

# Lecture 27

## How to use LLMs to turbo-charge your research productivity

---

Tyler Ransom

ECON 5253, University of Oklahoma

# Today's plan

1. Describe Large Language Models (LLMs, e.g. GPT-4)
2. Practical tips for getting the most out of LLMs
3. Walk through the workflow for producing a research paper
4. Go through an example step by step

# What are Large Language Models (LLMs)?

- LLMs are statistical models that can manipulate text (hence "language models")
- "Large" in the sense that they have billions (trillions?) of parameters
- Built on 5 steps (source: Cal Newport's podcast, [here](#)):
  1. **Word Guessing** - recursive text completion
  2. **Relevant Word Matching** - find similar words in training data ("relevant" is secret)
  3. **Voting** - probabilistic selection of next word
  4. **Feature Detection** - strengthen vote based on context of next word
  5. **Self-training** - iterative refinement of feature detection process

# LLM components and behavior

- Users interact with LLMs using **prompts** (divided into "tokens" [word chunks])
- **context window:** no. of tokens LLM can take as input when generating responses
- **temperature:** randomness of responses (0 = deterministic, 1 = random)
- **hallucination:** LLMs can generate text that is not in the training data
  - This is a feature, not a bug!
  - But it does mean you need to be careful in interpreting output if you want something that is factually correct
- Context window is currently a limiting factor in LLM performance (like computer RAM)

# What is the current state of the art?

As of late April 2024, there are 4 major players:

- OpenAI's GPT-4
- Google's Gemini
- Meta's Llama
- Anthropic's Claude

Each product may have differing derivative features

- Gemini can search YouTube; Bing Chat can use GPT-4
- Each product has a "freemium" business model; \$20/month unlocks full features

# Which chatbot should I use?

- Each product on the previous slide has strengths and weaknesses
  - You must experiment to find the best fit for your needs
- Each product delivers more or less the same baseline performance
- I use Gemini if I want a YouTube video summary
- GPT-4 allows for custom "GPTs" that can automate interactions
- Claude 3 Opus has large context windows and is my current favorite
- I expect OpenAI to release a major update within 6 months, at which point I will probably drop Anthropic and switch to OpenAI

# Other AI-based tools

- `Elicit.org` and `Consensus.app` for literature reviews
- GitHub Copilot (free for academics!) for code completion
  - LaTeX/RMarkdown files can be thought of as code
  - Code completion then becomes "writing completion" in these instances
- `lex.page` for writing and editing
- `Microsoft Copilot` for Word, PowerPoint, etc.
- `Perplexity` and `Bing Chat` for AI-augmented internet search

# Practical tips for using LLMs: prompts

- Prompt quality matters, e.g. "show me how to run a regression in R" versus

can you give me code to run a regression of mpg on weight, cylinders, and carburetor using the mtcars sample dataset in R? please use tidyverse packages where possible

- **Metaprompting** can be the easiest rule of thumb, e.g.
  - "give me a prompt for an LLM to get it to do the following task:"  
  
followed by
  - "what extra information do I need to include to best aid completion of the task?"



# Practical tips for using LLMs: variables in prompts

- It's also possible to put variables in your prompt, like so:

Write a personalized thank you letter for [customer] for buying [product]. The thank you letter is intended to be given with the product. Write the letter around how the product can help [customer] in a polite, glad, extremely authentic tone, and the reader should feel comfortable and connected to reach out to the company for feedback.

Product = "a graphic design software called Hue with integrated AI tools"

Customer = "name: Steve, a graphic designer"

# Steps for completing a research paper

- Choose a topic
- Conduct a literature review
- Develop a research design
  - Experimental, quasi-experimental, observational, ...
- Collect and analyze data
  - Data cleaning, statistical modeling, hypothesis testing, ...
- Communicate your results
  - Interpreting, writing, editing, visualizing, presenting, ...

