. However, academic work is only valuable if it adds a novel contribution to the existing state of knowledge. Intuitively, your contribution is the difference between the state of knowledge after your work and the state of knowledge before your work. This implies that you must properly cite existing knowledge and explain your contribution early on in your work (e.g., in the introduction).

Using output from an AI assistant based on an uploaded document is often problematic

Asking a ChatBot to write an introduction based on the uploaded thesis means that the student uses AI for an integral part of the work. The same holds if the student asks a ChatBot to write a literature summary based on uploaded papers. Also, letting a ChatBot conduct data analysis is problematic because this often constitutes an integral part of the thesis. Finally, uploading documents and data that are protected by intellectual property rights may infringe legal provisions.

Transparency

Output from AI assistants must be labeled with a citation According to the rules of scientific integrity, the work of others cannot be passed off as one's own. Therefore, students

should not use AI output as a source without citation (whether in original or paraphrased form). The University of Basel has good guidelines on how to cite such sources ([University of Basel, 2023](#)). When using AI assistants for secondary tasks (e.g. to improve language), it usually suffices to declare its use in an Appendix (see below).

When AI assistants were used for certain tasks, it must be declared in an Appendix AI assistants may significantly reduce grammar or spelling mistakes. Moreover, they may speed up other repetitive tasks. For transparency, students should declare the use of AI assistants. We show an example of such an AI checklist below.

Quality

Output from AI assistants is not regarded as a scientific source Even if information from AI assistants is appropriately cited, its direct use is not recommended. AI is trained based on documents created by other authors. Even if AI assistants use information without appropriate citation, students cannot do the same. Instead, students should find the information in the original scientific source and cite it accordingly.⁹ This increases the credibility of the information and attributes the idea to the correct author.

Students are responsible for the correctness and relevance of output generated by AI

The work students submit is regarded as their own. Any mistakes or irrelevant information will be attributed to the students, even if AI is correctly cited. This implies that students must verify the original source and vet the output from AI assistants for correctness and relevance before use. Remember that AI assistants only reproduce content on which they were trained. They will miss relevant research if they do not include the most recent working papers and work-in-progress. So, we should not only rely on AI assistants but also use some traditional academic tools (e.g. Scopus, Google Scholar, IDEAS/RePEc, for searching relevant literature).

4.2.2 For instructors

Decide whether and for what tasks to allow AI and communicate the rules There is a lot of uncertainty about whether and for what tasks AI assistants are allowed. Instructors should think about and communicate to students what is allowed and what is forbidden in their course. In addition, be aware that not all students have access to paid versions of these tools and

⁹For the same reason, a scientific paper will not cite Wikipedia articles but rather the original paper on which the Wikipedia article is based.

that some students may resist the use of these tools. Until these tools become more widely available and acceptable, make sure that students without access can follow the course as well.

Integrate AI assistants in teaching and show advantages and limitations For economic research, AI assistants will be an indispensable tool for more productive research. However, using AI assistants improperly may waste time and produce erroneous results. Let students experiment with whether AI assistants are helpful for a specific task or whether it is easier to solve the problem without them. Also, show examples where AI assistants produce erroneous results.

Remind students that rules of scientific integrity apply and that academic research is not about reiterating existing facts Many misuses of AI are covered by existing rules of scientific integrity. Therefore, it is helpful to remind students of these rules and explain how they affect the use of AI. In addition, remind students that AI assistants are good at reproducing existing knowledge but that economic research is about creating new knowledge that does not yet exist.

Account for the use of AI assistants in the evaluation mode It is difficult to reliably detect the use of AI in academic work. However, instructors can adapt the evaluation mode to ensure that students understand how they solved the main task of the problem asked. For example, a thesis could be accompanied by a presentation, where both the written thesis and the presentation must obtain a sufficient grade.

Account for AI assistants in the grading scheme Wrong citations, inconsistencies, and vague statements can be penalized more strongly.¹⁰ The weight of take-home problems can be reduced, and instead, more weight can be given to the evaluation in a controlled setting. Another possibility is to attribute more points for creativity in problem-solving rather than the solution itself.

Require students to document where and how they used AI assistants Let students fill in the documentation on how and which AI assistants they used (see an example for an AI checklist below).¹¹ This may also be useful for comparing the works of students with and

¹⁰AI often produces vague statements because the models are trained to avoid saying “I don’t know” and avoid making outright wrong statements.

