

Integrating ChatGPT Into Your Research

Ingar Haaland

Department of Economics, NHH Norwegian School of Economics

May 8, 2024

What is ChatGPT?

- ChatGPT (Generative Pre-trained Transformer) is a state-of-the-art language model developed by OpenAI for understanding and generating text.
- It predicts subsequent words in a given context to generate human-like text, making it useful for a variety of applications such as writing assistance and coding.
- ChatGPT employs the Transformer architecture, a deep learning approach that allows the model to understand patterns, context, and language structure from vast amounts of text data.
- To improve its performance, ChatGPT is fine-tuned using Reinforcement Learning from Human Feedback (RLHF), allowing the model to adapt its responses and become more effective over time.

Why should you start using ChatGPT?

- It can boost your **productivity** and **creativity**.
- By experimenting with ChatGPT, you can learn more about its strengths and limitations.
- Over time, you will become better at “prompting” and ChatGPT will become better at answering your prompts.
- Even if ChatGPT does not make you more productive today, it will make you more productive tomorrow.

ChatGPT: A transformer, not a search engine

- Many people think of ChatGPT as a search engine, but advanced search requests often lead to **disappointing results**:
 - Relies on historical training data
 - Generates responses based on learned patterns, not by accessing a database
 - Training data can reflect existing biases or contain inaccuracies
- Right mental model: **Transformers** that use input A to generate a relevant and contextually appropriate output B

Input	Output
English text	A German translation
A verbal explanation of a coding task	Python code
Text block	A list of grammatical errors
An academic paper	A discussion of its strengths and weaknesses

- Effective use of ChatGPT requires well-designed prompts:
 - Your job is to explain how you want the input transformed, enhancing accuracy and relevance.

Use Cases of ChatGPT in Economic Research

- ChatGPT can be integrated into your research workflow for a variety of tasks, including:
 - Idea generation and feedback on research ideas.
 - Language-related tasks, such as generating complete text from crude summaries, translating Qualtrics surveys, and proofreading your paper.
 - Coding-related tasks, including creating figures and tables in R and Python, TikZ figures in L^AT_EX, code debugging, and writing complex regular expressions.
 - Qualitative data analysis: Extracting insights and patterns from large amounts of text data and assisting in qualitative data analysis.
 - Qualitative data generation: Conduct qualitative interviews with human subjects.
 - Survey design and analysis: Suggesting survey responses, performing sentiment analysis, annotating open-ended survey responses, etc.
 - Providing critical feedback on drafts, ideas, grant applications, etc.

Use case example: Translating a Qualtrics survey

Prompt:

I will now copy paste content from a CSV file.

The CSV files is a language file from Qualtrics that contains original English instructions. I need these translated into Norwegian.

The CSV file has the following structure

PhraseID,EN,NO

For instance, it might say

Q1,"What is your gender?",""

In this case, you would output something like

Q1,"What is your gender?","Hva er ditt kjønn?"

Here are the responses:

QID938.QuestionText,"To what extent do you consider ESG ratings when buying stocks and stock mutual funds?",""

QID938.Choice1,"To a very large extent",""

QID938.Choice2,"To a large extent",""

QID938.Choice3,"To some extent",""

QID938.Choice4,"To a little extent",""

QID938.Choice5,"To no extent",""

QID937.QuestionText,"How concerned are you about climate change?",""

QID937.Choice1,"Extremely concerned",""

QID937.Choice2,"Very concerned",""

QID937.Choice3,"Somewhat concerned",""

QID937.Choice4,"Not very concerned",""

QID937.Choice5,"Not at all concerned",""

ChatGPT output:

Sure, here are the Norwegian translations for the phrases:

QID938.QuestionText,"To what extent do you consider ESG ratings when buying stocks and stock mutual funds?","I hvilken grad tar du hensyn til ESG-rangeringer når du kjøper aksjer og aksjefond?"

QID938.Choice1,"To a very large extent","I svært stor grad"

QID938.Choice2,"To a large extent","I stor grad"

QID938.Choice3,"To some extent","Til en viss grad"

QID938.Choice4,"To a little extent","I liten grad"

QID938.Choice5,"To no extent","Ingen grad"

QID937.QuestionText,"How concerned are you about climate change?","Hvor bekymret er du for klimaendringer?"

QID937.Choice1,"Extremely concerned","Ekstremt bekymret"

QID937.Choice2,"Very concerned","Svært bekymret"

QID937.Choice3,"Somewhat concerned","Noe bekymret"

QID937.Choice4,"Not very concerned","Ikke særlig bekymret"

QID937.Choice5,"Not at all concerned","Ikke bekymret i det hele tatt"

Opportunities and limitations

- It is not difficult to create examples where GPT provides nonsensical responses or makes simple mistakes.
- Continuous development and fine-tuning, along with advancements in AI research, will lead to better performance and fewer errors in the future
- Try to adopt a “glass half full” attitude and make use of ChatGPT sooner rather than later.

Overcoming ChatGPT limitations

- The (free version) of ChatGPT faces many limitations, including its inability to search the internet for information or integrate with other software without supplementary tools.
- Possible solutions to address these limitations include:
 - **Apps and API Integration:** Researchers can leverage apps like Wolfram Alpha to obtain external information or use a Python API integration to connect ChatGPT with other software and unlock additional features.
 - **Data Preprocessing:** Optimize data and prompts through preprocessing to minimize the need for external information and enhance the generated text's quality.
 - **Plus subscription:** Pay for ChatGPT Plus to get access to GPT-4, Internet search, and app integration.
 - **Alternative software:** Use Microsoft Bing's chat engine which is freely running on GPT-4 with internet access.

