Homework 1

Upload your assignment as a .pdf file on Canvas.

Question 1: AI-Driven Webpage Launch in 30 Minutes (25 points)

Brief summary: Create a creative exercise to discover different GenAI tools.

***Overview**:*

Inspired by Ethan Mollick's blog post on the immense capabilities of AI in marketing (please read the blog post here!), this exercise challenges you to select a favorite project idea for the course and create a landing page from scratch to advertise it. You will be in the driver's seat, exploring and leveraging various AI tools to enhance your project's visibility and appeal. This exercise demonstrates that although AI is superpowerful in generating content, you decide what to build and how. You are also the guide to the model. If this exercise is challenging for you, that's good news! It demonstrates that AI alone is insufficient and that many skills are needed to master a successful AI-human interaction!

***Objectives**:*

- ● *Create a Landing Page: Use ChatGPT to assist in generating content for a simple yet compelling landing page.*
- ● *Deploy Your Page: Make your landing page accessible online, allowing everyone to discover your idea.*
- ● *Incorporate AI-Generated Content: Enhance your page with AI-generated video, soundtrack, and image.*
- ● *Experiment and Reflect: After the initial 30-minute creation phase, spend an additional 30 minutes (or more) to refine your content. Reflect on the tools used, the process, and any challenges faced.*

***Questions**:*

1. (2 points) Choose a project idea you are interested in pushing further in this exercise. Create a picture to illustrate your project using an image generator: e.g., OpenAI DALL-E embedded in the Plus version, Midjourney (10$/month), Recraft.ai (free trial), Gemini, Bing Image generator. Comment on the process you used to get an image that is satisfying to you. If the first attempt gives you something you like, that's also fine, but keep exploring the generative tool(s) a bit to see how your prompt or feedback to the model modifies the output. Include the picture you generated in your .pdf handout.

Prompt 1: "Create a comical image of generative ai companies openai, perplexity, claude, and gemini battling each other"

- Did not like the images, too comical/kid-toon like

Prompt 2: “Instead of comical, make it scifi”

- Did not like images, too fantasy like for scifi

Prompt 3: “retry”

- Still too fantasy like, also the names were too jumbled

Prompt 4: Create an image of four futuristic robots battling each other. Label each robot as "OpenAI", "Perplexity", "Gemini", "Claude". Put them in a desolate environment

- Like the vibe, but it generated an image with more than four robots.

Prompt 5: There are too many robots in the image, I only want 4

- Approve of image. Used for Runway.ML

2. *(3 points) You will now generate a video to illustrate the project as well. You can use whatever video generator you like: e.g., Runway.ML (free trial), Pika (free), D-ID (free trial). Feel free to use a picture to kickstart the generation or start with a prompt. This video will appear on the website you create; there is no need to upload it on Canvas.*

3. *(10 points) Please create a webpage to display this content. You will create two .html files: one for the landing page named index.html and one for the additional content called content.html. The good news is that ChatGPT can do all of that for you directly with a high success rate! You won’t probably need to code a single line yourself. ChatGPT can implement most of your simple wishes (e.g., adding a button, embedding a video, an image, or music). With the right prompt, ChatGPT can code the two pages at once directly and accurately! I let you explore and don’t provide further guidance such that*

   you develop your own intuition and trial and error with the model.
   a. Content Requirement and mark breakdown: Ensure the landing page has an

automatically playing AI-generated video (5 points) and a button leading to a second page with more AI-generated content that is different than text (e.g., an image) (5 points).

4. *(5 points) Host your landing page on a platform like GitHub for full points, or use website builders like Wixsite, Google Sites, or Wordpress with a lower score (-3 if this is not based on your own .html code). I highly suggest to use GitHub pages (instructions here). Just create a repository, then choose to upload your content. Provide the public link to your repository so the teaching team can visit it, and provide two screenshots of your two web pages in the pdf handout.*
5. *(5 points) Enhance your webpage by making it more aesthetic and entertaining by asking ChatGPT to add elements or transformations of your choice. There is no need for anything fancy, just two minor improvements are enough for the full score. Share your learning experience, including the tools used, what was effective, and any obstacles encountered.*

Embrace this opportunity to harness the power of AI in bringing your project ideas to life. Good luck!

For inspiration and a deeper understanding, refer to Ethan Mollick's blog post here and explore my example project here I did this week in 30 minutes. Notice that I had never done this experience myself before, and I am not fluent in HTML. It is fine if you take more than 30 minutes to do it as you build your experience interacting with LLMs.