# Step 1: Choose a research question

- Find a topic that is interesting, relevant, and feasible
- Get some ideas from the literature, professors, peers, news, blogs, etc.
- Make sure your question is specific, clear, and answerable with data and analysis
- The key here is to **be curious**
- **LLM-based tools:**
  - Prompt iteration for brainstorming (start with metaprompt)

## Step 2: Conduct a literature review

- Find out what has been done before on your topic
- Summarize and synthesize the main findings and arguments of the literature
- Identify the gaps and controversies and how your research can contribute
- **LLM-based tools:**
  - `Elicit.org`
  - `Consensus.app`
  - Upload a PDF of a paper and ask for a summary
  - Upload a document of abstracts and ask for a synthesis

# Step 2a: Build a reference database

- Use BibTeX or similar to store references
- LLMs are great at creating BibTeX entries from copy/pasted metadata
- I created a "GPT" to do this, see [here](#) (requires paid OpenAI subscription)
- **LLM-based tools:**
  - Copy/paste messy article metadata and ask for a BibTeX entry in a code block, then copy output into your BibTeX file

# Step 3: Develop a research design

- Specify your data sources, variables, hypotheses, models, estimation methods, and tests
- Is this a causal or predictive model? Is there missing data? Measurement error?
- Explain how your data and methods can address your research question and test your hypotheses
- Discuss the strengths and limitations of your data and methods
- Consider data quality, sample size, measurement error, endogeneity, identification, robustness, etc.
- **LLM-based tools:** Prompt iteration (start with metaprompt)

# Step 4: Collect and analyze your data

- Use appropriate software tools, such as R, Python, Stata, etc. to collect and analyze your data
- Follow the steps of your research design and report your results in tables and graphs
- Interpret your results in light of your hypotheses and the literature
- Check for any errors or inconsistencies in your data and analysis
- Perform any sensitivity analyses or robustness checks as needed
- **LLM-based tools:** Prompt iteration (start with metaprompt); ask for code

# Step 5: Write your research paper

- Follow the structure and style of economics papers
- Include an abstract, an introduction, a literature review, a data and methods section, a results section, a discussion section, a conclusion section, and a reference list
- Follow the formatting and citation guidelines of your instructor or journal
- Write clearly, concisely, logically, and persuasively
- Use economic terminology and notation correctly and consistently
- Support your claims with evidence and references
- **LLM-based tools:** See the slide after next



# Step 6: Revise and edit your research paper

- Improve your paper's content, organization, clarity, accuracy, and presentation
- Proofread your paper for any spelling, grammar, punctuation, or formatting errors
- Check for any logical flaws, factual errors, or missing information in your paper
- Seek feedback from your instructor or peers on your paper
- Incorporate their suggestions as appropriate, but don't be too deferential
- **LLM-based tools:** See next slide

# LLM-based tools for writing and editing

- Prompt iteration (start with metaprompt); asking for help on the following:
  - outlining (from scratch)
  - distilling scattered ideas into cogent paragraphs
  - simplifying existing paragraphs; reducing word count
- [lex.page](#); GitHub Copilot; Microsoft Copilot (Word, PPT, etc.)
- **You should never blindly incorporate LLM writing into your own work**
  - Always read and edit the output to ensure it is accurate and appropriate
  - LLMs are tools, not substitutes for your own thinking and writing

# A step-by-step example with prompts

For all the following content, I used Claude 3 Opus (Anthropic's premium model), except where explicitly stated

## 1. Use a metaprompt to begin

can you give me a prompt for how to ask for an outline of an empirical research paper on the topic of "what americans believe to be healthy diet and exercise habits". what sorts of details should i provide to get the most out of the answer?

# Metaprompt result

Certainly! Here's a prompt you can use to ask for an outline of an empirical research paper on the topic "What Americans believe to be healthy diet and exercise habits":

"I'm working on an empirical research paper about what Americans believe to be healthy diet and exercise habits. Could you please provide me with a detailed outline for this paper? The outline should include the main sections typically found in an empirical research paper, such as the introduction, literature review, methodology, results, discussion, and conclusion.

