

# Using AI to Upskill Non-Technical Workers into Data Science: A Field Experiment

Emma Wiles      Mohamed Abbad      Francois Candelon      Daniel Sack  
MIT                      BCG                      BCG                      BCG

Pamela Mishkin      Lisa Kraye      Urvi Awasthi      Ryan Kennedy  
OpenAI                      BCG                      BCG                      BCG

Cristian Arnolds  
BCG

April 24, 2024

## Abstract

As firms integrate Large Language Models (LLMs) into their operations, concerns arise about their potential to automate tasks traditionally performed by workers. To remain employable, workers must adapt their skills to meet the changing demands of the labor market. While LLMs pose a risk to certain jobs, they also offer a means to “upskill” workers, by training workers in whatever new skills the market demands. In this study we investigate how LLMs can be used to help non data scientists upskill into data science roles. We run a randomized control trial on knowledge workers outside of data science to test whether LLMs are better at assisting workers in gaining technical skills than previously available tools. We give consultants at BCG data science problems, representative of what the data scientist role at the company demands, but which GPT-4 cannot solve on it’s own. We find that treated non-technical workers given access to and training in using ChatGPT are more likely to correctly solve all three tasks. After the experiment, treated workers are also more confident in their ability to contribute to data science projects in their jobs. These results suggest that workers can effectively utilize LLMs to update their skills to meet the evolving, more technical demands of the labor market.

## 1 Introduction

The rapid advances in Generative Artificial Intelligence (GenAI) and its widespread deployment has sparked both excitement and concern about its potential impact on the workforce. These models increasing capabilities of automating complex tasks and knowledge work

raises concerns about job displacement and the need for workers to adapt to the changing demands of the labor market. Reskilling or upskilling workers into new skills which complement emerging technology are key strategies for mitigating the negative effects of automation (Acemoglu and Restrepo, 2018; Djankov and Saliola, 2018). While recent studies have explored the effects of GenAI on worker performance in tasks within their existing skill set (Brynjolfsson et al., 2023; Dell’Acqua et al., 2023) there remains a gap in understanding how GenAI can be used to help workers acquire new skills and adapt to changing job demands. In this paper, we investigate the potential of GenAI as a tool for upskilling non-technical workers in the domain of data science.

Through a randomized experiment at a large management consulting firm, we demonstrate that providing non-technical workers with access to and training in GPT-4 can significantly improve their productivity and accuracy on data science tasks, even for problems that the AI alone cannot solve. We measure treated and control group workers on the accuracy of their output on a series of data science tasks. And we also compare their output to a benchmark set by the performance of data scientists at the firm for whom these tasks are a regular part of their job. This allows us to directly test how much non-technical workers armed with GenAI can do relative to workers in data science roles at their firm.

There is evidence from the literature that on-the-job training can help benefit both workers and firms. Acemoglu and Pischke (1998) argue that even in imperfect labor markets, employers have incentives to provide general skills training to their workers, as they can benefit from the resulting higher productivity. Heckman et al. (1998) shows that policies promoting skill formation, including on-the-job training, can significantly impact workers’ earnings and skill development throughout their careers. This suggests that as AI automation advances, firms in affected industries may benefit by increasing their training efforts to help workers adapt to new skills or even new roles.

However, in a landscape where the demands for skills are changing this rapidly, workers and firms may be weary about learning a new skill that will become obsolete as soon as the next generation of some language model is released. If GenAI is a general purpose technology (Eloundou et al., 2023) which can be used to allow workers to flexibly solve new types of problems as they emerge. There is evidence that GenAI is an effective and patient teacher (Mollick and Mollick, 2022). We therefore believe it is worth investigating whether GenAI itself can be used as a tool for on-the-job training in learning how to gain new skills.

We conducted a randomized controlled trial involving almost a thousand associates and consultants at the Boston Consulting Group (BCG), a large management consulting firm. Participants were randomly assigned to either a treatment group, which received access to and training in GPT-4, or a control group, which received training on using Stack Overflow

and other resources commonly used by data scientists. We surveyed workers both before and after the experiment. The experiment consisted a 20 minute interactive training session tailored to each treatment group followed by a series of data science tasks designed to be more technical than the participants' current roles. The tasks, developed in collaboration with OpenAI, were specifically designed to be challenging for GPT-4 to solve independently, requiring human input and reasoning. Each participant was randomly assigned two out of the three tasks, which included a statistics and machine learning task, prediction task, and a coding task. After grading the tasks, we also compare the performance of the participants was compared to a benchmark set by BCG data scientists who completed the same tasks without the use of ChatGPT.

We contribute to the new literature on the effects of Generative AI on worker productivity. In one study, customer service agents given access to LLM suggestions are able to resolve more customer complaints more efficiently (Brynjolfsson et al., 2023). In another example, consultants using GenAI on tasks that was in the model's range of ability were 25% faster at completing tasks , completed 12% more tasks and produced 40% higher quality output on average compared to their counterparts that did not use GenAI (Dell'Acqua et al., 2023). In this case, the tasks were within the skillset of the users (comparing interviews to excel data, and writing a Harvard Business Review style article), but they were able to use GenAI to produce output faster and on average of higher quality. However, on a task that the model reliably got the wrong answer on, the treated group was 25 percentage points less likely to come to the correct answer themselves. Our findings suggest that GenAI can improve the productivity of non-technical workers on complex tasks outside of their skillset, even for problems GenAI can not solve on its own.

Second, we contribute to a literature on job training and upskilling. Automation from AI is a serious concerns for academics and policy makers, with reskilling workers one of the primary strategies for workers to take to keep from being displaced (Djankov and Saliola, 2018; Acemoglu and Restrepo, 2018). In Deming and Noray (2020) the returns to work experience are a race between on-the-job learning and skill obsolescence. They show that the earnings premium for technology-intensive college subjects decline faster than more general subjects. This highlights the need for flexibility in skill acquisition, and show the benefit of training from a general purpose technology like GPT (Eloundou et al., 2023) by making it easier for workers to adapt to chaning job demands.

And lastly, we contribute to a literature on limitations of human-AI interaction. Prior work has shown that people given access to AI often are not able to judge the quality of AI's outputs. Radiologists paired with AI are worse than when AI does diagnostics alone, because the radiologists rely on the AI when they are most uncertain, even though when

they are uncertain is tightly correlated with when the AI is uncertain (Agarwal et al., 2023). In an online labor market, employers given access to AI-written first drafts of job posts produce more generic job posts which are less likely to make a hire (Wiles and Horton, 2024). Dell’Acqua (2022) finds that when recruiters have access to applicant recommendations by very high-quality AI, that they take the AI’s suggestions, even when its not correct. Our result that the workers with training in ChatGPT are overconfident in GPT-4 complement these findings, and we show that they even get worse than before at predicting the boundaries of GPT-4’s abilities.

The rest of the paper proceeds as follows. Section 2 describes the experimental design and tasks we administer to workers. Section 3 describes our methods and analysis. Section 4 reports the experimental results of the ChatGPT training on workers’ abilities to complete data science tasks as well as their perceptions about the technology and their own technical skills. Section 5 concludes.

## 2 Experimental Design

We report the results from a large randomized control trial run on associates and consultants of the Boston Consulting Group, a large managerial consulting firm, to test whether high skilled but non-technical workers can do data science work with the help of GPT-4<sup>1</sup>.

The experiment took place over the course three weeks in March of 2024. In the recruitment phase, all BCG associates and consultants and were sent an email inviting them to participate in a study on upskilling and GenAI. We indicate that participation in the study is voluntary, can be done during work hours, and the time will count as an "office contribution" to their career development committee, which has financial implications to their annual bonuses. We also provide additional incentives to the top 50% of performers in each treatment group to encourage an ‘honest effort’ in the tasks<sup>2</sup>. Those who registered were given a survey on their demographics, programming and ChatGPT skills, technology openness, creativity, and learning orientation (Agarwal and Prasad, 1998; Miron et al., 2004; Jha and Bhattacharyya, 2013). Demographic and other variables were later used for stratified random assignment, as described below. Details of the registration survey are available in the Appendix B.

Simultaneously, BCG data scientists were also invited to participated in a a similar exercise, where they simply completed the tasks used in the experiment. Their output from

---

<sup>1</sup>We pre-registered our study with Social Science Registry detailing the design structure, the experimental conditions, the dependent variables, and our main analytical approaches.

<sup>2</sup>Top performers in each treatment group received recognition among BCG leadership as well as an invitation to a small group chat with offers for OpenAI and OpenAI merchandise.

these tasks served as the benchmark<sup>3</sup> for the “typical performance of a data scientist.” 40 data scientists submitted tasks to serve as this benchmark.

After the registration survey, consultants were then randomly assigned to either be in treatment group or control group. Treatment was stratified across gender, location, role (i.e., associate or consultant), coding skills, college degree (i.e., bachelors, masters, Ph.D.), and experience with ChatGPT for coding. The experimental sample consisted of 986 consultants, with 493 each in the treatment and control group. Prior to participation in the experiment, all subjects were asked if they consented to participate in the experiment. The experiment contained four phases, a pre-experiment survey, a 15-20 minute training on effectively using ChatGPT (in the treatment group) or Stack Overflow (in the control group), a series of tasks more technical than their current role, and then a post-experiment survey.

The pre-experiment survey consisted of questions of the participants subjective coding skills, GenAI usage, professional identity, and career aspirations. Next, participants were provided with a 15-20 minute tailored training to each experimental condition – the treatment group will receive specialized instructions on ChatGPT prompting and using it for data science, while the control group will be get training on effective googling and how to leverage websites commonly utilized by data scientists like StackOverflow and Khan Academy. Both trainings involved a combination of videos and interactive practice to prove competence. Details of the pre-experiment survey and both trainings can be found in Appendix Section [D.1](#).

The main experiment involved participants completing three complicated tasks representative of work done by BCG data scientists. There were three possible tasks participants completed, to test their ability to do different types of data science work. The first task was on statistics and machine learning—participants were given data on home buyers and had to use this data to predict whether a couple will take a mortgage out on a house, and were given a series of graphs and machine learning output they had to interpret. The second was a problem-solving and predictions task, which requires the participants to use historical data on men’s international football games to develop a strategy for sports investing. The third was a coding task, where the participants had to write and submit python code to clean and merge data, and then answer questions about the data.

The tasks were created in collaboration with analysts from OpenAI to be unable for GPT-4 to correctly answer. For all three tasks, if the participant let ChatGPT answer the question on it’s own, the answer was incorrect. Each task was intended to take 90 minutes

---

<sup>3</sup>Data scientists were told not to use ChatGPT in their completion of the tasks. Therefore when comparing the results of the participants to the data scientists benchmark, it is likely the benchmark of a data scientist without access to AI.

to complete, so to avoid fatigue, we gave each participant a random two of the three tasks, with the task order randomized. Details on the tasks can be found in Appendix Section [D.2](#).

Following the task completion, a post-experiment survey similar to the pre-survey was sent to participants to measure any change in participants’ perceptions about GenAI. Details on the post-experiment survey can be found in Appendix Section [D.3](#).

Lastly, following the experiment, we conducted qualitative interviews with a select sample of participants. Particularly, we will target individuals who over-performed and under-performed in the data science tasks to better understand how they used their tools and their perceptions of what influenced their outcomes. Details on the interviews can be found in Appendix Section [E](#).

### 3 Methods

The main analysis is comparing scores on the three tasks across the treatment and control group. We use a benchmark of the performance of the actual data scientists on the same set of tasks to determine output quality.

#### 3.1 Primary outcomes

Our first set of outcomes are related to how the consultants perform on a statistics task, a problem solving and predictions task, and a coding task. Each task was allocated 90 minutes, and the data scientists who completed these tasks finished them in X, X, and X minutes, respectively. Our first set of outcomes is how the consultants perform on the data science tasks. For all but one task, there is a conservatively defined correct answer, which receives a ‘1,’ while anything else receives ‘0.’ For the problem solving and predictions task, there is one problem where the output is a vector of predictions which does not have a correct or incorrect answer. For this prediction problem, we compare the answer from each data scientist to the benchmark from the data scientists’ answers, with details in Section [3.2](#).

Our second main outcome is whether or not consultants appeared to “learn.” We measure this with a set of questions in the post-experiment survey that both groups are not allowed to use ChatGPT to answer. It is possible that in applying GPT-4 to technical problems the treated consultants will retain some knowledge. However it is also possible that even if the treated consultants are better at performing data science tasks, they may not have more knowledge about data science after the tool they learned with is taken away.

Our third main outcome is how well consultants are able to gauge whether or not a problem can be done by ChatGPT. In the in the pre and post experiment survey we provide

Table 1: Comparison of worker covariates, by treatment assignment

<i>Flow from initial allocation into analysis sample</i>				
	<i>Total (N)</i>	<i>Treatment (N)</i>	<i>Control (N)</i>	<i>P-value</i>
Total workers allocated	983	493	493	
↪ completed all of experiment	487	260	227	0.04
↪ completed coding task	321	171	150	0.1
↪ completed statistics task	333	178	155	0.07
↪ completed predictions task	320	171	149	0.09
<i>Pre-allocation attributes of the final analysis sample: N = 487</i>				
		<i>Treatment mean: <math>\bar{X}_{TRT}</math></i>	<i>Control mean: <math>\bar{X}_{CTL}</math></i>	<i>P-value</i>
Female		0.369	0.37	0.985
Bachelors Degree		0.238	0.291	0.192
Masters Degree		0.677	0.604	0.092
Doctorate		0.085	0.106	0.428
Consultant		0.515	0.493	0.361
Office in Africa		0.019	0.018	0.896
Office in Asia Pacific		0.135	0.115	0.505
Office in Central or South America		0.019	0.004	0.14
Office in Europe or Middle East		0.492	0.52	0.546
Office in North America		0.335	0.344	0.835

*Notes:* This table reports means and standard errors of various pre-treatment covariates for the treatment group and the control group. The reported p-values are for two-sided t-tests of the null hypothesis of no difference in means across groups. The first panel describes the flow of the sample from the allocation to the sample we use for our main experimental analysis. The complete allocated sample is described in the first line, with those who “completed all of the experiment” making up the main experimental sample. Each worker was only assigned two of the three tasks, and the following lines compare the number who submit any work for each of the tasks. The third panel looks at pre-allocation characteristics of the jobseekers in the sample we use for our analysis, N = 487. We report the fraction of workers on their self reported i) gender, ii) highest degree achieved, and iii) office location.

the consultants a series of problems and ask them “How likely is GPT-4 to solve this problem correctly?” We hypothesize that after completing these tasks, the consultants in the treatment group will be better at forecasting which types of problems AI is good at solving.

Our last set of main outcomes relate to the consultants confidence and expectations about their ability to do data science. We hypothesize that consultants who are in the treatment group will have more confidence in their ability to do data science and are more likely to consider moving into more data science heavy roles. We also hypothesize that the treat-

ment group will have higher confidence in their ability to use ChatGPT to help them learn new skills.

### 3.2 Task grading

Each task will be graded with quantifiable measures of correctness of answers and approach, depending on the hypothesis. Each task will be graded on both the correctness of answer, the steps the participant used to solve the problem, and the time taken to submit an answer<sup>4</sup>. Below we describe the main outcomes for correctness of answer and for the process scores.

#### Statistics and machine learning tasks

Each question in the statistics task will be graded against the rubric (shown in appendix section 8.3). The rubric scores are a weighted correctness score such that the final score will be determined by a weighted sum across all answers:

$$\text{Total correctness} = \sum_{i=1}^n (\text{Correctness of answer}_i \times \text{Complexity weight}_i) \quad (1)$$

where  $n$  is the total number of distinct questions, correctness of answer, and the complexity weighting is defined as the level of complexity of the question. The complexity weightings were determined by asking several lead data scientists, with greater than 5 years of experience, to rank the complexity of each question and averaging across their answers.

#### Predictions tasks

The problem-solving task is designed to have numerous possible answers, some of which are better than others. We will use the answers submitted by the data scientists as the baseline/benchmark by which to grade the results of the associates and consultants. The data scientists results will be manually graded by humans and categorically assigned a weight of good, better, or best to establish 3 baselines. Specifically, the participants are submitting a predictability score for each match. We will normalize the participants predictability scores for each match,  $\text{score}_i$ , and calculate a loss score for the answers submitted by the associates and consultants when compared to the data science benchmarks,  $\text{DS score}_i$ . For each participant we will create a loss score defined as follows:

$$\text{Loss Score} = \frac{1}{n} \sum_{i=0}^n |\text{score}_i - \text{DS score}_i| \quad (2)$$

where  $n$  is the number of football matches in the dataset.