¹¹Similarly, some publishers require a declaration of how AI was used when submitting scientific manuscripts (e.g. Elsevier, 2023). See also the declaration for this report in the Appendix.

without access to AI, at least during a transition phase where these tools are still not universally used.

Tools to detect the use of AI work poorly Do not use these tools to try to detect AI. In any case, it will be challenging to determine whether AI was used to generate the work itself (indeed forbidden) or only to correct language mistakes (could be allowed).

Wrong and spurious citations and other inconsistencies are a good indication that AI assistants have been used Check written work for errors and inconsistencies. AI assistants sometimes generate spurious output that looks correct to a non-expert. Such invented references or other inconsistencies can strongly indicate that scientific integrity rules have been violated. Another example of such an inconsistency is program code that still uses simulated data. Although the results look correct, the AI often does not directly have access to the actual data and, therefore, uses simulated data to illustrate the code.

Do not automatically assign grades based on AI Grading students' work is the sole responsibility of the instructor. The grading has to be the instructor's work and cannot be delegated to an AI. In addition, uploading students' work is problematic for intellectual property and privacy infringements.

Example: AI Checklist

The following document has to be submitted along with a term paper or master thesis. Please list all tasks for which AI was used, the section of the thesis, and the corresponding AI assistant (incl. version number).

Task	Section	Tool
Brainstorming		ChatGPT-4
Finding literature	2	Elicit
Reviewing literature		
Writing code		Github Copilot
Correcting code		
Analyzing data		
Writing text		
Correcting text	All	Grammarly
Formatting documents		
Other (please specify)		

The rules of the University of Neuchâtel on scientific integrity apply. In particular, using output from AI assistants without appropriate citation violates the rules of scientific integrity. In addition, AI uses information from large collections of documents. Appropriate citation implies that the original source is cited. Also, integral parts of a thesis cannot be delegated to AI. For example, writing a literature review based on uploaded documents and letting an AI assistant summarize the content is prohibited. Finally, the instructor may prohibit AI assistants entirely or for specific tasks. If you have any questions, please contact your instructor.

With your signature, you approve that you truthfully filled in this form.

Name:

Course:

Paper title:

Date:

Signature:

Appendix

A Declaration on the use of AI

Declaration of generative AI assistance in the writing process

During the preparation of this work, the authors used generative AI tools to copyedit the entire manuscript.^a After using these AI assistants, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

^aGrammarly Premium, app.grammarly.com and ChatGPT-4, chat.openai.com.

Literature

- Burri, M. (2023). Do daily lead texts help nowcasting GDP growth? IRENE Working Papers 23-02, IRENE Institute of Economic Research, Retrieved from <https://ideas.repec.org/p/irn/wpaper/23-02.html>.
- Elsevier (2023). The use of generative AI and AI-assisted technologies in writing for Elsevier. Retrieved from www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier. Accessed on 18 April 2024.
- EPF Lausanne (2023). Tips for the use of generative AI in research and education. Retrieved from www.epfl.ch/about/vice-presidencies/vice-presidency-for-academic-affairs-vpa/tips-for-the-use-of-generative-ai-in-research-and-education. Accessed on 24 April 2024.
- ETH Zürich (2024a). Academic integrity. Retrieved from ethz.ch/en/the-eth-zurich/education/ai-in-education/AcaIntegrity.html. Accessed on 24 April 2024.
- ETH Zürich (2024b). ChatGPT. Retrieved from ethz.ch/en/the-eth-zurich/education/ai-in-education/chatgpt.html. Accessed on 24 April 2024.
- ETH Zürich (2024c). Ethel – Virtual teaching assistant. Retrieved from ethz.ch/en/the-eth-zurich/education/ai-in-education/projects/ethel.html. Accessed on 31 May 2024.
- Felin, T. (2024). When prediction is not enough (with Teppo Felin). Retrieved from www.econtalk.org/when-prediction-is-not-enough-with-teppo-felin. Accessed on 16 April 2024.
- Fenley, M. (2023). Le travail académique original. Retrieved from www.unine.ch/blog/home/methodes/le-travail-academique-original.html. Accessed on 31 May 2024.
- Fowler, G. A. (2023). We tested a new ChatGPT-detector for teachers. it flagged an innocent student. Retrieved from www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin. Accessed on 22 April 2024.
- Garber, A. M., Weenick, M., and Jelinkova, K. (2023). Guidelines for using ChatGPT and other generative AI tools at Harvard. Retrieved from provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard. Accessed on 19 April 2024.
- Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4):1281–1317, DOI: [10.1257/jel.20231736](https://doi.org/10.1257/jel.20231736).
- Korinek, A. and Taliaferro, D. (2024). Generative AI for economic research. Retrieved from www.genaiforecon.org. Accessed on 31 May 2024.
- Miao, F. and Holmes, W. (2023). Guidance for generative AI in education and research. Report, UNESCO, Retrieved from unesdoc.unesco.org/ark:/48223/pf0000386693.

