NBS Working paper
13/2025

A Task-Based Approach to Generative AI: Evidence from a Field Experiment in Central Banking

Ales Marsal, Patryk Perkowski

NATIONAL BANK OF SLOVAKIA
EUROSYSTEM

www.nbs.sk

research@nbs.sk

This publication is available on the NBS website
www.nbs.sk/en/publications-issued-by-the-nbs/research-publications

ISSN 2585-9269 (online)

A Task-Based Approach to Generative AI: Evidence from a Field Experiment in Central Banking*

Aleš Maršál

National Bank of Slovakia

Patryk Perkowski

Yeshiva University

July 8, 2025

Abstract

We examine how generative AI impacts productivity across the task-based framework using a field experiment at the National Bank of Slovakia. In our experiment, we randomly assign generative AI access to central bank employees completing workplace tasks that mirror the theoretical task-based framework. Our results indicate that generative AI access leads to large improvements in both quality and efficiency for the majority of participants. We find a strong complementarity between generative AI and non-routine work, both on average and for most participants. We also find some support for generative AI as both cognitive-biased and specialist-biased, though smaller in magnitude than our tests of routine-biased. While workers in routine jobs experience larger individual performance gains, generative AI is less effective for the routine task content of their work. The mismatch between generative AI's task- versus worker-level impacts is economically large, and results from a simulation exercise suggest the organization can increase output by 7.3% by changing how workers are assigned to tasks in the presence of generative AI. Additionally, we find differences in how the benefits of generative AI relate to worker skills: low-skill workers benefit most in terms of quality while high-skill workers benefit in terms of efficiency. Our findings provide empirical support on generative AI and task-level complementarities, with important implications for how generative AI will impact workers, organizations, and labor markets more broadly.

Keywords: Generative AI, Worker Productivity, Central Banking, Field Experiment
JEL Classification: J24, M15, E58, C93, O33

*Aleš Maršál (ales.marsal@nbs.sk); Patryk Perkowski (patryk.perkowski@yu.edu). We are grateful to Pavel Gertler, Jan Klasco, Peter Kazimir, and Reiner Martin for their support in executing the field experiment. For helpful discussions and feedback, we thank seminar participants at the European Central Bank, the National Bank of Slovakia, the Czech National Bank, CUNY Queens College, CERGE-EI, and Wharton's AI and the Future of Work conference, as well as Pablo Hernández-Lagos, Philip R. Lane, Francesc Ortega, and Alev Yildirim. We thank Maria Rejdugova, Lubos Simoncic, Ha Anh Le, Sabina Marsal, Lea Gogova, Peter Lorinc, and Boris Fisera for research assistance and planning support. Maršál gratefully acknowledges research funding from the EU grant NPO "Systemic Risk Institute" number LX22NPO5101, and King Abdullah Petroleum Studies and Research Center (KAPSARC) for providing time and resources to work on this project during his employment. The views expressed herein are those of the authors and do not necessarily represent the views of the National Bank of Slovakia. IRB approval for this study was obtained from WCG IRB under protocol number 20241892. The experiment was pre-registered at https://osf.io/23jr6/.

1 Introduction

What impact will generative AI technologies have on labor demand, the design of jobs, and the labor market more generally? The canonical model for understanding how technological change impacts the labor market is the routine-biased technical change (RBTC) framework (Autor et al., 2003; Acemoglu and Autor, 2011a; Acemoglu and Restrepo, 2018). RBTC posits that new technologies enable the substitution of humans in "routine" tasks, which are repetitive, predictable, and codifiable, and complement worker productivity in "nonroutine" tasks, which require creativity, judgement, and interpersonal skills.[1] As a result, technological change leads to wage polarization, whereby demand falls for workers in routine jobs and increases for those in non-routine jobs. Empirical evidence from prior advances like the internet and communication technologies has aligned with predictions from the RBTC framework (Autor et al., 2003; Spitz-Oener, 2006; Goos et al., 2014, 2021; Michaels et al., 2014).

Despite the importance of the task-based framework in understanding the labor market impacts of new technologies, however, we lack empirical studies documenting task-based complementarities with generative AI. This is important given the wide-ranging capabilities of generative AI[2], the large percentage of workers that are likely to be impacted by the technology (Eloundou et al., 2023; Felten et al., 2023), and the rapidly improving capabilities of these models (Bubeck et al., 2023).

In this paper, we examine task-level complementarities between generative AI and various types of workplace tasks using a field experiment. Our setting is the National Bank of Slovakia (NBS), Slovakia's central bank. In the summer of 2024, bank employees completed a series of workplace tasks where we experimentally manipulated both access to generative AI and the task profile using the routine versus non-routine distinction for generalist tasks in Autor et al. (2003). We also included more specialist tasks that require domain expertise. Our design leverages both between-subjects variation (in the availability of generative AI conditional on a given task mix) and within-subjects variation (in the set of tasks where participants have access to generative AI) to test task-based complementarities

[1]Because routine tasks follow predictable, rule-based procedures, they are particularly amenable to substitution by machines. Once machines can replace such tasks reliably, firms can use machines to complete these tasks at a lower marginal cost than paying human wages, thereby substituting capital for labor. On the other hand, new technologies increase the productivity of non-routine workers by streamlining subtasks embedded in nonroutine roles that are more predictable and repetitive. New technologies have this effect because they help process large data sets, organize information, or handle repetitive tasks at speed and scale, while workers contribute higher-level thinking, insight, and flexibility. The resulting complementarity boosts overall productivity beyond what either the technology or the worker could accomplish in isolation.

[2]These include midlevel writing (Noy and Zhang, 2023), customer support (Brynjolfsson et al., 2023), consulting (Dell'Acqua et al., 2023), software engineering and (Cui et al., 2024), among others.

with generative AI.

Our experiment randomly assigned participants access to OpenAI's GPT-4o for a subset of tasks. We use GPT-4o because (i) OpenAI has the most widely adopted generative AI tools, as evidenced by its broad usage among workers both in general (Humlum and Vestergaard, 2024) and in our specific setting (Perkowski and Maršál, 2024); (ii) GPT-4o can be used across a diverse range of tasks, and (iii) GPT-4o permits a careful test of non-routine complementarity.[3] We find that providing access to GPT-4o significantly influenced tool uptake. Participants used GPT-4o for around 94 percent of the tasks when they had access to the tool, versus 0 percent without access. These results show that our treatment led to generative AI adoption across tasks, allowing us to evaluate the downstream impacts on task productivity.

Our results indicate that access to generative AI led to large productivity improvements across the entire bundle of tasks. In terms of the quality of responses, we find an average improvement of between 33 and 44 percent depending on the specification. In terms of efficiency, participants took on average 21 percent less time to complete the tasks when they had access to generative AI. Moreover, we find no correlation in generative AI's impacts on quality versus efficiency, indicating that improvements in quality did not come at the expense of lower improvements in efficiency. Leveraging our within-participant variation, we find that 94 percent of participants scored higher when they had access to generative AI versus when they did not. Access to generative AI had more heterogeneous impacts on efficiency: 80 percent of participants were more efficient when they had access to the tool, versus 13 percent who were less efficient.

We also find that the productivity impacts of generative AI are twice as large for specialist versus generalist tasks. Access to generative AI doubled performance in our specialized, domain-specific tasks, compared to 50 percent improvements for the generalist tasks. Lower-skilled workers and those with a larger share of routine job tasks experienced larger improvements on the generalist tasks, suggesting generative AI reduced inequality in productivity for these tasks. However, these patterns do not hold for specialist tasks. We find no differences in the impacts of generative AI on specialist tasks by baseline

[3]OpenAI's GPT-4o was launched in May 2024 – one month prior to the field experiment. Since then, OpenAI has launched newer generative AI tools such as GPT–o1. These versions introduce chain of thought reasoning, which systematically break down complex prompts into a sequence of intermediate steps or "thoughts" to bridge the query and the final output. Chain of thought reasoning is particularly helpful for non-routine tasks, which typically require higher-level reasoning. Thus, to the extent that we find a complementarity between generative AI and non-routine work, our results using OpenAI's GPT-4o will under-estimate the bias relative to newer generative AI tools, which would further increase productivity in non-routine tasks requiring higher-order reasoning.

skill or routine job share, and informal conversations with participants suggest high-skilled workers with domain expertise and prior generative AI exposure were able to effectively integrate generative AI to their work practices on the specialist tasks. Thus, while generative AI increased productivity for both generalist and specialist tasks, different types of participants benefited from the technology depending on the type of task.

Next, we leverage variation across our generalist tasks to shed light on the complementarity between generative AI and non-routine task mix. Our empirical test of complementarities comes from Brynjolfsson and Milgrom (2012), who show how a simple test statistic from performance equations captures the existence of complementarities. We estimate this test statistic using the experimental variation in generative AI and task types, both on average across our entire sample and also at the participant level.

We find empirical evidence of a complementarity between generative AI and non-routine work. Comparing sample averages, we find that access to generative AI increases performance by 24 percent on the set of routine tasks, versus 58 percent for non-routine tasks. When we apply the more formal test from Brynjolfsson and Milgrom (2012), the test statistic for complementarities is economically large and statistically significant. Across the entire sample, generative AI increases productivity by almost one more standard deviation for non-routine versus routine-tasks. At the participant level, this complementarity exists for around 80 percent of the sample, and only five percent of participants report a negative complementarity (i.e., higher performance increases for routine versus non-routine tasks) that is greater than one standard deviation.

We then benchmark this estimate using two other types of task variation. We test if generative AI is also cognitive-biased (versus manual) or specialist-biased (versus general) using the same test of complementarities. We find evidence of complementarities along both of these dimensions, though smaller in magnitude than our estimates using the routine versus non-routine distinction. Thus, while there is a complementarity between generative AI and both cognitive and specialist task mixes, the non-routine nature of work leads to the largest productivity improvements.

Lastly, we explore the distributional consequences and heterogeneous treatment impacts of generative AI access by participant characteristics. We find that the benefits of generative AI are more heterogeneous for quality than for efficiency. The benefits of generative AI on quality depend on the baseline skill and exposure to routine workplace tasks. Meanwhile, we find some evidence that those with prior experience with generative AI see larger efficiency improvements.

Our results also indicate that the benefits of generative AI depend on the baseline skill of participants. For generalist tasks, we find that lower-skilled participants experience larger increases in quality when given access to the tool. However, this is not the case for when considering efficiency as the outcome measure or for productivity in our specialist tasks. In fact, we find suggestive but noisy evidence that the highest-skilled workers experience the largest improvements in efficiency and larger increases in performance on the specialist tasks. These results indicate important interactions between generative AI and skills, though in more nuanced ways than previously thought.

Comparing our task-level versus participant-level estimates reveals a mismatch between the types of tasks where generative AI is most helpful and the types of workers who benefit most from generative AI. At the task level, generative AI leads to the largest productivity improvements in non-routine tasks. However, at the participant level, workers in roles with a high share of routine work benefit more from generative AI (across our collection of tasks). Taken together, these results indicate a mismatch between the types of workers who stand to benefit most from generative AI, and the types of tasks where generative AI is more helpful: workers in routine jobs derive larger benefits from generative AI, but generative AI is less effective for the task make-up of their work. We find the magnitude of this mismatch to be economically large. Results from a simulation exercise suggest that the organization can increase output by 7.3% through changes in how workers are allocated to tasks in the presence of generative AI.

Overall, our manuscript contributes to four related literatures. First, we contribute to the literature on the impacts of technological change on the labor market. Although various papers have used the task-based approach to theoretically model the impacts of technologies like generative AI (Acemoglu and Restrepo, 2019; Acemoglu, 2024; Acemoglu and Restrepo, 2024), the empirical literature on the labor market impact of generative AI is still growing.[4] Eloundou et al. (2023) and Felten et al. (2023) present estimates of the share of workers that are likely to be impacted by generative AI. Hui et al. (2024) and Demirci et al. (2024) analyze data from freelancer platforms to understand the impact of generative AI on demand in the freelancing labor market. We advance this literature by providing experimental evidence of the differential impacts of generative AI across the task-based framework. We document strong task-level complementarities between generative AI and non-routine work, a key

[4]These is also a corresponding literature on the labor market impact of robots (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020) and the impact of automation on team performance (Dell'Acqua et al., 2025).

parameter in models of how technological change impacts the labor market.[5]

Second, our results contribute to the literature on skill bias and the distributional consequences of generative AI. Prior work has documented that generative AI is skill-biased (but in the opposite direction of previous technological advancements) and leads to larger improvements for lower-skilled workers, thereby reducing inequality in performance (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Dell'Acqua et al., 2023). Our results reveal a more nuanced picture. In line with prior work, we find that lower-skilled workers see larger improvements in performance. These results, however, are limited to generalist tasks using the quality measure of productivity, and do not hold for efficiency or specialist tasks. Thus, our results suggest more subtle distinctions on skill bias and generative AI, and a less rosy picture on generative AI reducing skill-based inequality.

Third, our results also speak to a growing literature on the heterogeneous impacts of generative AI at work. Understanding how generative AI influences labor markets requires exploring which workers stand to gain with the adoption of generative AI. For example, previous work has documented a gender gap with generative AI, whereby women are less likely to know about or use generative AI (Humlum and Vestergaard, 2024; Kreacic and Stone, 2024; Otis et al., 2024). Our results rule out that this gap is due to unequal returns to generative AI by gender, as we find no gender differences in the impact of generative AI on either quality or efficiency for either generalist or specialist tasks. This suggests that the gender gap in generative AI is not driven by differences in the benefits of the technology, but rather barriers to initial adoption in the first place.

Lastly, we contribute to the growing literature that uses field experiments to understand the productivity impacts of generative AI across various occupations. These include midlevel writing professionals (Noy and Zhang, 2023), customer support agents (Brynjolfsson et al., 2023), consultants (Dell'Acqua et al., 2023), software engineers (Cui et al., 2024), and lawyers (Schwarcz et al., 2025). We contribute to this literature by examining the impacts of generative AI in central banking. Although central banks have been early adopters of machine learning techniques in areas such as financial supervision and forecasting (Doerr et al., 2021; Araujo et al., 2022; Aldasoro et al., 2024), we know less about the potential impacts of generative AI in this line of work. Perkowski and Maršál (2024) present survey

[5] We caution, however, that this complementarity is necessary but not sufficient for RBTC to occur at scale. For RBTC to fully materialize with generative AI, several market- and firm-level adjustments must occur, including reductions in the price of computing, changes in the relative demand for skills, changes in task composition and occupational structure, and complementary workplace adjustments from firms. The overall impact of generative AI on the labor market will depend on such adjustments, which are outside the scope of our field experiment.

data on how generative AI is used in three central banks. Our manuscript contributes to this literature by illustrating the positive productivity impacts of generative AI in this domain.

2 Experiment

2.1 Setting

Our field experiment was held at the National Bank of Slovakia (NBS), the country's central bank. The primary objective of the NBS is to maintain price stability in the Eurozone. To achieve this, the NBS formulates monetary policy, issues banknotes and coins, and ensures the smooth operation of payment and settlement systems. Additionally, the NBS manages Slovakia's foreign reserve assets, conducts foreign exchange operations and supervises financial markets. Since Slovakia's adoption of the euro in 2009, the NBS has been an integral part of the Eurosystem, collaborating with the European Central Bank and other eurozone national central banks to define and implement the monetary policy of the euro area.

The NBS employs around 1,100 employees across 13 departments who engage in a mix of routine and non-routine tasks. We display the organizational structure of the NBS in Appendix section B.1. Employees in every department typically engage in a mix of both routine and non-routine tasks. For instance, while the Cash Management and Statistics Departments primarily handle routine cognitive tasks, employees in these areas also perform analytical and interpersonal duties as needed. Similarly, smaller, specialized teams in departments like Research and Risk Management focus on non-routine cognitive and analytical tasks but may also undertake leadership or communication roles. Workers in departments like Facility Services perform a higher share of both routine and non-routine manual tasks.

2.2 Sample

With approval from the Board of Governors, we sent an email to all employees in March 2024. The email invited employees to participate in a two-stage research study on technology in central banking. The first stage (April 2024) included a detailed survey while the second stage (June 2024) included a two-hour workplace exercise where participants would complete job tasks. The email mentioned that participants would receive compensation for completing the study, and a bonus based on their performance in the tasks.

Overall, 160 employees completed stage one of the study and 101 completed both stages. Informal conversations with participants revealed that curiosity and the opportunity to apply generative AI to work-related tasks were the primary reasons for participation. This aligns with our survey evidence from phase one (Perkowski and Maršál, 2024), which identified the lack of concrete use cases as the main barrier to generative AI adoption within central banks. Many employees saw the experiment as a chance to explore practical applications of the technology in their jobs. Additionally, some participants were motivated by the monetary rewards offered. The 101 employees who completed both stages make up the sample for this field experiment, and Appendix section B.1 displays participation rates across the NBS's organizational structure.