---

<sup>4</sup>Overall time to complete each task will be measured and logged by Qualtrics.



### Coding task

There is one distinct correct answer for the coding assignment. Correctness is a binary measure where its 0 if wrong and 1 if correct. Code quality will be calculated through a coding effort score, where the higher the effort score means the lower the quality of the code:

$$\text{Coding Effort Score} = \sum (\text{Errors encountered} \times (1 - \frac{\varepsilon}{2})) \quad (3)$$

where  $\varepsilon$  is 0 if the bug was not resolved and 1 if the bug was resolved.

### 3.3 Estimating treatment effects

Across each of these metrics, we employ equation (4) to estimate the average treatment effects based on OLS (Ordinary Least Squares) regression where  $y_i$  is the dependent variable (e.g., representing a quantifiable measure of output quality in the coding task and efficiency of code), and  $T_{\text{GPT}}$  is the ChatGPT treatment dummy. Lastly,  $X_i$  is a set of covariates collected in the survey such as location, gender, years of experience and education.

$$y_i = \beta_0 + \beta_{\text{GPT}}T_{\text{GPT}} + \gamma X_i + \varepsilon_i \quad (4)$$

### 3.4 Heterogeneous treatment effects

In addition to the main outcomes above, we plan to explore various factors that may influence the performance and outcomes of the consultants. For example, we can see if consultants with more technical backgrounds or those who are better at guessing what types of problems are within GPT-4’s range of ability have larger treatment effects.

Table 2: Effects of AI to accuracy on statistical knowledge questions

	<i>Dependent variable:</i>		
	Stats Overall Correctness Score	Any missing	
	(1)	(2)	(3)
GenAI Treatment Assigned (Trt)	1.400*** (0.524)	3.458*** (0.940)	-0.040 (0.042)
Coding basics		2.110** (0.985)	
Proficient coder		3.630*** (0.935)	
Coding basics x Trt		-3.053** (1.288)	
Proficient coder x Trt		-2.217* (1.255)	
Mean Y in Control Group	12.47	12.47	0.48
Observations	311	311	573
R <sup>2</sup>	0.048	0.115	0.008

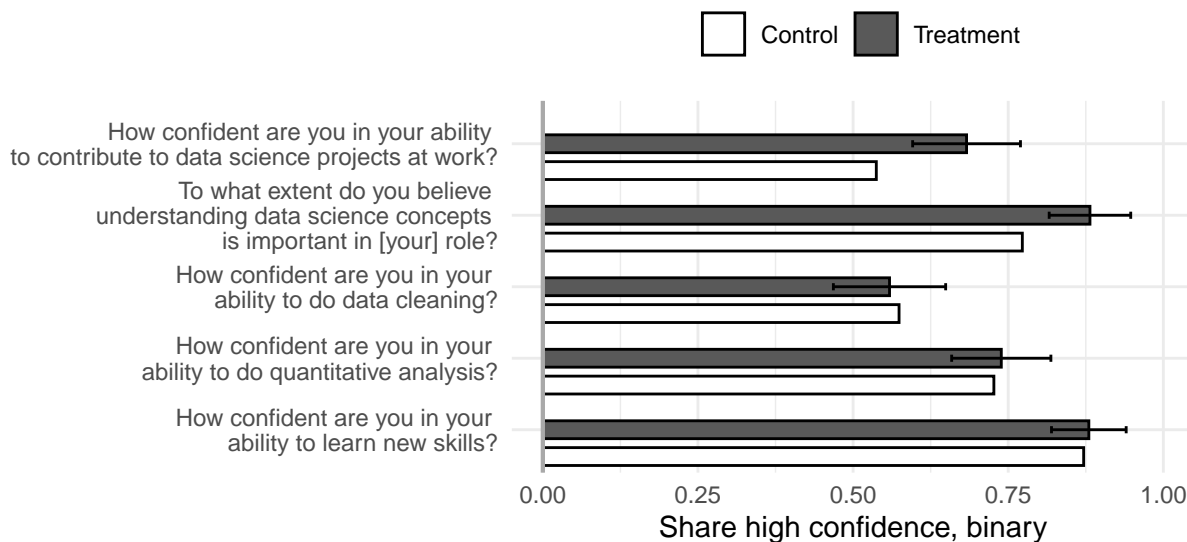
*Notes:* This table analyzes the effect of the treatment on the consultants ability to correctly answer questions about statistical knowledge. INSERT SENTENCE FROM URVY/Ryan ABOUT HOW STATSMCCCORRECTNESS IS CONSTRUCTED. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

Table 3: Effects of AI length of time to complete tasks

	<i>Dependent variable:</i>		
	Mins on Stats	Mins on Prediction	Mins on Coding
	(1)	(2)	(3)
GenAI Treatment Assigned (Trt)	2.439 (2.354)	-13.804*** (2.921)	-8.036*** (2.432)
Mean Y in Control Group	63.61	68.36	78.01
Observations	344	329	325
R <sup>2</sup>	0.022	0.087	0.052

*Notes:* This table analyzes the effect of the treatment on the length of time it took for consultants to finish each task, conditional on completion. The outcome in Column (1) is the number of minutes they spent on the Statistics task. The outcome in Column (2) is the number of minutes spent on the Problem Solving and Prediction task. And the outcome in Column (3) is the number of minutes spent on the Coding task. Consultants were randomly assigned two of the three tasks, and given 90 minutes maximum to complete each. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

Figure 1: Effect of AI treatment on consultant’s confidence in data science skills



*Notes:* This plot reports the effect of the treatment on the consultants self reported confidence in their own data science skills. Text of questions can be found in Appendix Section D.3. Regression details can be found in Table 5.

## 4 Results

### 4.1 Treated consultants performed better on data science tasks

#### 4.1.1 Treated consultants performed better at statistics knowledge questions

#### 4.1.2 Treatment effects are largest for most/least technical consultants

#### 4.1.3 Treated consultants completed tasks faster

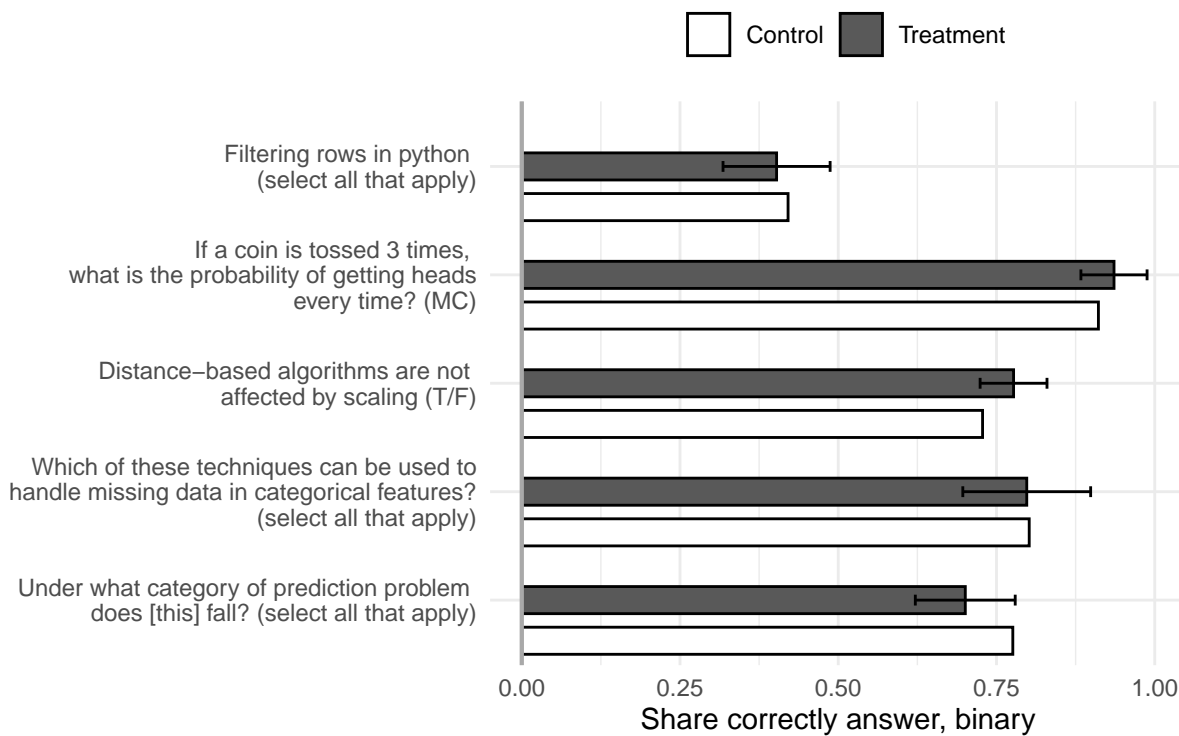
### 4.2 Treated consultants are more confident in their data science skills

### 4.3 No impact to post experiment tests of “learning”

Despite evidence that use of AI made consultants significantly better at solving data science problems, after the conclusion they were no more likely to be able to answer questions about probabilities, machine learning, or coding without the use of ChatGPT. In the post experiment survey, consultants were asked five questions on topics related to their tasks, for example<sup>5</sup>, “Distance-based algorithms are not affected by scaling” and “Which of these following code snippets will give us a dataframe filtered to only rows which correspond to

<sup>5</sup>Full text of questions can be found in Appendix Section D.3.

Figure 2: Effects of AI treatment to post experiment data science knowledge without use of GenAI



*Notes:* This analysis looks at the effect of treatment on consultants ability to correctly answer data science questions after the conclusion of the experiment. The x-axis is the mean probability of getting the correct answer for each treatment group. The y-axis has the text of each question, with the format of the answer. A 95% confidence interval is plotted around each estimate. Text of questions can be found in Appendix Section D.3. Regression details can be found in Table 6.

‘treatment’?” Consultants in both groups are instructed not to use ChatGPT to answer these questions.

In Figure 2 we report the effect of the treatment on whether the consultant correctly answered each question. In this table we define a correct answer conservatively— for “Select all that apply” questions, to receive a correct answer they must select all that are true and none that are false. The treated group performs no better than the control group on these questions<sup>6</sup>.

<sup>6</sup>Under less conservative definitions, the results to learning are even more precise nulls.

#### **4.4 Treated consultants exhibit overconfidence in AI’s current capabilities**

Consultants in the treatment group perform worse at guessing whether something is within GPT-4’s capabilities after the conclusion of the experiment. Before and after the experiment, consultants are posed seven problems and ask to give their opinion on the likelihood that GPT-4 is able to come to the correct conclusion for each problem<sup>7</sup>. Four of these problems GPT-4 reliably gets the answer wrong, while the other three it most often can solve. Prior to the experiment, consultants have comparable levels of confidence in GPT-4’s ability to correctly answer each question.

After the experiment, however, consultants in the treatment group become significantly more optimistic and significantly more wrong about GPT-4’s capabilities. In Figure 3 we show the difference between the treatment and control group in percentage points for consultants belief each problem can be correctly answered by GPT-4. Consultants in both groups were optimistic about GPT-4’s capabilities—the base rate for each problem is between 64 and 78%. For all four of the problems which it cannot solve, treated consultants report 5 to 10 percentage points higher likelihood’s that GPT-4 can correctly answer each problem. The only two problems which treated consultants are not more optimistic about GPT-4’s capabilities are two of the three problems it actually can correctly solve.

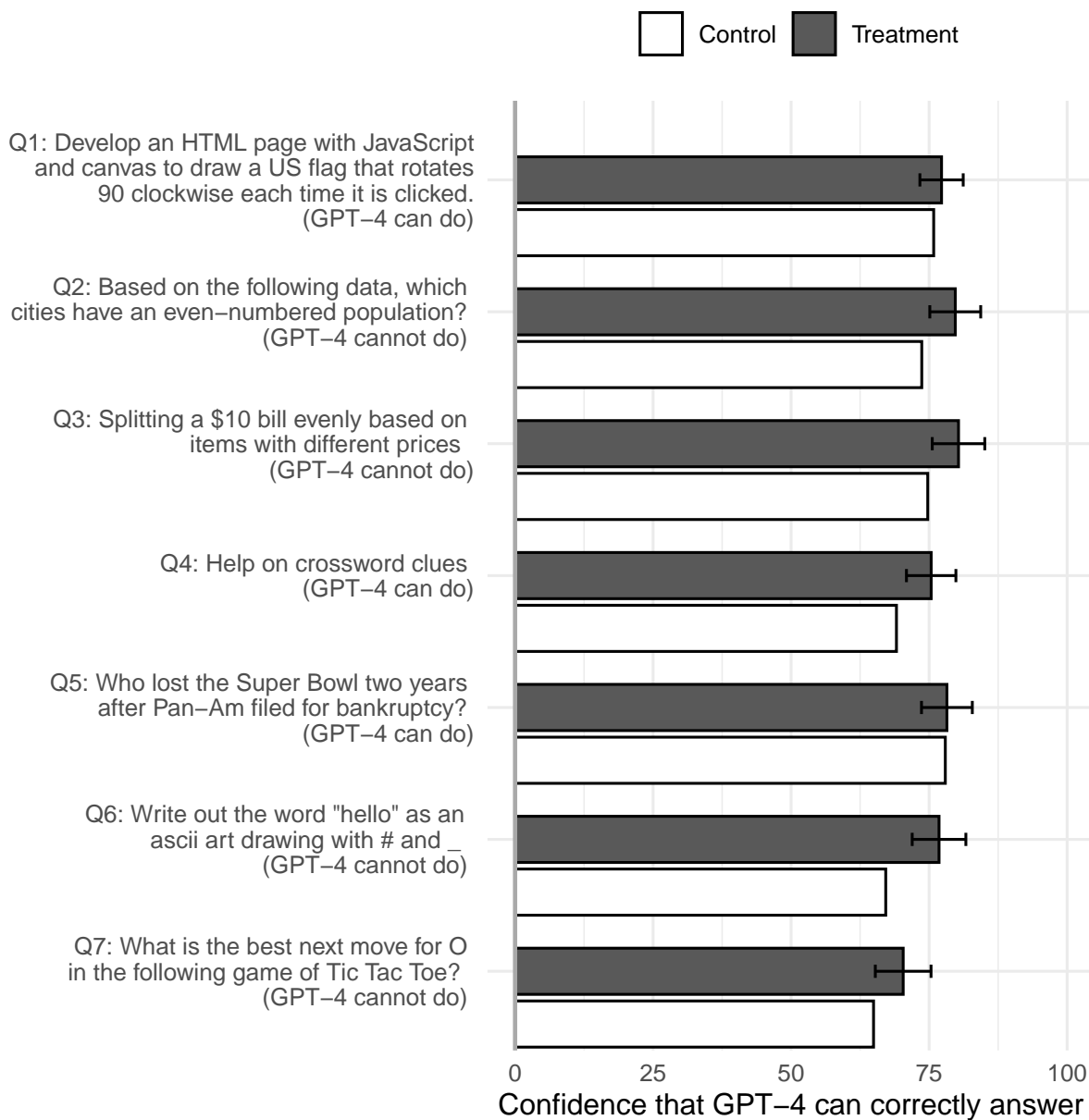
#### **4.5 Treated consultants are no more trusting of AI**

### **5 Discussion**

---

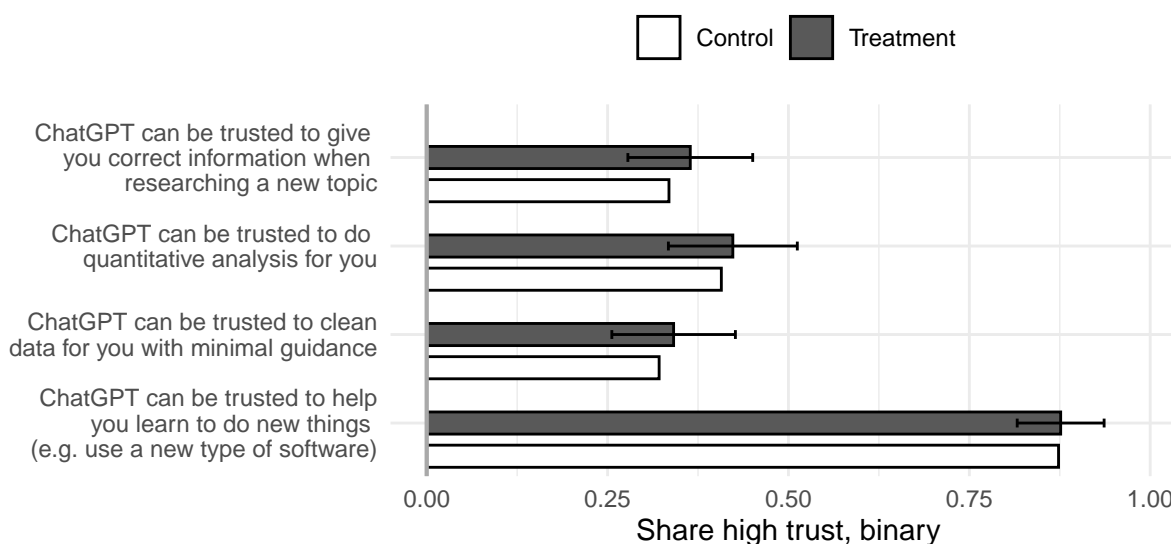
<sup>7</sup>Text of questions can be found in Appendix Section D.3.