Website: https://sdyuuw.github.io/

Question 2: Investigating search engines and RAG models (25 points)

Brief theme: Discover more about the interaction of search engines with GenAI

1. (5 points) Explore the answer of Perplexity.ai, Gemini Ultra, GPT4 equipped with RAG + Google search engine (or your favorite traditional search engine) when you ask the question: *What is retrieval augmented generation?*

How do the answers differ? Which one did you prefer? Which one was the most informative? How did you behave with the different answers? Did you explore something more, stop there? Describe your natural interactions.

- GPT provides a simple text response to the question outlining what RAG systems are and some generic use cases.
- Perplexity provides a text response as well, but slightly more in depth. A differentiator is that they provides sources from which they extract information to generate the response.
- Gemini provides a further in depth textual response than GPT, and similar to perplexity. However Gemini does not provide sources from which they generated the response, unless you select "Double-Check Response" at which point Gemini scans Google's search index for similar answers.
- My current toolbox consists of using Perplexity and Gemini, I like Perplexity as they provide sources from which I can explore to revise/validate my answers. As a research perplexity has a "Academic" feature which allows for querying papers which I enjoy.
- For AI-assisted coding, I prefer Gemini as I've had better answers compared to other platforms mentioned.

2. (5 marks) Do the same exercise with the following prompts:

- *who won superball 2024*
- *examples of biomimicry*
- *log(1996)*6/4+2023^9*exp(12) = ?*
- *what is the computing power behind one exaflop*
- *what is framingham heart study*
- *origin of valentine day*

How do the answers differ? Which ones look trustworthy? Be specific in your comments.

Prompts	Perplexity	Gemini	GPT4	Claude
Who won superball 2024	Able to transcribe to Super Bowl 58 in 2024 and provide succinct answer with final score include. Also, adds additional blurb about historical highlights.	Able to transcribe to Super Bowl 2024 and answer the question. If prompted further, generates highlights as well as final score.	Able to transcribe to Super Bowl 2024 and provides succinct answer with final scores included.	Able to transcribe to Super Bowl 2024, but unable to answer. Knowledge base not up to date.
Examples of biomimicry	Text list, with sources annotated as reference.	Text list, but source and image embeddings provided for further discovery.	Simple text list of examples of biomimicry	Simple text list of examples of biomimicry
log(1996)*6/4+ 2023^9*exp(12) = ?	~ 1.013 x 10^39 Miscomputes 2023^9 to 1.117 * 10^33. Even when using wolframalpha integration…	5674961831865 5131167125571 9463e12+log10 (795209593621 ^1/2) Provides step by step math solution as well as most technical, exact result	9.24 x 10^34 Deploys math package in python to deliver result. Correct!	2.24611*10^40 Interprets log as ln rather than base 10, thus miscomputes. Also, miscomputes carat (^)
what is the computing power behind one exaflop	1 quintillion	1 quintillion	1 quintillion	1 quintillion
what is framingham heart study	Generates text based on three sources: Wikipedia, NIH, and Boston Medical Center	Generates well-outlined summary of the study, in systematic, markdown formatting	Generates adequate response outlining the study as well as key contributions of the study.	Concise response outlining key points of the study.

Prompts	Perplexity	Gemini	GPT4	Claude
origin of valentine day	Wall of text summarizing the origins and progression of valentine's day. Sources somewhat credible: NPR, NYTimes, wikipedia	Breaks down three main theories starting with Romans, to Christians, to modern romance connotation. Primary sources from history.com	Wall of text albeit listed by numbers to indicate significant players in the origin story.	Concise summary indented and spaced between significant timeframes: Rome, Christians, Summary.

For trustworthiness, my preference still lies with using a mix of perplexity and gemini. Again, due to the inclusion of sources, I feel more confident in using the responses at face-value. We saw perplexity fail the math challenge due to misinterpretation of the carat (^) character. While I have yet to rely on Perplexity for maths, this is quite surprising to me as I expected their integration with wolfram alpha to provide the correct result.

3. (5 marks) First, interpret the meme in the image below. Then feed this image to the different LLM models (GPT4+Gemini Ultra, and perplexity if you happen to subscribe to it) and ask them: "explain this meme"
Comment on their answers. Did you interpret the meme similarly?

Fun fact: the picture behind the meme has been sold for $500,000 (source).