To help you create a more comprehensive outline, here are some additional details about my research:

# Outline of paper

Here is a suggested outline for your empirical research paper on what Americans believe to be healthy diet and exercise habits:

I. Introduction  
A. Background on the importance of healthy diet and exercise habits  
B. Research question: What do American adults believe constitutes healthy diet and exercise habits?  
C. Significance of the study  
D. Brief overview of methodology

II. Literature Review  
A. Overview of existing research on Americans' knowledge and beliefs about healthy eating and physical activity  
B. Key findings from previous studies  
C. Gaps in the literature that your study aims to address

III. Methodology  
A. Study design  
B. Sampling  
C. Data collection  
D. Data analysis

# Literature review: Elicit

**Elicit**

[FAQ](#) [Tasks](#) [Starred](#)

**what do americans believe to be healthy diet and exercise habits?**

SUMMARY OF TOP 4 PAPERS **BETA**

These papers collectively suggest that Americans have a general awareness of what constitutes a healthy diet and exercise habits, but there is still a gap between knowledge and behavior. [Morrow 2004](#) found that while Americans are aware of traditional physical activities that provide health benefits, they are less aware of specific exercise guidelines and lifestyle physical activities that can also result in health benefits. [Powell 2019](#) suggests that objective screening in the Electronic Health Record could improve awareness and education about nutrition and fitness, and help identify individuals who would benefit from interventional behavioral resources. [Kapsak 2013](#) emphasizes the importance of

Add information about all papers

Abstract summary

Intervention

Outcomes measured

Search for paper information

Has PDF


Filter

Sort by

Export as

Paper title	Abstract summary
<div></div> <div><b>Putting the dietary guidelines for americans into action: behavior-directed messages to motivate parents-phase III quantitative message testing and survey evaluation.</b></div> <div><div> Wendy Reinhardt Kapsak, Marianne Smith Edge, Christy White, Nancy M Childs, Constance J G...</div><div> <i>Journal of the Academy of Nutrition and Dietetics</i> </div></div> <div>2013 11 Citations <a href="#">DOI</a></div>	The consistent application of diet and physical activity recommendations to daily life is not widespread.
<div></div> <div><b>Screening for unhealthy diet and exercise habits: The electronic health record and a healthier population</b></div> <div><div> Heidi Sara Powell, Deborah Lynn Greenberg</div><div> <i>Preventive medicine reports</i> </div></div> <div>2019 11 Citations <a href="#">DOI</a> <a href="#">PDF</a></div>	A few relatively simple changes could prompt providers to assess and educate patients about nutrition and fitness and promote a healthier population.
<div></div> <div><b>American Adults' Knowledge of Exercise Recommendations</b></div> <div><div> J. Morrow, Jeanette A. Krzewinski-Malone, A. Jackson, T. Bungum, S. Fitzgerald</div><div> <i>Research Quarterly for Exercise and Sport</i> </div></div> <div>2004 134 Citations <a href="#">DOI</a></div>	Physical activity knowledge alone is not sufficient to elicit a behavior.
<div></div> <div><b>Health habits of U.S. adults, 1985: the "Alameda 7" revisited.</b></div> <div><div> C. Schoenborn</div><div> <i>Public health reports (1974)</i></div></div> <div>1986 Longitudinal 163 Citations</div>	The most likely to have unhealthy lifestyles are blacks, particularly black women.
<div></div> <div><b>Survey of Dietary and Exercise Habits and Perceived Barriers to Following a Healthy Lifestyle in a College Population</b></div>	Men exercised more frequently and at greater intensity than women.

# Literature review: Consensus

 beta

what do americans believe to be healthy diet and exercise habits?

☐ Synthesize

Since : 2015

Share This Page



The evidence shows that balanced diets and exercise could lead to healthy outcomes.

Published in Journal of National Black Nurses' Association : JNBNA | J. J. Hill et al. | 2018

The Healthy Eating and Exercise Lifestyle Program was an effective programme to achieve and sustain weight loss and increase exercise participation over 1 year.

Published in European Journal of Cardiovascular Nursing | Muaddi Alharbi et al. | 2016

Our review of the literature on determinants demonstrates that intentions, habits, self-regulatory skills, and the social and physical environment are the most important determinants of a healthy diet, which are in turn amenable to change by intervention strategies with varying levels of effectiveness.

 Systematic Review |  Highly Cited | Published in Psychology & Health | D. D. de Ridder et al. | 2017

Individuals' feasible opportunities for healthy diet and activity have clinically meaningful associations with intentions, achieved behaviors, and BMI.