Hallucination: A double-edged sword

- Hallucination is a common problem with language models, including ChatGPT, where the generated text is not supported by the input context, sometimes resulting in nonsensical or incorrect output.
- Hallucination can be both a strength and a weakness:
 - **Strength:** Instead of saying “I don’t know,” the model generates a response that seems correct, showcasing creativity and the ability to produce novel answers.
 - **Weakness:** Hallucination is a major problem in factual tasks, such as coding and literature reviews, where accuracy is critical.

Dealing with hallucination

- Researchers can mitigate the effects of hallucination by:
 - **Fine-tuning the model with additional data.** Training the model on more relevant and specific data helps generate more accurate and relevant text.
 - **Using prompt engineering.** Designing prompts that elicit desired responses guides the language model towards generating more relevant and accurate text.
 - **Applying temperature scaling.** Adjusting the "temperature" parameter during generation controls the randomness of the generated text and reduces the likelihood of hallucination.
- It is crucial to be aware of the potential for hallucination when using ChatGPT in research and take steps to minimize its impact on the generated text.

ChatGPT and domain-specific performance

- ChatGPT's performance varies across different domains and tasks due to differences in training data availability.
- Some areas show stronger performance, while others may require further improvements:
 - **Stronger Performance:** In widely covered domains, like Python and R programming and \LaTeX , ChatGPT excels at generating accurate code for reasonably advanced tasks.
 - **Weaker Performance:** In less represented domains, such as Stata programming, ChatGPT's performance is less reliable and may produce more hallucinated outputs.
- Although newer iterations like ChatGPT 4 have improved performance in underrepresented areas, hallucination still poses challenges for certain tasks.
 - Example: ChatGPT can easily generate publication style standard error bar charts for Python and R, but struggles with the same task for Stata.

Comparing ChatGPT-3 and ChatGPT-4

ChatGPT-3

- Smaller model: 175 billion parameters
- Limited context understanding
- Lower response accuracy and context awareness
- Faster processing time
- Available to free users
- Good for:
 - Simple language-related tasks
 - Basic coding assistance
 - Preliminary idea generation

ChatGPT-4

- Larger model: 300 billion parameters
- Improved context understanding
- Higher response accuracy and context awareness
- **Only available with a Plus subscription (\$20 month)**
- Preferred for:
 - Complex text generation
 - In-depth coding support
 - Qualitative data analysis
 - Advanced survey design and analysis
 - Providing critical feedback

Prompting: Setting the stage for better outputs

- Prompting is a crucial aspect of working with language models like ChatGPT.
- Effective prompts guide the model towards generating accurate, relevant, and contextually appropriate responses.
- Crafting good prompts can significantly reduce hallucination and improve the model's performance.
- Developing the skill of writing effective prompts often requires practice and experience to adapt the model to your specific needs and tasks.

Good and Bad Prompting Examples

- **Good Prompting:**

- Gives the LLM an appropriate context to understand the problem,
- Clear definition of relevant inputs (if relevant)
- Clear and specific about what the desired output should be and in what format

- **Bad Prompting:**

- Vague or ambiguous
- Insufficient context to understand the problem
- No guidance on desired output

- **Good Prompt Examples:**

- "Please provide a summary of the mtcars dataset in R, including the average miles per gallon (mpg), horsepower (hp), and weight (wt)."
- "Identify outliers in the mtcars dataset in R, particularly focusing on miles per gallon (mpg), horsepower (hp), and weight (wt)."
- "Describe the relationship between miles per gallon (mpg) and weight (wt) for the mtcars dataset in R."

- **Bad Prompt Examples:**

- "Tell me about cars." (*Lacks specificity*)
- "Which car in the mtcars dataset is the best?" (*Subjective without criteria*)
- "What's wrong with the mtcars data?" (*Ambiguous, lacks direction*)

Comparing vague and directed prompts

Vague prompt:

Tell me something about mtcars in R.

ChatGPT output:

The mtcars dataset is a built-in dataset in R that contains information on various attributes of 32 different automobile models, including the number of cylinders, horsepower, and miles per gallon (mpg) among others. Here's how you can access and view the mtcars dataset in R:

```
# Load the mtcars dataset
data(mtcars)

# View the first six rows of the dataset
head(mtcars)
```

This will display the first six rows of the mtcars dataset. You can also use various functions in R to explore and analyze the mtcars dataset, such as `summary()`, `str()`, `cor()`, and `plot()`, among others.

Directed prompt:

Describe the relationship between miles per gallon (mpg) and weight (wt) for the mtcars dataset in R.

ChatGPT output:

To describe the relationship between miles per gallon (mpg) and weight (wt) for the mtcars dataset in R, you can use the following code:

```
# Load the mtcars dataset
data(mtcars)

# Calculate the correlation between mpg and wt
correlation <- cor(mtcars$mpg, mtcars$wt)

# Print the correlation coefficient to the console
cat("Correlation between mpg and wt:", correlation, "\n")

# Load the ggplot2 package
library(ggplot2)

# Create a scatter plot of mpg against wt
ggplot(mtcars, aes(x = wt, y = mpg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(x = "Weight (wt)", y = "Miles per Gallon (mpg)",
       title = "Scatterplot of mpg vs wt") +
  theme_minimal()
```

This code calculates the correlation between mpg and wt and creates a scatterplot with a linear regression line to visualize the relationship.