My Interpretation	My interpretation of the meme alludes to the overhyped idea that chatgpt/genAI will replace intrinsic and empirical knowledge that humans have gathered over the last 5000 years.
Perplexity Pro	Able to process image and interpret text. In line with my interpretation, "meme is a humorous take on the idea that ChatGPT, as an AI language model, encapsulates a vast amount of human knowledge and expertise, which has been accumulated over thousands of years…"
Gemini Advanced	Unable to process images…
GPT4	Able to process image and interpret text. Answer seems have to better contextual knowledge about the meme itself. Explains the juxtaposition between the calm girl and chaos ensuing in the background.

4. (10 points) You're a product manager in a big tech company. You discover that your team has never heard of RAG systems or tried RAG LLM tools. Write a short memo explaining this technology and how your team can use the tools with recommendations and advice on the best search engine depending on the use case. A careful memo is important as it may increase or decrease their productivity and quality of outputs for your team! You can include pictures in your memo. Using LLMs is allowed to draft the memo but document the process in that case in a side paragraph.

I'm sure many if not all of you have heard and seen the hype around OpenAI's chatGPT. Perhaps some of you have played around with the technology and have incorporated into your workstreams already. Perhaps some of you have noticed, however, when asking it more and more granular questions it starts to make up things or "hallucinate" answers that sounds good, but contextually are quite awful.

RAG systems may be to blame, or rather the insufficient knowledge base of these RAG systems that resulted in your hallucinated answers. RAG means "retrieval augmented generation" and is quite explanatory in its naming. The system essentially **retrieves** contextual information either automatically from public online sources or manually via file or dataset upload to **augment** the **generated** answer to your question. Now the reason why GPT may break down with granular questions is because the RAG system at OpenAI may not have adequate contextual knowledge.

What does it mean for us?

We intend to implement chatGPT enterprise across the company and expect beta testing in the next quarter, with the intention of releasing it to the entire company by Q1 2025. As I have mentioned, we must ensure that the RAG system is well populated with our internal knowledge and database while being mindful of cybersecurity risks. We encourage those in beta testing to upload data and files while using the enterprise version to extract insights and aide in productivity. However, we ask that all GPT generated content be earmarked where possible.

I believe this is a truly remarkable evolutionary step in the way we will work and learn. But I ask that you suppress the hype and evaluate this tool thoroughly, just as you would anything else. We at Company X have always been front-runners in adopting new technology. We have earned that capability thanks to subject-matter experts such as yourselves. I look forward to seeing what we can do!

Question 3: Investigating the Jagged Frontier of GenAI (50 points)

Brief theme: Discover more about the limitations and strengths of GenAI and the human-AI interaction

1. (5 points) **The bird test**

Type this prompt in GPT3.5, GPT4, Perplexity, Gemini Ultra, Claude:
"give me 10 sentences that end with the word bird" What do you notice? Explain the behavior observed.

Perplexity	Gemini	GPT4	Claude
9/10 sentences end with the word "bird." Generated responses are like poems, very descriptive.	9/10 sentences end with the word "bird." Poetic like perplexity, but generated responses are different!	10/10 sentences end with the word "bird." Poetic but concise.	4/10 sentences end with the word "bird." Mix of poetic and literal answers

2. (20 points) ***ChatGPT4 is your personal data scientist!***
ChatGPT4 can perform standard data analytics and machine learning tasks autonomously from a csv or excel dataset you can upload.
Let's see how to leverage this with a real-world case study!

The World Health Organization estimates that 12 million deaths occur worldwide every year due to heart disease. In developed countries like the United States, half of all deaths are due to cardiovascular diseases. Early prognosis of cardiovascular diseases can aid in making decisions regarding lifestyle changes for high-risk patients, thereby reducing the likelihood of complications. This research aims to identify the most relevant risk factors for heart disease and predict overall risk using logistic regression.