Published in The Annals of Family Medicine | R. Ferrer et al. | 2016

# Help with data: metaprompt

can you give me a prompt? i'm interested in using NHANES to analyze perceptions and behaviors related to diet and exercise and how they correlate with body weight. i'd like to use the r package "NHANES" and use r to do some preliminary analysis. what sorts of details should I provide in my prompt?



# Help with data: suggested prompt

Using the NHANES package in R, I would like to analyze the relationship between perceptions and behaviors related to diet and exercise and how they correlate with body weight. Please focus on the following variables from the 2015-2016 and 2017-2018 survey cycles:

- Dietary variables: total calorie intake, macronutrient composition (% calories from carbohydrates, protein, and fat)
- Physical activity variables: minutes of moderate-to-vigorous physical activity per week
- Body weight variables: BMI, waist circumference
- Demographic variables: age, gender, race/ethnicity, socioeconomic status

# Help with data: output of suggested prompt

```
library(NHANES)
library(tidyverse)
library(rstatix)

# Load data from 2009-2010 and 2011-2012 cycles
data ← NHANES %>%
  filter(SurveyYr %in% c("2009_10", "2011_12"))

# Select relevant variables
selected_vars ← c("Age", "Gender", "Race1", "Education",
                  "Poverty", "BMI", "BMICatUnder20yrs", "Diabetes",
                  "PhysActiveDays", "AlcoholYear", "SmokeNow")

data ← data %>%
  select(all_of(selected_vars))

# Clean and preprocess data
data ← data %>%
  mutate(
    Gender = factor(Gender),
    Race1 = factor(Race1)
```

# Help with finding key variables: prompt

the list of variables in NHANES is below. Which ones do you think have to do with body weight, physical activity, and diet? [1] "ID" "SurveyYr" "Gender" "Age" "AgeDecade" [6] "AgeMonths" "Race1" "Race3" "Education" "MaritalStatus" [11] "HHIncome" "HHIncomeMid" "Poverty" "HomeRooms" "HomeOwn" [16] "Work" "Weight" "Length" "HeadCirc" "Height" [21] "BMI" "BMICatUnder20yrs" "BMI\_WHO" "Pulse" "BPSysAve" [26] "BPDiaAve" "BPSys1" "BPDia1" "BPSys2" "BPDia2" [31] "BPSys3" "BPDia3" "Testosterone" "DirectChol" "TotChol" [36] "UrineVol1" "UrineFlow1" "UrineVol2" "UrineFlow2" "Diabetes" [41] "DiabetesAge" "HealthGen" "DaysPhysHlthBad" "DaysMentHlthBad" "LittleInterest" [46] "Depressed" "nPregnancies" "nBabies" "Age1stBaby" "SleepHrsNight" [51] "SleepTrouble" "PhysActive" "PhysActiveDays" "TVHrsDay" "CompHrsDay" [56] "TVHrsDayChild" "CompHrsDayChild" "Alcohol12PlusYr" "AlcoholDay" "AlcoholYear" [61] "SmokeNow" "Smoke100" "Smoke100n" "SmokeAge" "Marijuana" [66] "AgeFirstMarij" "RegularMarij" "AgeRegMarij" "HardDrugs" "SexEver" [71] "SexAge" "SexNumPartnLife" "SexNumPartYear" "SameSex" "SexOrientation" [76] "PregnantNow"

- Note: I gave it the output of the following R code:

```
data(NHANES)
df ← NHANES
names(df)
```

# Help with finding key variables: output

Based on the list of variables provided, the following variables seem to be related to body weight, physical activity, and diet:

Body weight variables:

- Weight
- Height
- BMI
- BMICatUnder20yrs
- BMI\_WHO

Physical activity variables:

# Descriptive statistics: prompt

i've got some R code so far (at the very bottom of this prompt). can you please give me more code to do the following:

1. subset the data to remove anyone under age 18 or above age 75;
2. produce two separate summary statistics tables (one for numeric variables and one for categorical variables) of the following list of variables: weight, height, bmi, physactive, tvhrsday, comphrsday, alcoholyear, smokenow, smokeage, gender, age, race3, hhincome, education

