Stanford | Institute for Economic Policy Research (SIEPR)

AI AND THE FED

Sophia Kazinnik
Stanford University

Erik Brynjolfsson
Stanford University and NBER

JULY, 2025

Working Paper No. 25-25

NBER WORKING PAPER SERIES

AI AND THE FED

Sophia Kazinnik
Erik Brynjolfsson

Working Paper 33998
http://www.nber.org/papers/w33998

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025

The authors gratefully acknowledge the contributions of Anne L. Hansen and Seung Jung Lee, which greatly improved this paper. We also thank Thomas Cook, Ajit Desai, and Marco Macchiavelli for helpful comments and suggestions. We are grateful to the Stanford Digital Economy Lab for funding and support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33998

AI and the Fed
Sophia Kazinnik and Erik Brynjolfsson
NBER Working Paper No. 33998
July 2025
JEL No. C8, C9, G4

ABSTRACT

This paper examines how central banks can strategically integrate artificial intelligence (AI) to enhance their operations. Using a dual-framework approach, we demonstrate how AI can transform both strategic decision-making and daily operations within central banks, taking the Federal Reserve System (FRS) as a representative example. We first consider a top-down view, showing how AI can modernize key central banking functions. We then adopt a bottom-up approach focusing on the impact of generative AI on specific tasks and occupations within the Federal Reserve and find a significant potential for workforce augmentation and efficiency gains. We also address critical challenges associated with AI adoption, such as the need to upgrade data infrastructure and manage workforce transitions.

Sophia Kazinnik
Stanford University
Digital Economy Lab
kazinnik@stanford.edu

Erik Brynjolfsson
Stanford University
and NBER
erik.brynjolfsson@gmail.com

1 Introduction

We are in the midst of a technological revolution. While debates persist over timelines—be they two years or two decades[1] —the fundamental question is no longer *whether* machines will broadly match human intelligence, but *when*. This tipping point, often termed artificial general intelligence (AGI)[2], promises to redefine productivity, consumption, investment, income distribution, wealth, and labor markets (e.g., Aghion et al., 2017; Trammell and Korinek, 2023; Videgaray et al., 2024; Auer et al., 2024). But even today, integrating AI into established workflows can yield profound effects (Dell'Acqua et al., 2023; Wang et al., 2024).

How might central banks, keepers of monetary and financial stability, respond to these technological changes? While there is an extensive discussion about how central banks can respond to AI-driven changes in the wider economy (e.g., Cook, 2024; Harker, 2024; Hornstein, 2024; Aldasoro et al., 2024b), less attention has been paid to how they can use AI internally.[3] In this paper, we address this gap, focusing on the Federal Reserve System (henceforth, the Fed) as a representative central bank.[4]

We show how AI might be used to enhance Fed's core functions, including monetary policy, financial stability oversight, bank supervision, and economic research.[5] While

[1]Feng et al. (2024) conducted a survey at ICLR, a top ML conference, and found researchers expecting Artificial General Intelligence (AGI) in: 1–2 years (3.6%), 2–5 years (13.0%), 5–10 years (22.5%), 10–20 years (23.9%), or 20+ years (37.0%). Meanwhile, Metaculus's median forecast for human-level AI slipped from 2062 in 2020 to 2032 today, with 25% predicting it by 2026 (Metaculus, 2025). However, 76% of AI researchers in an AAAI survey judged that merely scaling current methods is unlikely to yield AGI (Association for the Advancement of Artificial Intelligence, 2025). The debate continues.

[2]As can be seen by the prevalence of many other equivalent terms, such as Strong AI, Full AI, Human-level AI or Transformative AI, this state of the world may occur gradually.

[3]The Bank for International Settlements (BIS) recently published a report on AI adoption in central banks through a governance lens (Bank for International Settlements, 2025), emphasizing ethical compliance and structured risk management. This manuscript differs in that the focus is on operational and strategic transformation.