Figure 3: Effect AI treatment on predictions about GPT-4's capabilities



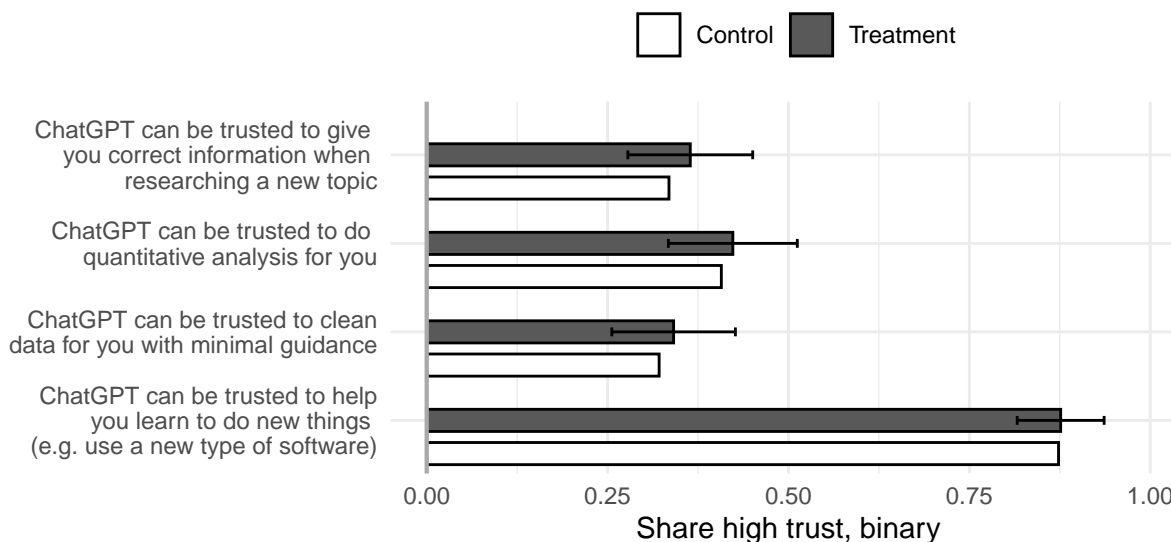
*Notes:* This analysis looks at the effect of treatment on consultants' confidence in AI's capabilities. The x-axis is the difference in the consultants' confidence out of 100 that GPT-4 would be able to answer the question correctly. The text, or a summary of the text, of the question is on the y-axis, with whether or not GPT-4 can actually get the correct answer. A 95% confidence interval is plotted around each estimate. Text of questions can be found in Appendix Section D.3. Regression details can be found in Table 7.

Figure 4: Effect of AI treatment on self reported trust in GenAI to do technical work



*Notes:* This plot reports the effect of the treatment on the consultants trust in GenAI to do technical work correctly. Each outcome is a binary variable for high trust or not. “High trust” is 1 if the consultant answered “Agree” or “Strongly agree” to the question, and 0 if else. A 95% confidence interval is plotted around each estimate. Text of questions can be found in Appendix Section D.3. Regression details can be found in Table 9.

Figure 5: Effects of AI treatment to self reported confidence in evaluating GenAI work



*Notes:* This plot reports the effect of the treatment on the consultants trust in GenAI to do technical work. Each answer is given on a scale of 1-7, where 1 = ‘Not at all’, 4 = ‘Neither’, and 7 = ‘Extremely’. We create a binary variable for “High confidence” which is 1 if the consultant answered 5,6, or 7 to the question, and 0 if else. A 95% confidence interval is plotted around each estimate. Text of questions can be found in Appendix Section D.3. Regression details can be found in Table 10.

## References

- Acemoglu, Daron and Jorn-Steffen Pischke**, “Why Do Firms Train? Theory and Evidence,” *The Quarterly Journal of Economics*, 1998, *113* (1), 79–119.
- **and Pascual Restrepo**, “The race between man and machine: Implications of technology for growth, factor shares, and employment,” *American economic review*, 2018, *108* (6), 1488–1542.
- Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz**, “Combining human expertise with artificial intelligence: Experimental evidence from radiology,” Technical Report, National Bureau of Economic Research 2023.
- Agarwal, R. and J. Prasad**, “A conceptual and operational definition of personal innovativeness in the domain of information technology,” *Information Systems Research*, 1998, *9* (2), 204–215.
- Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond**, “Generative AI at work,” Technical Report, National Bureau of Economic Research 2023.
- Dell’Acqua, F., E. McFowland, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, and K. R. Lakhani**, “Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality,” Technical Report, Harvard Business School Technology & Operations Mgt. Unit Working Paper 2023.
- Dell’Acqua, Fabrizio**, “Falling asleep at the wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters,” Technical Report, Working paper 2022.
- Deming, David J and Kadeem Noray**, “Earnings dynamics, changing job skills, and STEM careers,” *The Quarterly Journal of Economics*, 2020, *135* (4), 1965–2005.
- Djankov, Simeon and Federica Saliola**, “The changing nature of work,” *Journal of International Affairs*, 2018, *72* (1), 57–74.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock**, “Gpts are gpts: An early look at the labor market impact potential of large language models,” *arXiv preprint arXiv:2303.10130*, 2023.
- Heckman, James J, Lance Lochner, and Christopher Taber**, “Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents,” *Review of economic dynamics*, 1998, *1* (1), 1–58.



**Jha, S. and S. S. Bhattacharyya**, “Learning orientation and performance orientation: Scale development and its relationship with performance,” *Global Business Review*, 2013, 14 (1), 43–54.

**Miron, E., M. Erez, and E. Naveh**, “Do personal characteristics and cultural values that promote innovation, quality, and efficiency compete or complement each other?,” *Journal of Organizational Behavior*, 2004, 25 (2), 175–199.

**Mollick, Ethan R and Lilach Mollick**, “New Modes of Learning Enabled by AI Chatbots: Three Methods and Assignments,” 12 2022. Available at SSRN: <https://ssrn.com/abstract=4300783> or <http://dx.doi.org/10.2139/ssrn.4300783>.

**Wiles, Emma and John Horton**, “Algorithmic Writing Assistance on Jobseekers’ Resumes Increases Hires,” March 2024.

## A Appendix Tables & Figures

Table 4: P-value of difference between pre-allocations covariates, by treatment assignment

	<b>P-value</b>
Tenure years at BCG	0.797
English proficiency	0.120
Frequency of code for work	0.698
Know how to code	0.814
Python familiarity	0.505
Number of programming languages	0.189
ChatGPT for coding familiarity	0.618
Use ChatGPT or other LLM for work	0.599
Use ChatCGPTor other LLM for personal	0.766

*Notes:* This table reports means and standard errors of various pre-treatment covariates for the treatment group and the control group, in the final experimental sample where  $N = 487$ . The reported p-values are for two-sided t-tests of the null hypothesis of no difference in means across groups.

Table 5: Effects of AI to self reported confidence in data science skills, binary outcome

	<i>Dependent variable:</i>				
	Own DS Skills (1)	Importance of DS (2)	Various tech abilities (3)	(4)	(5)
GenAI Treatment Assigned (Trt)	0.146*** (0.044)	0.109*** (0.034)	-0.015 (0.046)	0.011 (0.041)	0.008 (0.031)
Mean Y in Control Group	0.54	0.77	0.57	0.73	0.87
Observations	465	479	461	466	470
$R^2$	0.057	0.055	0.030	0.038	0.006

*Notes:* This table analyzes the effect of the treatment on the share of consultants who report being confidence in their technical abilities. The outcome in Column (1) is a positive answer to the question “How confident are you in your ability to contribute to data science projects?” The outcome in Column (2) is their answer to “To what extent do you believe understanding data science concepts is important in the role of a BCG consultant?” The outcomes in Column’s (3) - (5) are on their confidence in their ability to do data cleaning, quantitative analysis, and learn new skills, respectively. Outcomes are on a scale from 1 to 7, with “High confidence” defined as 1 if they answer 5,6, or 7, and 0 otherwise. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

Table 6: Effects of AI treatment to post experiment data science knowledge without use of AI

	<i>Dependent variable:</i>				
	Data science or coding question				
	(1)	(2)	(3)	(4)	(5)
GenAI Treatment Assigned (Trt)	0.001 (0.039)	0.025 (0.027)	0.049* (0.027)	-0.004 (0.052)	-0.075* (0.040)
Mean Y in Control Group	0.35	0.91	0.73	0.80	0.78
Observations	573	399	418	253	408
R <sup>2</sup>	0.016	0.017	0.014	0.018	0.050

*Notes:* This table analyzes the effect of the treatment on the consultants ability to answer data science and coding questions, after the conclusion of the experiment. Text of questions can be found in Appendix Section [D.3](#). All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

Table 7: Effects of AI treatment to post experiment questions about GPT-4’s capabilities

	<i>Dependent variable:</i>						
	“Can GPT-4 answer [this question] correctly?”						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GenAI Treatment Assigned (Trt)	1.424 (2.001)	6.070** (2.352)	5.574** (2.427)	6.293*** (2.286)	0.287 (2.348)	9.653*** (2.480)	5.370** (2.576)
Mean Y in Control Group	75.82	73.66	74.75	69.08	77.93	67.14	64.93
Observations	454	475	473	464	465	431	451
R <sup>2</sup>	0.043	0.023	0.028	0.026	0.005	0.086	0.013

*Notes:* This table analyzes the effect of the treatment on the consultants confidence in GPT-4’s ability to get the right answer on various questions, after the conclusion of the experiment. For each question, the consultant gave a percentage confidence in GPT-4’s ability to answer the question correctly. The question in Columns 1 and 4, and 5 GPT-4 usually get correct. Questions 2,3,6 and7, GPT-4 almost never get correct. Text of questions can be found in Appendix Section D.3. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

Table 8: Effects of AI to whether or not they get the correct answer on various tasks

	<i>Dependent variable:</i>	
	Stats Task Score	Coding Task Correct
	(1)	(2)
GenAI Treatment Assigned (Trt)	1.663** (0.644)	0.047** (0.019)
Knowledge of GPT’s strengths	1.089 (0.811)	−0.007 (0.023)
Knowledge of GPT’s strengths x Trt	−0.754 (1.115)	−0.043 (0.031)
Mean Y in Control Group	12.47	0.00
Observations	311	296
R <sup>2</sup>	0.054	0.052

*Notes:* This table analyzes the effect of the treatment on the consultants ability to correctly answer questions. The first outcome is the sum of the consultant’s score on each statistics question, with a maximum possible score of 24.5. The second outcome is binary for whether or not they correctly answer the coding question. “Knowledge of GPT’s strengths” is 1 if the consultant got 4 out of 7, or more of questions correct on the pre-experiment survey asking about their guesses of whether or not GPT-4 can correctly answer a question. Exact definition of grading can be found in Appendix Section D.2. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

## B Pre-Experiment Survey: Registration for Generative AI Experiment with OpenAI

Thank you again for taking part in the Generative AI Experiment! The following questionnaire will take roughly 30 minutes to complete and contains questions about your background and your experiences. Please take the time to thoroughly and thoughtfully respond to these questions, as it is a crucial part of the overall experiment. We ask that you please take this questionnaire in one sitting, before February 16, 2024.

Please note that by submitting this questionnaire, you agree to not discuss the contents of the experiment to anyone, inside or outside of BCG. This is crucial for experimental integrity, to ensure robustness of the results for scientific publication.

### Data Use and Collection:

All data collected in this questionnaire will NOT be used for any other purposes other than this Generative AI experiment. Any data that is published internally to BCG, in scientific

Table 9: Effects of AI treatment to self reported trust in AI to do technical work, binary outcome

	<i>Dependent variable:</i>			
	High Trust			
	(1)	(2)	(3)	(4)
GenAI Treatment Assigned (Trt)	0.029 (0.044)	0.016 (0.045)	0.020 (0.044)	0.020 (0.044)
Mean Y in Control Group	0.33	0.41	0.32	0.87
Observations	480	480	480	480
R <sup>2</sup>	0.015	0.013	0.007	0.007

*Notes:* This table analyzes the effect of the treatment on the consultants trust in GenAI to do technical work. Each outcome is bucketed into high trust or not. “High trust” is 1 if the consultant answered “Agree” or “Strongly agree” to the question, and 0 if else. The outcome in Column (1) asks if ChatGPT can be trusted to give you correct information when doing research. The outcome in Column (2) asks if it can be trusted to do quantitative analysis for you. The outcome in Column (3) asks if it can be trusted to clean data with minimal guidance. The outcome in Column (4) asks if it can be trusted to help you learn new things. Full text of questions can be found in Appendix Section D.3. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$  : \*,  $p \leq 0.05$  : \*\* and  $p \leq .01$  : \*\*\*.

journals or alike will only be done so in aggregate, and personal information will never be released. This data will also only be shared with OpenAI in aggregate and personal information will not be released outside of BCG/BHI. Within the scope of this questionnaire, we will only collect your personal data, listed below.

- Name
- Email
- Location
- Gender
- Tenure
- Title
- English proficiency
- Education
- Proficiency and orientation towards tech

Table 10: Effects of AI treatment to self reported confidence in evaluating GenAI work, binary outcome

	<i>Dependent variable:</i>			
	High Confidence			
	(1)	(2)	(3)	(4)
GenAI Treatment Assigned (Trt)	-0.032 (0.046)	-0.028 (0.029)	-0.031 (0.029)	-0.054 (0.042)
Mean Y in Control Group	0.64	0.91	0.91	0.74
Observations	458	463	460	459
R <sup>2</sup>	0.012	0.010	0.027	0.045

*Notes:* This table analyzes the effect of the treatment on the consultants trust in GenAI to do technical work. Each answer is given on a scale of 1-7, where 1 = 'Not at all', 4 = 'Neither', and 7 = 'Extremely'. We create a binary variable for "High confidence" which is 1 if the consultant answered 5,6, or 7 to the question, and 0 if else. The outcome in Column (1) asks if the consultant is confident identifying factual inaccuracies in ChatGPT's responses. The outcome in Column (2) asks about their confidence in judging the relevance of ChatGPT's responses. The outcome in Column (3) asks about their confidence in assessing the understandability of ChatGPT's responses. The outcome in Column (4) asks about their confidence in evaluating the completeness of ChatGPT's responses. Full text of questions can be found in Appendix Section D.3. All regressions include controls for gender, location, native english status, and low tenure. Significance indicators:  $p \leq 0.10$ : \*,  $p \leq 0.05$ : \*\* and  $p \leq .01$ : \*\*\*.

Your personal data will only be used for testing the hypotheses of this Generative AI experiment, within the scope of your employment contract. We will process your personal data in accordance with applicable data protection laws and BCG's Privacy Policy [Link to internal policy]

## **CDC Contribution:**

As mentioned in the email, successful completion of participation in the study will count as an "office contribution" to your CDC to reflect our appreciation for your efforts. You will have the opportunity to provide your CDA details after completing the study. However, to avail of this opportunity, you must put in an "honest effort" throughout, as judged by the quality of your responses.