R code so far: `# Load packages library(tidyverse) library(magrittr)  
library(NHANES) library(modelsummary) # load data data(NHANES) df <-  
NHANES names(df) %>% print`

# Descriptive statistics: output (after some finagling)

```
library(NHANES)
library(tidyverse)
library(modelsummary)

# Load data from 2009-2010 and 2011-2012 cycles
data <- NHANES %>%
  filter(SurveyYr %in% c("2009_10", "2011_12"))

# Select relevant variables
vars <- c("Age", "Gender", "Race1", "Education", "Poverty", "BMI",
          "Diabetes", "PhysActive", "SleepHrsNight", "Smoke100",
          "SmokeNow", "Marijuana", "TVHrsDay", "CompHrsDay", "AlcoholYear")
```

**Tip:** if the AI keeps giving you code that won't work, it can be helpful to upload the documentation of whatever package you're using (e.g. `modelsummary`) and then ask it to refer to that documentation when making the code

# Cleaning data

- You will probably need to iterate on the descriptive statistics
- You might find some issues with them, ask about how to clean the data, and then repeat the process
- I'm going to assume the data is clean and move on to regression analysis

# Regression analysis: prompt

can you write me code (in a codeblock) for how to regress BMI on PhysActive, Age, Race3, HHIncome, and Education? Please use the modelsummary() function to print the output to the console in markdown

Note: it already gave us this code (or similar code) in the previous prompt even though we didn't explicitly ask for it!



# Regression analysis: output

```
library(NHANES)
library(modelsummary)

# Load data from 2009-2010 and 2011-2012 cycles
data ← NHANES %>%
  filter(SurveyYr %in% c("2009_10", "2011_12"))

# Select relevant variables
vars ← c("Age", "Race3", "Education", "HHIncome", "BMI", "PhysActive")

data ← data %>%
  select(all_of(vars))

# Clean and preprocess data
data ← data %>%
  mutate(
    Race3 = factor(Race3),
    Education = factor(Education),
    PhysActive = factor(PhysActive)
  ) %>%
  filter(Age > 18 & Age < 75)
```

# Interpreting regression output: prompt

can you help me interpret the coefficient on PhysActiveYes in the following regression output?

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.57999    1.135622  22.525 < 2e-16 ***
PhysActiveYes  -1.200173    0.246766  -4.864 1.21e-06 ***
Age             0.039872    0.007983   4.995 6.23e-07 ***
Race3Black      4.763254    0.608907   7.823 7.08e-15 ***
Race3Hispanic   3.144666    0.670377   4.691 2.84e-06 ***
Race3Mexican    3.718929    0.662087   5.617 2.12e-08 ***
Race3White      2.836051    0.511532   5.544 3.21e-08 ***
Race3Other      3.894542    0.869472   4.479 7.77e-06 ***
HHIncome 5000-9999 -0.087116    1.078976  -0.081  0.9357
HHIncome10000-14999 -0.857914    0.920440  -0.932  0.3514
HHIncome15000-19999 -1.000000    0.910000  -1.100  0.2705
```

# Interpreting regression output: response

In the regression output, the coefficient on PhysActiveYes is -1.200173. This means that, holding all other variables constant, individuals who are physically active (PhysActive = "Yes") are expected to have a BMI that is 1.200173 units lower compared to those who are not physically active (PhysActive = "No", which is the reference category).

The negative sign of the coefficient indicates that being physically active is associated with a decrease in BMI, after controlling for age, race, household income, and education level.


The p-value for the PhysActiveYes coefficient is  $1.21\text{e-}06$ , which is much smaller than the conventional significance levels (0.05 or 0.01). This suggests that the

# Formatting equation in LaTeX

**TR** can you put the regression model in mathematical format, in latex, and return in a code block?

Here is the regression model in mathematical format, written in LaTeX:

latex

 Copy code