[4]While each central bank functions within its own economic and political environment, the fundamental mandates and policy instruments remain broadly comparable across jurisdictions.

[5]In this paper, we use FRS and Fed interchangeably – unless stated otherwise, we refer to the system that includes all of the 12 reserve banks and the Federal Reserve Board.

the focus is on the Fed, our framework can be easily adapted by other central banks to strengthen their own policy, supervisory, and research functions. Our analysis finds that roughly a quarter of Fed occupations (particularly economists, data scientists, and financial analysts) stand to gain significant AI-driven productivity boosts. This suggests significant opportunity for augmentation. Assuming other central banks exhibit a similar workforce composition, they too could realize comparable opportunities for augmentation.

Early evidence suggests that the adoption of AI in government lags behind its implementation in the private sector (SAS Institute Inc., 2024), given the fundamental structural differences between these sectors. Unlike private firms that prioritize profit maximization or operational efficiency, a central bank is driven by public policy objectives, and often navigates ambiguous situations without a single "ground truth".[6] Beyond that, bureaucratic processes often clash with agile innovation cycles that drive adoption in the private sector.

Another major reason for slower adoption in government is outdated infrastructure. Many government systems rely on aging technologies, with some business processes still heavily dependent on manual data entry. The U.S. Department of Defense, for instance, only began transitioning from waterfall to Agile software development in 2023.[7] In addition, data silos can make it harder for different systems and departments to share and use information. Despite collecting petabytes of data, governments lack the infrastructure to prepare it for the AI era. In contrast, private-sector firms oftent circumvent these challenges through centralized data lakes.[8]

[6] For example, the Fed is tasked with promoting maximum employment, price stability, and moderate long-term interest rates, as outlined in the Federal Reserve Act, see `https://www.federalreserve.gov/aboutthefed/section2a.htm`.

[7] Efforts to modernize these systems, such as the TRACTOR project of the Defense Advanced Research Projects Agency (DARPA) to rewrite COBOL in Rust, highlight the scale of the challenge (see `https://www.darpa.mil/research/programs/translating-all-c-to-rust`).

[8] In essence, a data lake acts as a central library for all types of information, as opposed to separate filing cabinets that don't talk to each other. Salesforce's Einstein AI platform, for instance, integrates

AI is a general-purpose technology, and adopting it in a truly transformative way demands deep organizational changes. Bresnahan and Trajtenberg (1995) argue that most GPTs function as enabling technologies, creating new opportunities rather than offering complete solutions. For instance, the productivity gains from introducing electric motors in manufacturing extended beyond just reducing energy costs. Agrawal et al. (2023) similarly argue against use-case driven adoption, which assumes that organizations can plug new technologies into existing structures without making fundamental changes. Successful GPT adoption requires organizations to rethink processes, retrain workers, redesign workflows, and be willing, when necessary, to engage in *creative destruction* to fully achieve productivity gains.

In contrast to earlier GPTs that primarily impacted mostly manual, labor-intensive job functions, generative AI is expected to disrupt a new set of "cognitive" and "non-routine" tasks (e.g., Kinder et al., 2024). Generative AI excels at creative and complex reasoning by drawing inferences from associations, often outperforming humans in predictive tasks, even in areas once dominated by experts (Brynjolfsson and Mitchell, 2017).

At the same time, integrating generative AI into expert workflows presents significant challenges. Agarwal et al. (2023) study collaborations between an AI system and professional radiologists. They find that even though AI outperformed 75% of the radiologists in diagnostic accuracy, its assistance did not enhance the overall quality of diagnoses, suggesting that experts often struggle to reconcile AI-driven predictions with their contextual knowledge. This shows that simply adding AI for isolated use cases often fails, as experts tend to resist or ignore algorithmic advice. Instead, Agarwal et al. (2023) argue that organizations must redesign workflows so human expertise and

customer data from over 30 sources to enable real-time analytics. US banks are also extensively adopting and using data lakes. For example, JPMorgan Chase has invested in building a decentralized data lake that integrates data from across its global operations.