If there are any questions at all, please contact Lisa Kraymer (kramer.lisa@bcg.com)

## **Survey**

### **Demographics (Role and Location)**

1. Please Provide your Name First Name \_\_\_\_\_  
Last Name \_\_\_\_\_
2. Please Provide your BCG Email Address Below  
\_\_\_\_\_
3. Please Select Your Home BCG Office Location
  - Africa
  - Asia Pacific
  - Central & South America
  - Europe & The Middle East
  - North America
4. Please Select Your Home BCG Office (Conditionally Shown if: (2 = Africa))
  - Cairo
  - Casablanca
  - Johannesburg



- Lagos
- Luanda
- Nairobi
- Other (Please Elaborate) \_\_\_\_\_

5. Please Select Your Home BCG Office (Conditionally Shown if: (2 = Asia Pacific))

- Auckland
- Bangkok
- Beijing
- Bengaluru
- Canberra
- Chennai
- Fukuoka
- Ho Chi Minh City
- Hong Kong
- Jakarta
- Kuala Lumpur
- Kyoto
- Manila
- Melbourne
- Mumbai
- Nagoya
- Gurugram
- New Delhi
- Osaka
- Perth
- Seoul
- Shanghai
- Shenzhen

- Singapore
- Sydney
- Taipei
- Tokyo
- Other (Please Elaborate) \_\_\_\_\_

6. Please Select Your Home BCG Office (Conditionally Shown if: (2 = Central & South America))

- Bogota
- Buenos Aires
- Lima
- Panama City
- Rio De Janeiro
- Santiago
- San Jose
- Sao Paulo
- Other (Please Elaborate) \_\_\_\_\_

7. Please Select Your Home BCG Office (Conditionally Shown if: (2 = Europe & The Middle East))

- Abu Dhabi
- Amsterdam
- Athens
- Baku
- Barcelona
- Berlin
- Brussels
- Budapest
- Cologne
- Copenhagen

- Doha
- Dubai
- Dusseldorf
- Frankfurt
- Geneva
- Hamburg
- Helsinki
- Istanbul
- Lisbon
- London
- Madrid
- Milan
- Munich
- Oslo
- Paris
- Prague
- Riyadh
- Rome
- Stockholm
- Stuttgart
- Tel Aviv
- Vienna
- Warsaw
- Zurich
- Other (Please Elaborate) \_\_\_\_\_

8. Please Select Your BCG Affiliation Below

- Traditional BCG Consulting Team
- BCG X

- BCG Platinion
- Other (Please Specify) \_\_\_\_\_

9. Please Select Your Official Title at BCG

- Associate
- Consultant
- BCG X Data Scientist
- BCG X Senior Data Scientist
- Other (Please Specify) \_\_\_\_\_

10. Please Select Your Total Tenure at BCG (in Years)

- 0 to 1 Years
- 1 Years to 2 Years
- 2 Years to 3 Years
- 3 Years to 4 Years
- 4 Years to 5 Years
- 5+ Years

**Demographics (Education and Language)**

1. What is your gender?

- Female
- Male
- Prefer Not to Say
- Other

2. What is your English proficiency? (Reading, Written, and Spoken Combined)

- Beginner
- Intermediate
- Advanced
- Native

3. What is your highest education level?

- Bachelors
- Masters
- Professional Degree (e.g., MD, JD etc.)
- Doctorate

4. If you have a Bachelors degree, what was your major? Select the applicable categories and specify your degree in the text box.

- Science and Mathematics \_\_\_\_\_
- Engineering and Technology \_\_\_\_\_
- Health Sciences \_\_\_\_\_
- Social Sciences \_\_\_\_\_
- Business and Economics \_\_\_\_\_
- Arts and Humanities \_\_\_\_\_
- Education \_\_\_\_\_
- Agriculture and Environmental Studies \_\_\_\_\_
- Other \_\_\_\_\_

5. If you have a Masters degree, what was your major? Select the applicable categories and specify your degree in the text box. (Conditionally Hidden if: (12 = Bachelors))

- Science and Mathematics \_\_\_\_\_
- Engineering and Technology \_\_\_\_\_
- Health Sciences \_\_\_\_\_
- Social Sciences \_\_\_\_\_
- Business and Economics \_\_\_\_\_
- Arts and Humanities \_\_\_\_\_
- Education \_\_\_\_\_
- Agriculture and Environmental Studies \_\_\_\_\_
- Other \_\_\_\_\_

6. If you have a Professional degree, what was your major? Select the applicable categories and specify your degree in the text box. (Conditionally Hidden if: (12 = Bachelors OR 12 = Masters))

- Science and Mathematics \_\_\_\_\_
- Engineering and Technology \_\_\_\_\_
- Health Sciences \_\_\_\_\_
- Social Sciences \_\_\_\_\_
- Business and Economics \_\_\_\_\_
- Arts and Humanities \_\_\_\_\_
- Education \_\_\_\_\_
- Agriculture and Environmental Studies \_\_\_\_\_
- Other \_\_\_\_\_

7. If you have a Doctorate degree, what was your major? Select the applicable categories and specify your degree in the text box. (Conditionally Hidden if: (12 = Bachelors OR 12 = Masters))

- Science and Mathematics \_\_\_\_\_
- Engineering and Technology \_\_\_\_\_
- Health Sciences \_\_\_\_\_
- Social Sciences \_\_\_\_\_
- Business and Economics \_\_\_\_\_
- Arts and Humanities \_\_\_\_\_
- Education \_\_\_\_\_
- Agriculture and Environmental Studies \_\_\_\_\_
- Other \_\_\_\_\_

### **Programming Proficiency**

17. What tools do you currently use for data analysis?

- Excel
- Tableau

- Alteryx
- Programming (e.g. Python)
- ChatGPT
- Other LLMs / LLM based tools
- Other non-LLM based tools

18. Do you know how to code?

- Yes, I am an expert level coder
- I know how to code, but am not an expert
- I only know the basics of coding
- No, I do not know how to code

19. How often do you code for work? (Conditionally Hidden if: (17 = I know how to code, but am not an expert))

- I never code for work
- I code occasionally, but usually use other analytics tools
- I code every time I work on analytical projects, but this is only occasionally
- I code every time I work on analytical projects and I frequently am staffed on analytical projects
- Coding is a core part of my job

20. How many years of programming experience do you have? (Conditionally Hidden if: (17 = I know how to code, but am not an expert))

- 0-1
- 2-3
- 3-5
- 5-8
- 8+

21. How many programming languages are you familiar with? (Conditionally Hidden if: (17 = I know how to code, but am not an expert))

- 0

- 1
- 2-3
- 4+

22. How familiar are you with Python? (Conditionally Hidden if: (17 = I know how to code, but am not an expert))

- 0 = I Do Not Program
- 1 = Low Familiarity/Novice
- 2
- 3
- 4
- 5 = High Familiarity/Expert

### **ChatGPT Proficiency**

23. How often do you use ChatGPT or other LLMs for work?

- I have never used ChatGPT
- I have tried ChatGPT once or twice
- I use ChatGPT less than once per week
- I use ChatGPT at least once per week
- I use ChatGPT every day

24. How often do you use ChatGPT or other LLMs in your personal life?

- I have never used ChatGPT
- I have tried ChatGPT once or twice
- I use ChatGPT less than once per week
- I use ChatGPT at least once per week
- I use ChatGPT every day

25. Please rate the extent to which you agree or disagree with the following statements (1-7 Rating)

- I am familiar with GenAI for writing



- I am familiar with using GenAI for coding
- I am familiar with prompt engineering (i.e., crafting prompts to get a better answer from an AI model)
- I am familiar with more than 2 prompting strategies
- ChatGPT helps me become a better consultant
- I understand how large language models (LLMs), which underpin generative AI tools for writing, work
- I believe I can tell when ChatGPT is hallucinating
- I have created a specialized GPTs for my purposes
- I have used ChatGPT with Code Interpreter / Advanced Data Analytics
- I use ChatGPT for writing code

26. Please rate the extent to which you agree or disagree with the following statements (1-7 Rating)

- ChatGPT is primarily a Data Science tool
- ChatGPT is primarily a tool for writing
- ChatGPT helps me be more proficient at problem solving
- ChatGPT helps me be more efficient at creating slides

Here's the LaTeX version of the provided text:

### **Tech Openness and Playfulness**

27. Please rate the extent to which you agree or disagree with the following statements (1-7 Rating)

- If I hear about a new technology product or service, I will look for ways to experiment with it
- Among my peers, I am usually the first to try out new technology products and services
- In general, I am hesitant to try out new technology products and services
- I like to experiment with new technology products and services
- I am spontaneous when I interact with new technology products or services

- I am unimaginative when I interact with new technology products or services
- I am playful when I interact with new technology products or services
- I am flexible when I interact with new technology products or services
- I am uninventive when I interact with new technology products or services
- I am creative when I interact with new technology products or services
- I am unoriginal when I interact with new technology products or services

### **Creativity**

28. Please rate the extent to which you agree or disagree with the following statements (1-7 Rating)

- I try not to oppose team members
- I adapt myself to the system
- I adhere to accepted rules in my area of work
- I avoid cutting corners
- I am thorough when solving problems
- I address small details needed to perform the task
- I perform the task precisely over a long time
- I am good in tasks that require dealing with details
- I have a lot of creative ideas
- I prefer tasks that enable me to think creatively
- I am innovative
- I like to do things in an original way

### **Learning Orientation**

29. Please indicate the extent to which you agree or disagree with the following statements (1-7 Rating)

- I enjoy learning new topics
- I like to read diverse topics
- I find pleasure in learning

- I get intrinsically motivated to constantly expand my knowledge
- I seek deep-seated conceptual knowledge for the task assigned to me
- I spend a lot of time thinking about how my performance is in comparison to others
- I like to seek rewards in short term for my efforts
- I prefer to see tangible output as a reward for my effort
- I generally perform and undertake those tasks for which I get rewarded soon
- I feel very good when I know I have outperformed other colleagues
- I always try to communicate my achievements to my friends and supervisors

### **Concluding Remarks**

30. Do you agree not to discuss the contents of this experiment with anyone, inside or outside of BCG? This is crucial for experimental integrity, to ensure robustness of the results for scientific publication.

## **C Main experiment**

### **C.1 Survey**

#### **Introduction - consent**

Welcome to the ChatGPT / Generative AI Experiment!

We are thrilled to have you begin this study. Before you begin, please read through the following notes:

#### **Goal of Study:**

This is a scientific study conducted in collaboration with researchers from BCG and other institutions. We hope to publish the results from this experiment in a leading academic journal.

Due to the rigorous nature of academia, and the high standard needed for peer-reviewed publications, we ask for your full engagement and feedback. Please see the note about CDC contribution below for those that put in an honest effort.

This experiment will take you roughly 4 hours (or less) to complete.

## **Confidentiality:**

Please DO NOT discuss the details of this study with anyone, either among your peers or anyone inside or outside of BCG, even after they may have completed their participation. This seriously compromises the integrity of the full study. We want to absolutely avoid this.

## **Data Collected during Study:**

During the study, you will be given a short survey, series of tasks to perform and another short survey towards the end. For each task, you will type your answers in response. All data and information are fictional.

Your responses will be evaluated by a combination of humans and algorithms. All personal or identifying information will be scrubbed prior to this evaluation process.

## **Data Usage:**

Aggregate and deidentified information collected from this survey will be used for research purposes. All efforts will be made to keep your study-related information confidential. In particular, we will work to make sure that your responses are not accessed by anyone outside the research team.

Your personal data will ONLY be used to communicate your office contribution with your CDA and in case we need to have a follow-up interview or survey.

## **CDC Office Contribution and Other Incentives:**

As a token of our appreciation for your commitment, we are offering the following incentives for successful completion of this experiment:

1. CDC office contribution “recognition for anyone who puts in an honest effort” into all aspects of the experiment (including the follow-up interview, to be scheduled), as judged by the quality of their answers
2. In addition to the above, participants in the top 50th percentile as judged by quality of answers, with access to similar resources, will be noted as such to their CDA
3. In addition to the above, participants with access to similar resources with extraordinary performance will be commemorated with a BCG leadership recognition, a small group chat with OpenAI and OpenAI merchandise.

Note that participation is totally voluntary and there's no repercussions in case you decide to end your participation before finishing it. However, this would not count as an office contribution.

Once you have blocked 4 hours of uninterrupted time, you may start by continuing.

1. You cannot participate in this study on a phone, tablet etc. Please only proceed when you are logged in via your laptop/computer with a stable internet connection

- Yes, I am logged in via my laptop or my computer

Please type your BCG email address below to proceed \_\_\_\_\_

### **CDC Office Contribution**

If you'd like for your participation in this study to count as an "office contribution" as described above, please type in your CDA's BCG email address below.

If you do not want this, please type "N/A" \_\_\_\_\_

### **Overall flow of the study and expectations**

Your participation in this study will take approximately 4 hours and will consist of 7 sections:

- Filling out a short pre-survey (~10 min)
- Going over a short training (~15 min)
- Optional break (~10 min)
- Complete the first task (~90 min)
- Break (~10 min)
- Completing the second task (~90 min)
- Filling out post-survey (~15 min)

The tasks you will complete are independent of each other and unrelated to other survey components.

We highly encourage you to do your best to complete these tasks and while it might be challenging sometimes, we truly appreciate the effort. Don't forget that top 50th percentile, as judged by quality of answers, with access to similar resources, will be noted as such to their CDA.

## **D Main experiment**

### **D.1 Pre-Experiment Survey**

# GenAI\_DataScience\_Prod\_GPT

---

Start of Block: Welcome

Welcome **Welcome to the Upskilling Study!**

Thank you so much for taking your time to support this project. Your participation is critical to BCG's success as a thought leader in Generative AI.

---

Page Break

### Consent **Goal of Study:**

This is a scientific study conducted in collaboration with researchers from BCG, OpenAI, and other institutions. We hope to publish the results from this study in a leading academic journal.

Due to the rigorous nature of academia, and the high standard needed for peer-reviewed publications, we ask for your full engagement and feedback. Please see the note about CDC contribution below for those that put in an honest effort.

We anticipate that participation will take you roughly 4 hours (or less) to complete.

### **Confidentiality:**

**Please DO NOT discuss the details of this study with anyone, either among your peers or anyone inside or outside of BCG, even after they may have completed their participation.** This seriously compromises the integrity of the full study. We want to absolutely avoid this.

### **Data Collected during Study:**

During the study, you will be given a short survey, series of tasks to perform and another short survey towards the end. For each task, you will type your answers in response.

Your responses will be evaluated by a combination of humans and algorithms. All personal or identifying information will be scrubbed prior to this evaluation process.

All data will be aggregated and any personal identifiable information will be removed before sharing with any external collaborators, including OpenAI.

### **Data Usage:**

Aggregate and deidentified information collected from this survey will be used for research purposes. All efforts will be made to keep your study-related information confidential. In particular, we will work to make sure that your responses are not accessed by anyone outside the research team.

Your personal data will ONLY be used to communicate your office contribution with your CDA and in case we need to have a follow-up interview or survey.

All data collected in this questionnaire will NOT be used for any other purposes other than this Generative AI experiment. Any data that is published internally to BCG, in scientific journals or alike will only be presented in aggregate, and personal information will never be released. This data will also only be shared with OpenAI after it is aggregated and personal information will



not be released outside of BCG/BHI. Within the scope of this questionnaire, we will only collect your name, email and technical background.

Your personal data will only be used for testing the hypotheses of this Generative AI experiment, within the scope of your employment contract. We will process your personal data in accordance with applicable data protection laws and [BCG's Privacy Policy](#).

### **CDC Office Contribution and Other Incentives:**

As a token of our appreciation for your commitment, we are offering the following incentives for successful completion of your participation: CDC "office contribution" recognition for anyone who puts in an "honest effort" into all aspects of the study (including the follow-up interview, to be scheduled), as judged by the quality of their answers. In addition to the above, participants in the **top 50th percentile** as judged by quality of answers, with access to similar resources, will be noted as such to their CDA. In addition to the above, participants with access to similar resources **with extraordinary performance** will be commemorated with a BCG leadership recognition, and a small group chat with OpenAI. Note that participation is totally voluntary and there are no repercussions in case you decide to end your participation before finishing it. However, this would not count as an office contribution.

Once you have blocked 4 hours of uninterrupted time, you may start by continuing.

---

Page Break

**LaptopUse You cannot participate in this study on a phone, tablet etc. Please only proceed when you are logged in via your laptop/computer with a stable internet connection**

Yes, I am logged in via my laptop or computer (1)

---

**Email Please type your BCG email address below to proceed**

---

---

**CdcContribution CDC Office Contribution**

If you'd like for your participation in this study to count as an "office contribution" as described on the previous page, please type in your CDA's BCG email address below. If you do not want this, please type "N/A"

---

---

Page Break

Overview **Approximate flow of the study and what to expect.**

You can expect this study to take approximately 4 hours. It consists of 7 distinct sections:

Pre-survey (~10 min) Training (~15 min) Optional break ( ~10 min) First task (~90 min) Break (~10 min) Second task (~90 min) Post-survey (~15 min)

Note that the tasks are completely independent of each other and unrelated to other survey components.

**Please keep in mind that you cannot go backwards in this survey. Once you hit next, you will not be able to return. Please complete each page before moving on.**

We highly encourage you to do your best to complete these tasks and while it might be challenging sometimes, we truly appreciate the effort. Don't forget that **top 50th percentile** will be noted as such to their CDA.