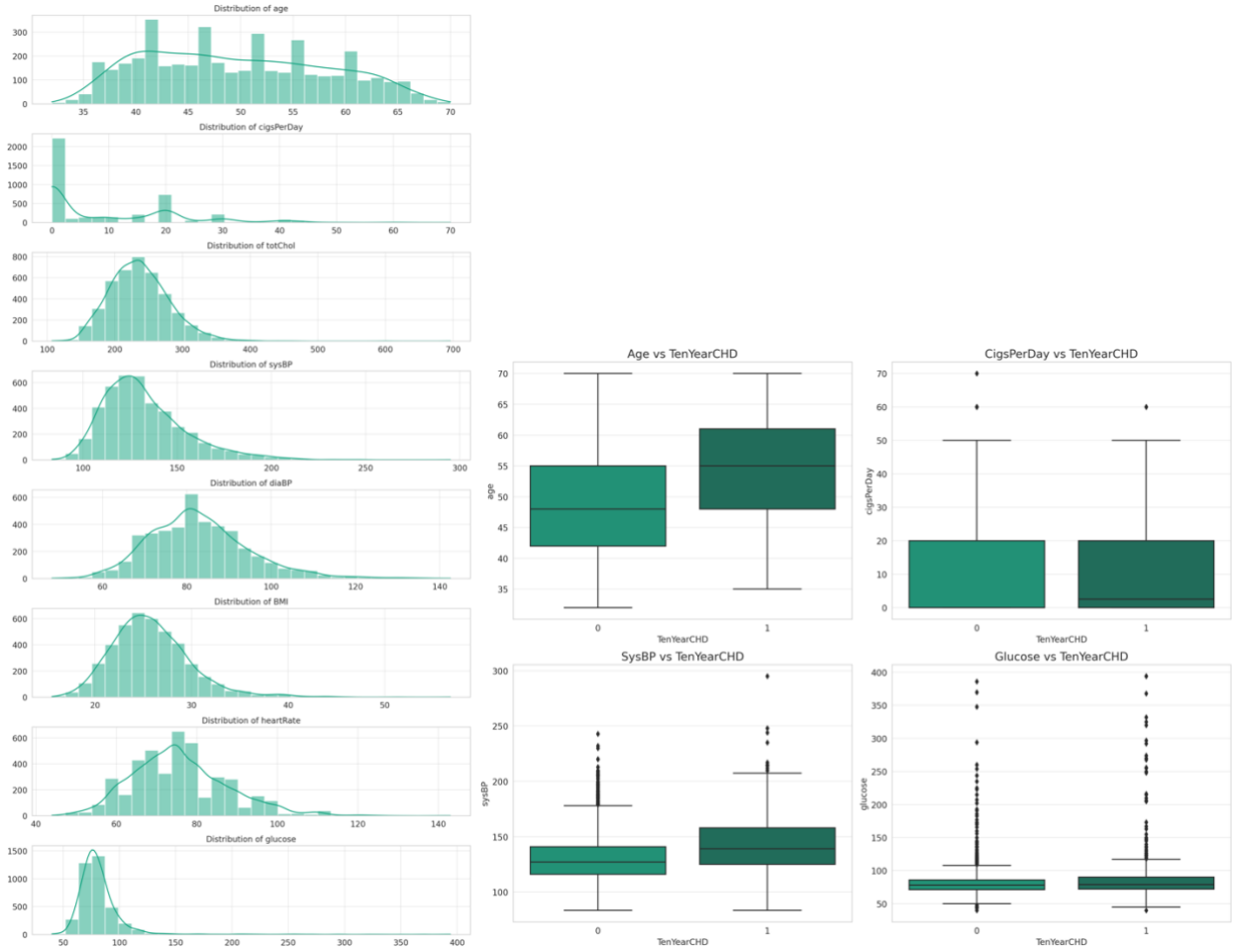
The dataset we are using is publicly available, and it is about an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. This study has been instrumental in advancing research in the field and is extremely famous.

Our goal with the data is to predict whether a given patient has a 10-year risk of future coronary heart disease (CHD). The dataset provides the patients' information. It includes over 4,000 records and 15 attributes.
Your data scientist talents have been spotted by doctors from UW Medicine, and they urgently need your help to get insights about this dataset for a report due tomorrow. Fortunately, you recently learned about the strong capabilities of ChatGPT4 to perform data analysis. You are confident you can provide the insights they want just by prompting and not typing a single line of code yourself! (hush, they don't need to know this). Note that this dataset is public and widely available, so no problem with uploading it to ChatGPT (in general, never upload any sensitive or proprietary data to ChatGPT!).

a. (5 marks) Open a new chat in ChatGPT4, and upload the data + type this prompt: "do a thorough exploratory data analysis"
Be a cyborg and keep the conversation going. Write a summary of the main insights, including figures.

GPT4 was able to deploy pandas to query df.head(), df.info(), and df.describe() as one would normally would to obtain high-level overview of the dataset including dtypes, null counts, mean, std, etc.

It then deployed pyplot and seaborn for distribution analysis of numerical variables as well as correlation analysis in relation to outcome variable "TenYearCHD" via box-whisker plots.