AI work hand in hand.

In addition, the full benefits of integrating AI into an organization's workflows are likely measured with a delay, which may generate additional resistance towards AI adoption. The "productivity J-curve" (Brynjolfsson et al., 2021) describes how new technologies, especially GPTs, deliver productivity gains only after a period of investment in complementary intangible assets, such as business processes and new skills. New technologies can even temporarily drag down measured productivity. As a result, earlier GPTs like electricity and the first wave of computers took decades to have a significant effect on productivity.

Faced with digital currencies, DeFi, cross-border flows, plus the vulnerabilities of real-time analytics, such as cyber threats and data integrity concerns, central banks face mounting pressure to modernize (Doerr et al., 2021; Berger et al., 2022; Doerr et al., 2022; Aldasoro et al., 2024a). Given these challenges, there may be extra incentives for central banks to integrate AI into their operations to maintain their effectiveness.[9] With these considerations in mind, we explore how central banks can effectively integrate AI with human expertise. In this partnership, people remain indispensable: they bear ultimate responsibility and handle challenges that AI cannot address.

The rest of the paper is organized as follows. Section 2 outlines the top-down approach, discussing how AI can contribute to existing policy workflows in central banks. Section 3 outlines components that might be part of a comprehensive data and IT strategy for AI integration. Section 4 discusses human capital development, up-skilling, and resource shifts that would likely be needed to prepare the bank's workforce for AI integration. Section 5 concludes.

[9] In a recent speech, Michael S. Barr presents two hypothetical scenarios for the AI evolution, both of which carry non-trivial implications for central banks and the broader economy and financial system(Barr, 2025).

2 AI & Data-Driven Policy Workflow

Although central banks may differ in mandates and specific organizational responsibilities, virtually all modern monetary authorities operate through a fundamentally similar, data-driven process. In practice, each responsibility follows a common, dynamic set of stages in which raw inputs are gathered, filtered, analyzed, summarized, reviewed by decision makers, and then communicated to a wider audience, with feedback from later stages often looping back to refine earlier ones. Figure 1 provides a high-level (and static) illustration of this process, showing how traditional and alternative data flow through a "data funnel" into analysis and reports, inform decisions, and ultimately become policy communications.

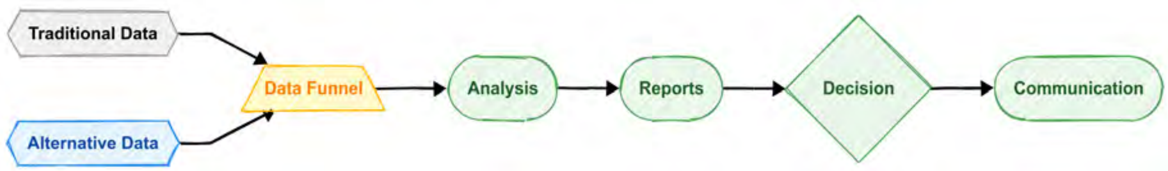

Figure 1: A simplified representation of data-driven policy workflow

In the section that follows, we describe how AI could help enhance each stage of this pipeline using the following core functions of the Federal Reserve as an applied example.

2.1 Monetary Policy

Sixty years ago, Fed economists needed days to gather reliable GDP figures. By the early 2000s, desktop software cut that process to minutes. Today, AI can do it almost instantly. With advances in computing power, the Fed's policy process has moved from slow, retrospective data tables to more timely, contemporaneous and even forward-looking analyses. AI is helping to refine that transition. This change isn't just about speed; it reshapes how the Federal Open Market Committee (FOMC) pursues its goals of price stability and maximum employment.

FOMC decisions are informed by quantitative assessments of inflation metrics, labor market statistics, and financial market indicators via econometric models and scenario evaluations. Forward-looking projections such as the Summary of Economic Projections (SEP) translate these model outputs into policy benchmarks, while qualitative information collected from regional contacts is synthesized to provide contextual nuance (e.g., Beige Book). The combination of quantitative data, market signals, forecasts, and qualitative information helps the committee in fulfilling its mandate under conditions of economic uncertainty.