Using guided examples to increase output quality

- Including examples in your prompts can provide context that helps the LLM understand the specifics of the task.
 - **Zero-Shot:** The model responds based on the prompt alone, without any prior examples.
 - **Few-Shot:** The prompt includes several examples to guide the model's understanding of the task.
 - **Fine-Tuning:** The model undergoes additional training on a specific dataset to enhance its performance on related tasks.
- Pros and cons of different approaches
 - **Zero-Shot:**
 - + Quick deployment, no need for example curation.
 - Less precise for complex queries, potential for higher error rates, higher variance.
 - **Few-Shot:**
 - + Better accuracy than zero-shot with minimal examples.
 - Performance heavily dependent on the quality and representativeness of examples.
 - **Fine-Tuning:**
 - + Highest performance and specificity to tasks.
 - Resource-intensive, risks overfitting, often limited benefit compared to few-shot prompting.

Applying prompting techniques: Sentiment analysis

- **Zero-shot prompt:** “Based on the content of this article, determine if the sentiment is positive, negative, or neutral.”
 - Quick deployment, requires no specific training data.
- **Few-shot prompt:** Additional context: “To help illustrate the difference between positive, negative, and neutral, consider these examples: ‘Article A’ is positive, ‘Article B’ is negative, and ‘Article C’ is neutral.”
 - Might improve accuracy by providing relevant examples, but performance depends heavily on the relevance of the examples provided.
- **Fine-tuning prompt:** The model is trained on a bigger data set with many more examples.
 - Delivers the highest performance and consistency, especially effective for complex and nuanced sentiment analysis.
 - Resource-intensive process, requiring significant time and data; risks overfitting to training data; often small gains compared to few-shot prompting.

Scaling with the API (application programming interface)

- The OpenAI API allows you to send external requests in which your data is processed using their models and responses are returned to you.
 - API requests do not contribute to training the models
- Example application: Sentiment analysis
 - Utilizes the API to analyze text data, identifying sentiments as positive, negative, or neutral.
 - Customizable prompts enable adaptation to specific datasets and contexts, enhancing the relevance and accuracy of analysis.
- Flexible implementation
 - **Data preparation:** Collect and preprocess textual data to ensure it is ready for analysis.
 - **API Setup:** Integrate the API using Python to dynamically handle and process large volumes of data.
 - **Execution:** Send text data to the API and receive analyzed sentiment evaluations based on GPT-4 (or a different model) from the OpenAI server.

Coding examples

Example Prompt and Output: String manipulation in Stata

Prompt:

I am working with the auto data in Stata.

Can you provide a string expression that replaces "Plym." with "Plymouth" in the string variable make?

ChatGPT output:

Yes, you can use the `subinstr()` function in Stata to replace "Plym." with "Plymouth" in the string variable `make`. Here's an example command:

```
replace make = subinstr(make, "Plym.", "Plymouth", .)
```

This command replaces all occurrences of "Plym." in the variable `make` with "Plymouth". The `subinstr()` function searches for the string "Plym." in the variable `make` and replaces it with "Plymouth". The dot at the end of "Plym." is included to ensure that only "Plym." is matched, and not "Plymouth" or any other string that contains "Plym." as a substring.

The `replace` command updates the values in the `make` variable with the new string expression. The last argument `."` specifies that the replacement should be done globally for all observations.

Example Prompt and Output: Coding assistance in Stata (1/2)

Prompt:

In Stata, I have a string variable with values such as "Expert", "Advice", "Impact", etc. I want to define a list along the lines of local "Export Advice Impact" for all options included in the variable. Is there a way I can automate this without reading the options of the tabulate?

ChatGPT output:

Yes, you can automate the creation of local macros from unique values of a string variable in Stata without having to manually read the options of tabulate. Here's a code snippet that should accomplish what you're looking to do:

```
// First, let's assume your variable is called str_var
// This is a blank local that will eventually hold all of your unique string variable values
local purpose

// This loop will go through each unique value in str_var and add it to the local macro "purpose"
levelsof str_var, local(levels)
foreach l of local levels {
    local purpose "`purpose' `l'"
}

// Now you can use the local macro "purpose" in other parts of your code
display "`purpose'"
```

The `levelsof` command gets the unique values of a variable and stores them in a local macro (`levels`). Then, the `foreach` loop goes through each value in `levels` and adds it to the `purpose` local macro.

Example Prompt and Output: Coding assistance in Stata (2/2)

Prompt:

I have four string variables code1, code2, code3, code4 created from a variable code and I currently have a loop to check whether they match the string expressions saved in my local macro purpose.

```
foreach x of local purpose {  
  gen purpose_`x' = code1=="`x'" | code2=="`x'" | code3=="`x'" | code4=="`x'"  
}
```

I don't like that I have to observe that the first code variable takes on 4 new variables (code1, code2, etc.) before I make the `foreach x of local purpose` loop. Is there a way I can avoid this to not make any assumptions about how many "code" variables I have to match?