Upon asking, “What other insights can you derive from this?”, it dives deeper into analysis of categorical variables education, currentSmoker, BPMeds, prevalentStroke, prevalentHyp, diabetes. It also provides potential outlier analysis of numerical variables as well.

b. (5 marks) Prompt the model to create a classification model to predict `TenYearCHD` *Comment on the output of the model.*

During data pre-processing, GPT4 decided to impute missing numerical values to its respective median, while dropping rows with null catagoricals. It split the data into 80 train:20 test which is common and used all 15 features for prediction. I suspect using all 15 features will cause collinearity issues with the model and result in poor performance.

For the logistic regression the overall metrics are as follows:

Accuracy: 85.5%, Precision: 50.0%, Recall: 6.5%, ROC-AUC Score: 70.9%

Decent accuracy, but abysmal precision and recall. Meaning only 6.5% of the patients would be properly identified as having CHD with this model and of those 6.5% the model would be correct only half the time. Unacceptable in healthcare settings in my opinion.

c. (5 marks) Keep interacting as a Cyborg to improve the performance, discover insights, and understand the significant features contributing to the prediction. Describe your interaction. Are you satisfied with how the conversation went? Did you manage to improve the performance?

Prompt: "Improve the model by feature engineering. Remove features with high covariance or collinearity."

Accuracy, Precision, ROC-AUC scores improve slightly. No improvements to Recall.

Prompt: "Improve the model by addressing the imbalance of outcome variable TenYearCHD"

As expected, addressing the imbalanced outcome variable increase Recall = 61.8%, however it dropped Accuracy to 66%, and Precision to 24%

Prompt: "Use DecisionTreeClassifier, SVC, and KNearestNeighbors to improve the model. Summarize all models into table format"

Model	Accuracy	Precision	Recall	ROC-AUC Score
Logistic Regression	66.0%	24.0%	61.8%	71.0%
Decision Tree	78.5%	24.3%	22.8%	55.4%
SVC	68.9%	24.5%	55.3%	67.0%
K-Nearest Neighbors	83.5%	27.0%	8.1%	58.9%

d. (5 marks) You are tasked to report to the doctors about the insights you discovered in the data. Write an executive summary to explain the quality of the data, the best model you could build (report the metrics like accuracy, AUC, precision, recall), and the significant factors to predict the 10-year risk of CHD. You are of course welcome to create some additional code yourself.

Using chatGPT alone, I was able to refine the logistic regression model such that Recall = 61.8%, Accuracy = 66.6%, and Precision = 24.0%. This indicates we are able to predict 61.8% of true positives of 10-year risk of CHD based on the features provided. While the model is not all telling, we believe it can be implemented by clinicians to assign high risk 10-year CHD patients and begin preventative measures sooner than later. This includes improving cardiovascular health's of patients via treatments such as regimented exercise and diet to mitigate diabetes, high BMI, and cigarette smoking patients.

3. The Centaur vs the Cyborg (25 points)

This exercise may be out of your comfort zone and technical skills, yet you have to provide an answer, if possible accurate! Let's see if it is within the jagged frontier of AI.

The state of Washington knows the legendary talent of the students of the MSIS program, and the COVID task force decided to call for your help. In case of a future pandemic, the state wants to be ready and plan for the deployment of mass vaccination clinics in the Seattle area.

Given available resources, you can build only 4 sites across the Seattle area. You have identified 10 potential sites, each with a fixed capacity (i.e., a maximum number of residents that can be vaccinated there). All residents will need to access one of the clinics you will construct. You aim to determine which sites to build to optimize the vaccination service provided to the population.

The area under consideration can be modeled as a 5 x 10 miles rectangle, with 50 major residential areas. We index the residential areas by i = 1... , 50 and the vaccination sites by j = 1,..., 10. We use the so-called "Manhattan distance": if area i's coordinates are (x_i, y_i) and site j's coordinates are in (x_j, y_j), the distance from i to j is given by $d_{ij} = |x_i - x_j| + |y_i - y_j|$.

You are provided with the following data:

Parameters:

- r_i : number of residents in area i =1,... , 50;

- C_j: capacity of vaccination site j =1, ..., 10;

- d_{ij}: distance between area i = 1, ..., 50 and site j = 1,... , 10.

Formulate an integer optimization model that optimizes the selection of sites to minimize the total distance across all residents required to access their assigned clinic.