The information that feeds into this process can be broadly categorized into two types: traditional data and alternative data. Traditional data refers to well-established sources typically used in policy analysis: macroeconomic indicators, financial market data, and official statistics. In contrast, alternative data includes less conventional but potentially more timely and/or granular information, such as consumer sentiment, social media activity, search trends, satellite imagery, and mobile location data. While both types of data are relevant for decision-making, central banks have historically relied more heavily on traditional data and often lack the infrastructure and expertise to effectively use alternative data sources despite the valuable signals they provide.

In the remainder of this section, we explore how AI could be integrated across key stages of the monetary policy workflow.

2.1.1 Data

Generative AI has significant potential to enhance macroeconomic forecasting and measurement by transforming two critical processes: *data gathering*, which involves sourcing and extracting insights from diverse (and often unstructured) data, and *signal filter-*

ing, where the most relevant indicators are identified and prioritized within the "data funnel" for downstream analysis.

Traditional economic metrics, such as inflation, output, and employment, often rely on data sources that are lagging and limited in granularity. Generative AI can address these limitations by automatically capturing and processing alternative, high-frequency data sources. For instance, inflation can be tracked in near-real-time through online product listings, social media sentiment, Google search trends, satellite imagery, and earnings call transcripts (Wu and Brynjolfsson, 2015; Angelico et al., 2022; Gosselin and Taskin, 2023; Cajner et al., 2024). The Bank for International Settlements (BIS) uses generative AI precisely for this purpose by extracting and interpreting online pricing information (Auer, 2024). More broadly, web-scraping combined with ML techniques has shown the ability to capture nuanced pricing dynamics often missed by traditional surveys, such as frequent price adjustments and seasonal fluctuations (Cavallo, 2017; Macias et al., 2023; Benchimol and Palumbo, 2024).

Similarly, economic output measurement benefits from alternative data sources; satellite images (e.g., nighttime illumination, retail parking lot photographs) and transactional data contain valuable signals that help estimate real-time economic activity (e.g., Antenucci et al., 2014; Gibson et al., 2020; Katona et al., 2024). In employment tracking, high-frequency payroll data from private providers like ADP with job postings and company announcements, delivers more timely signal (Cajner et al., 2023; Grigsby et al., 2021). More generally, measurement of GDP, designed nearly a century ago, inadequately reflects the value derived from modern digital goods and services, which are frequently free or low-cost yet significantly enhance consumer welfare. Brynjolfsson et al. (2019) propose a framework that address this measurement gap by leveraging online experiments to accurately capture the consumer surplus and welfare impact of digital goods and services.

Generative AI also allows for document processing at scale. While a single analyst cannot read thousands of news articles or hundreds of PDF reports in a way that is reasonably fast or useful, generative AI can. Moreover, many critical indicators are latent rather than explicitly stated, and generative AI models are good at interpreting these implicit signals within their broader context (Cook et al., 2025).[10]

Modern generative AI methods can also improve forecast accuracy across economic indicators. For inflation forecasting, large language models (LLMs) have sometimes been found to outperform traditional forecasts. For instance, the Czech National Bank (CNB) has explored LLM use in inflation forecasting following Faria-e Castro et al. (2023), with some success. CNB found that LLMs AI-based nowcasting substantially improved both data categorization and forecast accuracy (Michl et al., 2025). However, confirming these findings with out-of-sample tests is needed (Lopez-Lira et al., 2025) and will take some time, given the existing LLM knowledge cut-offs.[11]

More broadly, generative AI enables synthetic data generation, allowing researchers and policymakers to create synthetic microdata that can be used for simulations while preserving individual privacy and allow for scenario analysis and counterfactuals. Hansen et al. (2024) simulate the entire U.S. Survey of Professional Forecasters (SPF), finding that AI-generated forecasts not only mirror human expectations but often outperform them, especially at medium- and long-term horizons. This framework integrates real-time macroeconomic data, forecaster personas, and past human predictions to produce high-frequency, individualized inflation forecasts that reduce bias and improve accuracy. Similarly, Zarifhonarvar (2024) use LLMs to emulate Survey of Consumer Expectations respondents, generating realistic upward-biased inflation forecasts, and exhibiting heterogeneity based on partisan and demographic differences.