Table 1 contains summary statistics for the sample. The sample is highly educated– 92 percent have a master's degree or higher. The average age is 40 years old, and 35 percent of participants are female. Participants have been employees at NBS for on average 8 years, and 5 percent manage or lead a team. The average participant spends 31 percent of their time working on analytical tasks, 26 percent on non-routine cognitive tasks, 18 percent of their time on routine cognitive tasks, and 16 percent on leadership and management tasks. Overall, the sample exhibits awareness of generative AI, though workplace rates of usage are lower. 95 percent have heard of generative AI before and 75 percent have used generative AI before, but only around half have used generative AI at work before.

While our sample is admittedly more highly educated than the average labor force, the task profile and skill heterogeneity inside NBS mirror those in other organizations. Participants in our sample span a broad range of roles, from facility coordinators to research economists. Moreover, worker roles feature both codifiable routine tasks such as data entry and document formatting, and also open-ended non-routine tasks such as forecasting and stakeholder communication. This task profile mimics the distribution of task types from Autor (2013).

2.3 Experimental design

Figure 1 outlines the structure of the experiment. Across two stages of the experiment, participants completed a series of workplace tasks that mimicked the work done at the central bank. Part one included general workplace tasks, while part two featured specialized workplace tasks. Across both general and specialized workplace tasks, we randomly assigned participants access to a generative AI tool.

2.3.1 Tasks

The first stage of the experiment focused on understanding the impact of generative AI on performance in general workplace tasks. We used the task classification system from Autor (2013) and designed a series of seven types of general workplace tasks that are common to modern-day work. These task types included: (i) routine cognitive; (ii) non-routine cognitive; (iii) routine manual; (iv) nonroutine manual; (v) analytical; (vi) leadership; and (vii) interpersonal and communication tasks. For each of these general task types, we created two specific tasks to allow us to vary the use of generative AI. For example, each participant would complete one of the general analytical tasks with the use of generative AI, and complete the other one without the use of generative AI. This led to a total of 14 tasks.

Our goal with the design of these general workplace tasks was two-fold. First, we wanted to have general tasks so they are applicable to large parts of the organization, and modern-day work more generally. Second, as specified in our pre-analysis plan, we were interested in understanding the impact of generative AI depending on task type in the spirit of RBTC. To that end, we designed these tasks to be broadly representative, to correspond to the respective categories in the Autor (2013) task taxonomy, and to be similar to work done at the organization.

Appendix B.2 lists out the specific general tasks that participants completed. For example, one of the routine cognitive tasks asked participants to proofread and edit a paragraph for grammatical and spelling mistakes. One of the analytical tasks asked participants to calculate the percentage of the workforce that would be subject to a hypothetical new remote work policy. Given the experiment was run using laptops, we were unable to have participants do purely manual tasks. However, we created versions of manual tasks that were able to be completed online. For example, one of the non-routine manual tasks asked participants to walk through the steps they would take in order to fix a broken printer in the office. Overall, participants had one hour to complete all 14 general workplace tasks, such that each task averaged about four minutes.

We additionally included a second stage of the experiment that focused on more specialized tasks.

In early work, we document that employees are less likely to use generative AI for more specialized tasks versus more general workplace tasks (Perkowski and Maršál, 2024). This may reflect the inability of these technologies to help with workplace tasks that require more skills, or it may reflect participant aversion to using generative AI for more advanced tasks. Indeed, informal conversations with participants revealed that even among generative AI users at work, the second stage of the experiment was the first time they used the technology for more specialized job tasks. Additionally, some perspectives emphasize domain-specific knowledge in the returns to generative AI at the firm level (see, for example, Tambe (2025)).

One difficulty in studying the impacts of generative AI on specialized workplace tasks is the heterogeneous positions occupied by participants in our sample. As Table 1 indicates, our participants were spread across numerous divisions at the bank. For feasibility purposes, we used the organizational structure of NBS to assign employees to respective blocks given their division at the bank. This led to a total of five blocks of employees, who perform similar tasks within blocks at the organization. These blocks were: (i) supervision (which includes supervision and financial stability, and supervision and financial consumer protection); (ii) monetary policy and research (which includes monetary policy and market operations, and research and statistics); (iii) financial markets (which includes payment systems and cash management, and risk management and settlement; (iv) information technology (which includes information technology, financial management, and operation activities), and (v) other (which included strategy and development, the Office of the Governor, human resources, internal audit, legal services, and the NBS foundation). We then worked alongside industry professionals to design highly-specialized tasks for the first four blocks. Given the heterogeneous composition of those in the "other" block, these participants completed another set of generalist tasks.

The specialized tasks were designed to require high level subject-matter expertise. For example, one of the tasks for participants in the monetary policy and research block was to create a monetary policy report for a fictitious country. Participants were given a dataset and asked to apply the Hodrick-Prescott (HP filter) to decompose GDP into cyclical and trend components, fit an ARIMA model to the HP trend component, forecast the trend component beyond the sample period, and develop a strategic plan for monetary policy adjustments over the next year.[6] In Appendix B.3, we list out the specialized tasks that participants had to complete during the second stage of the experiment,

[6]The one exception to this is the specialized tasks for those in the other block. Given the heterogeneity in this group, we used general workplace tasks for this group.

depending on their block. Participants completed two specialized tasks during this stage: one with the use of generative AI and one without. They had 30 minutes for each of the specialized tasks.

Overall, the experiment took around two hours. Participants were paid a base rate of around 5 percent of the average monthly salary at the NBS, with an additional performance bonus of up to 5 percent.

2.3.2 Random assignment

Our experiment uses both within-subjects (List, 2025) and between-subjects variation. While all participants completed 14 tasks, we randomly allowed the use of generative AI for a subset of these tasks. We then randomized access to generative AI across task categories within each subject. Some participants were eligible to use generative AI for the first set of the seven general tasks and one specialized workplace task, while others were eligible to use it for the second set and other specialized task. This design ensured that participants completed each of the task types from the task-based framework, both with and without generative AI, with variation across participants regarding who completed which task with or without generative AI.

The generative AI tool used in the experiment was OpenAI's GPT-4o, which was launched in May 2024 – one month prior to the field experiment. We use this generative AI tool for three primary reasons. First, OpenAI's ChatGPT is the most widely adopted generative AI tool, as evidenced by its broad usage among both workers in general (Humlum and Vestergaard, 2024) and workers in our setting specifically (Perkowski and Maršál, 2024). Using OpenAI's latest model helps ensure our results are relevant to how workers engage with the current landscape of generative AI tools and makes use of participants' existing familiarity with the technology. Second, and relatedly, GPT-4o is a generalist generative AI tool that can be used across various settings. While other specialised tools exist, GPT-4o has shown to be useful in diverse task applications (OpenAI, 2024). Third, this version of GPT permits a careful test of a complementarity between generative AI and non-routine work. Since our experiment, OpenAI has launched newer generative AI tools such as GPT–o1. These versions introduce chain of thought reasoning, which systematically break down complex prompts into a sequence of intermediate steps or "thoughts" to bridge the query and the final output. Chain of thought reasoning is particularly helpful for non-routine tasks, which often require higher-level reasoning. Thus, to the extent that we find a complementarity between generative AI and non-routine work, our results using OpenAI's GPT-

4o will under-estimate the bias relative to newer generative AI tools, which would further increase productivity in non-routine tasks requiring higher-order reasoning.

The experimental manipulation randomly assigned *access* to OpenAI's GPT-4o for a subset of participants and tasks. We collaborated with the information technology department of NBS to ensure that GPT-4o was accessible or inaccessible depending on the treatment condition. Participants in the treatment condition were informed that they had access to GPT-4o for the relevant tasks and were encouraged to use it if they found it helpful. We find strong rates of uptake for those in the treatment condition: 94 percent of those who gained access to GPT-4o for a given task used the tool during the task and sent on average 1.5 prompts (versus none in the control group; see section 3.1). Meanwhile, the IT department blocked GPT-4o access for those in the control condition. These participants completed the tasks relying solely on their own knowledge and resources, such as prior training, the internet, and other standard tools available in their typical work environment. One of the co-authors was present during the experiment to further ensure compliance in the control group.

2.4 Data

Our data come from three sources. First, we use the data collected from the tasks during the actual experiment itself. This includes the result submissions for each of the tasks plus the amount of time it took to complete the tasks. Second, we have detailed chat transcripts for each participant on their use of generative AI. This was collected directly through OpenAI's GPT-4o. Lastly, for each participant, we have detailed survey data collected over a month prior to the experiment itself, plus a few questions collected following the experiment. These include exposure to generative AI, workplace characteristics, and beliefs regarding generative AI.[7]

Our data is at the participant-task level, and two key outcomes capture performance in our setting. The first captures a subjective evaluation of task performance by evaluators. For each of the 14 tasks, we devised a zero to ten rubric to measure performance using the rubrics in Appendix sections B.2 and B.3. Each general task response was evaluated by three individuals (a PhD student, a post-doctoral fellow, and an employee at the bank), and we took the average of the evaluations across all three evaluators. Meanwhile industry professionals evaluated the responses to the specialized task. All evaluators were blind to treatment assignment. Our first performance measure is the quality of

[7] See Perkowski and Maršál (2024) for additional information on this survey.

submissions using these evaluation of task performance.

The second is an efficiency measure, which captures how long it took participants to complete each task. We use timestamps from submitting the tasks to generate the time it took participants to complete each task. Given that the general tasks were given as bundles of seven tasks, we cannot estimate efficiency for each general task. Instead, we have the amount of time it took to complete all seven tasks.

2.5 Empirical strategy

Our goal is to understand the impact of generative AI on productivity, both on average, across tasks, and across participants. We leverage variation across participants and task-types to estimate treatment effects at several levels.

2.5.1 Overall impact of generative AI on productivity

First, we can estimate the impact of generative AI on productivity across all participants and tasks using within-participant variation and the following equation:

$$y_{p,t} = \beta_0 + \beta_1 * GAI_{p,t} + \theta * P_p + \lambda * T_t + \epsilon_{p,t} \tag{1}$$

In this equation, p indexes participants and t indexes tasks. $y_{p,t}$ is a performance measure for participant p on task m. $GAI_{p,t}$ is a binary indicator that equals one if participant p had access to generative AI for task t, and zero otherwise. P_p is a matrix of participant-level fixed effects, T_t is a matrix of task-level fixed effects, and $\epsilon_{p,t}$ is the error term. We estimate equation 1 using robust standard errors clustered at the participant level. The coefficient of interest is β_1, which captures the average impact of generative AI on performance across all participants and tasks.

We can further use this within-participant variation to estimate participant-level impacts of generative AI on productivity. To do so, we drop the participant-level fixed effects from equation 1 and instead run the equation for each participant in our sample.[8] These participant-level estimates are noisier than those estimated across the entire sample, but provide some insights into the distributional effects

[8]This is possible because we have 14 observations per participant, half of which are with generative AI and half without. This allows us to estimate the productivity impact of generative AI for each participant in our sample.

of generative AI (for example, by plotting the distribution of treatment effects and its correlates).

We can additionally estimate the impact of generative AI on productivity by using between-participant variation. Some participants receive access to generative AI during the first set of tasks, while others do so during the second set of tasks. We can subset the data to only the first set of seven tasks, and compare outcomes for participants who did versus did not have access to generative AI for these tasks. This between-participant estimation is given by:

$$y_{p,t} = \beta_0 + \beta_1 * GAI_{p,t} + \theta * X_p + \lambda * T_t + \epsilon_{p,t} \tag{2}$$

This equation is similar to equation 1 except for two changes. First, we replace the matrix of participant fixed effects (P_p) with a vector of participant-level controls (X_p). These controls include participant demographics, previous exposure to generative AI, beliefs in the productivity impact, among others. Second, we only include observations for the first set of tasks. The coefficient of interest here is β_1, which measures the average difference in productivity across the first set of tasks, for participants who receive generative AI versus those who do not.

2.5.2 Complementarity tests

In addition to the overall impact of generative AI access on productivity in our setting, we test for complementarities between generative AI and task-type, namely non-routine tasks. Our empirical test of complementarities comes from Brynjolfsson and Milgrom (2012), who show how a simple test statistic from performance equations captures the existence of complementarities.

Let $f(a, b)$ be productivity in the organization, where a captures generative AI access and b captures non-routine task mix. In this set-up, $f(1, 1)$ captures the productivity on non-routine tasks when generative AI is accessible, $f(1, 0)$ captures the productivity on routine tasks when generative AI is accessible, $f(0, 1)$ captures the productivity on non-routine tasks when generative AI is not accessible, and $f(0, 0)$ captures the productivity on routine tasks when generative AI is not accessible. A complementarity between generative AI access and non-routine task content exists if $f(1, 1) - f(0, 0) \geq f(1, 0) - f(0, 0) + f(0, 1) - f(0, 0)$. Equation 10 from Brynjolfsson and Milgrom (2012) suggests using the test statistic $\kappa_p = f(1, 1) + f(0, 0) - f(1, 0) - f(0, 1)$ to test for complementarities.

We can estimate κ_p using regression. To do so, we estimate the following equation:

$$y_{p,t} = \beta_0 + \beta_1 * GAI_{p,t} + \beta_2 * NonRoutine_t + \beta_3 * GAI_{p,t} * NonRoutine_t + \theta * P_p + \epsilon_{p,t} \qquad (3)$$

In this equation, $NonRoutine_t$ is a binary indicator that equals one if task t is a non-routine task, and zero otherwise. β_1 captures the average impact of generative AI on productivity in routine tasks, while β_2 captures the average difference in performance for participants on non-routine versus routine tasks (without the presence of generative AI). The coefficient of interest is β_3, which captures the additional treatment effect of generative AI on non-routine versus routine tasks. β_3 corresponds to the test statistic κ_p[9]; a test finding $\beta_3 > 0$ provides evidence of a complementarity between generative AI and non-routine tasks. We estimate equation 3 with robust standard errors clustered at the participant level.

In addition to testing for this complementarity across the entire sample, we can also estimate it for each participant. While the previous equation estimates κ_p across the entire sample, we can also estimate κ_p for each participant. To do so, we drop the participant-level fixed effects ($\theta * P_p$) in equation 3, and instead estimate equation 3 for each participant. This returns an estimate of the test statistic for each participant, which allows us to examine the distribution of κ_p across the sample.

Additionally, we test for complementarities between generative AI and two other task attributes: whether the task is cognitive or manual, and whether the task requires specialized subject-matter expertise. These tests are similar to our tests using the routine versus non-routine distinction, except we replace the non-routine indicator and interaction with those for cognitive tasks and specialist tasks, respectively.

2.5.3 Heterogeneous impacts of generative AI access

Lastly, we can estimate the heterogeneous impacts of generative AI access by using the following equation:

[9] This follows from replacing $f(a, b)$ with the predicted values from the regression.

$$y_{p,t} = \beta_0 + \beta_1 * GAI_{p,t} + \theta * P_p + \lambda * T_t + \beta_2 * GAI_{p,t} * C_p + \epsilon_{p,t} \tag{4}$$

In equation 4, C_p is a chosen covariate at the participant level. In accordance with our pre-analysis plan, we are especially interested in participant heterogeneous treatment effects by gender, tenure, the percent of workload in English, and previous exposure to generative AI. We estimate equation 4 using robust standard errors clustered at the participant level. The coefficient of interest here is β_2, which captures how much higher or lower the impact of AI is for a participant with attribute C_p.

3 Results

We break out the results section into four sections. In our first subsection, we examine how access to generative AI impacts uptake across the tasks. In the second, we examine the productivity impact of generative AI on both generalist and specialized central banking tasks. Subsection three contains our analysis of complementarities between generative AI and task mix. Lastly, we examine the distributional impacts of generative AI and heterogeneous treatment effects depending on participant characteristics in subsection four.

3.1 Impact of generative AI access on usage

We first examine whether access to generative AI influenced uptake in our experiment in Table 2. We do so using three different types of outcome measures. Columns 1 and 2 use a binary indicator for whether the participant used the generative AI tool during the task. Columns 3 and 4 use the number of prompts sent to the generative AI tool. Columns 5 and 6 use a binary measure for the amount of overlap between the prompt response and submitted response. The odd columns use a within-participant analysis and regress each outcome on a binary indicator for whether the participant is in the AI treatment condition, participant-level fixed effects, and task-type fixed effects. The even columns conduct the between-subjects design by subsetting the sample to only the first seven tasks. It displays the results of a regression of each outcome on the AI indicator, fixed effects for task types, and participant prior exposure to generative AI.

The results in Table 2 indicate that access to GPT-4o significantly influenced tool uptake in the experiment. Participants used GPT-4o for around 94 percent of the tasks where they had access to the tool, versus 0 percent without access. Participants sent, on average, 1.5 prompts to GPT-4o when they had access to it, versus zero when they did not have access. Additionally, there was substantial engagement with GPT-4o-generated content, as evidenced by a high degree of overlap between prompt responses and submitted work. These results show that our treatment led to generative AI adoption across tasks, allowing us to evaluate the downstream impacts on task productivity.