ChatGPT output:

Yes, you can make your code more dynamic by using a wildcard (*) to match all variables that start with "code". Then you can loop through all of these variables when generating your purpose variables. Here's how to do it:

```
* Loop over each category in 'purpose'  
foreach x of local purpose {  
  * Initialize the new variable as 0  
  gen purpose_`x' = 0  
  
  * Loop over all code variables and update the new variable  
  foreach var of varlist code* {  
    replace purpose_`x' = 1 if `var' == "`x'"  
  }  
}
```

In this version of the code, the second `foreach` loop iterates over each of the "code" variables (code1, code2, etc.) and updates the relevant "purpose" variable if a match is found. This allows the code to work even if more code variables are added in the future.

Example Prompt and Output: Standard error bar chart in R

Prompt:

I want to create a figure in R. Provide code for the following:

1. Generate fake data for two treatment groups with 100 observations each and a mean of 0.5 for the first group and a mean of 0.7 for the second.
2. Create a standard error bar chart (that is, two bars with 95% confidence intervals indicated in the figure). Use modern packages and make the figure look as good as possible.

ChatGPT output:

Here is the R code to generate a fake dataset and create a standard error bar chart using the **ggplot2** package:

```
# Load required packages
library(ggplot2)
library(dplyr)

# Set seed for reproducibility
set.seed(42)

# Generate fake data
n <- 100
group1 <- data.frame(value = rnorm(n, mean = 0.5, sd = 0.2), group = "Group 1")
group2 <- data.frame(value = rnorm(n, mean = 0.7, sd = 0.2), group = "Group 2")
data <- rbind(group1, group2)

# Calculate means and standard errors for each group
summary_data <- data %>%
  group_by(group) %>%
  summarize(mean = mean(value),
            sd = sd(value),
            n = n(),
            se = sd / sqrt(n),
            ci = qt(0.975, n - 1) * se)

# Create the standard error bar chart
ggplot(summary_data, aes(x = group, y = mean, fill = group)) +
  geom_bar(stat = "identity", color = "black", width = 0.6) +
  geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci), width = 0.2) +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(title = "Standard Error Bar Chart",
       x = "Treatment Group",
       y = "Mean Value") +
  scale_fill_manual(values = c("Group 1" = "#FC8D62", "Group 2" = "#66C2A5"))
```

This code first loads the required packages, sets a seed for reproducibility, and generates the fake data with 100 observations for each group. Then, it calculates the means and standard errors for each group using the **dplyr** package. Finally, it creates a standard error bar chart with 95% confidence intervals using the **ggplot2** package, applying a minimal theme and custom colors for a modern look.

Running the code in R

R code:

```
# Load required packages
library(ggplot2)
library(dplyr)

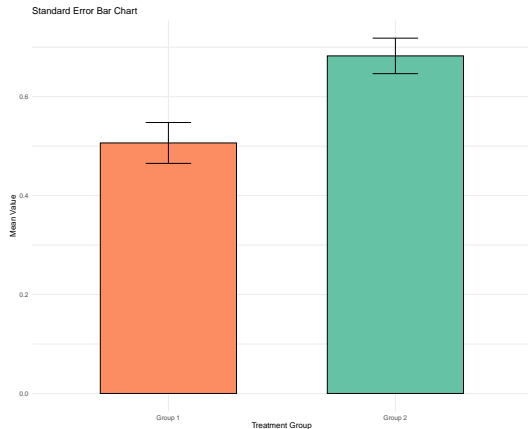
# Set seed for reproducibility
set.seed(42)

# Generate fake data
n <- 100
group1 <- data.frame(value = rnorm(n, mean = 0.5, sd = 0.2), group = "Group 1")
group2 <- data.frame(value = rnorm(n, mean = 0.7, sd = 0.2), group = "Group 2")
data <- rbind(group1, group2)

# Calculate means and standard errors for each group
summary_data <- data %>%
  group_by(group) %>%
  summarize(mean = mean(value),
            sd = sd(value),
            n = n(),
            se = sd / sqrt(n),
            ci = qt(0.975, n - 1) * se)

# Create the standard error bar chart
ggplot(summary_data, aes(x = group, y = mean, fill = group)) +
  geom_bar(stat = "identity", color = "black", width = 0.6) +
  geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci), width = 0.2) +
  theme_minimal() +
  theme(legend.position = "none") +
  labs(title = "Standard Error Bar Chart",
       x = "Treatment Group",
       y = "Mean Value") +
  scale_fill_manual(values = c("Group 1" = "#FC8D62", "Group 2" = "#66C2A5"))
```

R output:



Example Prompt and Output: Creating Beamer slides

First prompt:

I now want you to assist me with creating slides in Beamer.
I will give you the title of the slide and then use bullets and sub-bullets.
Bullets are indicated with * and sub-bullets are indicated with **.
Use the spacing `\medskip` for normal bullets and `\smallskip` for sub-bullets.
Confirm that you have understood.

ChatGPT output:

Understood. Please go ahead and provide me with the title of the slide and the bullet points with sub-bullets.

Second prompt:

Title: This is an example

- * It is always a hassle to create slides in beamer
- ** Especially when you need sub-bullets
- * With ChatGPT, you can make the process become easier
- ** You no longer have to stress about the sub-items..