[10] For example, the concept of "inflation expectations" implied through price discussions in earnings calls, or "supply chain stress" subtly referenced in central bank surveys.

[11] He et al. (2025) provide a solution to this by training chronologically consistent large language models which incorporate only the text data that would have been available at each point in time.

Finally, generative AI can address the "data funnel" problem, which involves selecting and interpreting extensive and diverse data in downstream tasks (Hewitt et al., 2024; Manning et al., 2024). Data extraction and filtering can be automated and scaled with the help of generative AI. Realizing these benefits depends on robust validation processes, such as back-testing, output audits and explainability analyses, so that AI complements rather than replaces established methods.

2.1.2 Analysis and Reports

In recent years, substantial progress has been made in incorporating machine learning into economic analysis. As Desai (2023) highlights, methods like topic modeling, convolutional neural networks, and ensemble learning have become key for uncovering hidden relationships, automating feature extraction, and improving predictive accuracy (Cafarella et al., 2023). Generative AI doesn't replace these tools so much as extend them, bringing together text, time-series, and other data types within a single model architecture. Another important advancement is the few-shot learning that allows generative AI models to quickly adapt to new tasks like sentiment analysis in specialized markets with minimal additional training.

Equally transformative is how generative AI is lowering the barrier to entry for non-experts. Traditionally, economic data analysis required specialized software, statistical expertise, and deep domain knowledge, creating barriers for many. Now, generative AI lets users type plain-language questions to run complex analyses without specialized knowledge. For example, the ECB uses AI to translate natural language queries into code, thus simplifying data retrieval for non-technical staff (Vagen, 2024). Bias and explainability remain concerns, but this approach makes data access and analysis more accessible.

Beyond analysis, generative AI could help with uniform data presentation across sources to improve consistency and reduce bias in data inputs.[12] By aligning formats and harmonizing terminology, it removes a common source of bias that arises when models over- or under-weight inputs simply because they arrive in different forms. For example, Gupta et al. (2025) use LLMs to automate government report generation by having the models read graphs, extract data, search for updates, and create revised visualizations. Testing on UN-GDP reports showed lower error rates and 80-90% time savings compared to manual methods (though the system struggles with infographics and charts lacking clear labels).

2.1.3 Decision Making

While modern policy analysis relies heavily on data, effective policymaking ultimately depends on human judgment. As Dreyfus and Dreyfus (2005) highlight, technical models reach their fullest potential only when complemented by human expertise, especially in complex or uncertain environments.

Policy challenges often have no single "right" answer, and blind reliance on algorithmic prescriptions can lock in existing biases or rule out unconventional but viable options (Loaiza and Rigobon, 2024). Take the Taylor rule: in March 2020 the Federal Reserve sharply cut rates to near zero, a move well beyond what the rule would have prescribed, as policymakers drew on expert judgment to navigate the COVID-19 crisis. This episode showcases the role of human discretion in complementing and, when needed, overriding model outputs.

Similarly, central banks do not rely solely on quantitative models when forecasting key macroeconomic variables. A substantial body of research shows that combining model projections, timely data, and expert judgment produces more reliable forecasts than any single method alone (see, e.g., Faust and Wright, 2009; Croushore, 2010). In-

[12] Or, alternatively, make data presentation hyper-customized.

tegrating expert judgment with AI tools also helps to address ethical and accountability concerns associated with automated decision-making, given the opacity and potential biases of machine models (Loaiza and Rigobon, 2024; Cook and Kazinnik, 2024). As Governor Cook observes, "*like the Mechanical Turk, ultimately the human inside the machine is still in charge*" (Cook, 2024).