3.2 Impact of generative AI access on productivity in job tasks

We next examine the overall impact of generative AI on productivity across both our generalist and specialist central banking tasks. We begin by visually exploring differences in performance when participants had access to generative AI, versus when they did not. In Figure 2, we plot the distribution of our two productivity measures for the generalist tasks by treatment condition. In Panel A, we display the total score across all generalist seven tasks in a given round, while Panel B displays the total number of minutes taken to complete all seven generalist tasks. In Panel C, we display the total score across the specialist job tasks.[10]

The results in Figure 2 indicate that generative AI led to large average improvements in both our productivity measures. In terms of the aggregate score on the generalist tasks, participants scored on average 60 points when they had access to generative AI versus 41 points when they did not have access to generative AI. For time spent on the bundle of generalist tasks, those who had access to generative AI spent on average 23 minutes to complete all seven tasks, versus 30 minutes for those who did not have access to generative AI. For our specialist tasks, participants scored on average 13 with access to generative AI versus 6 when they did not. Thus, the graphical evidence in Figure 2 provides strong but suggestive evidence that generative AI led to improvements in quality and completion time.

We next formally examine the impact of generative AI on productivity across all tasks in the experiment. Table 3 displays the results of equations 1 and 2 from section 2.5. Columns 1 and 2 use the score on the task as our measure of productivity, while columns 3 and 4 use the time spent on the task as our measure of task performance. The odd columns use the within-person design to estimate the average participant-level impact of generative AI on productivity, while the even columns

[10]We do not have a measure of efficiency for the specialist tasks because participants struggled to complete them within the 30 minute timeframe.

use our between-participant design and compares performance for the first seven tasks for participants who had or did not have access to generative AI.

The results in Table 3 indicate that generative AI led to large improvements in productivity on the generalist tasks as measured by task quality. Using the within-subjects design in column 1, we find that participants scored 2.6 points higher on an average task when they had access to generative AI versus when they did not. This represents an average improvement of 48 percent (= 2.615/5.401) relative to when participants did not have access to generative AI. We find qualitatively similar results when we limit our sample to the first seven tasks and use the between-subjects design in column 2. Participants who had access to generative AI in the first seven tasks scored 2.3 points higher than those who did not have access in the first seven tasks, representing an average increase of 39 percent relative to the control mean.

We find that generative AI also led to similar improvements using our time measure of productivity to capture efficiency. The results in columns 3 and 4 indicate that participants with access to generative AI took up to seven fewer minutes to complete the tasks versus those who did not have access to generative AI. Participants had on average 30 minutes to complete the tasks, so this represents an efficiency of improvement of about 23 percent relative to the control mean. Overall, the results indicate that generative AI led to large improvements in both of our measures of productivity across the aggregate set of generalist tasks. The effect sizes, relative to the control mean, were about twice as large for improvements in quality versus improvements in time.

We next explore the impacts of generative AI on productivity in generalist versus specialist tasks. We do so in Table 4, where we separately estimate equations 1 and 2 in generalist (columns 1 and 2) versus specialist (columns 3 and 4) tasks. Our outcome measure is the quality of the task submission.

Our results indicate that generative AI led to large productivity improvements for generalist tasks and even larger improvements for specialized ones. Our specifications indicate generative AI increased performance in generalist tasks: our within-subjects design estimates a treatment effect of 50 percent (= 2.710/5.401) while our between-subjects design estimates an effect of 36 percent (= 2.318/6.421) relative to the control means. Moreover, the productivity impacts of generative AI are positive for the vast majority of participants (see Panel A in Figure 3). Meanwhile, performance increases are even larger for specialized tasks. While the point estimates in columns 3 and 4 are similar in magnitude, the specialist tasks were more difficult and had lower average performance when participants did not have access to generative AI (5.4 for generalist tasks versus 2.2 for specialist tasks). Relative to the control mean, generative AI increased performance in specialized tasks by between 108 and 117 percent, slightly more than twice the impact on generalist tasks.

Moreover. the effects of generative AI were more heterogeneous for specialist tasks– the standard errors in columns 3 and 4 are more than twice those in columns 1 and 2 despite qualitatively similar point estimates. In Appendix A.1, we examine heterogeneous effects for generalist versus specialist tasks. Interestingly, we find that while generative AI helps lower-quality participants on generalist tasks, the same is not true for specialist tasks. This suggests that the equalizing effects of generative may be limited to generalist tasks.

3.3 Generative AI and task-level complementarities

We next explore the impact of generative AI on productivity across the various generalist tasks types. We are especially interested in testing whether there exists a complementarity between generative AI and non-routine tasks. To do so, we compare productivity with and without generative AI across non-routine and routine tasks.[11]

We begin by displaying average productivity, both with and without access to GPT-4o, for routine versus non-routine tasks. We display this in Table 5. Across the sample, access to generative AI led to large productivity improvements for both sets of tasks. For routine tasks, workers scored, on average, 24 percent higher when they had access to generative AI than when they did not. For non-routine tasks, access to generative AI increased productivity by 58 percent across the sample, or 34 percent higher than the impact for routine tasks. Taken together, these results suggest the presence of a complementarity between generative AI and non-routine tasks.

In order to estimate this complementarity more formally, we use the empirical strategy in section 2.5.2. We estimate κ_p using equation 3 and display the results in Table 6. To allow us to compare treatment effects across the task types, we first standardize performance in each task type to follow a normal distribution with mean zero and standard deviation one. Our treatment effects thus represent the difference in performance in standard deviations in each task type, relative to a control condition where participants do not have access to generative AI for the task. We then estimate κ_p both on

[11] Given that we do not have the time spent on each individual task, we cannot use the efficiency measure of productivity for this analysis. Instead, we focus on the quality score for each of the generalist task types.

average and for each participant.

The results in Table 6 provide evidence of a complementarity between generative AI and non-routine tasks across the entire sample. On average, generative AI increases productivity in routine tasks by 0.24 standard deviations, but this rises to 1.12 (= 0.24 + 0.88) standard deviations for non-routine tasks. Moreover, κ_p is statistically significant across the entire sample with $p < 0.01$. Moreover, this pattern holds for the majority of the sample. In Figure 4, we plot estimates of κ_p by participant.[12] Around 80 percent of the sample has $\kappa_p > 0$, and only 5 percent report a negative complementarity that is greater than one standard deviation. Thus, the majority of participants experience a positive complementarity between generative AI access and non-routine task share.

We can additionally compare the complementarity between generative AI and two other task-level attributes. The first is whether the task requires cognitive (versus manual) skills. The second is whether the task required subject-matter expertise, versus being a generalist task. We test these in columns two and three of Table 6. Our test of complementarities here is similar to that in column 1, except we change the interaction terms. In column 2, we interact generative AI access with a binary indicator for being a cognitive task; in column 3, we do so for specialist tasks.

The results in Table 6 also provide some support for generative AI as both cognitive-biased and specialist-biased. The results in column 2 indicate that the performance impacts of generative AI are 0.25 standard deviations larger for tasks that require more cognitive complexity, while column 3 shows that generative AI leads to 0.67 standard deviation in specialist tasks versus generalist tasks. We find similar patterns with our within-participant estimates, where around 60 percent of participants report positive interaction effects (see Appendix section A.2).

Overall, these results indicate that the returns to generative AI are larger for non-routine tasks, for cognitive tasks, and for more specialised tasks. Which of these factors lead to the most heterogeneous productivity improvements? To answer this question, we compare the size of κ_p for each attribute relative to the impact of generative AI on the omitted category for each test. For example, generative AI leads to a 0.93 standard deviation larger increase in non-routine tasks, relative to an increase of 0.26 standard deviations in routine tasks. This means that the returns to generative AI in specialist tasks is 3.6 times larger (= 0.92/0.26) than that in routine tasks. The corresponding estimates are 0.33 for cognitive tasks (= 0.25/0.75) and 0.71 for specialist tasks (= 0.67/0.95). This indicates that

[12] We do so by estimating 3 for each participant. This is possible because we have 14 observations per participant.

while there is a complementarity between generative AI and both cognitive and specialist tasks, the non-routine nature of work leads to the largest productivity improvements.

3.4 The distributional impact of generative AI access

We lastly explore the distributional consequences of generative AI. While our prior analysis reports a large increase in productivity on average and across tasks, which participants stand to gain more versus less through the diffusion of generative AI? Given that we have 14 observations per participant for the generalist tasks, half of which were with generative AI and half of which were without, we can estimate a participant-specific treatment effect. To do so, we loop through each participant, and regress that participant's task score on a binary indicator for whether they had access to generative AI. The coefficient of interest captures the difference in average productivity across all 14 tasks, when the participant uses generative AI versus does not.

We plot the distribution of individual-level treatment effects in Figure 3. The top panel of the figure presents the treatment effects on our measure of quality, while the bottom panel does so for our measure of efficiency. Overlaid on each figure is a normal distribution. The results in Figure 3 illustrate that the productivity benefits of generative AI on quality are large and positive for the majority of the distribution. Across the entire sample, 94 percent of participants scored higher when they had access to generative AI versus when they did not. In terms of the impacts of generative AI on efficiency, we see more heterogeneous effects. Generative AI leads to efficiency improvements for 80 percent of participants, while 13 percent see efficiency decrease. Moreover, we find no correlation in generative AI's impacts on quality versus efficiency (see Appendix section A.3).

Which participants benefit more versus less from the use of generative AI, and across which outcome? To investigate this point, we estimate heterogeneous treatment effects from our experiment. We rerun the specifications from the odd columns in Table 3 and additionally include interactions between our treatment indicator and various participant characteristics. We display the results in Table 7 by bundling all generalist and specialist tasks, with specific versions for generalist versus specialist tasks in Appendix section A.1.

Four important patterns emerge from these results. First, the benefits of generative AI are more heterogeneous for quality than for efficiency. The results in columns 1, 5, and 9 indicate there are heterogeneous benefits to generative AI depending on the baseline quality, exposure to routine workplace tasks, and tenure at the organization. Meanwhile, only one of our estimated interactions for efficiency is marginally significant at conventional levels. Part of this may reflect the difference in sample sizes across the two estimators, whereby we are more powered to detect heterogeneous effects for quality than time. Our results indicate that workers who have prior experience with generative AI see larger decreases in time to completion versus those who do not (column 4; $p = 0.08$). There is also suggestive evidence that those who spend more of their time working in English report larger efficiency improvements (column 12; $p = 0.11$).

Second, the benefits of generative AI on quality vary depending on the baseline skill of the participant. Our results in column 1 indicate that lower-skilled workers see the largest improvements in quality with the use of generative AI. These results suggest that generative AI has an "equalizing" effect on the quality dimension of performance, which is consistent with previous field experiments on generative AI (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Dell'Acqua et al., 2023). However, this pattern is only limited to performance in generalist tasks using the quality metric of productivity. We find no average differences in efficiency for generalist tasks depending on baseline quality (column 2), and the point estimate suggests that higher-skilled participants experience larger (though noisy) increases in efficiency. In Appendix section A.4, we provide suggestive evidence that while low-baseline productivity participants benefit most in terms of quality, high baseline productivity participants benefit most in terms of efficiency. Moreover, lower-skilled workers do not experience larger productivity increases than higher-skilled workers in our specialist tasks (see Appendix A.1). Informal conversations with participants following the experiment revealed that high-skill workers with prior generative AI experience saw the largest increases, though we are not powered to detect this effect. Thus, our results point to more subtle distinctions on how skills interact with generative AI to impact workplace productivity.

Third, we find no evidence of heterogeneous impacts by gender. Women experience the same improvements in quality and efficiency as men (columns 7 and 8), and this is true when we focus only on the specialist tasks as well (Appendix A.1). Thus, while prior work has documented a gender gap in generative AI adoption and use (Humlum and Vestergaard, 2024; Kreacic and Stone, 2024; Otis et al., 2024), our results indicate this gap is not explained by unequal returns to the technology AI by gender. Instead, it is likely driven by barriers to initial adoption.

Lastly, the impacts of generative AI relate to the task content of one's job at baseline. Our results indicate that workers in roles featuring a larger share of routine tasks see larger improvements in quality (column 5). This aligns with the earlier result where lower-skilled workers benefit more from generative AI. These results, however, pose an interesting comparison to our results regarding task-level complementarities. In section 3.3, we document that generative AI leads to larger productivity improvements in nonroutine versus routine tasks. Thus, while workers in routine jobs derive larger benefits from generative AI, generative AI is less effective for the task content of their work. This suggests a mismatch between the types of workers who stand to benefit most from generative AI, and the types of tasks where generative AI is most helpful.

What impact does this mismatch have on aggregate output? To answer this question, we develop a simple task-based production framework that builds upon the work of Autor et al. (2003) and Acemoglu and Autor (2011b). In our framework, workers have heterogeneous productivity in routine and non-routine tasks, both with and without the presence of generative AI. Each worker allocates their effort across task types and aggregate output is determined by how much time is assigned to each task. Using data from our field experiment, it is possible to trace out the production possibility frontier by solving a linear programming problem. In Appendix Section A.5, we offer more details about this method. Conceptually, our set-up mirrors Hsieh and Klenow (2009) in treating micro-allocation wedges as a source of aggregate TFP losses, but we extend their framework to a two-task environment and to a technology shock that changes the distribution of routine and non-routine shares.

In Figure 5, we trace out the production possibility frontier under two scenarios: one where no one has access to generative AI, and one where everyone has access. The x-axis captures the share of routine output while the y-axis captures the share of non-routine output. Comparing the no-AI and all-AI production possibility frontiers isolates the task bias of generative AI: an outward shift biased toward the x-axis signals a routine-AI complementarity, whereas a tilt toward the y-axis indicates a non-routine complementarity. As with our results in section 3.3, the figure illustrates a complementarity between generative AI and non-routine work, whereby generative AI leads to larger increases in the output of non-routine versus routine work.

We can then compare this frontier to aggregate output where task shares remain constant but generative AI is introduced. This is captured by the green triangle in Figure 5. As the figure illustrates, this point is below the production possibility frontier under generative AI. The vertical difference represents an output loss of 7.3%. Coming to the frontier and increasing aggregate output requires reassigning workers to routine- versus non-routine tasks based upon their comparative advantage. Overall, our back-of-the-envelope calculations suggest sizable productivity gains through improved employee-task matching in the presence of generative AI.

4 Conclusion

In this paper, we present the results of a field experiment designed to test task-level complementarities with generative AI. We find that access to generative AI leads to large improvements in both quality and efficiency for the vast majority of participants. Furthermore, we find evidence of a complementarity between generative AI and non-routine work, both on average across the entire sample, and for the vast majority of individual participants in our study.

Understanding how generative AI will impact the labor market is of paramount importance, and our manuscript provides novel field-experimental evidence of task-level complementarities with regard to generative AI. We caution, however, that this complementarity between generative AI and non-routine work is a necessary but not sufficient condition for routine-biased technological change to occur at scale. For routine-biased technological change to fully materialize at scale, several market- and firm-level adjustments must occur. These include reductions in the price of computing, changes in the relative demand for skills, changes in task composition and occupational structure, and complementary workplace adjustments from firms. The overall impact of generative AI on the labor market will depend on such adjustments. Our manuscript provides micro-level evidence on task-level complementarities that can inform such models of technological change.

Our manuscript suggests various avenues for future work in this space. For example, while prior work has documented that generative AI leads to larger improvements for lower-skilled individuals, we find this is not the case for efficiency or performance on specialist tasks that require domain expertise. Understanding how worker skills relate to generative AI, and whether generative AI will expand or contract skill inequality, is paramount to understanding the labor market impacts of the technology. Similarly, the design of jobs will likely play an important role in this regard. Our results suggest a mismatch between the types of workers who stand to benefit most from generative AI, and the types of tasks where generative AI is more helpful, suggesting a prominent role for job design decisions and upskilling opportunities. Lastly, special attention should be paid to the macro-level adjustments described above. Previous waves of RBTC have been accompanied by changes in organizational practices (Bresnahan et al., 2002) and differences in workforce composition (Autor et al., 2003), among others. Such responses from workers and firms will ultimately shape how generative AI impacts the labor market.

References

Acemoglu, Daron, "The Simple Macroeconomics of AI," *Working paper*, 2024.

— **and David Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, pp. 1043–1171.

— **and —** , "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171.

— **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.

— **and —** , "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, Spring 2019, *33* (2), 3–30.

— **and —** , "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

— **and —** , "Automation and Rent Dissipation: Implications for Wages, Inequality, and Productivity," Working Paper 32536, National Bureau of Economic Research June 2024.

Aldasoro, I. et al., "Generative Artificial Intelligence and Cyber Security in Central Banking," BIS Papers 145, Bank for International Settlements 2024.

Araujo, D. et al., "Machine Learning in Central Banking," IFC Bulletins 57, Irving Fisher Committee on Central Bank Statistics 2022.