ChatGPT output:

```
\begin{frame}{This is an example}
  \begin{itemize}
    \item It is always a hassle to create slides in beamer
    \begin{itemize}
      \smallskip
      \item Especially when you need sub-bullets
    \end{itemize}
    \medskip
    \item With ChatGPT, you can make the process become easier
    \begin{itemize}
      \smallskip
      \item You no longer have to stress about the sub-items.
    \end{itemize}
  \end{itemize}
\end{frame}
```

Example Prompt and Output: Creating BibTeX references

Prompt:

Please convert this reference into BibTeX format:

Alberto Abadie, Susan Athey, Guido W Imbens, Jeffrey M Wooldridge, When Should You Adjust Standard Errors for Clustering?, The Quarterly Journal of Economics, Volume 138, Issue 1, February 2023, Pages 1-35, <https://doi.org/10.1093/qje/qjac038>

ChatGPT output:

```
@article{abadie2013when,  
  title={When should you adjust standard errors for clustering?},  
  author={Abadie, Alberto and Athey, Susan and Imbens, Guido W and Wooldridge, Jeffrey M},  
  journal={The Quarterly Journal of Economics},  
  volume={138},  
  number={1},  
  pages={1--35},  
  year={2023},  
  month={February},  
  doi={10.1093/qje/qjac038}  
}
```

Example Prompt and Output: Regular expression in Qualtrics

Prompt:

I need a regular expression to validate that a text response in Qualtrics is a valid URL.
Provide a regular expression and explain briefly how I implement it in Qualtrics.

ChatGPT Output

To validate that a text response in Qualtrics is a valid URL, use the following regular expression:

```
^(https?|ftp):\\/[^\s/$. ?#] . [^\s]*$
```

To implement this regular expression in Qualtrics:

1. Add a **Text Entry** question in your survey.
2. Click the question to reveal editing options, and click the gear icon to open the **Question Settings** menu.
3. Select **Validation Type** and choose **Custom Validation**.
4. In the **Custom Validation** dialog, click the **Add a Rule** button.
5. Under **If**, select the text entry question, then choose **Matches Regex** as the condition.
6. Copy and paste the regular expression provided above into the **Pattern** field.
7. Under **Then**, customize the error message (e.g., "Please enter a valid URL.").
8. Click **Save** to apply the custom validation rule.

Now, the regular expression will validate that the response is a valid URL. If the input does not match the pattern, the customized error message will appear, prompting the respondent to correct their entry.

Examples of language-related tasks

Example Prompt and Output: Suggesting survey questions

Prompt:

I am creating a survey and want to measure political polarization in the United States.
I have thought about the following three questions:

1. The federal government has too much power
2. Gun laws in the U.S. should be stricter
3. Global warming is a serious problem

I want at least five more questions. Could you provide some good suggestions that include the most important topics in the US debate?

ChatGPT output:

Here are five additional questions to measure political polarization in the United States:

4. Immigration policy should be more restrictive.
5. The Affordable Care Act (Obamacare) should be repealed and replaced.
6. Abortion should be legal in most or all cases.
7. The U.S. should increase taxes on the wealthy to fund social programs.
8. The U.S. should prioritize renewable energy sources over fossil fuels.

Example Prompt and Output: Classifying open-ended responses

Prompt:

I asked survey respondents whether they would like to subscribe to a newsletter. Among respondents who said “No” to subscribing to the newsletter, I asked the following open-ended question: “Can you briefly explain why you did not want to subscribe to the newsletter?”

Can you please analyze the open-ended responses copied in below and then to the following:

- Suggest a classification scheme with the most common reasons people mention for not wanting to sign up for the newsletter. Use about 5 categories.
- Classify the open-ended responses according to this scheme.
- Show the results of the classification using a LaTeX percentage frequency table that includes one column with a representative example response.

I do not like having any newsletters because they feel like they are pandering to a specific audience so I instead try to look into multiple sources myself
I subscribe to enough newsletters on economic policy as it is and do not believe this one would benefit me.

[...]

ChatGPT output:

Based on the open-ended responses, I propose the following classification scheme for the reasons people did not want to sign up for the newsletter:

- Overwhelmed by existing subscriptions or avoiding email clutter
- Preference for multiple or self-selected news sources
- Lack of interest in the subject matter (economic policy)
- Mistrust of the newsletter’s credibility, bias, or accuracy
- Time constraints or disinterest in reading newsletters in general

Here is a LaTeX percentage frequency table representing the classification results, including a representative example response for each category:

Category	Percentage	Example Response
Overwhelmed by existing subscriptions or email clutter	31.3%	I already get a lot of newsletters so trying to cut back and I can check economic news manually by going to my usual news sites
Preference for multiple or self-selected news sources	19.3%	I like to do my own research on news as well as news sources. Comparing multiple sources in order to understand a perspective.
Lack of interest in the subject matter (economic policy)	21.3%	I’m not very interested in economic policy.
Mistrust of the newsletter’s credibility, bias, or accuracy	16%	If you don’t plan to run all the information and pick and choose, what good are you as a news source.
Time constraints or disinterest in reading newsletters in general	24.7%	I don’t have time nor do I want to read another newsletter, but thank you.

Example Prompt and Output: Getting a mock report on a paper

Prompt:

Write a referee report based on the introduction copied in below. Be sure to make a clear recommendation about whether the article should be published and point out strength and weaknesses of the paper.

Mounting empirical evidence shows that news outlets report the news in a politically biased way and that readers tend to consume like-minded news [...]

ChatGPT output:

Recommendation: Consider for publication after addressing concerns.

Strengths:

- Novel experimental design to quantify the relative importance of accuracy concerns and belief confirmation motives in driving the demand for news.
- Comprehensive methodology with multiple experiments, a structural estimation approach, and a separate mechanism experiment.
- Clear and well-motivated research question, relevant for understanding the dynamics between media bias, regulation, and political polarization.