```
\begin{aligned}
\text{BMI} &= \beta_0 + \beta_1 \text{PhysActiveYes} + \beta_2 \text{Age} \\
&+ \beta_3 \text{Race3Black} + \beta_4 \text{Race3Hispanic} \\
&+ \beta_5 \text{Race3Mexican} + \beta_6 \text{Race3White} \\
&+ \beta_7 \text{Race3Other} \\
&+ \sum_{i=1}^{10} \gamma_i \text{HHIncome}_i \\
&+ \delta_1 \text{Education9-11thGrade} \\
&+ \delta_2 \text{EducationHighSchool} \\
&+ \delta_3 \text{EducationSomeCollege} \\
&+ \delta_4 \text{EducationCollegeGrad} \\
&+ \varepsilon
\end{aligned}
```

# Writeup: prompt

can you write up a three-paragraph analysis based on these regression results? I'd like you to consider the following questions:

- 1) is exercise negatively correlated with BMI?
- 2) is the effect practically significant (as opposed to statistically significant)?
- 3) does this model show that the relationship between physical activity and BMI is causal?
- 4) are there any other interesting correlations in the regression output that would be worth discussing?

# Writeup: output

Based on the regression results, there is a statistically significant negative association between physical activity and BMI. The coefficient for PhysActiveYes is -1.200173, indicating that, on average, individuals who engage in physical activity have a BMI that is 1.200173 units lower compared to those who do not, holding all other variables constant. The p-value for this coefficient is  $1.21 \times 10^{-6}$ , which is much smaller than the conventional significance levels of 0.05 or 0.01. This suggests that the relationship between physical activity and BMI is not likely to be due to chance alone.

While the relationship between physical activity and BMI is statistically significant, the practical significance of the effect should also be considered. A decrease of 1.200173 units in BMI may be meaningful at the population level.

# Asking for help finding motivational facts

**Me:** what sort of motivating facts could I put in the introduction of my paper that would set up why my analysis is important?

## Claude 3 Opus:

In the introduction of your paper, you can include several motivating facts to emphasize the importance of your analysis on the relationship between physical activity, BMI, and other demographic and socioeconomic factors. Here are some ideas:

1. Prevalence of obesity: Mention the high prevalence of obesity in the United States (or globally, depending on the context of your paper). You can cite

# Asking for help discussing policy conclusions

**Me:** what sort of policy conclusions do you think would follow from my research findings?

## **Claude 3 Opus:**

Based on your research findings, several policy conclusions and recommendations could be made to address the relationship between physical activity, BMI, and other demographic and socioeconomic factors:

1. Promote physical activity: Given the significant negative association between physical activity and BMI, policymakers should prioritize initiatives that encourage and facilitate increased physical activity among the



# Suggestions for title and abstract

**Me:** could you suggest for me a title and abstract of this report? please make sure the abstract is less than 150 words long

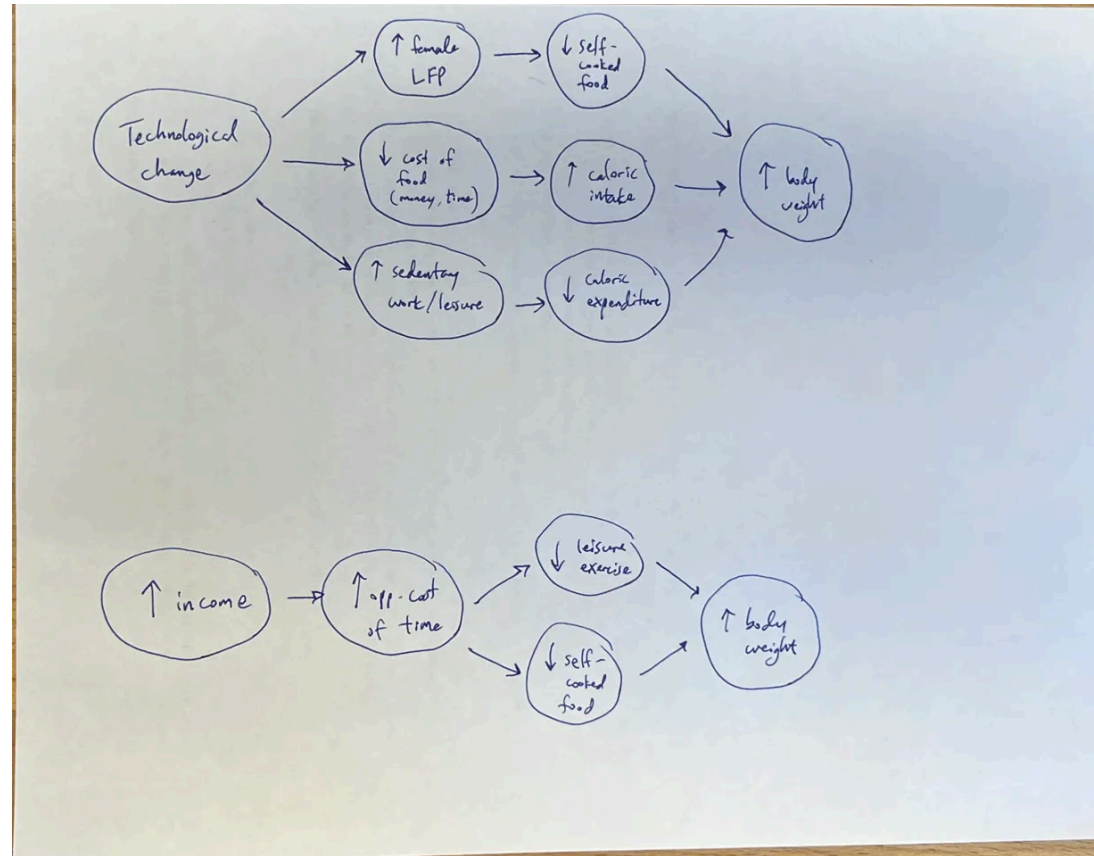
**Claude 3 Opus:**

Title: The Association Between Physical Activity, BMI, and Demographic and Socioeconomic Factors: An Analysis of NHANES Data

Abstract: This study investigates the relationship between physical activity, Body Mass Index (BMI), and various demographic and socioeconomic factors using data from the National Health and Nutrition Examination Survey (NHANES) 2009-2012. Linear regression analysis reveals a statistically

Note: actually 146 words (LLMs aren't good at word-counting)

# Killer feature: LaTeX from a picture (1/2)



becomes ...

# Killer feature: LaTeX from a picture (2/2)