- Myers, A. (2023). AI-detectors biased against non-native English writers. Retrieved from hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers. Accessed on 22 April 2024.
- Nikolov, P. (2023). Writing Tips for Crafting Effective Economics Research Papers – 2023-2024 Edition. IZA Discussion Papers 16276, Institute of Labor Economics (IZA), Retrieved from ideas.repec.org/p/iza/izadps/dp16276.html.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.
- SNSF (2024). The SNSF's approach to the use of artificial intelligence in funding applications . Retrieved from www.snf.ch/en/Gz1VrJlkLyay4H7z/news/the-snsfs-approach-to-the-use-of-artificial-intelligence-in-funding-applications. Accessed on 17 June 2024.
- Solis, T. (2023). Die ChatGPT-Richtlinien der 23 Schweizer Hochschulen. Retrieved from www.scribbr.ch/ki-tools-nutzen-ch/chatgpt-richtlinien-hochschulen-schweiz. Accessed on 15 April 2024.
- The Economist (2024). Generative AI is a marvel. is it also built on theft? Retrieved from www.economist.com/business/2024/04/14/generative-ai-is-a-marvel-is-it-also-built-on-theft. Accessed on 18 April 2024.
- UC Berkeley (2023). Understanding ai writing tools and their uses for teaching and learning at UC Berkeley. Retrieved from teaching.berkeley.edu/understanding-ai-writing-tools-and-their-uses-teaching-and-learning-uc-berkeley. Accessed on 19 April 2024.
- University of Basel (2023). Guidelines «Citing AI tools »– How to use tools based on artificial intelligence. Retrieved from www.unibas.ch/dam/jcr:6d51f230-8ed1-499b-8c6e-d9ba9f60c346/guidelines-citing-AI-Tools_Apr-2023.pdf. Accessed on 16 April 2024.
- Università della Svizzera italiana (2023). ChatGPT & Generative AI @ USI. Retrieved from www.desk.usi.ch/en/node/13584. Accessed on 24 April 2024.
- Universität Basel (2024). AI in learning and teaching. Retrieved from www.unibas.ch/en/Studies/My-Studies/AI-in-learning-and-teaching.html. Accessed on 24 April 2024.
- Universität Bern (2023). Leitlinien zum Umgang mit künstlicher Intelligenz in der Lehre. Retrieved from www.unibe.ch/unibe/portal/content/e809/e878/e880/e915/e36136/e1368980/e1383230/20230516_LeitlinienKI_datiert_ger.pdf. Accessed on 24 April 2024.
- Universität Luzern (2023). Guidelines for the use of Artificial Intelligence (AI). Retrieved from www.unilu.ch/fileadmin/fakultaeten/wf/institute/hrm/dok/Studium/ENG_Merkblatt_KI_20231019_MM.pdf. Accessed on 24 April 2024.
- Universität Zürich (2023). Generative künstliche Intelligenz. Retrieved from www.uzh.ch/de/explore/basics/ai. Accessed on 24 April 2024.
- Université de Fribourg (2024). Génération de textes et d'images avec une intelligence artificielle. Retrieved from www.unifr.ch/it/de/generative-ai.html. Accessed on 24 April 2024.
- Université de Genève (2024). Use of generative artificial intelligence at UNIGE. Retrieved from www.unige.ch/en/university/politique-generale/use-generative-artificial-intelligence-unige. Accessed on 24 April 2024.
- Université de Lausanne (2024). FAQ IA. Retrieved from unil.ch/numerique/home.html. Accessed on 24 April 2024.

Université de Neuchâtel (2014). Règlement en matière de respect de l'intégrité scientifique. Retrieved from rsn.ne.ch/DATA/program/books/rsne/pdf/41610103.pdf. Accessed on 22 April 2024.

Université de Neuchâtel (2024). ChatGPT: Frequently asked questions about artificial intelligence and how it could impact teaching. Retrieved from www.unine.ch/files/live/sites/sp/files/doc/ChatGPT/AI%20and%20ChatGPT%20Frequently%20Asked%20Questions_SP2023_EN.pdf. Accessed on 24 April 2024.