Weaknesses:

- Limited generalizability due to the use of a single news outlet (The Boston Herald) and the focus on specific political issues.
- Experiments conducted on an online survey platform (Prolific) may not capture the behavior of the general population.
- The results may not account for the potential influence of other factors, such as readers' cognitive biases or social preferences, on the demand for news.

Recommendations:

- Discuss the generalizability of the results and consider replicating the experiments with different news outlets and political issues.
- Provide additional robustness checks to ensure that the sample is representative of the general population.
- Consider exploring alternative explanations for the observed demand patterns, such as cognitive biases or social preferences.

This report recommends considering the article for publication after addressing the concerns mentioned. The paper has a novel experimental design and comprehensive methodology. However, it has some weaknesses, such as limited generalizability, reliance on an online survey platform, and potential influence of other factors. The authors should address these concerns before the article can be considered for publication.

Example Prompt and Output (1/2): Proofreading assistance

Prompt (1):

You are now my copy editor. I am writing an academic paper in the field of economics.

I will give you a few paragraphs from the paper. I want you to point out each sentence that includes clear examples of unclear writing or awkward phrasing.

Specifically, for sentences with unclear writing or awkward phrasing, create a bullet list for each sentence that you identify with a suggestion for how to improve the sentence. I also want a brief justification for the change. Each suggestion based on a sentence should thus follow the following format: Old sentence, new sentence, and justification.

I also want you to create a separate bullet list of all grammatical errors that you find in the paragraphs.

Let me know if these instructions are clear. Only start copy editing when I actually give you a paragraph.

ChatGPT output (1):

Yes, the instructions are clear. Please provide the paragraphs you would like me to copy edit, and I will identify sentences with unclear writing or awkward phrasing and provide suggestions for improvement, along with justifications for the changes. I will also create a separate bullet list of grammatical errors found in the paragraphs.

Example Prompt and Output (2/2): Proofreading assistance

Prompt (2):

Mounting empirical evidence show that news outlets report the news in a politically biased way and that readers tend to consume like-minded news (Durante and Knight, 2012; Gentzkow and Shapiro, 2010). This has led to growing concerns about the news media contributing to increasing political polarization (Campante, Durante and Tesei, 2022; Durante, Pinotti and Tesei, 2019; Levy, 2021). Furthermore, since biased news could lead to less informed voters and increase social fragmentation, Sunstein (2018) warns that media bias and the emergence of political echo chambers could threaten the functioning of democracies [...]

ChatGPT output (2):

Here are the suggested improvements for sentences with unclear writing or awkward phrasing:

- Old sentence: Mounting empirical evidence show that news outlets report the news in a politically biased way and that readers tend to consume like-minded news (Durante and Knight, 2012; Gentzkow and Shapiro, 2010).
New sentence: Mounting empirical evidence indicates that news outlets exhibit political bias in their reporting, and readers tend to consume news that aligns with their existing beliefs (Durante and Knight, 2012; Gentzkow and Shapiro, 2010).
Justification: The new sentence provides a clearer and more concise description of the empirical evidence related to news reporting and consumption.
- Old sentence: This has led to growing concerns about the news media contributing to increasing political polarization (Campante, Durante and Tesei, 2022; Durante, Pinotti and Tesei, 2019; Levy, 2021).
New sentence: Consequently, concerns have grown about the news media's role in contributing to increased political polarization (Campante, Durante, and Tesei, 2022; Durante, Pinotti, and Tesei, 2019; Levy, 2021).
Justification: The new sentence more clearly attributes the cause of growing concerns and avoids the awkward repetition of "increasing" twice.
- Old sentence: An incredible identification challenge when trying to quantify the relative importance of the two motives is that theories based on belief confirmation motives often make predictions that are observationally equivalent with Bayesian updating about source quality (Gentzkow and Shapiro, 2006).
New sentence: A significant challenge in quantifying the relative importance of the two motives is that theories based on belief confirmation motives often produce predictions observationally equivalent to those from Bayesian updating about source quality (Gentzkow and Shapiro, 2006).
Justification: The new sentence uses a more appropriate word ("significant" instead of "incredible") and rearranges the sentence structure for better clarity.

Grammatical errors:

- "Mounting empirical evidence show" should be "Mounting empirical evidence shows" to agree with the singular subject "evidence."
- "Campante, Durante and Tesei, 2022; Durante, Pinotti and Tesei, 2019; Levy, 2021" should be "Campante, Durante, and Tesei, 2022; Durante, Pinotti, and Tesei, 2019; Levy, 2021" to maintain consistency with the Oxford comma.
- "This makes it challenging" should be "This makes it challenging" to correct the spelling error.

The making of a research project

First step: Literature review

Prompt:

I am planning a research project where I want to give people information about the extent of income inequality and see how it affects preferences for redistribution. Can you provide a quick literature review where you summarize the key insights from the five most relevant papers in economics?

ChatGPT output:

Certainly! Here is a brief literature review of five relevant papers in economics that examine the relationship between income inequality, preferences for redistribution, and the provision of information about inequality. Keep in mind that my knowledge is limited to articles published up until September 2021.

1. Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review*, 105(4), 1478-1508.
This paper investigates the elasticity of preferences for redistribution using randomized survey experiments. The authors find that providing information about income inequality increases support for redistribution. However, they also note that the effect is smaller for individuals who perceive higher social mobility, suggesting that beliefs about social mobility play a crucial role in shaping preferences for redistribution.
2. Alesina, A., & La Ferrara, E. (2005). Preferences for Redistribution in the Land of Opportunities. *Journal of Public Economics*, 89(5-6), 897-931.
This study examines the determinants of preferences for redistribution in the United States, using data from the General Social Survey. The authors find that individuals who perceive higher income inequality are more likely to support redistribution. They also highlight the role of racial composition and historical factors in shaping these preferences.
3. Cruces, G., Perez-Truglia, R., & Tetaz, M. (2013). Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment. *Journal of Public Economics*, 98, 100-112.
This paper uses survey experiments in Argentina to analyze how providing information on income inequality affects preferences for redistribution. The authors find that respondents who were informed about the actual extent of inequality showed a stronger preference for redistribution than those who were not. This effect was more pronounced for individuals who had underestimated the level of inequality.
4. Karadja, M., Mollerstrom, J., & Seim, D. (2017). Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution. *Review of Economics and Statistics*, 99(2), 201-212.
This study investigates how relative income improvements affect preferences for redistribution using a natural experiment in Sweden. The authors find that an increase in relative income leads to a decrease in support for redistribution, suggesting that people's preferences for redistribution are influenced by their own economic situation and social comparisons.
5. Trump, K. S. (2018). Income Inequality Influences Perceptions of Legitimate Income Differences. *British Journal of Political Science*, 48(4), 929-952.
This paper examines how income inequality affects perceptions of legitimate income differences. The author finds that individuals in countries with higher income inequality perceive larger income differences as legitimate. This suggests that, in societies with higher income inequality, people might be less inclined to support redistribution because they perceive the existing income differences as fair.

These papers collectively highlight that providing information about income inequality can influence preferences for redistribution. However, these preferences are also shaped by factors such as perceptions of social mobility, relative income, and beliefs about the fairness of income differences.

Second step: Generate research ideas

Prompt:

It seems there has been a lot of work done in this space already. I need some good ideas for how to differentiate my project from the papers you listed above and make sure it's novel and ambitious enough to be published in a good economics journal.

I still want the paper to take this general flavor of giving people information about income inequality and look at how it affects their redistributive preferences. Can you help me brainstorm about a potential project? It would be especially helpful if you could make some concrete suggestions for possible treatment interventions (that is, the treatment text that I should provide to respondents) as well as which outcomes and populations I should target.

ChatGPT output:

Certainly! Here are three potential project ideas that build on the existing literature but introduce novel elements to differentiate your research. Each idea includes suggestions for treatment interventions, targeted outcomes, and populations to focus on.

1. Investigate the impact of regional and global inequality comparisons on preferences for redistribution.
Treatment interventions: Provide respondents with information about income inequality within their region (e.g., city or state) as well as comparisons with national and global levels of inequality. You may also include visual aids, such as graphs or charts, to help participants understand the comparisons.
Targeted outcomes: Assess changes in preferences for redistribution and explore whether regional or global comparisons have a greater impact on these preferences. Investigate if the respondents' own income level plays a role in their responses.
Populations: Consider focusing on countries or regions with distinct levels of income inequality or social mobility. You may also target specific income groups to understand how their preferences for redistribution change based on regional and global inequality information.
2. Examine the role of information framing and perceived fairness in shaping preferences for redistribution.
Treatment interventions: Present participants with information about income inequality using different frames (e.g., highlighting the top 1% earners, focusing on poverty rates, or emphasizing the role of inherited wealth). You may also include scenarios that emphasize fairness or unfairness in income distribution (e.g., opportunities for education or the influence of corruption).
Targeted outcomes: Analyze how different framings of income inequality affect preferences for redistribution and whether these effects are moderated by participants' perceptions of fairness or unfairness. Assess whether the participants' own income level or political ideology influences their responses.
Populations: Focus on populations with diverse political affiliations and socioeconomic backgrounds to better understand the interaction between information framing, perceived fairness, and preferences for redistribution.
3. Investigate the influence of perceived social mobility and policy alternatives on preferences for redistribution. [...]

These project ideas should help you create a novel and ambitious research project that builds upon the existing literature while exploring new aspects of how information about income inequality affects preferences for redistribution.

Third step: Make one idea concrete

Prompt:

I like your second idea of examining role of information framing and perceived fairness in shaping preferences for redistribution. To get started on this idea, could you please sketch a survey where you include

1. Which beliefs, if any, I should measure pre-treatment and how to formulate the questions (provide formulations I could copy directly into Qualtrics)
2. The exact treatment texts I could provide to respondents in the different groups
3. Survey instruments for the post-treatment outcomes (provide formulations I could copy directly into Qualtrics)

ChatGPT output:

Here's a sketch of a survey based on the second idea, examining the role of information framing and perceived fairness in shaping preferences for redistribution. Please note that you may need to adjust the questions and treatment texts to fit your specific research context and population.

Pre-treatment questions: Measuring beliefs. To assess participants' initial beliefs, consider asking the following questions:

On a scale of 1 to 10, how would you rate the level of income inequality in your country? (1: Very low, 10: Very high)

On a scale of 1 to 10, how would you rate the current income distribution in your country? (1: Completely unfair, 10: Completely fair)

What do you think is the primary driver of income inequality in your country? (Options: Educational opportunities, Inherited wealth, Hard work and talent, Corruption and cronyism, Labor market conditions, Other)

Treatment texts: Information framing. Create different treatment groups, each receiving a distinct framing of income inequality information. Here are four examples:

A: Top 1% earners: "In our country, the top 1% of earners have captured a significant share of income growth over the past few decades. Their incomes have grown much faster than the rest of the population, contributing to higher income inequality."