Autor, David H, "The "task approach" to labor markets: an overview," *Journal of Labour Market Research*, 2013, *46* (3), 185–199.

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 11 2003, *118* (4), 1279–1333.

Banker, Rajiv D., Abraham Charnes, and William W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, 1984, *30* (9), 1078–1092.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Brynjolfsson, Erik and Paul Milgrom, "Complementarity in Organizations," Draft chapter for the Handbook of Organizational Economics 2012. August 2008; Revised December 2010 and January 2012.

_ , Danielle Li, and Lindsey R Raymond, "Generative AI at Work," Working Paper 31161, National Bureau of Economic Research April 2023.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Yuanzhi Li, Xi Wang, Yuhuai Zhang, Shun Garg, Preetum Nakkiran, Sham Narayanan, Lance Ansell, Joseph Merullo, Sergio Boixo, Karen Simonyan, Oriol Vinyals, Demis Hassabis, and Ilya Sutskever, "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," 2023. arXiv preprint arXiv:2303.12712.

Charnes, Abraham, William W. Cooper, and Edward Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 1978, *2* (6), 429–444.

Costinot, Arnaud and Jonathan Vogel, "Matching and Inequality in the World Economy," *Journal of Political Economy*, 2010, *118* (4), 747–786.

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," Working Paper 2024.

Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski, "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination," *The Review of Economics and Statistics*, 05 2025, pp. 1–16.

_ , Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper No. 24-013, Harvard Business School Technology & Operations Mgt. Unit September 2023.

Dell'Acqua, Fabrizio, Edward III McFowland, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," 2023. Working paper, Harvard Business School.

Demirci, Ozge, Jonas Hannane, and Xinrong Zhu, "Who is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms," CESifo Working Paper 11276 2024.

Doerr, S. et al., "Big Data and Machine Learning in Central Banking," BIS Working Papers 930, Bank for International Settlements 2021.

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," 2023.

Felten, Edward W., Manav Raj, and Robert Seamans, "Occupational Heterogeneity in Exposure to Generative AI," April 2023. Available at SSRN: https://ssrn.com/abstract=4414065 or http://dx.doi.org/10.2139/ssrn.4414065.

Goos, Maarten, Alan Manning, and Anna Salomons, "Explaining job polarization: Routine-biased technological change and offshoring," *The American Economic Review*, 2014, *104* (8), 2509–2526.

_ , Emilie Rademakers, and Ronja Röttger, "Routine-Biased technical change: Individual-Level evidence from a plant closure," *Research Policy*, 2021, *50* (7), 104002.

Graetz, Georg and Guy Michaels, "Robots at Work," *The Review of Economics and Statistics*, 12 2018, *100* (5), 753–768.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Hui, Xiang, Oren Reshef, and Luofeng Zhou, "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," *Organization Science*, 2024, *0* (0).

Humlum, Anders and Emilie Vestergaard, "The Adoption of ChatGPT," Working Paper, University of Chicago, Becker Friedman Institute for Economics 2024.

Kreacic, Ana and Terry Stone, "Women are Falling Behind on Generative AI in the Workplace. Here's How to Change That," World Economic Forum 2024. Available at: https://www.weforum.org/agenda/2024/04/women-generative-ai-workplace/.

List, John A., "The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs," NBER Working Paper 33456, National Bureau of Economic Research February 2025.

Michaels, Guy, Ashwini Natraj, and John Van Reenen, "Has ICT polarized skill demand? evidence from eleven countries over twenty-five years," *The Review of Economics and Statistics*, 2014, *96* (1), 60–77.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

OpenAI, "Hello GPT-4o," Technical Report May 2024.

Otis, Nicholas G., Katelyn Cranney, Solène Delecourt, and Rembrand Koning, "Global Evidence on Gender Gaps and Generative AI," 2024. Working Paper.

Perkowski, Patryk and Aleš Maršál, "Unlocking the productiviyty benefits of generative AI: Evidence from three central banks," *Working paper*, 2024.

Schwarcz, Daniel, Sam Manning, Patrick James Barry, David R. Cleveland, J. J. Prescott, and Beverly Rich, "AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice," Minnesota Legal Studies Research Paper No. 25-16; University of Michigan Public Law Research Paper (forthcoming) March 2025. Available at SSRN.

Spitz-Oener, Alexandra, "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 2006, *24* (2), 235–270.

Tambe, Prasanna, "Reskilling the Workforce for AI: Domain Experts, Algorithms, and Organizational Performance," *Management Science*, 02 2025.

Tables

Table 1: **Summary statistics**

	N	Mean	SD
Education, high school or associate's degree	100	0.03	0.17
Education, university	100	0.02	0.14
Education, masters degree or higher	100	0.92	0.27
Age, in years	100	39.96	8.46
Female	100	0.34	0.48
Tenure at bank, in years	100	8.67	10.57
Manages or leads team	100	0.05	0.22
Has a laptop	100	0.98	0.14
% of work in English	100	0.33	0.26
Total weekly hours	100	46.73	11.43
Percent of work week, routine cognitive	100	0.18	0.12
Percent of work week, non-routine cognitive	100	0.26	0.16
Percent of work week, routine manual	100	0.03	0.06
Percent of work week, non-routine manual	100	0.02	0.04
Percent of work week, analytical	100	0.31	0.19
Percent of work week, leadership and management	100	0.16	0.10
Percent of work week, interpersonal and communication	100	0.03	0.07
Heard of LLM	100	0.95	0.22
Used LLM for either personal or work use	100	0.73	0.45
Used LLM for personal use	100	0.66	0.48
Used LLM for workplace use	100	0.51	0.50
Score on task	1575	6.75	3.31
AI treatment condition	1575	0.50	0.50
Time (minutes)	196	27.11	5.85

Notes: This table displays summary statistics of the sample.

Table 2: **Impact of generative AI access on generative AI uptake**

	Used AI		# of AI prompts		AI retention	
	(1)	(2)	(3)	(4)	(5)	(6)
AI treatment condition	0.937***	0.939***	1.495***	1.411***	0.770***	0.816***
	(0.0199)	(0.0205)	(0.0680)	(0.0784)	(0.0243)	(0.0273)
R2	0.918	0.902	0.587	0.527	0.738	0.735
Observations	1,148	567	1,148	567	1,148	567
Control mean	0.000	0.000	0.000	0.000	0.000	0.000
Fixed effects	Yes	No	Yes	No	Yes	No
Controls	No	Yes	No	Yes	No	Yes
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines the impact of generative AI access on generative AI uptake in the experiment. Columns 1 and 2 use a binary indicator for whether the participant used the generative AI tool during the task. Columns 3 and 4 use the number of prompts sent to the generative AI tool. Columns 5 and 6 use a binary measure for the amount of overlap between the prompt response and submitted response. Odd columns use a within-participant analysis and regress each outcome on a binary indicator for whether the participant is in the AI treatment condition, participant-level fixed effects, and task-type fixed effects. Even columns conduct the between-subjects design by subsetting the sample to only the first seven tasks. It displays the results of a regression of each outcome on the AI indicator, fixed effects for task types, and participant prior exposure to generative AI. All regressions include robust standard errors clustered at the participant level.

Table 3: **Impact of generative AI access on productivity across all tasks**

	Score on task		Time to complete	
	(1)	(2)	(3)	(4)
AI treatment condition	2.615***	2.304***	-6.298***	-6.951***
	(0.143)	(0.205)	(0.647)	(0.990)
R2	0.394	0.468	0.302	0.381
Observations	1,575	777	196	98
Control mean	5.401	5.942	30.224	31.233
Fixed effects	Yes	No	Yes	No
Controls	No	Yes	No	Yes
p-value	<0.001	<0.001	<0.001	<0.001

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines the impact of generative AI on productivity across all of our tasks. The first two columns use the quality score on the task as the measure of productivity, while the second two columns use the time to complete the task (in minutes). Column 1 displays the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, participant-level fixed effects, and task-type fixed effects. Column 2 conducts the between-subjects design by subsetting the sample to only the first seven tasks. It displays the results of a regression of task score on the AI indicator, fixed effects for task types, and participant prior exposure to generative AI. Column 3 displays the results of a regression of the time to complete the task on a binary indicator for whether the participant is in the AI treatment condition and a fixed effect for the phase of the experiment. Column 4 conducts the between-subjects design by subsetting the sample to only the first seven tasks. It displays the results of a regression of time on the AI indicator and participant prior exposure to generative AI. All regressions include robust standard errors clustered at the participant level.

Table 4: **Impact of generative AI access on productivity for generalist versus specialist tasks**

	Generalist tasks		Specialist tasks	
	(1)	(2)	(3)	(4)
AI treatment condition	2.710***	2.318***	2.613***	2.230***
	(0.168)	(0.212)	(0.450)	(0.512)
R2	0.300	0.358	0.710	0.284
Observations	1,407	693	168	84
Control mean	5.401	6.421	2.240	2.058
Fixed effects	Yes	No	Yes	No
Controls	No	Yes	No	Yes
p-value	<0.001	<0.001	<0.001	<0.001

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

Notes: This table examines the impact of generative AI on productivity for generalist versus specialist tasks. The outcome of interest is the quality measure of productivity. Columns 1 and 2 subset the sample to our generalist tasks, while columns 3 and 4 subset the sample to the specialist tasks. Columns 1 and 3 display the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, participant-level fixed effects, and task-type fixed effects. Columns 2 and 4 conduct the between-subjects design by subsetting the sample to only the first set of tasks in each batch. It displays the results of a regression of task score on the AI indicator, fixed effects for task types, and participant prior exposure to generative AI. All regressions include robust standard errors clustered at the participant level.

Table 5: **Average performance by task type and treatment condition**

	Average performance		%
Task type	Control	AI	change
	(1)	(2)	(3)
Generalist tasks	5.8	8.5	47
Routine	6.3	7.8	24
Nonroutine	5.7	9.0	58
Specialist tasks	2.2	4.9	113

Notes: This table displays average productivity across the generalist task categories. Column 1 displays average productivity for those in the control condition, while column 2 does so for those in the treatment condition. Column 3 gives the percentage improvement in productivity for each task type.

Table 6: **Task-level complementarities**

	(1)	(2)	(3)
	Task score, z-score		
AI treatment condition	0.26***	0.75***	0.95***
	(0.09)	(0.10)	(0.05)
Nonroutine task	0.01	0.47***	
	(0.07)	(0.04)	
Cognitive task	-0.02	-0.14*	
	(0.05)	(0.08)	
Specialist task			0.13
			(0.15)
AI * Nonroutine task	0.93***		
	(0.11)		
AI * Cognitive task		0.25**	
		(0.11)	
AI * Specialist task			0.67***
			(0.18)
R2	0.345	0.312	0.287
Observations	1,407	1,407	1,575
Fixed effects	Yes	Yes	Yes
Controls	No	No	No
κ_p p-value	<0.01	0.02	<0.01

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table tests various complementarities regarding GAI. Columns 1 and 2 test for complementarities within the generalist tasks, with column 1 exploring the complementarity between GAI and nonroutine task mix and column 2 between GAI and cognitive task mix. Column 1 displays the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, an interaction between the AI treatment condition and a binary indicator for non-routine tasks, participant-level fixed effects, and task-type fixed effects. Column 2 does the same analysis but uses interactions for cognitive tasks. Columns 1 and 2 limit the sample to only the generalist tasks. Column 3 explores the complementarity between GAI and specialist tasks. All outcomes are normalized to have mean zero and standard deviation one per task. All regressions include robust standard errors clustered at the participant level.

Table 7: **Heterogeneous impact of generative AI access on task productivity, by participant characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Quality	Time	Quality	Time	Quality	Time	Quality	Time	Quality	Time	Quality	Time
AI treatment condition	4.40***	-4.66***	2.58***	-5.40***	2.23***	-7.05***	2.66***	-6.24***	2.35***	-6.55***	2.79***	-4.40***
	(0.29)	(1.77)	(0.17)	(1.02)	(0.23)	(1.51)	(0.19)	(1.15)	(0.19)	(1.12)	(0.23)	(1.40)
AI * Baseline quality	-0.03***	-0.03										
	(0.00)	(0.03)										
AI * GAI exposure			0.01	-0.28*								
			(0.02)	(0.16)								
AI * Routine task percent					1.82**	3.80						
					(0.81)	(5.80)						
AI * Female							-0.11	0.00				
							(0.29)	(1.86)				
AI * Tenure									0.03**	0.04		
									(0.01)	(0.08)		
AI * English language											-0.52	-5.52
											(0.62)	(3.47)
R2	0.412	0.727	0.393	0.742	0.394	0.723	0.393	0.721	0.395	0.722	0.393	0.736
Observations	1,575	196	1,561	196	1,561	196	1,561	196	1,561	196	1,561	196
Outcome	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency
Category	Quality		GAI exposure		Routine		Gender		Tenure		Language	
p-value	0.00	0.33	0.60	0.08	0.03	0.51	0.70	1.00	0.03	0.62	0.40	0.11

Standard errors in parentheses
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

Notes: This table examines the heterogeneous impacts of generative AI on productivity in generalist central banking tasks. Odd columns use the score on the task as the measure of productivity, while even ones use the time to complete the task (in minutes). The odd columns displays the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, interactions between this indicator and various participant characteristics, participant-level fixed effects, and task-type fixed effects. The even columns display the results of a regression of the time to complete the task on a binary indicator for whether the participant is in the AI treatment condition, an interaction between the AI condition and participant characteristics, and a fixed effect for the phase of the experiment. Quality is measured as the baseline productivity percentile from the quality metric. GAI exposure is measured by number of hours spent using GAI at work. Routine measures the percent of ones workplace tasks that would classify as routine. Tenure is measured in years at the organization. Language captures the percent of working time that is spent in English. The p-value at the bottom of the table is for the reported interaction effect. All regressions include robust standard errors clustered at the participant level.

Figures

Figure 1: **Structure of experiment**

		Participant				
	Task	**1**	**2**	**3**	**4**	**5**
Generalist tasks	1	R_C	R_C	R_C	R_C	R_C
	2	NR_C	NR_C	NR_C	NR_C	NR_C
	3	R_M	R_M	R_M	R_M	R_M
	4	NR_M	NR_M	NR_M	NR_M	NR_M
	5	NR_A	NR_A	NR_A	NR_A	NR_A
	6	NR_L	NR_L	NR_L	NR_L	NR_L
	7	NR_T	NR_T	NR_T	NR_T	NR_T
	8	R_C	R_C	R_C	R_C	R_C
	9	NR_C	NR_C	NR_C	NR_C	NR_C
	10	R_M	R_M	R_M	R_M	R_M
	11	NR_M	NR_M	NR_M	NR_M	NR_M
	12	NR_A	NR_A	NR_A	NR_A	NR_A
	13	NR_L	NR_L	NR_L	NR_L	NR_L
	14	NR_T	NR_T	NR_T	NR_T	NR_T
Specialist tasks	15	S	S	S	S	S
	16	S	S	S	S	S

Treatment conditions

- Access to Generative AI
- Control Group

Task categories:

Routine tasks

- R_C: Routine cognitive
- R_M: Routine manual

Nonroutine tasks

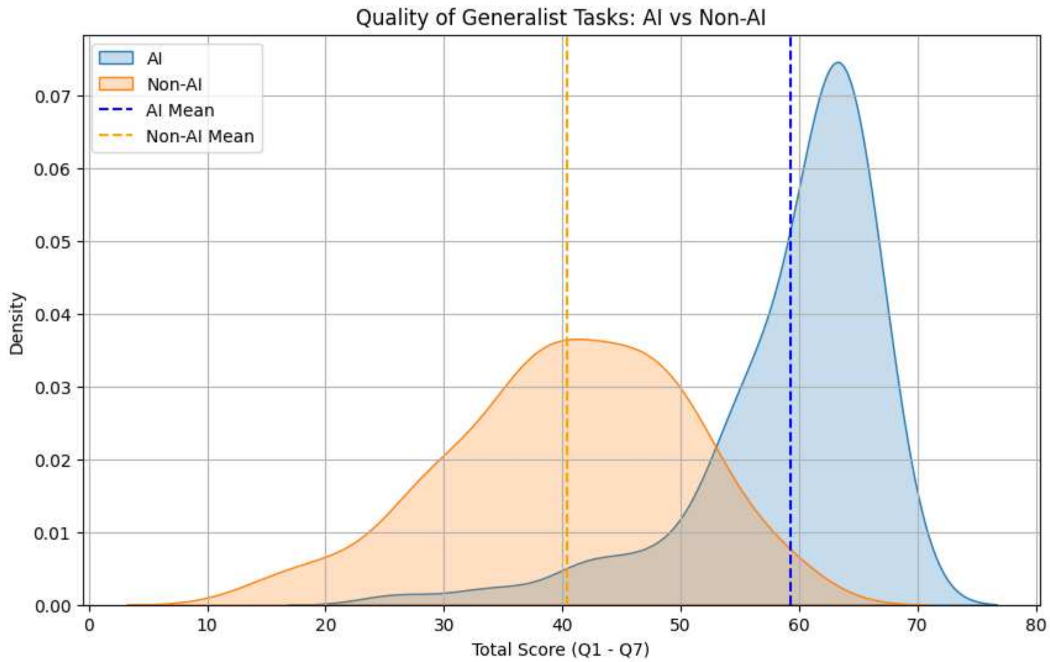
- NR_C: Nonroutine cognitive
- NR_M: Nonroutine manual
- NR_A: Nonroutine analytical
- NR_L: Nonroutine leadership
- NR_T: Nonroutine communication