End of Block: Welcome

---

Start of Block: Pre-Survey



PreSurTaskOnLoadTime PreSurveyTaskOnLoadTimeTracker

---

PreSurvey **Pre-Survey**

**First, we would like you to answer a few survey questions**

---



NeedCognition **Please indicate the extent to which you agree or disagree with the following statements:**

	Strongly disagree (0)	Disagree (1)	Neither agree nor disagree (2)	Agree (3)	Strongly agree (4)
I would prefer complex to simple problems (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I like to have the responsibility of handling a situation that requires a lot of thinking (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Thinking is not my idea of fun (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would rather do something that requires little thought than something that is sure to challenge my thinking abilities (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I really enjoy a task that involves coming up with new solutions to problems (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



ConsistencyInterest **Please indicate the extent to which you agree or disagree with the following statements:**

	Strongly disagree (0)	Disagree (1)	Neither agree nor disagree (2)	Agree (3)	Strongly agree (4)
I often set a goal but later choose to pursue a different one (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been obsessed with a certain idea or project for a short time but later lost interest (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have difficulty maintaining my focus on projects that take more than a few months to complete (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
New ideas and projects sometimes distract me from previous ones (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My interests change from year to year. I become interested in new pursuits every few months (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



**Perseverance** Please indicate the extent to which you agree or disagree with the following statements:

	Strongly disagree (0)	Disagree (1)	Neither agree nor disagree (2)	Agree (3)	Strongly agree (4)
I finish whatever I begin (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Setbacks don't discourage me (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am diligent (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am a hard worker (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have achieved a goal that took years of work (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have overcome setbacks to conquer an important challenge (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>





**DataScienceSkills In what aspects of data science do you have experience?**

	No experience (0)	Somewhat experienced (1)	Very experienced (2)
Data visualization (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Machine learning models (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Statistical analysis (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Data cleaning and preparation (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



**DSCconfidence On a scale of 1-7, where 1 = "Not at all" and 7 = "Extremely", please rate the following.**

	1	2	3	4	5	6	7
How confident are you in your ability to contribute to data science projects? ()							
To what extent do you believe understanding data science concepts is important in the role of a BCG A/C? ()							

DataScienceTools **What tools do you currently use for data analysis? Select all that apply.**

- Excel (1)
  - Tableau (2)
  - Alteryx (3)
  - Programming (4)
  - ChatGPT (5)
  - Other LLMs / LLM based tools (6)
  - Other non-LLM based tools (7)
- 



ExcelFrequency **How frequently do you use Excel, Tableau or Alteryx in your day-to-day work?**

- Daily (5)
  - Several times a week (4)
  - Once a week (3)
  - A few times a month (2)
  - Rarely (once a month or slightly less) (1)
  - Never (0)
- 



QuantExpertise **Please indicate the extent to which you agree or disagree with the following statements:**

	Strongly agree (4)	Somewhat agree (3)	Neutral (2)	Somewhat disagree (1)	Strongly disagree (0)
I consider myself an expert in Excel (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I consider myself an expert in Tableau (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I consider myself an expert in Alteryx (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Based on prior CDC reviews, PSI (problem solving and insights) has been a core strength (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



CodingPre **Do you know how to code?**

- Yes, I am an expert level coder (3)
- I know how to code, but am not an expert (2)
- I only know the basics of coding (1)
- No, I do not know how to code (-1)

Page Break



ProfessionalIdPre1 **Please indicate the extent to which you agree or disagree with the following statements:**

	Strongly Agree (4)	Somewhat Agree (3)	Neutral (2)	Somewhat Disagree (1)	Strongly Disagree (0)
Generative AI helps me feel valuable in my role (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI elevates how important I feel my job is for society (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI elevates my professional status and level of influence within my organization (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI helps me feel more competent in my role (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI enables my ability to execute tasks and reach desired outcomes (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI enables my ability to execute data analytics tasks and reach desired outcomes (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Generative AI increases the value I place on my expertise and skill cultivation (7)

Generative AI increases my level of autonomy in making individual decisions in my role (8)

Generative AI helps me be more confident that I will meet my project managers expectations (9)

Generative AI enables me to do what I really want to do in my role (11)

Generative AI will change the dynamic in my team (14)

Generative AI improved how I perceive my role in the organization (15)





ProfessionalIdPre2 **Please indicate the extent to which you agree or disagree with the following statements:**

	Strongly Agree (4)	Somewhat Agree (3)	Neutral (2)	Somewhat Disagree (1)	Strongly Disagree (0)
Using Generative AI helps me stay aligned with my project managers expectations (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe using Generative AI will contribute to the betterment of others in my work (12)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I see Generative AI as my coworker (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would recommend Generative AI to other consultants (16)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am proud of BCG's approach to Generative AI adoption within the firm (17)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I believe BCG is at the leading edge of the Generative AI revolution (18)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My managers and supervisors will expect more output from me because of Gen AI (19)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Problem solving and insights' dimension (20)

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Communication and Presence' dimension (21)

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Practicality and Effectiveness' dimension (22)



**GenAIUsage Rate how helpful you think Generative AI tools are for these use cases  
(Rating 1-7, where 1 = Not at all helpful; 7 = Extremely helpful; with ability to say "I don't know")**

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	I don't know (-1)
Write a first draft for simple texts (e.g., emails) (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write a first draft for complex texts (e.g., reports) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write my final version for simple texts (e.g., emails) (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Write my final version for complex texts (e.g., reports) (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review my writing (grammar, typos, etc.) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Be more persuasive (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Brainstorm ideas (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Be more creative (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing code for data analytics (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing code for data visualizations (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Learning how to use excel for data analysis and visualizations (11)

Identifying which machine learning models to use for a project (12)

Understanding the statistical significance of a result (13)

Writing code for data cleaning and preparation (14)



GenAllImpactPre **Since implementing Generative AI, how have your project teams been affected? Mark the position of the team relative to the description on the left and the description on the right**

	1	2	3	4	5	6	7	
--	---	---	---	---	---	---	---	--

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	
Decreased collaboration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Increased collaboration
Decreased efficiency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Increased efficiency
Decreased clarity of responsibilities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Increased clarity of responsibilities
Decreased learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Increased learning
Decreased decision quality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Increased decision quality
Reduced team morale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Improved team morale

GenAIBenefitsPre **In a few words, what do you think will be the biggest benefits of Generative AI for you?**

---

GenAIRisksPre **In a few words, what do you think will be the biggest risks of Generative AI for you?**

---

GenAIRolePre **Given the capabilities of Generative AI, do you see the role of associates and consultants evolving in the next 5 years? If so, how?**

---



---



---

---

---

---

Page Break 

---



GenAICalPre0 Rate how likely you think it is that ChatGPT will give a correct answer to the following prompts.

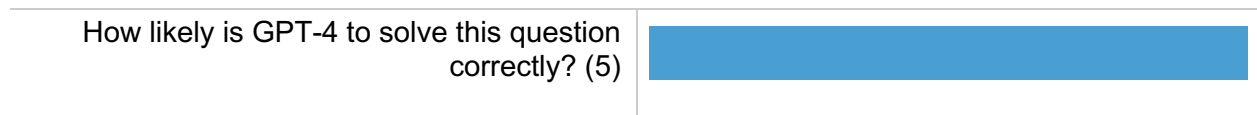
**PLEASE DO NOT USE ChatGPT, OTHER LLMs OR ANY OTHER SEARCH ENGINE (e.g., Google) TO ANSWER THESE QUESTIONS**

---

GenAICalPre1 Develop an HTML page with JavaScript and canvas to draw a representation of the US flag that rotates 90 degrees clockwise each time it is clicked.

Extremely unlikely   Somewhat unlikely   Neither likely nor unlikely   Somewhat likely   Extremely likely   I don't know

0 10 20 30 40 50 60 70 80 90 100



---

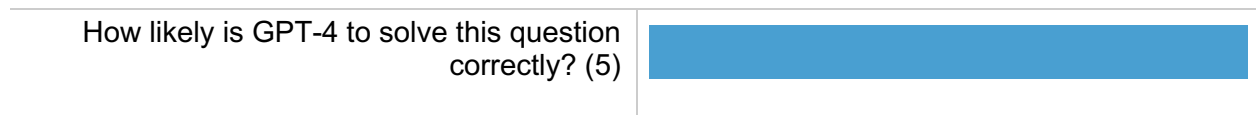
GenAICalPre2 Here is some data about Australian cities that I copied from Wikipedia. Based on this data, which cities had an odd-numbered population in 2011?

Australian Capital City Statistical Areas Population Table

City Statistical Area	Pop. June 2022	Pop. June 2011	Growth	Included SUAs
Greater Sydney	5,297,089	4,608,949	+14.93%	Sydney, Central Coast
Greater Melbourne	5,031,195	4,169,366	+20.67%	Melbourne, Bacchus Mars
Greater Brisbane	2,628,083	2,147,436	+22.38%	Brisbane
Greater Perth	2,224,475	1,833,567	+21.32%	Perth
Greater Adelaide	1,418,455	1,264,091	+12.21%	Adelaide
Australian Capital Territory	456,692	367,985	+24.11%	Canberra, Queanbeyan (A
Greater Hobart	252,693	216,273	+16.84%	Hobart
Greater Darwin	149,582	129,106	+15.86%	Darwin

Extremely unlikely   Somewhat unlikely   Neither likely nor unlikely   Somewhat likely   Extremely likely   I don't know

0 10 20 30 40 50 60 70 80 90 100



GenAICalPre3 Imagine you have a large box filled with small identical cubes. The box is completely full, and the dimensions of the box are 10 cubes long, 5 cubes wide, and 2 cubes high. You decide to take out all the cubes and rearrange them to form a new box that is 5 cubes long, 4 cubes wide, and 4 cubes high. How many cubes do you have left over after filling the new box?

Extremely unlikely   Somewhat unlikely   Neither likely nor unlikely   Somewhat likely   Extremely likely   I don't know

0 10 20 30 40 50 60 70 80 90 100

How likely is GPT-4 to solve this question correctly? (5)



GenAICalPre4 I'm playing wordle. My guesses so far are 1. CRANE (only the last E present, but in the wrong location) 2. POURS (only the first P present, but in the wrong location) 3. MIGHT (no letters present) 4. DEARY (only the E present, but in the wrong location) What do you think the word actually is?

Extremely unlikely   Somewhat unlikely   Neither likely nor unlikely   Somewhat likely   Extremely likely   I don't know

0 10 20 30 40 50 60 70 80 90 100

How likely is GPT-4 to solve this question correctly? (5)



GenAICalPre5 Write a webpage that shows a drawing of a cake and plays "happy birthday" when the page loads. Both should be generated with javascript. Make sure the cake looks right and the melody and note duration are correct in the music.

Extremely unlikely   Somewhat unlikely   Neither likely nor unlikely   Somewhat likely   Extremely likely   I don't know

0 10 20 30 40 50 60 70 80 90 100

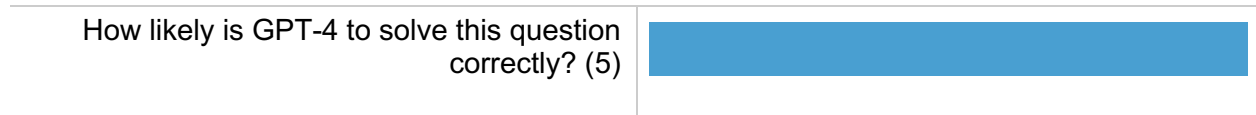
How likely is GPT-4 to solve this question correctly? (5)



GenAICalPre6 Write a single HTML file that has a javascript program that uses a canvas2d to draw "hello" with individual lines and curves. Do not use fillText.

ExtremelySomewhat Neither SomewhatExtremely I don't  
unlikely unlikely likely likely likely know  
nor  
unlikely

0 10 20 30 40 50 60 70 80 90 100



GenAICalPre7 Capitalize each sentence beginning with ""Input: ""

Input: darcy, she left Elizabeth to walk by herself.

Output: Darcy, she left Elizabeth to walk by herself.

Input: funny little Roo, said Kanga, as she got the bath-water ready.

Output: Funny little Roo, said Kanga, as she got the bath-water ready.

Input: hello this is a string.

Output: Hello this is a string.

Thank you for your help with this. From now on you will count the number of words in a sentence.

Input: This is an example sentence.

Output: 5

Input: Now another sentence.

Output: 3

Input: How long is this much longer sentence that has many words?

ExtremelySomewhat Neither SomewhatExtremely I don't  
unlikely unlikely likely likely likely know  
nor  
unlikely

0 10 20 30 40 50 60 70 80 90 100

How likely is GPT-4 to solve this question correctly? (5)



---

Page Break



FatiguePre **How would you rate your current level of focus and energy for completing this survey?**

- Very high – I'm fully focused and ready (5)
- Somewhat high – I feel quite prepared and alert (4)
- Neutral – I'm neither tired nor particularly energized (3)
- Somewhat low – I'm a bit tired or distracted (2)
- Very low – I'm already feeling quite fatigued or unfocused (1)

End of Block: Pre-Survey

---

Start of Block: Training\_GPT

TimerTrainingGPT Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

---

TrainingGPTInt **Training**  
**Introduction to ChatGPT Enterprise**

**This training should take you 15-20 minutes and will auto-advance to the next section in 30 minutes.**

This training program is designed to equip you with advanced skills in talking to ChatGPT. Through a series of interactive modules, you will learn how to effectively use ChatGPT to your advantage.

---

TrainerGPTVideo 1. Please start by watching the 2 videos below  
Getting ChatGPT to do what you want

Signing Into ChatGPT



Page Break

TrainingGPT2 2. ChatGPT is a guide, not an individual contributor

Throughout this assignment and in general when working with ChatGPT, you may be tempted to have ChatGPT do all your work. The outputs look very convincing!

In our study last year, we saw that ChatGPT hurt performance by 23% for individuals who over-relied on it for problem solving. Therefore, we encourage you to do the assignments alongside ChatGPT – using ChatGPT as your guide.

**Use your own rigor and intuition to quality check ChatGPT's output.**

---

Page Break



## TrainingGPT3 3. Introduction to talking to ChatGPT

This training will describe process of designing and refining instructions (i.e. prompts) given to a large language model (e.g. ChatGPT) to get better results and elicit desired behaviors. Note that combining these methods can sometimes have greater effect.

---

### TrainingGPT3.1 **Standard Prompting**

Most users of ChatGPT use standard prompting (also known as “naïve” prompting or “zero-shot” prompting). This is when the model is given a task without prior examples; it must deduce what to do from the prompt and its existing training. For example, just asking ChatGPT “What are the best practices for talking to ChatGPT?”.

Standard prompting is often sufficient if you are following a few best practices:

Be clear and concise with common language. Avoid confusing consulting jargon!

Provide context such as the purpose of the ask and details behind the instructions

Be specific by clearly stating what you are trying to accomplish      Clarify the output format – e.g. bullets, tables, paragraphs, etc

**Worse**

**Better**

Analyze this data a CSV file containing sales data from the last quarter, including columns for date, product ID, and sales volume. Can you provide a Python script using pandas to read this file and calculate the total sales volume for each product? Please include comments in the script explaining each step of the process.	I have What's the best machine learning model? Given a dataset with 1000 rows of customer demographic information (age, income, number of purchases) and a binary target variable indicating whether each customer subscribed to a service, which machine learning model would be most appropriate for predicting subscription likelihood? Please explain your recommendation based on the data characteristics.
---	---

Make a graph from this data series data showing the daily number of visitors to my website over the past year. Could you guide me on how to use Python to plot this data, including a moving average line to highlight trends? Additionally, could you explain how to customize the plot to add labels for the x-axis ("Date") and y-axis ("Number of Visitors"), and a title for the chart ("Daily Website Visitors")?	I have time-
--	--------------

---

### TrainingGPT3.2

#### **Ask the model to adopt a persona**

Asking the model to adopt a persona can allow you to get more specific answers compared to

just asking the question. This can either be a certain individual (e.g. Elon Musk), or a specific qualification. Adapting the concept of adopting a persona for data science tasks can help in obtaining more specialized and nuanced responses. Let's try it!

---

TrainingGPTPrompt3.3

Type the following into ChatGPT: "What's the best way to analyze large datasets?" and copy the answer below:

---

---

---

---

---

---

TrainingGPTPrompt3.4      Now tell it to adopt a persona: "Acting as a data scientist, tell me the best way to analyze large datasets."      Now copy the answer below and take a mental note of how the answer has changed:

---

---

---

---

---

---

TrainingGPTPrompt3.5      Finally, get more specific in your persona: "You'll act as a data scientist who specializes in big data analytics, with extensive experience in Python and Spark. Explain the most efficient method to process and analyze multi-terabyte datasets, including step-by-step instructions on setting up the environment, loading the data, and performing

exploratory data analysis.”