B: Poverty rates: "In our country, poverty rates have remained stubbornly high over the past few decades, with many people struggling to make ends meet. This has resulted in a growing gap between the poor and the rest of the population, leading to increased income inequality."

C: Inherited wealth: "In our country, a large proportion of wealth is inherited, meaning that those born into wealthy families are more likely to remain wealthy throughout their lives. This concentration of wealth has contributed to the high levels of income inequality."

D: Corruption and cronyism: "In our country, corruption and cronyism play a significant role in shaping income distribution. Those with political connections or who engage in corrupt practices often enjoy higher incomes, leading to a more unequal distribution of income."

Post-treatment outcome questions. After providing the treatment texts, ask participants the following questions:

On a scale of 1 to 10, how would you now rate the level of income inequality in your country? (1: Very low, 10: Very high)

On a scale of 1 to 10, how would you now rate the current income distribution in your country? (1: Completely unfair, 10: Completely fair)

Given the information provided, how important do you think it is for the government to implement policies aimed at reducing income inequality? (1: Not at all important, 10: Extremely important)

Which policy would be most effective in addressing income inequality? (Options: Increasing taxes on the wealthy, Expanding access to quality education, Implementing policies to reduce corruption, Increasing the minimum wage, Implementing a universal basic income)

Fourth step: Aspirational abstract

Prompt:

Thanks for these excellent suggestions. Now assume I have collected the data and need to write up an abstract to send to a top economics journal. I want to get a feeling for how the final paper could look like. Make up some plausible results and write the abstract. Also suggest a good title.

ChatGPT output:

Title: “Framing Inequality: How Information Presentation and Perceived Fairness Shape Preferences for Redistribution”

Abstract: This paper investigates how information framing and perceptions of fairness influence preferences for redistribution. We conducted an online survey experiment with a diverse sample of 2,000 participants, randomly assigning them to one of four treatment groups. Each group received distinct information on income inequality in their country, framed by either the top 1% earners, poverty rates, inherited wealth, or corruption and cronyism. Participants’ beliefs about income inequality and fairness were measured both pre- and post-treatment, along with their preferences for redistributive policies.

Our findings reveal that the framing of income inequality significantly affects respondents’ preferences for redistribution. The inherited wealth frame (Treatment C) produced the largest increase in support for redistributive policies, followed by the corruption and cronyism frame (Treatment D). The top 1% earners and poverty rate frames (Treatments A and B) had more modest effects. We also find that respondents’ perceptions of fairness play a crucial role in moderating the relationship between information framing and preferences for redistribution. Participants who perceived the income distribution as unfair were more sensitive to the framing effects and demonstrated stronger support for redistribution.

These results highlight the importance of considering both the content and the presentation of information when studying preferences for redistribution. By understanding how individuals’ perceptions of fairness interact with different framings of income inequality, policymakers can better design communication strategies and policy proposals to foster public support for reducing income disparities.

Final step: Evaluate your idea critically

Prompt:

Assume we have the finished paper written up. Make a list of strengths and weaknesses of the paper that referees are likely to point out.

ChatGPT output:

Strengths:

1. Novel approach: The paper investigates a relatively underexplored aspect of preferences for redistribution by focusing on the role of information framing and perceived fairness, providing valuable insights into how these factors shape individual preferences.
2. Robust experimental design: The use of a randomized survey experiment with a diverse sample ensures that the causal effects of information framing on preferences for redistribution can be reliably estimated.
3. Rich data collection: The survey measures beliefs about income inequality and fairness both pre- and post-treatment, allowing for a comprehensive analysis of the changes in preferences for redistributive policies.
4. Policy relevance: The paper's findings have important implications for policymakers aiming to design effective communication strategies and policy proposals to address income inequality.

Weaknesses:

1. Generalizability: The results might be specific to the country or population studied, limiting the generalizability of the findings to other contexts. Further research in different countries or with different demographic groups may be necessary to establish the external validity of the results.
2. Potential for social desirability bias: Respondents might be influenced by social desirability bias when answering questions about perceived fairness and preferences for redistribution, which could affect the validity of the results.
3. Limited number of frames: The study focuses on four specific frames of income inequality, leaving room for additional research on other potential framings and their effects on preferences for redistribution.
4. Lack of exploration of potential mechanisms: The paper could benefit from a more in-depth exploration of the psychological or cognitive mechanisms that drive the observed effects of information framing and perceived fairness on preferences for redistribution.

Concluding remarks

Conclusion

- **Versatile assistant:** ChatGPT is a powerful AI tool that can significantly enhance your research and productivity in economics, offering support in various tasks such as idea generation, academic writing, and coding.
- **Effective usage:** After gaining experience with prompting, you can optimize ChatGPT's performance, mitigating issues like hallucination and obtaining more targeted results.
- **Continuous improvement:** As AI technology progresses, ChatGPT will become even more capable and reliable, further expanding its potential applications and benefits in the field of economics. At the same time, you will become better at using it with experience.
- **Tailoring to your needs:** To maximize the value of ChatGPT, it is crucial to adapt it to your specific workflow, learning to leverage its strengths and overcome its limitations for your unique research goals and projects.