Specialist tasks

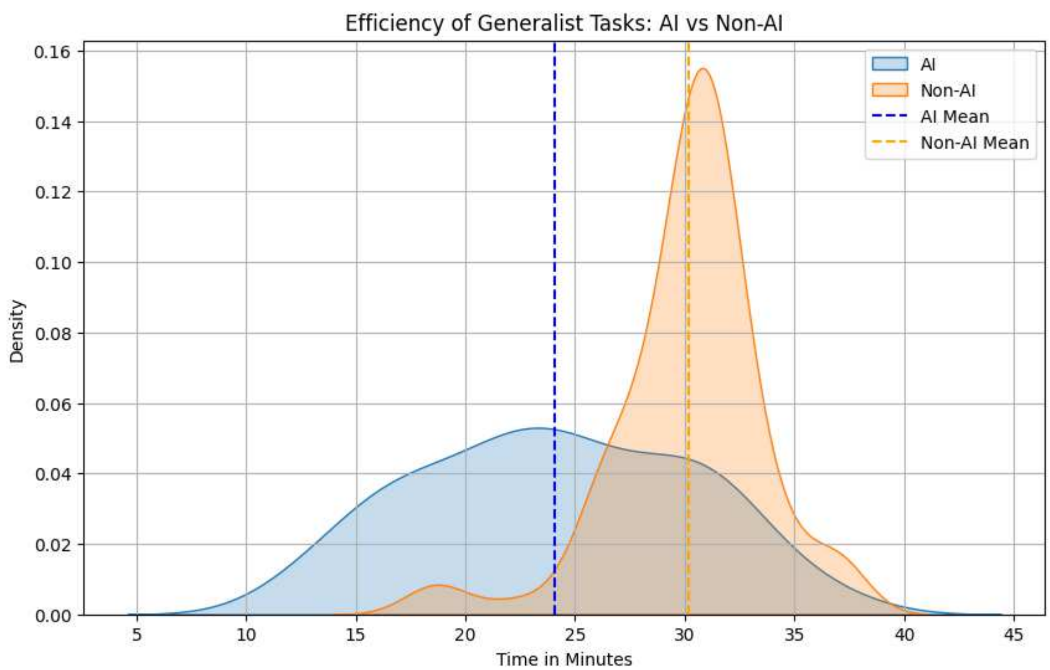
- S: Specialist task

Notes: This figure displays the structure of the experiment

Figure 2: **Distribution of aggregate productivity measures, by generative AI treatment condition and task type**

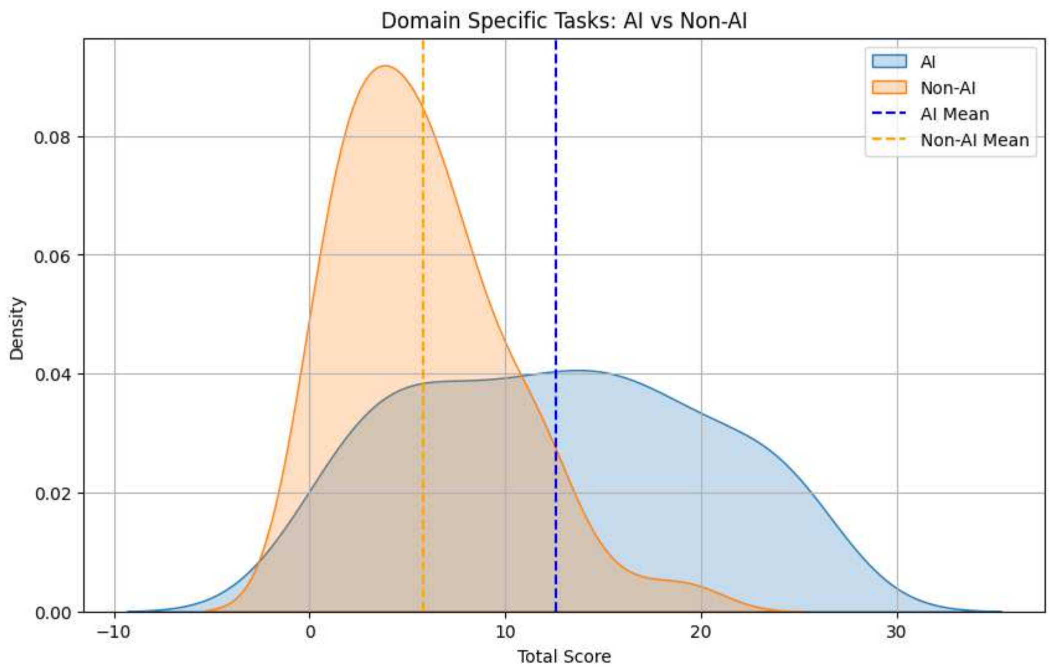


(a) Total task score, generalist tasks



(b) Total time to complete the task, generalist tasks

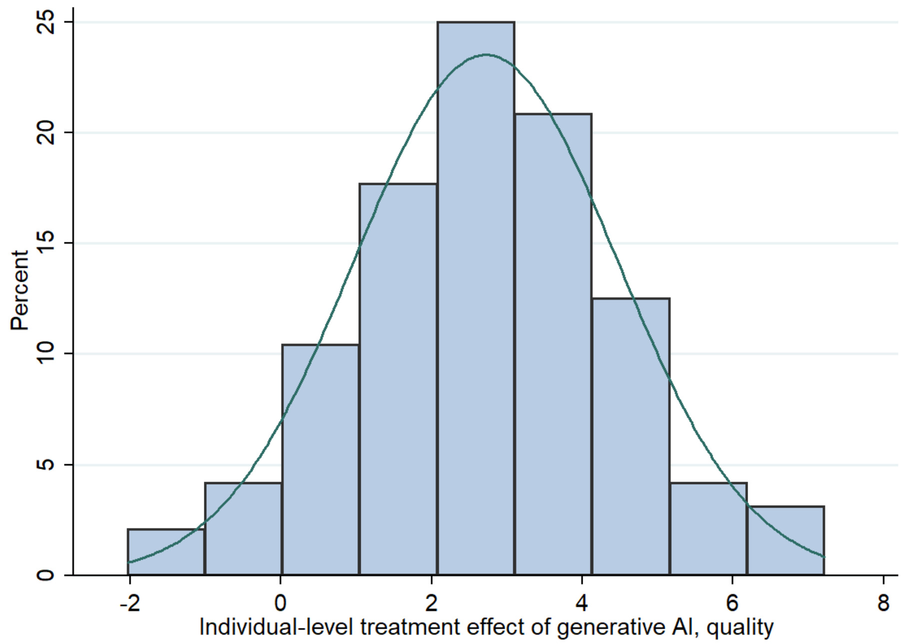
Figure 2 (continued): Distribution of aggregate productivity measures, by generative AI treatment condition and task type



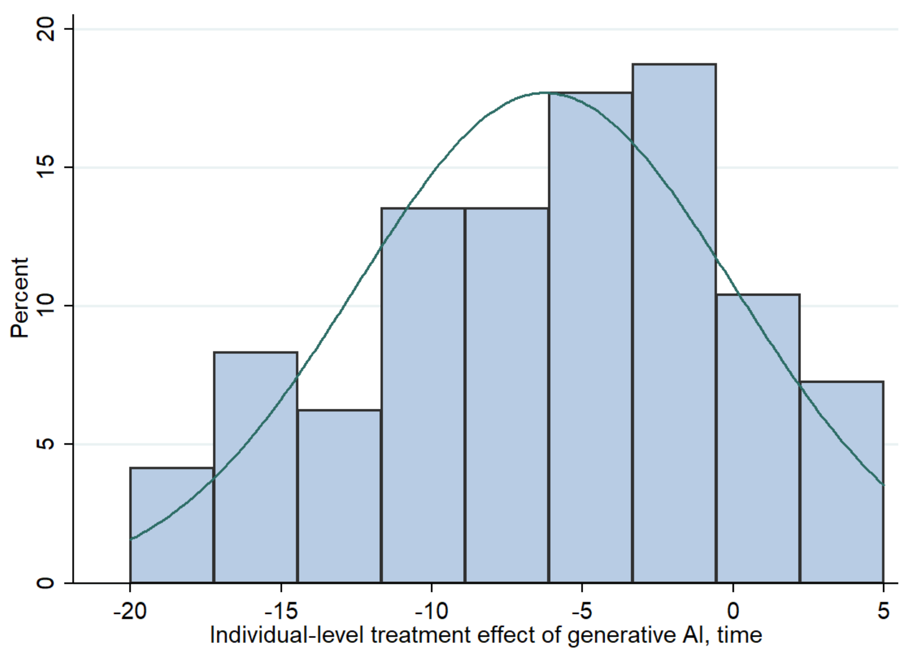
(c) Total task score, specialist tasks

Notes: This figure displays histograms of productivity in our setting when participants have access to generative AI versus they do not. In Panel A, we examine differences in submission quality for our generalist tasks. In Panel B, we examine differences in submission efficiency for our generalist tasks. In Panel C, we examine differences in submission quality for our specialist tasks.

Figure 3: **Distributional impact of generative AI on productivity in generalist tasks, by participant type**



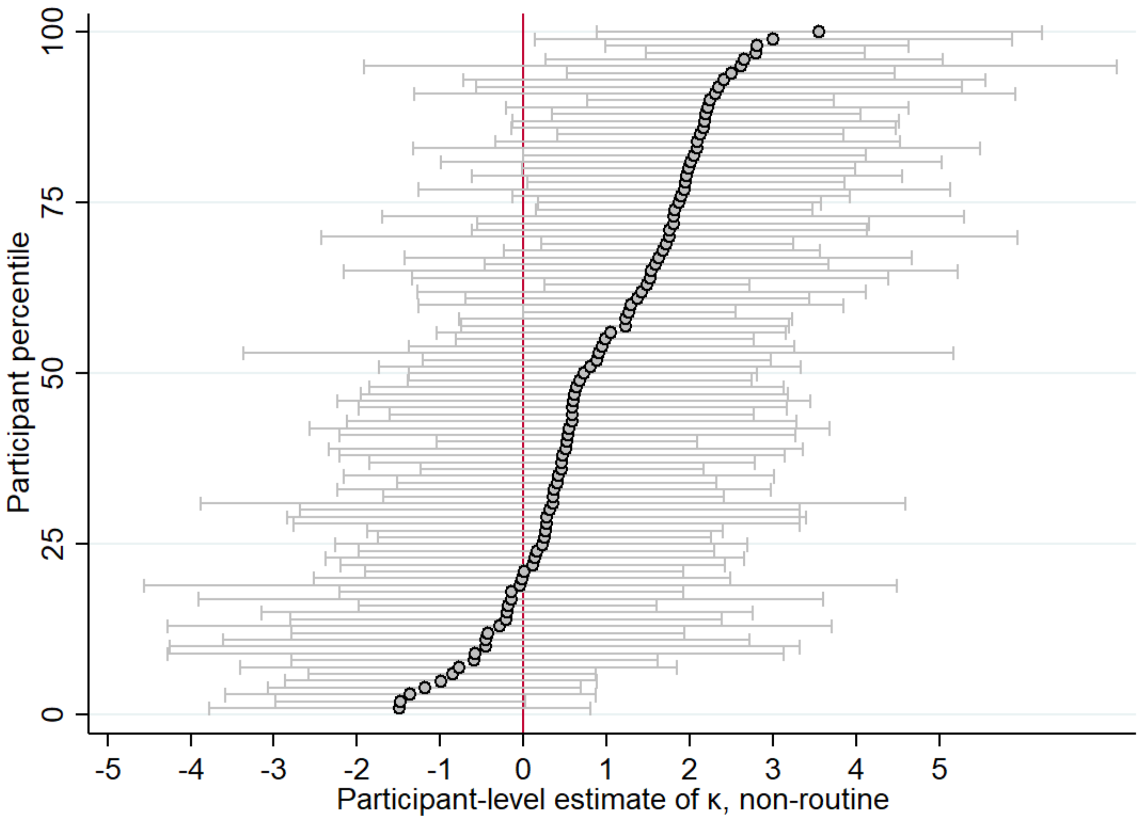
(a) Total task score



(b) Total time to complete the task

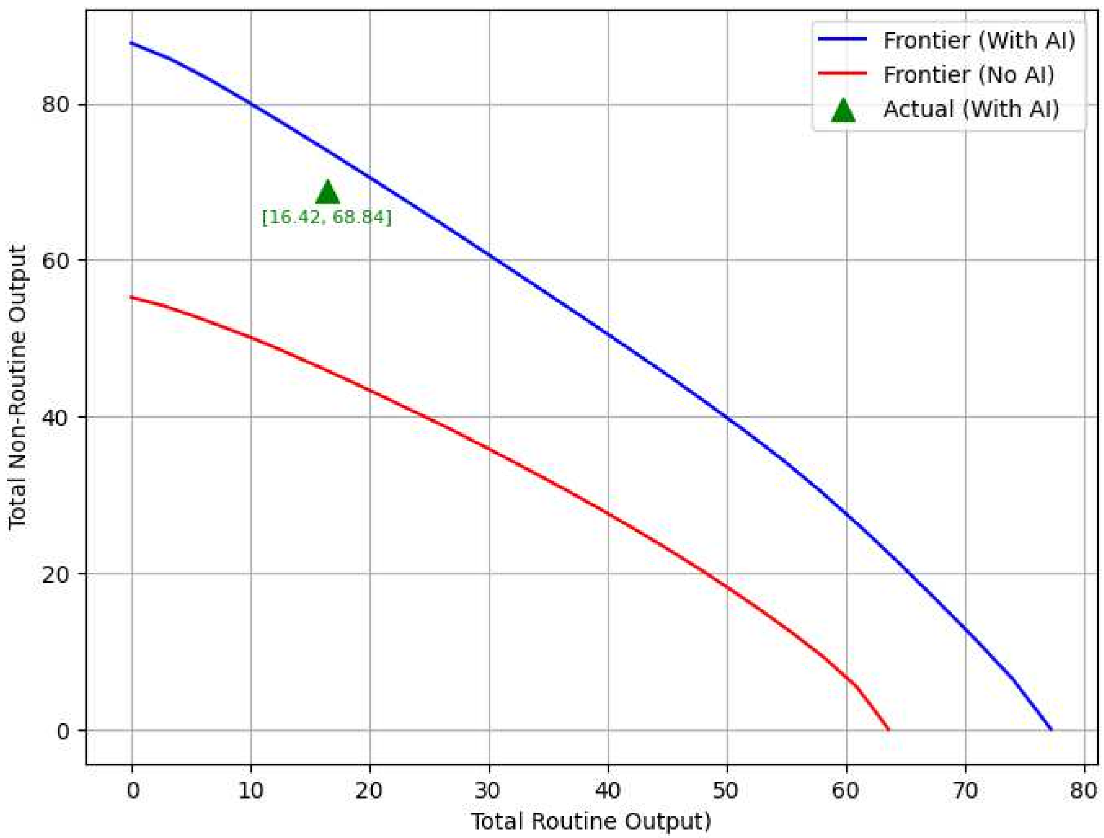
Notes: In this figure, we display individual-level treatment effects of generative AI across the generalist tasks. We leverage our within-participant variation, estimate equation 1 for each participant, and plot the distribution of coefficients. Panel A does this analysis for the quality measure while Panel B does so for efficiency.

Figure 4: **Distribution of κ across the sample**



Notes: This figure examines the complementarity between generative AI and non-routine tasks at the participant level. For each participant, We estimate equation 3. The figure displays the distribution of κ_p from these estimations. The coefficients are sorted by size.

Figure 5: **Production possibility frontier, with and without generative AI**



Notes: This figure displays the production possibility frontier with and without generative AI using the statistical procedure described in section 3.4.

Appendix: For Online Publication Only

A Additional experimental results

A.1 Heterogeneous treatment effects, for generalist versus specialist tasks

In this subsection, we estimate heterogeneous treatment effects across task types. Table A.1 estimates these for our generalist tasks, while Table A.2 does it for our specialist tasks.

A.2 Within-participant estimates of the complementarity between generative AI and various task attributes

In this subsection, we report our within-participant estimates of κ_p for two attributes: the level of cognitive complexity and the level of specialized skills required for the task. We do so by estimating equation 3 with interactions for the attribute of interest, for each participant.