1. *(3 points) Use the centaur method, i.e., copy-paste the problem formulation directly into your favorite LLM. Report your answer. How confident are you it is correct? Explain why. Use only one LLM of your choice, and do not use any other tool. Explain why you chose this LLM, among others. Note: you will probably realize that the math won't appear with subscripts if you copy-paste the text from the assignment pdf. Is it a problem?*

Using GPT4:

The integer optimization model can be formulated as follows:

Objective Function:
Minimize $Z = \sum_{i=1}^{50} \sum_{j=1}^{10} r_i \cdot d_{ij} \cdot y_{ij}$

Where r = number of residents, d = "Manhattan distance", y = binary decision variable

I feel comfortable with the answer generated by GPT4.0, as a starting point for developing my model. It does a great job of incorporating binary decision variables, as well as the constraints provided by the parameters.

2. *(3 points) Start a new chat and use the cyborg method to provide an answer again, i.e., interact with the model, think step-by-step, and break down the task into sub-parts to solve the previous problem again. What is your final answer? How confident are you now? Use the same LLM, and do not use any other tool.*

Prompt: How to deploy integer optimization in python?
Answer highlight: Use PuLP package along with example outlining problem variable, decision variable, objective function, and constraints.

Step by step copy-paste information from assignment pdf and assign them to the variables described in PuLP example. Returns essentially the same objective function as before:

The objective is to minimize the total distance across all residents required to access their assigned clinic. This can be expressed as:
Minimize $\sum_{i=1}^{50} \sum_{j=1}^{10} r_i \cdot d_{ij} \cdot x_{ij}$

Where r = number of residents, d = distance, x = binary decision variable

3. *(5 points) You can now use as many tools as you want and interact with several LLMs. You are free to explore whatever you want. What is your final answer? Describe your interaction with the tools and your method. How much money would you be willing to bet that your latest answer is correct?*

Ideally, I would like to gather the data for number of residents and clinic capacity and develop the model. I am comfortable implementing this in python or using excel Solver. We explored integer optimization problems in Professor Hillier's MSIS 504 class last year. I am willing to bet the answer is correct as it considers mixed-integer programming optimization

d. (10 points) Comment on your experience with this question and relate it to the jagged frontier. Read this paper https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321 or this blog post https://www.oneusefulthing.org/p/centaurs-and-cyborgs-on-the-jagged before answering. You have to use ChatGPT to draft your answer, but the answer should reflect YOUR opinions and

experience. Write the best answer possible to look informative, compelling, and exciting for me to read. I am expecting 2-5 paragraphs. Describe in particular how your trust in your final answer evolved across a,b, and c and how you proceeded to increase it.

Prompt: Upload the SSRN paper as pdf into GPT. “Summarize this paper for me”
Response: “…The study involved 758 consultants and introduced them to tasks with and without AI support, also offering a prompt engineering overview to some. It found that AI significantly increased productivity and quality within its capabilities but decreased performance outside its capabilities. The study also identified two successful human-AI collaboration strategies: "Centaurs," who strategically divide tasks between humans and AI, and "Cyborgs," who integrate AI into their workflow more seamlessly.

The paper essentially outlines the “Cyborg” vs “Centaur” idea of utilizing AI tools, where “Cyborgs” fully adopt AI as tool for essentially every task, whereas “Centaurs” are a bit cautious about the “jagged frontier” of AI. Centaurs are quick to strategize what tasks to give to AI while keeping some tasks for themselves as humans, playing to their respective strengths.

Currently, I am aligned with the “Centaur” strategy in using AI tools. I use it on the premise of mapping problem domains, gathering methods, and refining content. Especially in the realm of being a student, the Centaur strategy provides high-level, conceptual knowledge and suggestions on how to solve a ML or data science problem. I believe I owe it to myself to then execute the sub-task, such as programming and understanding of statistics, to develop a solution that I know I am 100% confident about.

e. (4 points) Comment and describe how you drafted this previous commentary in d. Do NOT use ChatGPT to comment on how you used ChatGPT in d.

- Read the paper manually, no assistance.
- Download the paper in pdf format. Upload to GPT4. Prompt it to summarize the paper.
- Additional prompt to “dive deeper into the paper”
- Extract highlights that pertain to my strategy in using AI tools
- Relate the content of the paper to this homework assignment, and why I prefer to use “Centaur” strategy than Cyborg as of this moment.

Note that we won’t grade whether your optimization problem formulations are correct. We are grading your capacity to reflect on the tools' use, limitations, and creativity to overcome challenges. However, as an incentive to get a correct answer, we will add 2 bonus points if your final answer in c is correct.