2.1.4 Communication

Even the best analysis falls flat without clear communication, and AI is transforming how communication is crafted and delivered (Bricongne et al., 2024). Recent research shows that LLMs can capture subtle nuances in central bank communications. Woodhouse and Charlesworth (2023) shows GPT-3.5's ability to accurately classify Bank of England speeches, closely matching market sentiment and forecasting policy moves. Beyond sentiment analysis, advanced models (e.g., GPT-4) outperform former SOTA technologies like BERT in interpreting complex narratives within FOMC announcements (Hansen and Kazinnik, 2023). Specialized models like `CentralBankRoBERTa` are trained on the unique terminology of economic policy statement (Pfeifer and Marohl, 2023; Gambacorta et al., 2024).

Not surprisingly, some central banks have started using these tools in production. For example, the Bundesbank's MILA (Monetary-Intelligent Language Agent) analyzes central bank statements by classifying their tone and context, aiding human experts with clearer messaging (Deutsche Bundesbank, 2025). Similarly, AI-driven translation helps the ECB effectively communicate across the EU's 24 languages (Cipollone, 2024).

Nonverbal cues play an equally important role in central bank communication. Studies by Curti and Kazinnik (2023), and Gorodnichenko et al. (2023) show that facial expressions and vocal tone during central bank press conferences can significantly impact market reactions. Combining analysis of both verbal content and nonverbal signals promises a more complete understanding of how different audiences interpret policy decisions. Today's AI makes this type of analysis significantly more accessible.

Finally, generative AI makes it possible to craft messages that adapt tone and detail to specific audiences. AI-powered digital avatars and interactive formats can translate abstract policy concepts into engaging, accessible presentations. For example, Zoom's personalized AI avatars illustrate how such tools can render complex economic ideas more tangible (Parks et al., 2014).

2.2 Financial Stability

Text analysis and machine learning have long helped track economic conditions, predict crises, and map links between financial actors (Manela and Moreira, 2017; Bybee et al., 2024; Chen et al., 2023; Bartel et al., 2024; Kazinnik et al., 2022). Generative AI promises to enhance the accuracy and depth of these methods. Some central banks already use AI to watch for risks in the financial system. For example, the Hong Kong Monetary Authority (HKMA) built an in-house generative AI tool to sift through thousands of bank earnings call transcripts, extracting signals of emerging risks (Wong et al., 2025). The tool identifies key risk factors (e.g. credit quality concerns, geopolitical tensions) that align with actual stress events like the Russian war on Ukraine and the 2023 US bank turmoil.

At the same time, AI itself may introduce systemic risks if not properly managed. The Financial Stability Oversight Council (FSOC) identified AI as a financial stability risk, noting potential amplification of market correlation, liquidity stress, and increased volatility from widespread AI adoption (Financial Stability Oversight Council, 2023). The IMF's 2024 Global Financial Stability Report similarly cautions on AI-related volatility (International Monetary Fund, 2024). Because AI models tend to be trained on "ordinary" historical patterns, they might miss or misjudge the buildup of systemic risk. Additionally, as AI tools proliferate, many financial institutions will rely on just a handful of large cloud and model providers. As noted by Aldasoro et al. (2024c), market-wide adoption of AI may concentrate data and cyber risk in a few large providers. This concentration creates single points of failure that can amplify systemic shocks when those vendors experience outages or breaches.

McLemore and Mihov (2025) find that banks spending more on AI often face higher operational losses. Although this new technology promises greater efficiency in banking operations, it also introduces new vulnerabilities. While AI can streamline processes, its complexity can lead to unexpected failures, fraud, and compliance issues, especially during rollout. Strong governance practices could help mitigate these risks.