Figure A.1 displays within-participant estimates for cognitive bias, and Figure A.2 does so for specialist bias.

A.3 Correlation between generative AI's impact on quality versus efficiency

In this subsection, we examine whether the impact of generative AI on participant quality is correlated with its impact on time. To do so, Figure A.3 displays a binned scatter plot, where the x-axis captures the percent improvement in quality, while the y-axis captures the percent improvement in time. The figure illustrates that there is no correlation between the two.

A.4 Correlation between generative AI's impact on quality versus efficiency, versus baseline productivity

In this subsection, we examine whether the impact of generative AI on participant quality is correlated with its impact on efficiency and with baseline productivity. To do so, Figure A.4 This figure displays

Table A.1: **Heterogeneous treatment effects of generative AI on productivity in generalist tasks**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Quality	Time	Quality	Time	Quality	Time	Quality	Time	Quality	Time	Quality	Time
AI treatment condition	4.48***	-4.66***	2.59***	-5.40***	2.18***	-7.05***	2.66***	-6.24***	2.32***	-6.55***	2.68***	-4.40***
	(0.29)	(1.77)	(0.17)	(1.02)	(0.24)	(1.51)	(0.19)	(1.15)	(0.19)	(1.12)	(0.23)	(1.40)
AI * Baseline quality	-0.04***	-0.03										
	(0.00)	(0.03)										
AI * GAI exposure			0.01	-0.28*								
			(0.03)	(0.16)								
AI * Routine task percent					2.02**	3.80						
					(0.81)	(5.80)						
AI * Female							-0.12	0.00				
							(0.30)	(1.86)				
AI * Tenure									0.04**	0.04		
									(0.02)	(0.08)		
AI * English language											-0.19	-5.52
											(0.63)	(3.47)
R2	0.338	0.727	0.316	0.742	0.318	0.723	0.316	0.721	0.319	0.722	0.316	0.736
Observations	1,407	196	1,393	196	1,393	196	1,393	196	1,393	196	1,393	196
Outcome	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency	Quality	Efficiency
Category	Quality		GAI exposure		Routine		Gender		Tenure		Language	
p-value	0.00	0.33	0.78	0.08	0.01	0.51	0.68	1.00	0.02	0.62	0.76	0.11

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

Notes: This table examines the heterogeneous impacts of generative AI on productivity in generalist central banking tasks. Odd columns use the score on the task as the measure of productivity, while even ones use the time to complete the task (in minutes). The odd columns displays the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, interactions between this indicator and various participant characteristics, participant-level fixed effects, and task-type fixed effects. The even columns display the results of a regression of the time to complete the task on a binary indicator for whether the participant is in the AI treatment condition, an interaction between the AI condition and participant characteristics, and a fixed effect for the phase of the experiment. Quality is measured as the baseline productivity percentile from the quality metric. GAI exposure is measured by number of hours spent using GAI at work. Routine measures the percent of ones workplace tasks that would classify as routine. Tenure is measured in years at the organization. Language captures the percent of working time that is spent in English. The p-value at the bottom of the table is for the reported interaction effect. All regressions include robust standard errors clustered at the participant level.

Table A.2: **Heterogeneous treatment effects of generative AI on productivity in specialist tasks**

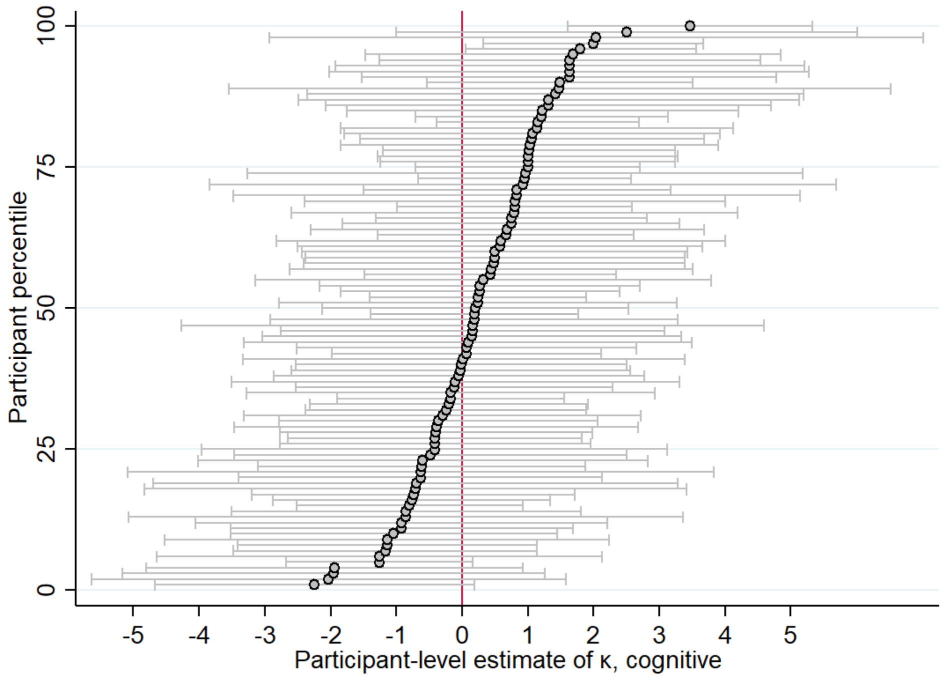
	(1)	(2)	(3)	(4)	(5)	(6)
			Score on task			
AI treatment condition	3.88***	2.50***	2.80***	2.67***	2.55***	3.73***
	(0.96)	(0.51)	(0.83)	(0.58)	(0.59)	(0.73)
AI * Baseline quality	-0.02					
	(0.02)					
AI * GAI exposure		0.05				
		(0.08)				
AI * Routine task percent			-0.70			
			(3.56)			
AI * Female				-0.02		
				(0.88)		
AI * Tenure					0.01	
					(0.05)	
AI * English language						-3.22*
						(1.66)
R2	0.734	0.722	0.720	0.719	0.720	0.745
Observations	168	168	168	168	168	168
Outcome	Quality	Quality	Quality	Quality	Quality	Quality
Category	Quality	GAI exposure	Routine	Gender	Tenure	Language
p-value	0.13	0.56	0.84	0.98	0.80	0.06

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

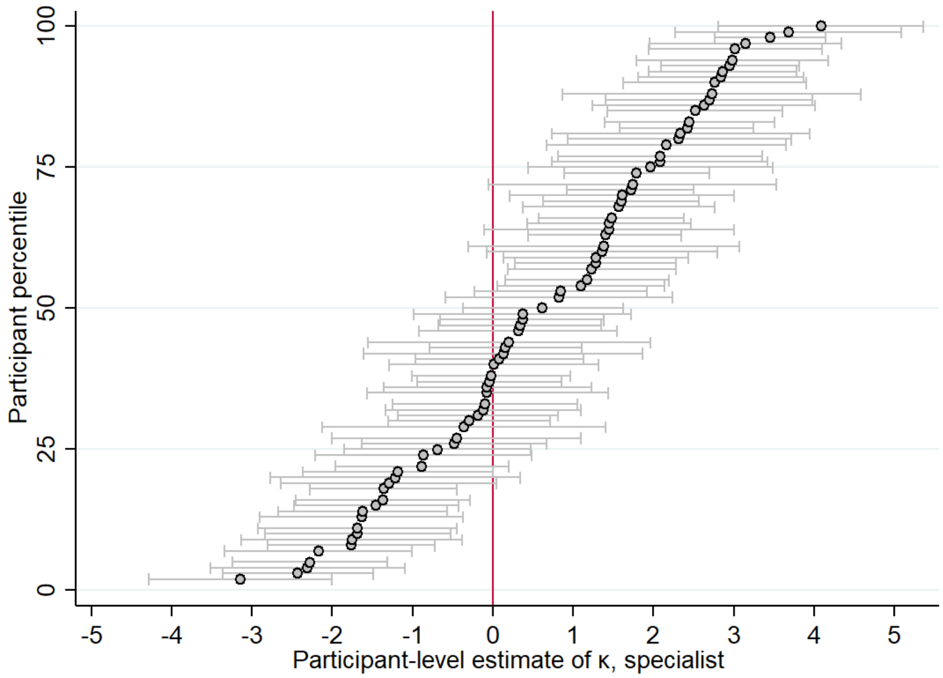
Notes: This table examines the heterogeneous impacts of generative AI on productivity in generalist central banking tasks. We use the score on the task as the measure of productivity and display the results of a regression of the task score, on a binary indicator for whether the participant is in the AI treatment condition, interactions between this indicator and various participant characteristics, participant-level fixed effects, and task-type fixed effects. Quality is measured as the baseline productivity percentile from the quality metric. GAI exposure is measured by number of hours spent using GAI at work. Routine measures the percent of ones workplace tasks that would classify as routine. Tenure is measured in years at the organization. Language captures the percent of working time that is spent in English. The p-value at the bottom of the table is for the reported interaction effect. All regressions include robust standard errors clustered at the participant level.

Figure A.1: **Distribution of κ across the sample: cognitive**



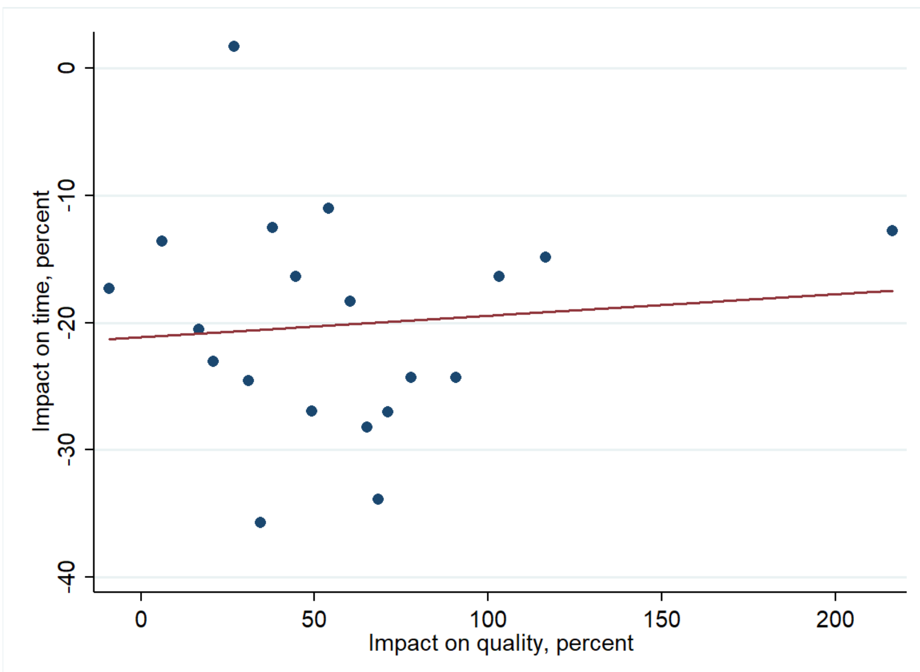
Notes: This figure displays the distribution of κ_p across the entire sample for cognitive bias. The coefficients come from running equation 3 for each participant and plotting the interaction between the treatment indicator and our binary indicator for cognitive complexity.

Figure A.2: **Distribution of κ across the sample: specialist**



Notes: This figure displays the distribution of κ_p across the entire sample for specialist bias. The coefficients come from running equation 3 for each participant and plotting the interaction between the treatment indicator and our binary indicator for specialist tasks.

Figure A.3: **Correlation between generative AI's impact on quality versus efficiency**

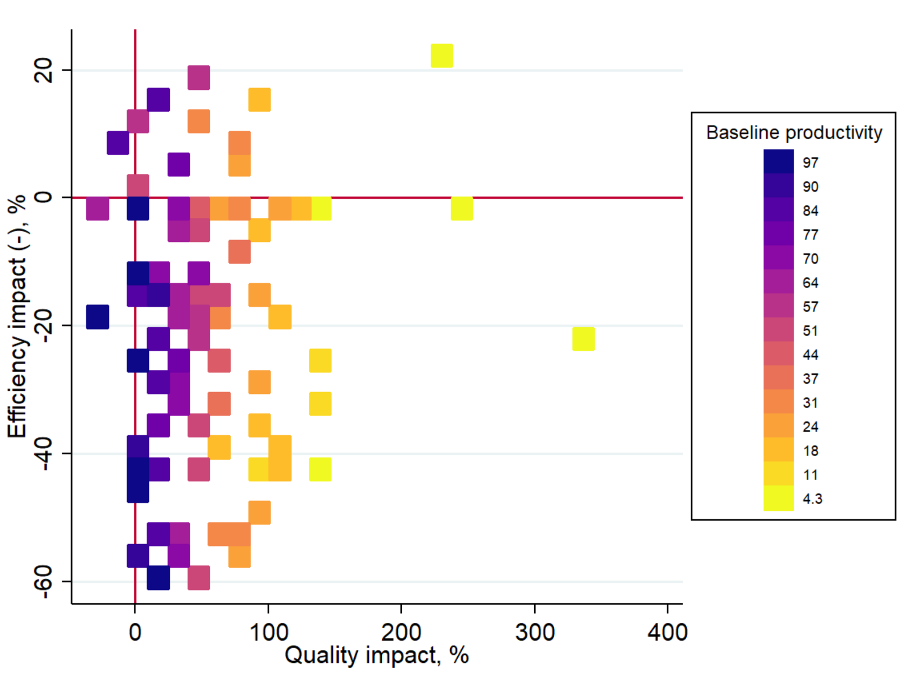


Notes: This figure displays a binned scatter plot, where the x-axis captures the percent improvement in quality, while the y-axis captures the percent improvement in time.

a heatplot of the impacts of generative AI on quality versus efficiency versus baseline productivity. The x-axis captures the percent improvement in quality, while the y-axis captures the percent improvement in efficiency (negated). The z-axis measures the participant's baseline productivity percentile.

The figure illustrates that while low-productivity participants benefit most in terms of quality, high-productivity participants benefit most in terms of efficiency.

Figure A.4: **Correlation between generative AI's impact on quality versus efficiency, versus baseline productivity**



Notes: This figure displays a heatplot of the impacts of generative AI on quality versus efficiency versus baseline productivity. The x-axis captures the percent improvement in quality, while the y-axis captures the percent improvement in efficiency (negated). The z-axis measures the participant's baseline productivity percentile.

A.5 Quantifying the impact of mismatch on aggregate output

A.5.1 Task-based Production Framework

We adopt the task-assignment view of technical change in which labor is allocated across routine and non-routine activities that differ in their routine nature (Autor et al., 2003; Acemoglu and Autor, 2011b). Each worker i is characterised by a productivity vector

$$(r_i, n_i) \in R^2_+,$$

where r_i (routine) and n_i (non-routine) are proxied by the scores obtained in our controlled field experiment at the National Bank of Slovakia (NBS).[1]

Let $x_i \in [0,1]$ denote the fraction of worker i's time assigned to routine work. Aggregate routine and non-routine output are then

$$R = \sum_i x_i r_i, \qquad N = \sum_i (1 - x_i) n_i. \tag{5}$$

The economy's *production-possibility frontier* (PPF) is the upper envelope of all feasible (R, N) pairs generated by admissible allocations $\{x_i\}_i$. Under the discrete rule $x_i \in \{0,1\}$, workers are assigned to the task in which they have a comparative advantage; the resulting frontier is the two-factor analogue of the Roy model studied by Costinot and Vogel (2010). Allowing x_i to vary continuously produces a *partial-allocation frontier* which, by construction, bounds any discrete allocation from above and is formally identical to the efficiency frontier in data-envelopment analysis (Charnes et al., 1978).

A.5.2 Empirical Construction of the Frontier

Experiment scores are normalised and merged with survey data on task hours; observations with missing entries or zero total hours are excluded. To trace the continuous PPF we solve, for a grid of α values,

$$\max_{\{x_i \in [0,1]\}} \sum_i (1 - x_i) n_i \quad \text{s.t.} \quad \sum_i x_i r_i \geq \alpha, \tag{6}$$

where $\alpha \in [0, \sum_i r_i]$. Each linear programme is of moderate dimension ($N \leq 200$) and is solved with the `PuLP` Python library; repeating (6) at 25 grid points yields a smooth frontier (Figure 5).[2]

Actual aggregate output is anchored by replacing the decision variables in (5) with observed time

[1] Scores were recorded both with and without access to generative-AI tools, giving two technology draws $(r_i^{\text{AI}}, n_i^{\text{AI}})$ and (r_i^0, n_i^0). The experiment therefore yields ex-ante, worker-specific production functions that embody the heterogeneous AI effects documented by Dell'Acqua et al. (2023).

[2] Mathematically, (6) is the dual of a constant-returns DEA problem (Banker et al., 1984).

shares from the NBS experiment. Comparing the pre-AI and post-AI PPFs therefore isolates the *task bias* of generative AI: an outward shift biased toward the R-axis signals routine-augmenting (routine-biased) change, whereas a tilt toward the N-axis indicates non-routine complementarity.