Copy the answer below and take a mental note of how the answer has changed again:

---

---

---

---

---

---

**TrainingGPTExample Provide examples (i.e. one-shot or few shot prompting)**

Providing general instructions that apply to all examples is generally more efficient than demonstrating all permutations of a task by example, but in some cases providing examples may be easier. For example, if you intend for the model to copy a particular style of responding to user queries which is difficult to describe explicitly. Incorporating the concept of one-shot or few-shot prompting into data science or data cleaning tasks can effectively guide the model to understand and replicate a specific answering or problem-solving style. This is known as "few-shot" prompting. Let's try it!

---

**TrainingGPTExample1** Type the following into ChatGPT: “How do I extract key phrases from text?” and copy the answer below

---

---

---

---

---

---

**TrainingGPTExample2** Now try giving it an example of how to respond: “Answer in a consistent style as this example. Question: How can I identify the sentiment of user reviews? Answer: Sentiment analysis of user reviews can be efficiently performed using Natural

Language Processing (NLP) techniques. The first step involves preprocessing the text by removing stop words and punctuation, followed by tokenization. Next, applying a pretrained model like VADER (Valence Aware Dictionary and Sentiment Reasoner) or a fine-tuned BERT model can classify the sentiment of each review into categories such as positive, negative, or neutral. This process enables an automated and scalable way to gauge customer sentiment from textual data. Now that I have given you an example – How do I extract key phrases from text?” Now copy the answer below and take a mental note of how the answer has changed:

---

---

---

---

---

---

TrainingGPTSteps

**Specify the steps required to complete a task**

Some tasks are best specified as a sequence of steps. Writing the steps out explicitly can make it easier for the model to follow them. This is known as “Chain-of-thought” prompting. Let’s try it!

---

TrainingGPTSteps1 Ask ChatGPT: “How do I classify images using a deep learning model?” and copy the answer below:

---

---

---

---

---

---

TrainingGPTSteps2 Now try a chain-of-thought approach and type the following into ChatGPT:

“Consider and include the following elements in your project to classify images using a deep learning model:

**Data Collection:** Identify and gather images for your dataset. Mention the source of your images and how many categories or classes of images you plan to classify.

**Data Preprocessing:** Describe the steps for resizing images to a uniform size, normalizing pixel values, and splitting the dataset into training, validation, and test sets.

**Model Architecture Design:** Choose a deep learning model architecture suitable for image classification. Consider whether to use a pre-trained model for transfer learning or to design a model from scratch.

**Model Training:** Outline the process for compiling the model with an appropriate optimizer and loss function. Mention how you will use data augmentation to improve model generalization.

**Hyperparameter Tuning:** Discuss the approach for tuning hyperparameters, such as learning rate, batch size, and the number of epochs, to improve model performance.

**Model Evaluation:** Explain how to evaluate the model's performance on the test set using metrics such as accuracy and precision. Consider plotting a confusion matrix to understand the model's classification behavior across different classes.

**Model Deployment:** Describe how you would deploy the model for real-world use, including converting the model to a suitable format for deployment and integrating it with an application for image classification.

**Performance Monitoring and Updating:** Consider how you will monitor the model's performance in production and the steps for retraining the model with new data or adjusting it based on performance feedback.

Let us think step by step. How do I tackle this image classification project with a deep learning model?” Now copy the answer below and take a mental note of how the answer has changed:

---

---

---

---

---

---

TrainingGPTTime

**Give the model time to “think”**

If asked to multiply 17 by 28, you might not know it instantly, but can still work it out with time. Similarly, models make more reasoning errors when trying to answer right away, rather than taking time to work out an answer. Asking for a "chain of thought" before an answer can help the model reason its way toward correct answers more reliably.

---

### TrainingGPTTime1

Suppose for example we want a model to evaluate a student's solution to a math problem. The most obvious way to approach this is to simply ask the model if the student's solution is correct or not. However, this can sometimes lead to ChatGPT giving you the wrong answer. The following is an example prompt that is more likely to give an accurate answer: "First work out your own solution to the problem. Then compare your solution to the student's solution and evaluate if the student's solution is correct or not. Don't decide if the student's solution is correct until you have done the problem yourself."

---

TrainingGPTTime2 You can also ask the model to check it's own work or identify if it missed anything.

For example, suppose that we are using a model to list excerpts from a json file. If the file is very large, it is common for a model to stop too early and fail to list all relevant excerpts. In that case, you can ask the model to find any excerpts it missed on previous passes.

Are there more relevant excerpts? Take care not to repeat excerpts. Also ensure that excerpts contain all relevant context needed to interpret them - in other words don't extract small snippets that are missing important context.

---

Page Break

---

TrainingGPTDataA 4. Analyzing Data with ChatGPT's Data Analyst When working with data, especially if using code like Python, using ChatGPT's Data Analyst can significantly enhance your ability to analyze and interpret data directly within the chat. The Data Analyst allows for the execution of Python code, enabling data analysis, visualization, and more. **Keep in mind that ChatGPT's Data Analyst is still being refined and it sometimes makes mistakes! Its critical to do the work alongside ChatGPT to check your work!**

Here's an example of how to get started with Data Analyst and how to make sure it shows you the code it uses so you can put that code into your own notebooks for checking its work:

---

TrainingGPTDataA1 Best Practices for Using ChatGPT's Data Analyst: **Do the work alongside ChatGPT:** ChatGPT can make errors, even when using Data Analyst! If you are asking ChatGPT to do analysis for you, test what it is doing somewhere outside of ChatGPT. For example, if you ask it to write code for you, copy and paste the code it uses into a Jupyter notebook or other IDE. Run the code and test that it is doing what you expect!

**Clear Definition of the Task:** Before asking ChatGPT to use Data Analyst, clearly define what you aim to achieve with your data. Whether it's data cleaning, visualization, statistical analysis, or machine learning - having a clear objective will guide the code you write and the questions you ask the ChatGPT Data Analyst.

**Break Down the Task:** Divide your overall task into smaller, manageable steps. This could include data importation, preprocessing, exploratory data analysis (EDA), model building, and evaluation. Addressing each step individually can simplify complex analyses.

**Provide Context:** When prompting the ChatGPT Data Analyst, provide as much context as possible. This includes the structure of your dataset, the libraries you wish to use, and any specific methods or techniques you're interested in.

**Specify the Output Format:** Indicate how you'd like the results to be presented. For instance, if you're visualizing data, specify the type of plot you need. For statistical analyses, mention how you'd like the results to be summarized.

**Ask ChatGPT about the errors you see when testing its work:** Always test ChatGPT's work outside of ChatGPT! If you run into errors, ask ChatGPT to help you solve the problem.

**Iterative Exploration:** Data analysis is often exploratory and iterative. Don't hesitate to refine your questions/prompts based on the output you receive. If an analysis doesn't provide the insight you were hoping for, adjust your approach and try again.

**Use ChatGPT to Help:** Try asking ChatGPT for help refining your prompts to get the outcome you are looking for!

---

TrainingGPTDataA2 **Try some prompts that will get the data analysis process started:**

---

TrainingGPTDataA3 **Example: Time Series Analysis**      **Initial Task:** Analyze seasonal patterns in time series data.      **Prompt:** "I have a time series dataset stored in a pandas DataFrame df with two columns: 'Date' (datetime) and 'Daily\_Sales'. I'd like to analyze seasonal patterns in daily sales over the year. Write Python code using pandas and statsmodels to decompose the time series into trend, seasonal, and residual components. Then, plot these components using matplotlib to visualize the seasonal patterns."

---

TrainingGPTDataA4 **Example 2: Natural Language Processing (NLP) for Sentiment Analysis**      **Initial Task:** Perform sentiment analysis on customer reviews.      **Prompt:** "Given a list of customer reviews stored in a pandas DataFrame reviews\_df with a column 'Review\_Text', use Python's Natural Language Toolkit (NLTK) or another NLP library to preprocess the text (tokenization, removing stopwords, and lemmatization). Then, apply a pre-trained sentiment analysis model from the library to classify each review as positive, negative, or neutral. Summarize the overall sentiment distribution among the reviews."

---

Page Break

---



TrainingGPTIssues 5. Common issues and their solutions

**Context memory is overloaded** – ChatGPT can only remember so much! If you find that your chat is getting stuck in a concept you don't want– make a new chat and restart with more specific instructions up front. If you do this during the study, make sure to provide us with links to every chat you create to answer the question. **A file is overloading the context** – If you load a file into ChatGPT it will sometimes read the file directly into the chat context memory and overload the chat creating weird results. When uploading large files, you tell ChatGPT not to read it into the chat's memory by saying something like: "Only open this file when running python code using Data Analyst and don't store it in the context memory of this chat".

Alternatively, you can tell ChatGPT to only read a certain portion of the file, e.g. "Only read the first 10 rows of this file." **ChatGPT's Data Analyst keeps having errors** – Ask ChatGPT not to run the code and instead just generate the code. Then run the code yourself – e.g, by using Jupyter.

---

TrainingGPTOtherRes 6. Other Resources:

**Stack Overflow** (<https://stackoverflow.com>) is one of the largest, most trusted online communities for developers to learn, share their programming knowledge, and build their careers. It features a vast repository of questions and answers on a wide array of programming and data science topics.

Feel free to watch this video on how to leverage Stack Overflow for coding and problem solving.

---

TrainingGPTOAI Finally, read through OpenAI's prompt engineering documentation if you want some additional examples of techniques and strategies. Or even if you just want to reference some of these strategies again during the study:

<https://platform.openai.com/docs/guides/prompt-engineering/strategy-use-external-tools>

---

Page Break

## **D.2 Task details**

TrainingGPTOffDoc Please download the following pdf to have access to all of these resources while working on the rest of the survey

[Training Document](#)

End of Block: Training\_GPT

---

Start of Block: Begin Tasks - Optional Break

TimerOptionalBreak Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

---

OptionalBreak1 **Please feel free to take an optional break if you desire. Please keep breaks to under 10 minutes, if possible, to not disturb the flow of your engagement. The next section (once you advance) will be a timed 90-minute section focusing on a data science task.**

**Once you are ready to proceed to the next task, you may continue by using the arrow below.**

End of Block: Begin Tasks - Optional Break

---

Start of Block: Coding Task - Instructions

CTInstruct1 **Coding Task**

This next Task is a coding assignment. You will be asked to use a Google Colab notebook to write Python code to process some data and arrive at a solution. Python is a common programming language used for data science tasks. We will walk you through what all this means.

You will be limited to 90 minutes to complete the assignment. But BEFORE the 90-minute timer starts, you will need to watch some Colab introduction videos and verify Colab is setup properly.

1. Colab Instructions (~15 minutes)

Setup Google & Colab

Setup Coding Task Notebook

Colab Videos to help you with setting up the Coding Task Notebook  
2. Coding Task (90 minutes)

---

CTInstruct2 **1. Colab Instructions (~15 minutes)**

---

CTInstruct3 **Setup Google & Colab**

You should have received instructions to set up a Google Account with your BCG Email and Install Google Colab on your Google Account. The email was titled “[GenAI Experiment with OpenAI] Getting Started Details”. If you have not gone through the instructions in that email, please do that now.

In case you can't find the email, the instructions are in this file for your convenience: [Setup Google & Colab](#)

If you are having issues, please check this Troubleshooting Guide, with solutions to common issues that come up during this process: [Troubleshooting](#)

---

CTInstruct4 **Setup Coding Task Notebook:**

For the Coding Task, you will be working in a shared Google Folder located at the following link. You will need to copy and paste this URL into a new tab: `{e://Field/google_drive_link}`

The instructions for opening the Colab document and getting started can be found in the "Colab\_Setup\_Instructions.docx" file located within this shared folder.

Once you have the Colab document open, you need to follow the instructions inside the notebook itself, to update and run the "Setup Logging Before The Task" Cell.

If these instructions are confusing, don't worry! The videos below will also walk you through these instructions step-by-step.

---

CTInstruct5 Please Confirm that you have Colab installed on your Google Account, and are able to open the shared notebook file in Colab ([\\${e://Field/google\\_drive\\_link}](#)):

Yes, I have installed Colab on Google Account and am able to open the shared notebook. (1)

---

CTInstruct6 **Colab Videos:** Watch the following videos to familiarize yourself with Colab.

---

CTInstruct7 Confirm you have watched the Colab videos above.

Yes (1)

---

CTInstruct8 Once you have successfully run the "Setup Logging Before The Task" cell, you are ready to advance to the Task instructions on the next page. You will know that it is successful if you see the words "Successfully mounted your Drive! Continue below to the task" below the cell.

Here is what it will look like if you are successful:

Once you advance to the next page, your 90 minute task timer will start.

---

CTInstruct9 **In Case of Errors** If you run into errors during the setup, you will not be able to complete the Coding Assignment. In this case, please refer to the [Troubleshooting](#) and see if any of the solutions apply to your case.

Here is a common error message. If you see this - DO NOT proceed!

If you get an "Access Blocked : Authorization Error" while running the setup cell, please switch to a personal Google account to complete this task.

If you have tried everything and you still see an error, please reach out to Ryan Kennedy ([kennedy.ryan@bcgfd.com](mailto:kennedy.ryan@bcgfd.com)) on Slack. Please do not reach out with any questions regarding how to complete the task itself, but rather only questions regarding setup errors.

**DO NOT Proceed to the Task if you have not completed this step or if you are seeing any errors**

---

CTInstruct10 Confirm that you have gone to the shared Google Folder, were able to Mount your Google Drive using the "Setup Logging Before The Task" cell, and see "Successfully mounted your Drive! Continue below to the task" in the space below the "Setup Logging Before The Task" cell.

Yes (1)

---

CTInstruct11 You are now ready to move on to the Coding Task. The official 90-minute task timer will start once you advance.

#### End of Block: Coding Task - Instructions

---

#### Start of Block: Coding Task

CodingInstructionGPT **Please read through the instructions to complete the next task!**

You are **not expected to have any prior coding experience**. Please try your best to complete the assignment **with code in Colab**, and get as far as you can.

You may spend **up to 90 minutes** on this task. At the 90-minute mark, the form will **automatically submit and move you to the next portion of the study, regardless of your progress**. You may choose to advance prior to the 90 minutes if you have fully completed the task.

**Use ChatGPT Enterprise** to perform this task **in whatever way you like (uploading files, copying and pasting the questions or errors and results)**. Access ChatGPT Enterprise by going to <https://chat.openai.com/> and log in using your BCG login.

**We highly encourage you to do your best to complete this task.** It might be challenging sometimes! But you're going to get CDC credit just for putting in an honest effort and if you are

in the **top 50th percentile of your group** we'll notify your CDA about how well you did! We truly appreciate your effort.

-----

Page Break

---

CodingTaskTimer Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)



CodingTaskOnLoadTime CodingTaskOnLoadTimeTracker

---

---

CodingTaskIntro **The clock has started! You now have 90 minutes to complete the coding task. In case you need it again, here's the link to the Google Colab**

**Documents:** [\\${e://Field/google\\_drive\\_link}](#)

**Use ChatGPT enterprise** to perform this task **in whatever way you like (uploading files, copying and pasting the questions or errors and results)**. Access ChatGPT Enterprise by going to <https://chat.openai.com/> and log in using your BCG login.

No matter how you use ChatGPT or other resources to complete this task, **make sure any code you use is run in your Google Colab notebook so that we can review your solution.** If you used ChatGPT to help generate and execute your code, please copy the code back to your Colab notebook, run it, and troubleshoot for errors. We will not be able to review your solution if the code is not run in Colab so you will receive no credit if your notebook file is empty. Return to this survey when you are finished with the task in Colab.

Here are the Intro Colab Video Links provided again for reference:

[Intro to Colab](#)

[Using Colab Features](#)

---

CodingTaskInstruct Below you will find the details of the question. You will be provided with the overview of the data sets and an overview of the data cleaning steps you will need to take. The steps you need to take are also noted in your Colab Notebook.

---

CodingTask Assignment

**Use the datasets found in your Google Folder to answer the question: Which 5 customer IDs had the highest average order by total price in May 2022?**



## Overview of Datasets

### Dataset 1: Orders Data (orders.csv)

File Type: CSV

Contents:

customer\_id: The unique identifier for each customer.

order\_info: Information about each order in the format order number ;

date and time. The order number in this dataset (once decoupled from the date and time) corresponds with that in the next.