```
\begin{frame}  
\frametitle{Diagram of economic pathways to obesity}  
\resizebox{\textwidth}{!}{%  
\begin{tikzpicture}[node distance=2cm and 5cm, on grid]  
  \node[draw, rectangle] (tech) {Tech.\,change};  
  \node[draw, rectangle, above right=of tech] (lfp) {$\uparrow$ women LFP};  
  \node[draw, rectangle, right=of tech] (OC) {$\uparrow$ (money/time) cost of food};  
  \node[draw, rectangle, right=of OC] (calin) {$\downarrow$ calorie intake};  
  \node[draw, rectangle, below=of tech, yshift=-4cm] (income) {$\uparrow$ income};  
  \node[draw, rectangle, below right=of tech] (sedentary) {$\uparrow$ sedentary work/leisure};  
  \node[draw, rectangle, right=of sedentary] (calout) {$\downarrow$ calorie expenditure};  
  \node[draw, rectangle, right=of income] (opp) {$\uparrow$ opp cost of time};  
  \node[draw, rectangle, above right=of opp] (leisure) {$\downarrow$ leisure exercise};  
  \node[draw, rectangle, below right=of opp] (self) {$\downarrow$ self-prep food};  
  \node[draw, rectangle, right=of lfp] (self2) {$\downarrow$ self-prep food};  
  \node[draw, rectangle, right=of calin, yshift=-3cm] (weight) {$\uparrow$ body weight};  
  
  \draw[→] (tech) to[out=45,in=180] (lfp);  
  \draw[→] (tech) to[out=0,in=180] (OC);  
  \draw[→] (tech) to[out=270,in=90] (income);
```

# Killer feature: BibTeX entry from listed reference

**TR** please make a bibtex entry out of the following information. put as the bibtex key the name of the first author (in lower case) followed by "\_aYYYY" where YYYY is the year of publication. use full names for all authors. include the doi as a field but not the url. remove any digits from the authors' names. don't include the month field or the season field (like summer). use an endash ("--") to separate page ranges. capitalize the title according to chicago style. return the result in a code block.

====

Winston, Gordon, C. 1999. "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education." *Journal of Economic Perspectives*, 13 (1): 13-36\*\*\*DOI: 10.1257/jep.13.1.13

====

Here is the bibtex entry in a code block:

```
bibtex Copy code

@article{winston_a1999,
  author = {Gordon C. Winston},
  title = {Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education},
  journal = {Journal of Economic Perspectives},
  year = {1999},
  volume = {13},
  number = {1},
  pages = {13--36},
  doi = {10.1257/jep.13.1.13}
}
```

Also works well for garbled metadata

# Killer feature: Summarizing multiple PDFs at once

Courtemanche\_al\_2015\_Econ\_Journal\_WPversion

PDF

Cutler\_al\_2003\_JEP

PDF

Lakdawalla\_al\_2005\_AER-P&P

PDF

Phillipson\_Posner\_2003\_Per spBioMed

PDF

**TR** please create an additional slide called "existing economic theories of obesity" that summarizes the ideas of these four documents that i've attached. again, please keep each \item in the tex code to 15-20 tokens. you do not need to recreate the entire latex file, just give me the latex code in a code block for the slide i've requested

Here is the LaTeX code for the "Existing economic theories of obesity" slide:

latex

Copy code

```
\begin{frame}
\frametitle{Existing economic theories of obesity}
\begin{itemize}
\item Cutler et al. (2003): \\
-- Technological change reduced food prices and preparation time \\
-- Increased calorie intake is the primary driver of obesity
\item Phillipson & Posner (2003): \\
-- Technological change has made both work and leisure sedentary \\
-- Lowered cost of calories and raised cost of physical activity
\item Chou et al. (2004): \\
-- Lower food prices and increased labor force participation of women \\
-- Demand for restaurants and calorie-dense foods rose
\item Lakdawalla et al. (2005): \\
-- Technological change lowered food production costs \\
-- Made agricultural production more efficient and work more sedentary
\end{itemize}
```

# Killer feature: Refactoring code (GPT-4) (1/2)

**Me:** what ideas do you have to abstract this code? there are some ways in which it's not "loop-able" but ways in which it indeed is:

```
#-----  
# load and clean exercise data from "What We Eat in America" surveys  
#   (source: NHANES, via USDA website)  
#-----  
  
# 2001-2002  
  
# Table 1. Nutrient Intakes, One Day, 2001-2002  
fpath ← "../..../data/wweia/raw/Table_1_BIA_GEN_01.pdf"  
dc01 ← pdfcleaner(fpath,1)  
for (i in 2:7){  
  dc01 ← left_join(dc01, pdfcleaner(fpath,i), by = c("id"))  
}  
dc01 %>% mutate(year = 2001) %>%  
  select(-starts_with("Range"))%>%  
  filter(!is.na(year))
```

# Killer feature: Refactoring code (GPT-4) (2/2)

## GPT-4:

To abstract and streamline the provided R code, you can leverage functions and loops to handle repetitive tasks efficiently. This approach will make your code more compact, flexible, and easier to maintain, especially when dealing with similar data processing tasks across different datasets. Here's a structured approach:

1. Define a Function for Data Processing: Create a function that encapsulates the repetitive steps of loading, cleaning, and preparing the data. This function should accept parameters for the specific parts that change with each dataset, such as the file path, the year, and the range of pages to

# So far, I've used LLMs to help me ...

- Fill out bureaucratic forms
- Write code that automates grading
- Write code that systematizes data analysis (for ease of replication)
- Write code to create data visualizations
- Write unit tests of code
- Prepare discussion slides for a conference
- Prepare peer review reports
- Prepare this slide deck
- Reduce word count of an abstract
- Improve sentence clarity in a paper
- Write survey questions that a survey methodologist would approve of
- Explain poorly written abstracts / papers in simpler terms
- Invert mathematical functions
- ... not to mention a bunch of stuff in my personal life



# Staying on top of new developments

The following sources are helpful for keeping on top of new developments:

- `One Useful Thing` Substack by Ethan Mollick
- `Marginal Revolution` blog by Tyler Cowen & Alex Tabarrok