Our approach translates micro-level heterogeneity in AI productivity into macro-level production sets, providing a quantitative counterpart to the "jagged frontier" metaphor.. Conceptually, we mirror Hsieh and Klenow (2009) in treating micro-allocation wedges as a source of aggregate TFP losses, but extend their one-dimensional framework to a two-task environment and to a technology shock that changes the distribution of (r_i, n_i). The resulting envelope offers a transparent, non-parametric benchmark against which alternative structural models of AI adoption can be evaluated.

A.5.3 Distance to the Frontier

To quantify how far the observed allocation lies below the production-possibility frontier, we use an *output-oriented* efficiency measure. Let $(R_{\text{act}}, N_{\text{act}})$ be the actual aggregate outputs and let (r_i, n_i) denote the normalised routine and non-routine productivities of worker i in the post-AI technology. Holding routine output fixed at its realised level, we maximise potential non-routine output by solving

$$\begin{aligned} \max_{\{x_i \in [0,1]\}} \quad & \sum_i (1 - x_i)\, n_i \\ \text{s.t.} \quad & \sum_i x_i\, r_i \;\geq\; R_{\text{act}}, \end{aligned} \tag{7}$$

where x_i is the share of worker i's effort allocated to routine tasks. Problem (7) is a linear programme of dimension N that we solve with the PuLP optimiser; its optimum attains the frontier value $N_{\max}(R_{\text{act}})$.

The vertical gap

$$\Delta N \;=\; N_{\max}(R_{\text{act}}) \;-\; N_{\text{act}}$$

measures the *additional* non-routine output that could be produced through optimal re-assignment without sacrificing routine production. We report the percentage efficiency shortfall

$$\text{Eff. loss} \;=\; \frac{N_{\max}(R_{\text{act}}) - N_{\text{act}}}{N_{\text{act}}} \times 100,$$

A value of 7.33% indicates that non-routine production could rise by that amount if workers were reallocated along the frontier while keeping routine output unchanged. Next, we use the labor shares for the US economy to find that the potential percentage improvement in weighted output: 10.34%

B Additional experimental materials

B.1 Sample

In Figure B.5, we display the distribution of study participants across the organizational hierarchy at NBS.

B.2 General workplace tasks

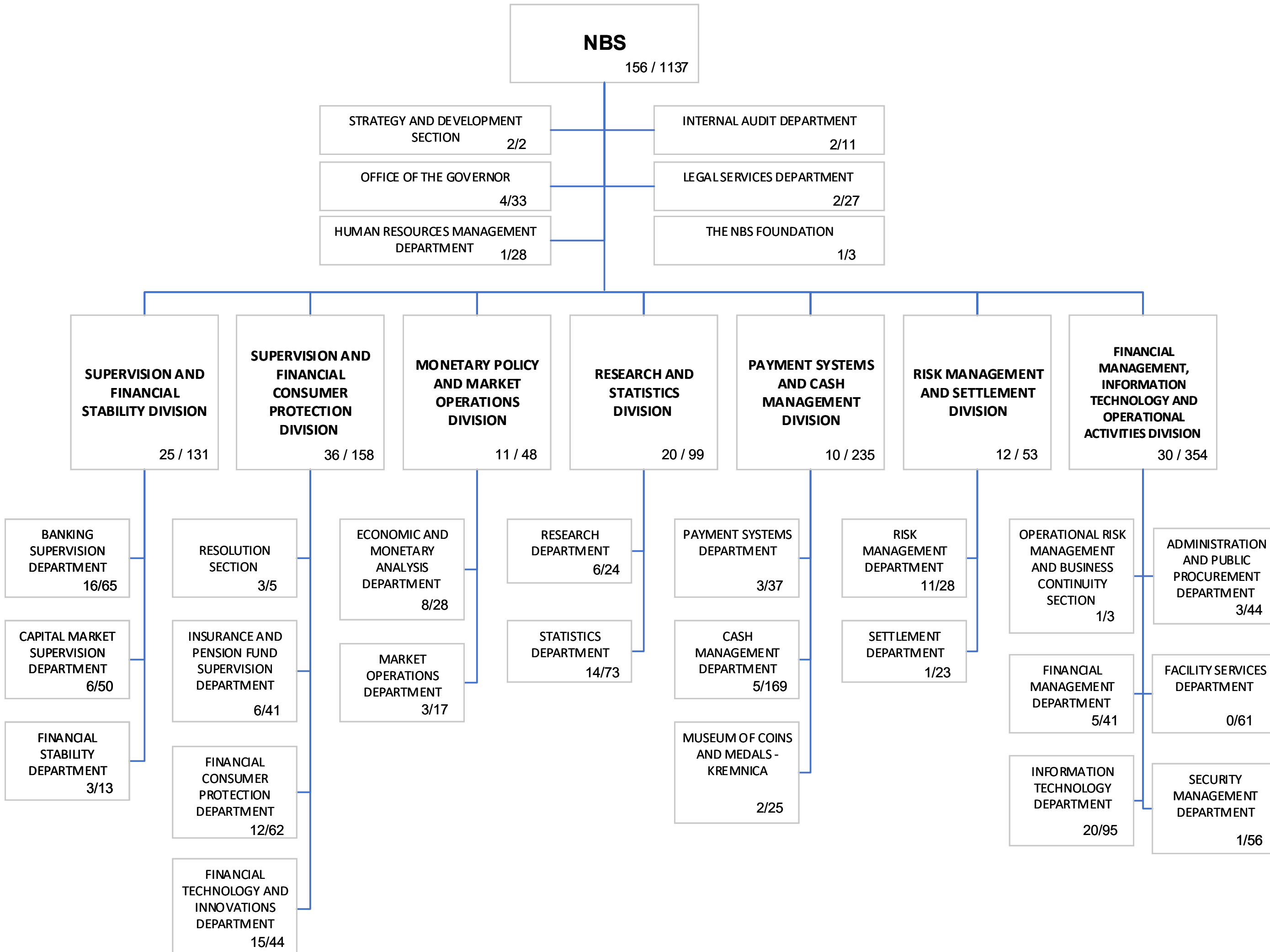
B.2.1 Routine-cognitive tasks

Task one Please proofread the following paragraph. Fix any grammatical or spelling issues. Submit the new paragraph in the text box below.

The global economy is a complex and interconnected system that encompasses the economic activities of countries worldwide. international trade plays a crucial role in the world economy, allowing nations to exchange goods, services, and capitol. Economic growth is influenced by various factors, including technological advancements, population growth, and government polcies. developed countries, such as the United States, Japan, and nations in Western Europe, have historically been major contributors to the world economy. However, emerging markets like china, India, and Brazil have increasingly gained economic prominence in recent decades. Economic crisises, such as the 2008 global financial crisis, have highlighted the importance of international cooperation and regulation in maintaining Global economic stability.

Task two Please fill in the table below Slovakia's seasonally adjusted unemployment rate by year.

Figure B.5: **Distribution of participants across NBS' hierarchy**



Notes: This figure displays the distribution of study participants across the organizational hierarchy at NBS.

Year	Unemployment rate
2000	
2005	
2010	
2015	
2020	

Table B.3: Unemployment Rates Over Time

B.2.2 Nonroutine-cognitive tasks

Task one You wake up one morning and discover you can talk to animals. What happens next? Write a narrative demonstrating your ability to think creatively and generate novel ideas. Please paste your submission in the following space.

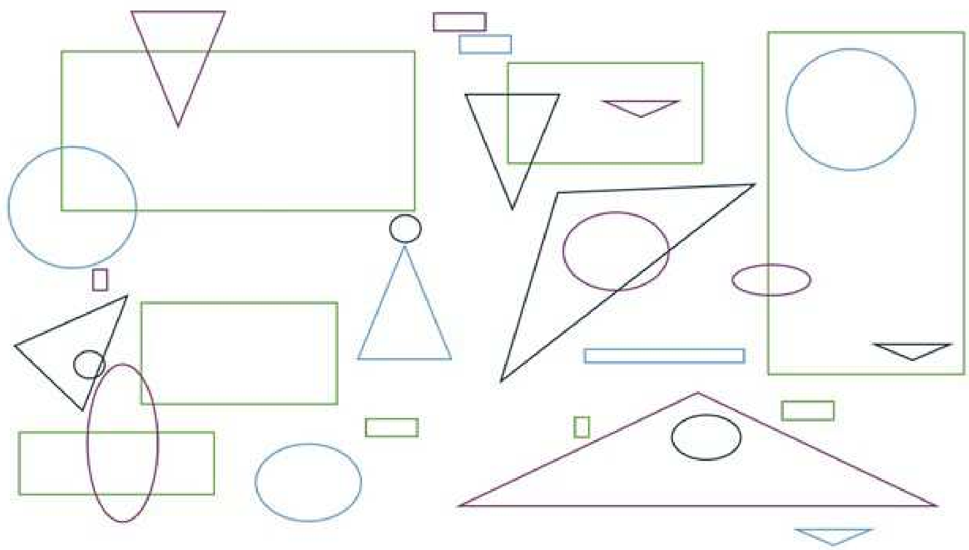
Task two Paperclips are usually used to hold paper today. List as many creative and innovative uses for a paperclip at the bank that you can think of. These should be alternative uses, beyond its intended purpose of holding papers together. Please paste your submission in the following table.

B.2.3 Routine-manual tasks

Task one You are a clerk at the bank responsible for entering user information into an electronic records system. In the next 5 minutes, accurately input the following details for 10 new users.

User	ID	Access Code
1	Dpah7638	8255
2	JSJPZ01	4813
3	X45PJ8mNL	5971
4	FPSM38	2358
5	bsh3279	16749
6	wqpbQg	59823
7	Djah8PP	187796
8	jCD8761z	154
9	83uisl09	135977
10	k3GbtWr	9402

Task two In the image below, you will find a picture of various shapes. Please fill in the total number of shapes, plus the total number of rectangles, circles/ovals, and triangles. Please paste your submission in the following table.



B.2.4 Nonroutine-manual tasks

Task one The printer at the bank is jammed, and you need to fix it. Please walk through the steps you would take to troubleshoot the issue and repair it.

Task two You need to give an improvised presentation, but don't have any visual aids prepared. How would you use typical office supplies (paper, markers, sticky notes, etc.) to create a simple but effective visual aid or model to support your presentation? Please paste your submission in the following space.

B.2.5 Analytical tasks

Task one The attached table contains daily sales data for FirmCo. Please analyze the data. On average, which day of the week does FirmCo have the largest sales? Enter your submission in the table below.

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	85	75	92	83	90
Week 2	87	92	86	93	83
Week 3	83	92	82	95	87
Week 4	90	85	89	98	94

Task two A bank is considering requiring all employees to work from the office at least three days a week. The bank's management team has provided the following statements and information:

- Employees who live more than 30 kilometers away from the office will be exempt from the new policy.
- Employees with children under the age of 5 will also be exempt from the new policy.
- 40% of the company's employees live within 30 km of the office.
- 25% of the company's employees have at least one child under the age of 5.
- There are no employees who both live more than 30 km away and have a child under 5.

What percentage of employees will be required to follow the new policy and work from the office at least three days a week? If the company has a total of 500 employees, how many employees will be exempt from the new policy?

B.2.6 Leadership tasks

Task one As a project manager, you have the following tasks and initiatives that need to be prioritized:

- Launching a new product feature
- Conducting team training on a new software tool
- Updating the company's website
- Preparing quarterly reports for stakeholders

- Hiring and onboarding new team members

Rank these items in order of priority and provide a justification for your ranking.

Task two You are the team lead for a group of IT employees at the bank. One of your team members, who was previously a strong performer, has been consistently submitting work with errors and missing deadlines over the past month. Prepare a script for how you would approach a one-on-one feedback conversation with this team member to address their performance issues

B.2.7 Communication tasks

Task one You've recently completed a project with your colleague, Alex. Alex played a crucial role in the project's success, going above and beyond to ensure everything was completed on time and to a high standard. Your manager has praised the project's outcome and has asked you to inform the rest of the team about the project's success.

Draft an email to Alex with the following objectives:

- Express your gratitude for their hard work and dedication to the project.
- Highlight that Alex's contributions were instrumental to the project's success.
- Inform Alex about the positive feedback received from your manager.
- Suggest celebrating the project's success together, either in-person or virtually.

Please paste your submission in the following table.

Task two You have been outperforming at work and would like to ask your manager for a raise. Draft an email to your manager with the following objectives:

- Clearly explain the work that you have been doing and how you have been outperforming
- Ask for a pay raise of 5%
- Explain why you deserve the raise

- Request a follow-up meeting with your manager to further discuss

Please enter your submission in the following space.

B.3 Specialized workplace tasks

B.3.1 Economist tasks

Task one Welcome to the Econoland Monetary Policy Report exercise. As part of this case study, you will be tasked with creating a Monetary Policy Report for the fictitious country, Econoland. This report is crucial for the Monetary Policy Council (MPC) in setting and guiding monetary policies of the country.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic stability of the country. Quarterly data for GDP and Inflation for the last 15 years are in the excel you received with this document, including GDP, Consumption, Investment, Unemployment, Inflation and Central bank interest rates.

1. Part 1
   (a) Use the provided data to generate basic descriptive statistics for each variable (mean, median, standard deviation). If any variable needs to be adjusted do so. What other statistics could be beneficial for the MPC meeting? Create simple graphs (line charts) showing the trends of GDP, unemployment, and inflation over the 15 years.
   (b) To work out the report you must schedule a series of meetings with the Team. Using a calendar application schedule necessary meetings with innovatewithai@nbs.sk for the following two weeks. Steps: *i*) Identify potential meeting dates and times, *ii*) send out meeting invitations using the calendar application. *iii*) Ensure there are no scheduling conflicts
   (c) During the compilation of the final report, there is a miscommunication within the team regarding the latest inflation data, leading to inconsistencies in the initial summary statistics and charts. Email 1:
   From: Economic Analyst
   Subject: Discrepancy in Inflation Data
   Body: "Hi Team, I've noticed that the inflation data for the past 5 years in our summary report doesn't match the raw data provided. Can someone double-check the entries? I believe there might be a mistake. Thanks, [Analyst 1]"
   Email 2:
   From: Team Leader
   Subject: Urgent: Inflation Data Correction Needed
   Body: "Hello Team, We've identified a potential error in the inflation data entries for the past 5 years. This needs immediate correction as it impacts our summary statistics and charts. Please review and correct the data at your earliest convenience. Once corrected, ensure that all charts and statistics are updated accordingly. Best, [Team Leader]
   As part of the team, you need to formulate a response to address the discrepancy and outline steps for correction.

Steps: Read the Emails and formulate response email to the team acknowledging the error. Outline the steps to correct the data calculated in (a).

2. Part 2
    - On the data set from 2010 till 2024 find the cyclical and trend component of GDP using HP filter (two sided or one sided, comment on your decision).
    - Fit the ARIMA model to the HP trend.
    - Use the fitted model to forecast the trend component beyond the sample period.
    - Develop a strategic plan for monetary policy adjustments over the next year based on projected economic trends.
3. Part 3
    - Perform a regression analysis to explore the relationship between inflation and unemployment rates (Phillips curve) using the provided data.
    - Write up a paragraph about what your estimates imply for the current monetary policy.
4. Part 4
    - Write up summary report based on your Monetary Policy Report which should targeted to the general public.

Task two **Background:** You work as a researchers in the Monetary Policy Department of Econobank, your insights and analyses play a crucial role in shaping our country's economic policies. At Econobank, we are committed to staying at the forefront of economic research and policy development. To maintain this commitment, we continuously explore new innovative ideas and technologies that can support our applied work. Your expertise and dedication are vital to the success of Econobank.

Objective: This case study consists of two main tasks structured to simulate real-world scenarios where you will:

1. Summarize a provided research article and answer specific questions related to the paper.
2. Write a short literature review on a given topic related to monetary policy.

---

1. Article Summary and Specific Questions
    - **Task**
        (a) Read the article you received together with this task sheet: Why Do We Dislike Inflation by Stefanie Stantcheva. Your task is to summarize the key points, findings, and implications of the article concisely and accurately.
    - **Task**
        (a) How does the paper defines "shrinkflation" ?
        (b) How do individuals adapt their purchasing behaviors in response to escalating prices? Specifically, do they preemptively increase their purchases of durable goods to avoid future even higher prices, or do they reduce their expenditure on such items due to diminished disposable income after covering essential nondurables?

(c) One result of Calvo style sticky price model is that in times of increased inflation firms' costs are growing faster than firms' revenue. Does this finding align with the perspectives of the respondents?

(d) Does the Phillips curve - suggesting an inverse relationship between unemployment and inflation - reflect the beliefs of the respondents?

2. Literature Review

- Conduct a search for relevant academic papers, reports, and articles on the people believes about what drives cost of inflation.
- Write short literature review summarizing the key findings, debates, and gaps in the existing literature.