### Dataset 2: Products Data (products.csv)

File Type: CSV

Contents:

customer\_id: The unique identifier for each customer associated with an order.

order\_id: ID of each order in the format order number. The order number in this dataset corresponds with that in the previous once decoupled from the date and time.

order\_products: Details about the products in each order in the following format: {product\_id: [product\_price, product\_quantity], ...}.

Each product is sold either at its original price or a 20% discount.

### Commonalities

Both datasets share the customer\_id field and order\_id information with the order number.

Each combination of order ID and customer ID is unique. This is because each order ID is unique, whereas customer IDs may be repeated across multiple orders.

Note that the order and customer IDs across the two files are consistent.

Whenever you have information about either one of the IDs, it is correct.

### Overview of Data Cleaning Steps

#### Data Quality and Cleaning Guidelines

Order and Customer IDs: Entries are always correct when not NULL, and NULL values should be tried to be filled in wherever possible using data from elsewhere.

Date and Time Fields: Entries with incorrect values should be removed.

Product Quantities and Product IDs: Always correct unless marked as NULL, which indicates missing values.

Product Prices: Each product ID is associated with a unique price. For some orders, the original unique price for a given product ID is discounted at 20% so that the discounted price is what is shown for those orders. However, for orders where the price is not discounted, sometimes there are junk or NULL values instead of the correct original price. Junk or NULL values in the product

prices should be replaced with the original price (the discounted price should be left as is wherever it is shown but not added in elsewhere).

Tips for Handling Junk and NULL Values; duplicates

Examine common values in each column to identify patterns and potential corrections.

Attempt to fix junk or NULL values using information elsewhere in the data before considering row deletion.

Date time fields can have incorrect fields that are not correctable, discard the affected rows and values to maintain data integrity.

Check for duplicates at every stage

---

ConfirmColabCode **Please confirm that you used Colab to complete this assignment and that all the code is in the Colab Notebook.**

Yes, all my code for this assignment is in the Colab Notebook. (1)

---

CodingTaskAnswers **Enter your answer from the task in a comma separated list here (Ex : "193738, 129490, 102948, 109812, 892738") or leave blank if you did not finish**

---

---

Page Break

*Display This Question:*

*If If Enter your answer from the task in a comma separated list here (Ex : "193738, 129490, 102948, 109812, 892738") or leave blank if you did not finish Text Response Is Empty*

**CodingTaskPostAnswer If you were not able to enter your answer before the timer, please enter your answer from the task in a comma separated list here (Ex : "193738, 129490, 102948, 109812, 892738") or leave blank if you did not finish**

---

---

Page Break

---



CodingGPTConvConf **Confirm you used ChatGPT to help complete your task**

- Yes I used ChatGPT (1)
- No I did not use ChatGPT (0)

---

*Display This Question:*

*If Confirm you used ChatGPT to help complete your task = No I did not use ChatGPT*

CodingGPTConvWhy Explain why you did not use ChatGPT when you were instructed to. Make sure to use it for any remaining task where you are instructed to.

---

---

CodingGPTConvLink **"Share" your ChatGPT conversations with us, so we can better understand how ChatGPT assisted you with the tasks. Only include conversations that you used for the task you just completed. If you used multiple ChatGPT conversations for the task, please share a link to each one.**

***NOTE:*** Please do not delete your conversations for a week or so, so we can make sure to collect the data we need to from them. If you delete the conversation on your side, we will no longer be able to view your shared links.

The screenshot below shows how you can Share your conversation. Select the 3 dots on the Individual Conversation tab, and select "Share", then copy the link and paste it below.

---

---

---

---



CodingGoogleConf **Did you use Google to help complete your task?**

- Yes (1)
  - No (0)
- 

CodingOtherTools **Please explain any other tools you used to complete your tasks. Include the name of the tool used, and how you used it to assist you.**

*Be as specific as you can.*

---

---

---

---

---

---

End of Block: Coding Task

---

Start of Block: Break1

TimerTaskBreak Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)

---

*Display This Question:*

*If Task\_Counter = 1*

TimerBreakText **Please take at least a 10 minute break. Note that the survey will auto-advance to the next section in 30 minutes. When you are ready to start the next task,**

**click the arrow below to continue.**

**The next Task should be completed within 90 minutes**

End of Block: Break1

---

Start of Block: Statistics Task

StatsInstructGPT **Please read through the instructions to complete the next task!**

You may spend **up to 90 minutes** on this task. At the 90-minute mark, the form will **automatically submit and move you to the next portion of the study, regardless of your progress**. You may choose to advance prior to the 90 minutes if you have fully completed the task.

**Use ChatGPT Enterprise** to perform this task **in whatever way you like (uploading files, copying and pasting the questions or errors and results)**. Access ChatGPT Enterprise by going to <https://chat.openai.com/> and log in using your BCG login.

**We highly encourage you to do your best to complete this task.** It might be challenging sometimes! But you're going to get CDC credit just for putting in an honest effort and if you are in the **top 50th percentile of your group** we'll notify your CDA about how well you did! We truly appreciate your effort.

-----  
Page Break

---

TimerStats Timing

First Click (1)

Last Click (2)

Page Submit (3)

Click Count (4)



StatsTaskOnLoadTime StatsTaskOnLoadTimeTracker

---

---

StatsInstructions **Instructions:**

For this task **USE ChatGPT** enterprise version by accessing <https://chat.openai.com/> using your BCG login to perform the task below and answer all the questions. **However, do not send any images to GPT and refrain from copying and pasting the exact question.**

---

StatsQ1 **Question 1**

**The following is the first five rows of data containing financial and demographic information about domestic partners who have co-purchased a home in the last several years.**

**Please note that the following table is illustrative and represents a snapshot sample of the data to solve this problem. All the information you need to solve the problem is contained within this snapshot.**

Age 1	Age 2	Income 1	Income 2	Borough	ZIP Code	Date	Price	Mortgage
39	37	270000	180000	Manhattan	10076	1 January 2016	1,125,000	Yes
NULL	38	445000	670000	Manhattan	10025	1 January 2016	2,249,000	Yes
27	29	145000	225000	Queens	11106	2 January 2016	900,000	Yes
33	NULL	90000	76000	Brooklyn	11203	2 January 2016	415,000	Yes
68	55	78000	450000	Bronx	10474	2 January 2016	3,399,000	No

---

StatsQ1.1 You have been tasked with predicting based on demographics and price whether a mortgage was taken out to by the house. You prompt ChatGPT for detailed instructions on how to do this, and ChatGPT recommend using a logistic regression model. It recommends the following steps (the text in blue is the ChatGPT output we are referring to).

---



StatsQ1.1.A 1. **Understand Your Dataset Explore and Preprocess:** Start by exploring your dataset to understand the features available and their types (numerical, categorical). Clean the data by handling outliers and possibly irrelevant features. Preprocessing steps like encoding techniques (e.g., one-hot encoding) might be necessary for categorical data. Ensure that your dataset does not have missing values. You can either fill them in with a strategy (mean, median, mode) or remove the rows/columns with missing values, depending on the situation.



**Which of the following are among the steps you could take to address this point?  
Select all that apply.**

- Plot the distribution of each of the numerical variables and remove rows with outliers from this dataset (1)
- One-hot encode the "Borough" variable (2)
- Investigate relationships between variables (3)
- Convert date to a numerical variable (4)
- One-hot encode the ZIP code variable (5)
- One-hot encode the age variables (6)
- Bin the ZIP codes by neighborhoods and do not process further (7)
- Bin the ZIP codes by neighborhoods and one-hot encode (8)
- Check columns with null values and remove those with >80% missing values (9)
- Impute NULL values by using a summary statistics or by developing a simple model that predicts those values based (10)

---

**StatsQ1.1.B 2. Split the Data**      **Train-Test Split:** Divide your dataset into a training set and a testing set (commonly a 70-30 or 80-20 split) to evaluate the model's performance on unseen data. **3. Train the Model**      **Training:** Use the training dataset to train your model, adjusting parameters as needed. For complex models, consider using cross-validation to fine-tune hyperparameters and prevent overfitting. **4. Evaluate the Model**      **Performance Metrics:** Evaluate your model on the test set using appropriate metrics such as accuracy, precision, recall, F1 score, and the ROC-AUC curve. These metrics will help you understand how well your model is performing in terms of both its ability to predict mortgages correctly and its robustness against false positives or negatives.      **What issue necessitates**

**using all these metrics? Which of the above steps is affected by this issue and how?**  
(Answer in 100 words or less – bullet points ok)

---

StatsQ1.1.C **Would you change the order of any of the above steps (i.e., steps 1-4)? Why or why not?**

(Answer in 100 words or less – bullet points ok)

---



StatsQ1.2 **You want to try a k-Nearest Neighbors model. Which of the following are not required (although recommended) for logistic regression, but absolutely necessary for k-Nearest Neighbors? Select all that apply.**

- Transform numerical variables (e.g. log) (4)
- Make sure there are only two classes to predict (5)
- Convert Mortgage column from string to binary (6)
- Standardize numerical variables (7)
- Impute the missing age with the other age in the same row (8)
- One-hot encode the appropriate variables (9)



StatsQ1.3 **You also try a decision tree model for the same classification problem, to compare performance. You realize your model is performing quite poorly on both**

**training and validation sets. You double-check the code and there are no bugs. What could be causing this problem? Select all that apply**

- Your model is underfit (4)
- Your model is overfit (5)
- The learning rate hyperparameter is too small (6)
- The learning rate hyperparameter is too large (7)
- The decision tree is too shallow (8)
- The decision tree is too deep (9)
- None of the above (10)



**StatsQ1.4 Next, you have been instructed to predict the price based on the other variables, and this time you have been instructed to use linear regression. Following instructions from ChatGPT, you perform a basic linear regression. You notice that your R2 value is too low. You prompt ChatGPT for suggestions on how to diagnose the**

**problem, and it is recommended that you check the residual plots. You notice that the residual plot does not appear random. What could this mean? Select all that apply.**

- The observed values of your dependent variable are independent from each other (4)
  - Your model is missing an important variable (5)
  - There is some interaction between your variables (6)
  - A higher order term might be required in your regression (7)
  - Variance of the residual is the same for any value of X (8)
- 

**StatsQ1.5 For the following residual plots, what could be the characteristics of or issues with the data or model that are corresponding with these results (choose from the list provided for each image)? It is possible that more than one characteristic or issue applies to any given image, and it is possible that a characteristic or issue may apply to more than one image.**

---



StatsQ1.5.A **Plot A**

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
- 

StatsQ1.5.B **Plot B**

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
-

StatsQ1.5.C **Plot C**

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
- 

StatsQ1.5.D **Plot D**

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
-

StatsQ1.5.E Plot E

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
- 

StatsQ1.5.F Plot F

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
-

StatsQ1.5.G **Plot G**

**Characteristic or issues choices:**

- No characteristic or issue is apparent (1)
  - Heteroscedastic data (2)
  - Outliers (3)
  - Response variable requires transformation (4)
  - A higher order variable might be required (5)
- 

StatsQ1.6 **You are asked to train a new model to predict price on the newest version of the dataset. In this version, there are several more fields collected with demographic information and financial information of the couples. However, this data is only from the last month. Which of the following steps recommended by ChatGPT could be beneficial to take to address some of the issues that are likely to arise because of this? Select all that apply.**

- Perform PCA (4)
  - Use a neural network instead of linear regression (5)
  - Use a regularized model instead of linear regression (6)
  - None of the above (7)
- 

StatsQ2.A **Question 2**

**You are asked to prepare a simple linear model to classify the following points into class 1 (black dots) and class 2 (white dots). What is the best empirical risk of this model that you can achieve with 0-1 loss?**



*Justify your answer and show your working steps.*

---

---

---

---

---

---

StatsQ2.B ChatGPT has run 3 classifiers on your data and provided a visual output, but not specified which models yielded which output. For each of the three images, name a classifier that could create the boundary represented by the solid black line, and one that could not (class 1 is the orange dots, and class 2 is the blue dots). You can ignore the dashed line, you can use the metrics on the bottom-left but you do not need them.

*Justify your answer.*

---

---

---

---

---

---

StatsQ3 Question 3:

Imagine you're a logistics manager and one of your delivery trucks has gone missing. You believe it lost its signal while on either Route A or Route B, with a 65% and 35% chance of being on each route respectively. Based on the coverage area of these routes, if the truck is on Route A and you search for a day, there's a 45% chance you'll find it. However, if it's on Route B and you search for a day, the probability of locating it is 75%.

---

**StatsQ3.A** If you only had one day to search for the truck, on which route would you focus your search efforts in order to maximize your chances of finding it?

*Explain your choice and break down your calculations.*

---

---

---

---

---

---

**StatsQ3.B** Assume that you made the rational decision on the first day, but didn't manage to locate the truck. The truck remains at the position that it was originally lost at and has not been moved. You have another day committed for search - has your initial idea of which route the truck is on changed? Where should you search now?

*Explain your choice and break down your calculations.*

---

---

---

---

---

---

Page Break



StatsGPTConvConf **Confirm you used ChatGPT to help complete your task**

- Yes I used ChatGPT (1)
- No I did not use ChatGPT (0)

*Display This Question:*

*If Confirm you used ChatGPT to help complete your task = No I did not use ChatGPT*

StatsGPTConvWhy Explain why you did not use ChatGPT when you were instructed to. Make sure to use it for any remaining task where you are instructed to.

---

StatsGPTConvLink **"Share" your ChatGPT conversations with us, so we can better understand how ChatGPT assisted you with the tasks. Only include conversations that you used for the task you just completed. If you used multiple ChatGPT conversations for the task, please share a link to each one.**

***NOTE:*** Please do not delete your conversations for a week or so, so we can make sure to collect the data we need to from them. If you delete the conversation on your side, we will no longer be able to view your shared links.

The screenshot below shows how you can Share your conversation. Select the 3 dots on the Individual Conversation tab, and select "Share", then copy the link and paste it below.

---

---

---

---



StatsGoogleConf **Did you use Google to help complete your task?**

Yes (1)

No (0)

---

StatsOtherTools **Please explain any other tools you used to complete your tasks. Include the name of the tool used, and how you used it to assist you.**

*Be as specific as you can.*

---

---

---

---

---

End of Block: Statistics Task

---

Start of Block: Problem Solving Task

PSInstructionsGPT **Please read through the instructions to complete the next task!**

You may spend **up to 90 minutes** on this task. At the 90-minute mark, the form will **automatically submit and move you to the next portion of the study, regardless of your progress**. You may choose to advance prior to the 90 minutes if you have fully completed the task.

**Use ChatGPT Enterprise** to perform this task in whatever way you like (uploading files, copying and pasting the questions or errors and results). Access ChatGPT Enterprise by going to <https://chat.openai.com/> and log in using your BCG login.

**We highly encourage you to do your best to complete this task.** It might be challenging

sometimes! But you're going to get CDC credit just for putting in an honest effort and if you are in the **top 50th percentile of your group** we'll notify your CDA about how well you did! We truly appreciate your effort.



Page Break



TimerPS Timing  
First Click (1)  
Last Click (2)  
Page Submit (3)  
Click Count (4)

---



ProbsTaskOnLoadTime ProbsTaskOnLoadTimeTracker

---

PSInstructions **Instructions:**

For this task, **USE ChatGPT** enterprise version by accessing <https://chat.openai.com/> using your BCG login. Feel free to work with ChatGPT in whatever way you like (uploading files, copying and pasting the question or results).

---

PSQuestion **Problem Solving Task**

**QUESTION**

Imagine you are staffed on a case where you must implement a data-driven strategy for sports investing. You are given a dataset containing records of 45,360 international football matches, spanning from the inaugural official match in 1872 through to the year 2024. The competitions range from the FIFA World Cup, FIFI Wild Cup, to ordinary friendly games. All matches are men's senior internationals, excluding Olympic Games, matches involving B-teams, U-23, or league select teams.

Your task (make sure to describe and document your approach and your findings):

**Develop and implement a method for quantifying how predictable each match result was.** *You can solve this problem however you like, using any analytics platforms at your disposal (e.g. Excel, Alteryx, Python).* Explain in detail each step you took for your approach and justify. What was the most surprising match result in this dataset, based on your method? Return a .csv or Excel file containing four columns – the match date, the home team, the away team, and **your numerically determined match result predictability using the above method** for each match in the dataset

Keep in mind you have 90 minutes to complete this task. Time box and make sure you return a final answer.

**DATASET INFORMATION**

The `results.csv` file encompasses columns for:

- `date` - the match date
- `home\_team` - the home team's name
- `away\_team` - the away team's name
- `home\_score` - home team's score at the end of the match, including extra time but excluding penalties
- `away\_score` - away team's score at the end of the match, including extra time but excluding penalties
- `tournament` - tournament name
- `city` - the city or locality of the match
- `country` - the country hosting the match
- `neutral` - a TRUE/FALSE indicator of whether the venue was neutral

**Assume that the result as shown in this dataset (win or tie) is the entire result – there are some cases of penalty shootouts and goals scored from penalties, but for complexity, we will ignore those for this exercise.**

For clarity, current names are used for both teams and countries in historical matches. For example, an 1882 match featuring the team then known as Ireland against England is listed under Northern Ireland, reflecting the modern successor of the 1882 team. Country names are recorded as they were at the time of the match, but discrepancies between team and country names (e.g., Ghana vs. Gold Coast) are accounted for, with the `neutral` column marking such matches as non-neutral to clarify they were played at home.

Data: [results.csv](#)

---

---

---

---

---



PSUpload **Upload your csv or Excel file here:**

-----

Page Break

\_\_\_\_\_





ProbsGPTConvConf **Confirm you used ChatGPT to help complete your task**

- Yes I used ChatGPT (1)
- No I did not use ChatGPT (0)

*Display This Question:*

*If Confirm you used ChatGPT to help complete your task = No I did not use ChatGPT*

ProbsGPTConvWhy Explain why you did not use ChatGPT when you were instructed to. Make sure to use it for any remaining task where you are instructed to.

---

ProbsGPTConvLink **"Share" your ChatGPT conversations with us, so we can better understand how ChatGPT assisted you with the tasks. Only include conversations that you used for the task you just completed. If you used multiple ChatGPT conversations for the task, please share a link to each one.**

***NOTE:*** Please do not delete your conversations for a week or so, so we can make sure to collect the data we need to from them. If you delete the conversation on your side, we will no longer be able to view your shared links.

The screenshot below shows how you can Share your conversation. Select the 3 dots on the Individual Conversation tab, and select "Share", then copy the link and paste it below.

---

---

---

---

### **D.3 Post-Experiment Survey**

## (a) Post-survey

Now that you've completed the tasks, how would you rate your current level of focus and energy for completing this survey?

- Very high – I'm fully focused and ready
- Somewhat high – I feel quite prepared and alert
- Neutral – I'm neither tired nor particularly energized
- Somewhat low – I'm a bit tired or distracted
- Very low – I'm already feeling quite fatigued or unfocused

Please answer the below questions to the best of your knowledge

***PLEASE DO NOT USE CHATGPT, OTHER LLM OR ANY OTHER SEARCH ENGINE (e.g., Google) TO ANSWER THESE QUESTIONS***

1. Suppose we have a 'test\_group' column in our dataframe (df) which has the values 'treatment' and 'control'. Which of the following code snippets will give us a dataframe filtered only to have the rows which correspond to 'treatment'? Select all that apply.
  - `df = df['treatment']`
  - `df = df[df['treatment']]`
  - `condition = df['test_group'] = 'treatment'`
  - `df = df[condition]`
  - `df = df['test_group'] = 'treatment'`
  - `df = df['test_group'] == 'treatment'`
  - `condition = df['test_group'] == 'treatment'`
  - `df = df[condition]`
  - `condition = df['test_group'] != 'control'`
  - `df = df[condition]`
2. If a coin is tossed 3 times, what is the probability of getting heads every time?
  - 1 out of 2
  - 1 out of 4
  - 1 out of 6
  - 1 out of 8
3. Distance-based algorithms are not affected by scaling
  - True
  - False
4. Which of these techniques can be used to handle missing data in categorical features? Select all that apply.
  - Removing rows having missing data
  - Replacing missing values with the most frequent category
  - Replacing missing values with the mean
  - Replacing missing values using predictive algorithms like classifiers
  - Replacing missing values using predictive algorithms like regressors

5. You are given a dataset of logos of famous companies , and you have to predict whether the review contains alphabets or not. Under which category does this problem fall? Select all that apply.
- Classification
  - Regression
  - Clustering
  - Natural language processing

1. This set of questions tests your ability to predict ("forecast") how well GPT-4 will perform at various types of questions. (In case you've been living under a rock these last few months, GPT-4 is a state-of-the-art "AI" language model that can solve all kinds of tasks.)

How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

Develop an HTML page with JavaScript and canvas to draw a representation of the US flag that rotates 90 degrees clockwise each time it is clicked. ()



How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

Here is some data about cities in Japan that I copied from Wikipedia. Based on this data, which cities have an even-numbered population?



City (Special Ward)	Prefecture	Population	Area (km <sup>2</sup> )	Density (per km <sup>2</sup> )	Founded
Special wards of Tokyo	Tokyo	9,375,104	621.81	13,890	
Yokohama	Kanagawa	3,732,616	437.38	8,500	1889-04-01
Osaka	Osaka	2,691,185	222.30	11,900	1889-04-01
Nagoya	Aichi	2,327,557	326.45	6,860	1889-10-01
Sapporo	Hokkaido	1,976,257	1,121.26	1,763	1922-08-01
Fukuoka	Fukuoka	1,588,924	340.96	4,515	1889-04-01
Kawasaki	Kanagawa	1,531,646	142.70	9,626	1924-07-01
Kobe	Hyōgo	1,524,601	552.23	2,772	1889-04-01
Kyoto	Kyoto	1,464,890	827.90	1,800	1889-04-01
Saitama	Saitama	1,324,854	217.49	5,483	2001-05-01
Hiroshima	Hiroshima	1,199,391	905.13	1,286	1889-04-01

How likely is GPT-4 to solve this question correctly?


0 10 20 30 40 50 60 70 80 90 100

I'm at a restaurant with a \$10 bill and want to use it exactly on some of the following items. Which ones should I buy: steak \$5.23 fries \$1.24 shake \$2.48 salad \$4.87 salmon \$4.13 cake \$1.00 ()




How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

<p>Can you help me answer the following crossword clues. 1. "Lamented, in a way" (4 letters) 2. "Princess's irritant in a classic fairy tale" (3 letters) 3. "Bobbie Gentry's "___ to Billie Joe"" (3 letters) 4. "Leave no way out" (4 letters) 5. "Expression of false modesty from a texter" (4 letters) ()</p>	
--	--

How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

<p>Who lost the Super Bowl two years after Pan-Am filed for bankruptcy? ()</p>	
--	--


How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

<p>Write out the word "hello" as an ascii art drawing with # and _ ()</p>	
---	---

How likely is GPT-4 to solve this question correctly?

0 10 20 30 40 50 60 70 80 90 100

<p>What is the best next move for O in the following game of Tic Tac Toe?</p> <pre> - .   O ----- .  O   X ----- X .   X </pre>	
---	--

**2. How did you find the use of Generative AI? Was it easy or difficult? Did it give you the answers you were looking for?**

- The use of ChatGPT was easy and provided me with all the answers I was looking for
- The use of ChatGPT was easy and provided me with most the answers I was looking for
- The use of ChatGPT was easy, but did not provide me with most the answers I was looking for
- The use of ChatGPT was difficult, but provided me with all the answers I was looking for
- The use of ChatGPT was difficult, but provided me with most the answers I was looking for
- The use of ChatGPT was difficult and did not provide me with most the answers I was looking for

**Next, please indicate the extent to which you agree or disagree with the following statements :**

	Strongly Agree	Somewhat Agree	Neutral	Somewhat Disagree	Strongly Disagree
Generative AI helps me feel valuable in my role	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI elevates how important I feel my job is for society	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI elevates my professional status and level of influence within my organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI helps me feel more competent in my role	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI enables my ability to execute tasks and reach desired outcomes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI enables my ability to execute data analytics tasks and reach desired outcomes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI increases the value I place on my expertise and skill cultivation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generative AI increases my level of autonomy in making individual decisions in my role	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Generative AI helps me be more confident that I will meet my project managers expectations

Using Generative AI helps me stay aligned with my project managers expectations

Generative AI enables me to do what I really want to do in my role

I believe using Generative AI will contribute to the betterment of others in my work

I see Generative AI as my coworker

Generative AI will change the dynamic in my team

Generative AI improved how I perceive my role in the organization

I would recommend Generative AI to other consultants

I am proud of BCG's approach to Generative AI adoption within the firm

I believe BCG is at the leading edge of the Generative AI revolution

My managers and supervisors will expect more output from me because of Gen AI

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Problem solving and insights' dimension

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Communication and Presence' dimension

Sustained use of ChatGPT for data science would have the potential to make me a better consultant in the 'Practicality and Effectiveness' dimension

**Rate how helpful you think Generative AI tools are for these use cases (Rating 1-7; with ability to say "I don't know")**

**Experience with GenAI**



	1	2	3	4	5	6	7	I don't know
Brainstorm ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing code for data analytics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing code for data visualizations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learning how to use excel for data analysis and visualizations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identifying which machine learning models to use for a project	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understanding the statistical significance of a result	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing code for data cleaning and preparation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**3. GenAI benefits - In a few words, what do you think will be the biggest benefits of Generative AI for you?**

**4. GenAI risks - In a few words, what do you think will be the biggest risks of Generative AI for you?**





**5. On a scale of 1-7, where 1 = "Not at all" and 7 = "Extremely", please rate the following ...**

	Not at all		Neither		Extremely		
0	1	2	3	4	5	6	7

How confident are you in your ability to contribute to data science projects?	
To what extent do you believe understanding data science concepts is important in the role of a BCG A/C?	

6. Given the capabilities of Generative AI, do you see the role of associates and consultants evolving in the next 5 years? If so, how?

7. Finally, answer the following questions on a scale of 0 to 10, where 0 is "Do not enjoy at all" and 10 is "enjoy to a great extent"

	Do not enjoy at all	Neutral	Enjoy to a great extent								
	0	1	2	3	4	5	6	7	8	9	10
How much do you think your coworkers enjoy their work?											
How much do you think your coworkers enjoy using ChatGPT for their work?											
How much would you enjoy doing more data analysis at work with the help fo ChatGPT?											
How much would you enjoy being tasked with data science tasks with the help of ChatGPT?											

Generally speaking, would you say that most people can be trusted, or that you need to be very careful in dealing with people?








- Most people can be trusted
- You need to be very careful in dealing with people
- Don't know

**Please indicate the extent to which you agree or disagree with the following statements :**

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
ChatGPT can be trusted to give you correct information when researching a new topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ChatGPT can be trusted to do quantitative analysis for you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ChatGPT can be trusted to clean data for you with minimal guidance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ChatGPT can be trusted to help you learn to do new things (e.g. use a new type of software)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

On a scale of 1-7, where 1 = "Not at all" and 7 = "Extremely", please rate the following ...

Not at all		Neither			Extremely	
1	2	3	4	5	6	7

How confident are you in your ability to do data cleaning in python using ChatGPT as your guide?	
How confident are you in your ability to do quantitative analysis using ChatGPT as your guide?	
How confident are you in your ability to learn new skills with ChatGPT as your guide?	
How confident are you in identifying factual inaccuracies in ChatGPT's responses?	
How confident are you in judging the relevance of ChatGPT's responses to your questions?	
How confident are you in assessing the clarity and understandability of ChatGPT's output?	
How confident are you in evaluating the completeness of ChatGPT's answers to your queries?	

## 8. Investment Game

Next, you'll have an **exciting opportunity** to play a game with ChatGPT as a second player (**yes, ChatGPT can play games and make decisions!**) and **earn points that you will be able to redeem for exciting rewards!**

Note that **you will not see ChatGPT for this question**. Instead, we will email you with ChatGPT's response and with the reward options! The more points you have at the end of the game – the better your options will be!

---

Here's how the game works:

You (the Investor) and ChatGPT (the Responder) **will be given 100 tokens each**

As the Investor, you will have the opportunity to **pass some, all or none of your tokens to ChatGPT (the Responder), as you like**

Whatever amount you decide to give, **we will triple it and pass it to ChatGPT**

For example, if you decide to pass 0 tokens, we will give ChatGPT  $3 \times 0 = 0$  and so it will have 100 tokens;

if you decide to pass 50 tokens, we will give ChatGPT  $3 \times 50 = 150$  tokens so it will have a total of  $150 + 100 = 250$  tokens;

if you decide to pass 100 tokens, we will give ChatGPT  $3 \times 100 = 300$  tokens so it will have a total of  $300 + 100 = 400$  tokens

After that, **ChatGPT (the Responder) will decide how many of the tokens it will give back to you**

**In case this impacts your answer – here are the instructions we've given to ChatGPT:**

*"I would like to play the investment game with you. I'll be the investor and you will be the responder. As a starting point, you and I will each have 100 tokens. Whatever amount I decide to invest on you, there's a middle person who will triple that amount before passing it over to you. At that time, you will decide how much to pass back to me based on how much I invested and your total endowment. As a response, I need 4 numbers, the number I decided to pass over to you (label it "Investment amount"), the number that you will return back to me (label it "ChatGPT return"), total I have (label it "Total amount investor has") and total you've (label it "Total amount ChatGPT has")."*

*Please make sure that the total amount you and I have at the end of the game sums up to the amount I've left, the amount that was tripled by the middle person and the amount you had at the beginning of the game. Also ensure that every time we play, we start with a fresh endowment of 100 tokens each."*

To make sure, you have a handle of the game, let's assume you decided to pass 20 tokens, how many total tokens does ChatGPT have after the researchers tripled the amount?

To make sure, you have a handle of the game, let's assume you decided to pass 50 tokens, how many total tokens does ChatGPT have after the researchers tripled the amount?

To make sure, you have a handle of the game, let's assume you decided to pass 80 tokens, how many tokens does ChatGPT have after the researchers tripled the amount?

Okay, let's play! **(Please do not use ChatGPT for this game, we will email you the results with the reward options)**

You now have 100 tokens and ChatGPT has 100 tokens. Don't forget, whatever number you decide to invest, we will triple it and pass it to ChatGPT. At that point, ChatGPT will decide how much to return back to you.

**How many tokens would you like to pass to ChatGPT? (Please enter a number between 0 and 100)**

- Next, could you guess how many tokens ChatGPT will give back to you. You will receive a bonus for a good guess (if your guess is within 5 tokens of the actual number), you will earn additional 10 tokens to redeem.

**How much do you think ChatGPT has decided to give back to you? (The number should not be more than ChatGPT's endowment -  $[3 * (\text{how much you decided to send})] + 100$ )**

End of survey

We thank you for your time spend participating in this study. Your response has been recorded. You might be selected to participate in a short follow-up interview.



## **E Post-Experiment Interview**

### **E.1 Understanding ChatGPT and LLM Tools**

- Basic Familiarity with ChatGPT
  - Can you explain what an LLM tool (e.g., ChatGPT) is and how it works at a high level?
  - How do you stay updated with the advancements in LLM tools like ChatGPT?
- Practical Applications in Data Science
  - Can you share a specific instance where you used ChatGPT or any other LLM tool to solve a data science problem? What was the outcome?
  - How do you evaluate the reliability of solutions provided by the tool in data science tasks?

### **E.2 Upskilling and Continuous Learning**

- Learning New Concepts and Techniques
  - How have you used ChatGPT or any other LLM tool to learn new concepts or techniques in data science or other topics?
  - Discuss a concept you better understood thanks to an explanation from ChatGPT. How did it compare to other learning resources?
- Staying Current with Industry Trends
  - In what ways have LLM tools such as ChatGPT helped you stay current with data science trends and industry best practices?
  - Can you give an example of an industry trend or technology you explored through interactions with ChatGPT (or other tools)?

### **E.3 Problem-solving and Project Acceleration**

- Enhancing Productivity
  - How do you integrate ChatGPT into your data science workflow to increase productivity or efficiency?
  - What ChatGPT tools did you find helpful or unhelpful? Why or why not?

- Can you describe a scenario where ChatGPT significantly accelerated your project delivery or improved the outcome?
- Collaboration and Communication
  - Have you used ChatGPT to facilitate collaboration among team members or to communicate complex data science concepts effectively? Please provide examples.
  - How do you ensure that the use of ChatGPT in communication and collaboration adds value without compromising the quality of interaction?

## **E.4 Conclusion and Future Perspectives**

- Future of AI Tools in Data Science
  - How do you envision the role of AI tools like ChatGPT evolving in the data science landscape?
  - What skills or areas of knowledge do you think are crucial for data scientists to develop in light of the growing integration of AI tools in the field?
  - What other topics or areas outside data science do you think GenAI could be used to upskill? How?
  - Any last minute thoughts that you want to share?