B.3.2 Finance tasks

Task one Welcome to the Econoland Finance exercise. As part of this exercise, you will be working for the fictitious central bank in the country Econoland.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task: Verify the banking transactions you received together with this task document.
  1. Verify the banking transactions you received together with this task document.
  2. Check for any discrepancies between the document details and the entries.
  3. Report any discrepancies you find
- Task: Suppose that you are considering investing in two stocks. After analyzing the two stocks you think that the are two possible states for the economy over the next year: *Good* and *Bad*. Each state is equally likely (probability of 0.5). The returns of the two securities in each state are as follows.

State	return to stock 1	return to stock 2
Good	30%	5%
Bad	10%	10%

  1. What is the expected return and standard deviation of each stock return?
  2. What is the covariance and the correlation between the two stock returns?
  3. Draw a picture to illustrate the tradeoff between risk and return that is available by investing in these two stocks
  4. Suppose that a risk-free investment of 5% is also available. Does this present a profit opportunity to you? Why or why not?
- Task
  1. Read through the attached case study

2. Develop a robust response strategy for potential financial crises, drawing lessons from Bear Stearns' collapse.
3. Instructions:

   Case Study Review: Analyze the case study to understand the key factors leading to Bear Stearns' failure, focusing on leverage, liquidity, and trust.
   Simulation Development: Create a simulation scenario where similar crisis factors begin to affect the market. Use historical data and hypothetical stressors (e.g., sudden liquidity dry-up, rapid decline in asset values).
   Strategy Formulation: Formulate a crisis response strategy that includes immediate actions, communication plans, and longer-term measures to restore stability and confidence.
   Presentation: Present the crisis simulation and response strategy to the trading department, highlighting potential warning signs and action steps.
   Deliverables: A detailed description of the crisis simulation scenario. A crisis response strategy document. A report outlining the scenario, potential impacts, and response strategy.
   Evaluation Criteria:

   – Relevance and realism of the crisis simulation.
   – Effectiveness and comprehensiveness of the crisis response strategy.
   – Clarity and persuasiveness of the presentation.

Task two **Background:** Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task

Table B.4: SEPA Transaction Documents

Document ID	**Transaction Date**	**Counterparty Name**	**Transaction Amount**	**Reference Number**
001	2023-06-15	Alpha Corp	€5,000	RF783456
002	2023-06-17	Beta Ltd	€20,000	RF783457
003	2023-05-30	Gamma Inc	€9,500	RF783458
004	2023-06-20	Alpha Corp	€15,000	RF783459
005	2023-05-25	Delta PLC	€8,000	RF783460
006	2023-06-18	Beta Ltd	€3,000	RF783461
007	2023-05-28	Gamma Inc	€12,000	RF783462
008	2023-06-15	Alpha Corp	€7,000	RF783463
009	2023-05-22	Delta PLC	€6,500	RF783464

1. Create a spreadsheet entering the data from the table.
2. Include filters in the spreadsheet to sort by date and counterparty. Check for any discrepancies between the document details and the entries.
3. paste the table into the text editor of your choice to be included in the final pdf you submit

- Task

1. You are at lunch one day with your friend and the conversation turns to investment in stocks and bonds. Your friend has little experience with investment in stocks. You start

by explaining the benefits of diversification. Your friend then says, "This is great!" That means if I invest in 1,000 stocks I will be able to avoid risk altogether. Why have I been keeping my money in the bank all this time?" Under what situation would your friend be correct? Do you think that this situation is realistic? What further "instruction" would you give your friend?

- Task
  1. Read through the case study and provide short summary
  2. What ideas from the case study could be implemented at NBS to manage its portfolio of assets. How does the case relates to the portfolio management at the Central Bank?

B.3.3 IT tasks

Task one Welcome to the IT exercise. Imagine you are working at fictitious central bank in the country Econoland. As part of this case study, you will be tasked with several tasks you might encounter as an IT expert.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task: You can use any programming language you want. In the PDF you submit please present the code and name of the programming language it can be verified that the code is running correctly. You should also show in the PDF (as a table) the first 10 rows and 10 columns of the resulting matrix
  1. create a 5000 times 5000 matrix with random numbers from the standard normal distribution, time your code using relevant timing instruction at the stat and end of our code. Next, write a loop filling the 5000 by 5000 matrix
     - the first option has an outer loop over rows and an inner loop over columns
     - the second has an outer loop over columns, and an inner over rows
     - which one is faster? Why?
- Task
  1. Please comment on the steps and actions you would prefer in the below scenario. Please provide as much detail as possible.

     A system for recording company expenses was developed internally in your organization many years ago. The system is hosted on premise. It has considerable technology debt associated with security gaps. The company is not investing in the resolution of the debt, since the replacement is on the go. However due to major delay on the new technology implementation the legacy system needs to be prolonged. The legacy system cannot be upgraded due availability of resources. What would you need to consider on tech debt, security gaps, risks and stakeholders‘s agreements. What would be your course of action as system owner to prolong the lifecycle of the application.
- Task

1. The system from previous example is heavily integrated with other applications on your company‘s landscape. The system needs to be decommissioned. What would such decommissioning plans document contain. Please be specific on all the action items you would need to take for the decommissioning.

Task two Welcome to the IT exercise. Imagine you are working at fictitious central bank in the country Econoland. As part of this case study, you will be tasked with several tasks you might encounter as an IT expert.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task: You can use any programming language you want. In the PDF you submit please present the code and name of the programming language it can be verified that the code is running correctly. You should also show in the PDF (as a table) the first 10 rows and 10 columns of the resulting matrix
  1. Write a loop to create a matrix A where $A = 1/(row + column - 1)$
- Task: The internal customer has reached out to the IT department with a request to replace a finance system. You are in the middle of a major upgrade for that system. The finance colleagues have already mapped systems on the market which would replace this system.
  1. Please write a detailed plan on the selection of such system. What would be the technology criteria and questions to assess such selection. Who would be the people involved into the process and what aspects of technology they should assess.
  2. Once the system is selected, assume a cost is provided by the vendor. How would you assess the total cost.
- Task: Continue with the example above.
  1. You have completed the assessment of several proposed solutions on the market. Some came with different strength and gaps. Assume the integration capabilities of that system is not up to the standards of your organisation however you believe that the system is scoring best among others and that the risk associated with lacking integration capabilities could be mitigated by using older standards. At the same time, given that the upgrade on the legacy system is in progress and the new system needs to be implemented asap due to compliance functional needs, assess the clash of new implementation and the upgrade. Please produce a business case and include all relevant aspects of decision making process. Include the findings of the selections criteria. Your output should be a formal business case document requesting endorsement form the leadership , covering the proposed solutions, costs, risks, assessment alternatives , clash with the upgrade and potential change in upgrade approach , resourcing needs and more.

B.3.4 Payments tasks

Task one Welcome to the Econoland Payment System Report exercise. As part of this case study, you will be tasked with creating a Payment System Report for the fictitious central bank in the country Econoland.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task

Table B.5: SEPA Transaction Documents

Document ID	Transaction Date	Counterparty Name	Transaction Amount	Reference Number
001	2023-06-15	Alpha Corp	€5,000	RF783456
002	2023-06-17	Beta Ltd	€20,000	RF783457
003	2023-05-30	Gamma Inc	€9,500	RF783458
004	2023-06-20	Alpha Corp	€15,000	RF783459
005	2023-05-25	Delta PLC	€8,000	RF783460
006	2023-06-18	Beta Ltd	€3,000	RF783461
007	2023-05-28	Gamma Inc	€12,000	RF783462
008	2023-06-15	Alpha Corp	€7,000	RF783463
009	2023-05-22	Delta PLC	€6,500	RF783464

1. Create a spreadsheet entering the data from the table.
2. Include filters in the spreadsheet to sort by date and counterparty. Check for any discrepancies between the document details and the entries.
3. paste the table into the text editor of your choice to be included in the final pdf you submit

- Task

1. In this task you should write a comprehensive report detailing each step of a SEPA transaction, including diagrams or flowcharts that illustrate the transaction process. More specifically,
   - **Understanding and Describing SEPA Transactions**. This task is designed for employees within the Payments System Department at a central bank, focusing on understanding and detailing the Single Euro Payments Area (SEPA) transaction processes. SEPA aims to simplify and streamline bank transfers denominated in euro, ensuring that cross-border payments are as straightforward as domestic payments. The objective of this task is to deepen the employees' understanding of the SEPA transaction framework and the bank's role in ensuring immediate settlement of payments by examining the SEPA rule book. In the first step you should describe the technical and regulatory processes involved in SEPA transactions, highlighting the bank's responsibilities. Reveiw SEPA rule book relevant to SEPA Credit Transfers and SEPA Instant Credit Transfers. Focus should be on the rules and guidelines that govern transaction processing and settlement mechanisms.
   - In the Step 2 describe SEPA Transaction Processes. This should include: how a SEPA transaction is initiated by the payer, including the information required to process the payment. Processing: The steps taken by the bank to verify and forward the transaction details to the payee's bank. Settlement: How the funds are settled between banks to ensure that the payee receives the payment within the specified time frame.
   - Step 3 relates to behind the Scenes of SEPA Transactions: Explain the technical backend processes that support SEPA transactions, such as: the role of interbank systems in facilitating transactions. How banks ensure compliance with SEPA requirements, including data security and anti-fraud measures. The impact of real-time processing in SEPA Instant Credit Transfers and the technologies that support it.

  - In step 4 explain Bank's Role in Immediate Settlement, detail the specific actions and systems a bank must have in place to ensure the immediate settlement of SEPA payments, such as: Liquidity management to guarantee funds availability for immediate transfer. Participation in the pan-European automated clearing house (ACH) systems and Ensuring 24/7 availability of systems to process instant payments.

- Task

  1. Develop a strategic plan detailing the phased adoption of the digital euro within national banks and propose integration strategies with existing financial infrastructures.
  2. Engage in a simulation exercise to address a hypothetical cybersecurity breach affecting the digital euro transactions. Describe the potential breach and possible measures which should be taken.

Task two Welcome to the Econoland Payment System Report exercise. As part of this case study, you will be tasked with creating a Payment System Report for the fictitious central bank in the country Econoland.

Background: Econoland is a small country with a diverse economy. Over the past 15 years, it has experienced various economic cycles, including a significant recession. The central bank of Econoland, Econobank, has played a critical role in managing the economic and financial stability of the country.

- Task: Verify the banking transactions you received together with this task document.

  1. Verify the banking transactions you received together with this task document.
  2. Check for any discrepancies between the document details and the entries.
  3. report discrepancies if any

- Task

  1. describe the difference between standard payment settlement and immediate settlement in terms of processes and impact on banking system

- Task

  1. Imagine that an unidentified threat actor has gained unauthorized access to the TARGET2 system, potentially compromising transaction integrity and confidentiality. The breach was detected by anomaly detection systems that noticed unusual transaction patterns and unauthorized access attempts to the system database. Create Incident Report: A comprehensive report detailing the breach, investigation findings, impact assessment, and recommended corrective actions. Include the questions below.

     Task

     (a) Outline the immediate steps to be taken following the detection of the breach. This includes isolating affected systems, notifying relevant stakeholders, and initiating a preliminary investigation.

     (b) Conduct a detailed investigation to determine the source and extent of the breach. You need to work with cybersecurity teams to analyze logs, transaction records, and system access patterns.

(c) Communication: Develop a communication strategy for internal stakeholders, regulatory bodies, and the public to explain what happened, what is being done, and how such incidents will be mitigated in the future.

(d) Recovery and Strengthening: Propose measures to recover any compromised data and secure the system against future attacks. This might involve updating security protocols, enhancing monitoring systems, and conducting regular security audits.

(e) Policy Update: Review and update the existing cybersecurity policies based on the lessons learned from the incident. Include recommendations for preventing similar breaches, focusing on both technological upgrades and training for relevant personnel.

References

Acemoglu, Daron, "The Simple Macroeconomics of AI," *Working paper*, 2024.

_ **and David Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, pp. 1043–1171.

_ **and** _ , "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171.

_ **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.

_ **and** _ , "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, Spring 2019, *33* (2), 3–30.

_ **and** _ , "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

_ **and** _ , "Automation and Rent Dissipation: Implications for Wages, Inequality, and Productivity," Working Paper 32536, National Bureau of Economic Research June 2024.

Aldasoro, I. et al., "Generative Artificial Intelligence and Cyber Security in Central Banking," BIS Papers 145, Bank for International Settlements 2024.

Araujo, D. et al., "Machine Learning in Central Banking," IFC Bulletins 57, Irving Fisher Committee on Central Bank Statistics 2022.

Autor, David H, "The "task approach" to labor markets: an overview," *Journal of Labour Market Research*, 2013, *46* (3), 185–199.

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 11 2003, *118* (4), 1279–1333.

Banker, Rajiv D., Abraham Charnes, and William W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, 1984, *30* (9), 1078–1092.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Brynjolfsson, Erik and Paul Milgrom, "Complementarity in Organizations," Draft chapter for the Handbook of Organizational Economics 2012. August 2008; Revised December 2010 and January 2012.

_ **, Danielle Li, and Lindsey R Raymond**, "Generative AI at Work," Working Paper 31161, National Bureau of Economic Research April 2023.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Yuanzhi Li, Xi Wang, Yuhuai Zhang, Shun Garg, Preetum Nakkiran, Sham Narayanan, Lance Ansell, Joseph Merullo, Sergio Boixo, Karen Simonyan, Oriol Vinyals, Demis Hassabis, and Ilya Sutskever, "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," 2023. arXiv preprint arXiv:2303.12712.

Charnes, Abraham, William W. Cooper, and Edward Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 1978, *2* (6), 429–444.

Costinot, Arnaud and Jonathan Vogel, "Matching and Inequality in the World Economy," *Journal of Political Economy*, 2010, *118* (4), 747–786.

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," Working Paper 2024.

Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski, "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination," *The Review of Economics and Statistics*, 05 2025, pp. 1–16.

_ **, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani**, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper No. 24-013, Harvard Business School Technology & Operations Mgt. Unit September 2023.

Dell'Acqua, Fabrizio, Edward III McFowland, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," 2023. Working paper, Harvard Business School.

Demirci, Ozge, Jonas Hannane, and Xinrong Zhu, "Who is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms," CESifo Working Paper 11276 2024.

Doerr, S. et al., "Big Data and Machine Learning in Central Banking," BIS Working Papers 930, Bank for International Settlements 2021.

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," 2023.

Felten, Edward W., Manav Raj, and Robert Seamans, "Occupational Heterogeneity in Exposure to Generative AI," April 2023. Available at SSRN: https://ssrn.com/abstract=4414065 or http://dx.doi.org/10.2139/ssrn.4414065.

Goos, Maarten, Alan Manning, and Anna Salomons, "Explaining job polarization: Routine-biased technological change and offshoring," *The American Economic Review*, 2014, *104* (8), 2509–2526.

— **, Emilie Rademakers, and Ronja Röttger**, "Routine-Biased technical change: Individual-Level evidence from a plant closure," *Research Policy*, 2021, *50* (7), 104002.

Graetz, Georg and Guy Michaels, "Robots at Work," *The Review of Economics and Statistics*, 12 2018, *100* (5), 753–768.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Hui, Xiang, Oren Reshef, and Luofeng Zhou, "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," *Organization Science*, 2024, *0* (0).

Humlum, Anders and Emilie Vestergaard, "The Adoption of ChatGPT," Working Paper, University of Chicago, Becker Friedman Institute for Economics 2024.

Kreacic, Ana and Terry Stone, "Women are Falling Behind on Generative AI in the Workplace. Here's How to Change That," World Economic Forum 2024. Available at: https://www.weforum.org/agenda/2024/04/women-generative-ai-workplace/.

List, John A., "The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs," NBER Working Paper 33456, National Bureau of Economic Research February 2025.

Michaels, Guy, Ashwini Natraj, and John Van Reenen, "Has ICT polarized skill demand? evidence from eleven countries over twenty-five years," *The Review of Economics and Statistics*, 2014, *96* (1), 60–77.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

OpenAI, "Hello GPT-4o," Technical Report May 2024.

Otis, Nicholas G., Katelyn Cranney, Solène Delecourt, and Rembrand Koning, "Global Evidence on Gender Gaps and Generative AI," 2024. Working Paper.

Perkowski, Patryk and Aleš Maršál, "Unlocking the productiviyty benefits of generative AI: Evidence from three central banks," *Working paper*, 2024.

Schwarcz, Daniel, Sam Manning, Patrick James Barry, David R. Cleveland, J. J. Prescott, and Beverly Rich, "AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice," Minnesota Legal Studies Research Paper No. 25-16; University of Michigan Public Law Research Paper (forthcoming) March 2025. Available at SSRN.

Spitz-Oener, Alexandra, "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 2006, *24* (2), 235–270.

Tambe, Prasanna, "Reskilling the Workforce for AI: Domain Experts, Algorithms, and Organizational Performance," *Management Science*, 02 2025.