2.3 Supervision and Regulation

Another key function of the Fed is the supervision and regulation of financial institutions to ensure their safety and soundness, an area particularly ripe for transformation by AI. Primarily, AI could enhance decision-making by efficiently processing the large amounts of structured and unstructured data that bank examiners rely on during assessments.[13]

Supervisors begin each examination by defining its scope, timing, and required resources, tailored to a bank's business model, complexity, and risk profile (Hirtle and Kovner, 2022). AI can enhance this planning stage by ingesting structured inputs (e.g., financial ratios, transaction histories) together with unstructured materials (examiner notes, past reports) to compute vulnerability scores and recommend optimal exam schedules and staffing. In the data-collection phase, examiners review internal reports, interview auditors and senior managers, and conduct independent compliance checks. AI-based parsing tools can automatically flag governance or control gaps by

[13] For instance, Figure 1 illustrates an approach broadly relevant to the Federal Reserve's supervisory processes.

scanning bank documentation, meeting minutes, and examiner notes to uncover emerging risks more efficiently. In the final reporting phase, examiners produce large volumes of text as outputs of their supervisory activities, e.g., bank examination letters and Matter Requiring (Immediate) Attention (MR(I)A) letters (Goldsmith-Pinkham et al., 2016). These must adhere to regulatory templates and maintain a consistent tone, and generative language models can help with that.

Some central banks are increasingly turning to AI technologies to strengthen this function. For example, the ECB has introduced a comprehensive AI-driven suite of tools tailored to supervisory activities.[14] The ECB also employs a dedicated "data lake" for banking supervision, where generative AI converts natural-language queries into executable scripts, making data retrieval more efficient (Prenio, 2024).

A second reason for adding AI technologies to the supervision and regulation toolbox is that familiarity with AI helps regulators and supervisors evaluate its risks. Mullin (2023) explores the increasing role of AI in the U.S. banking sector, and finds that adoption of AI in banking has been gradual but persistent, with uses ranging from fraud detection to customer service chatbots and credit evaluation. Banks have leveraged AI for cost savings, fraud prevention, and regulatory compliance. However, concerns on model transparency, potential biases, and data privacy remain. Bowman (2024) advocates a balanced regulatory approach that capitalizes on AI's benefits, while mitigating operational threats. In a similar vein, the Bank of England AI Consortium provides a platform for the public to give feedback on AI development, use, and safety in the U.K. financial sector.[15]

[14]These include Athena (for textual analysis), GABI (for big data analytics), NAVI (for mapping complex ownership structures), Heimdall (for management due diligence), and Medusa (designed to streamline internal model reporting). The use range from more accessible data analysis to improvements in the consistency of supervisory reports.

[15]More information is available at the BoE's website: https://www.bankofengland.co.uk/research/fintech/artificial-intelligence-consortium.

2.4 Payment Systems

Operating secure and efficient payment systems is a core responsibility of the Federal Reserve. Modern payment networks generate vast amounts of real-time data, creating significant opportunities for AI to enhance speed, security, resilience, and user experience. Major networks like Visa and Mastercard already employ AI to identify fraudulent transactions instantaneously; similarly, the Federal Reserve's FedNow real-time payment service could leverage AI-driven fraud monitoring and anomaly detection. Recognizing this potential, the ECB has initiated an AI action plan emphasizing the development of secure, responsible AI tools, infrastructure, and capabilities (Cipollone, 2024).

Desai et al. (2024) highlight several practical machine learning applications in payments, including real-time fraud detection, dynamic credit scoring, personalized customer recommendations, and predictive analytics for cash-flow forecasting. These ML techniques also drive operational efficiencies by automating payment reconciliation, detecting errors swiftly, and optimizing payment routing processes. Looking ahead, the integration of AI with quantum computing holds promising possibilities for further advancement of payment systems (McMahon et al., 2024).

2.5 Consumer Protection and Community Development

Keeping an eye on fair lending and other consumer finance rules across thousands of banks is a significant challenge. AI can help by monitoring loans in real time and alerting regulators to trouble spots. Predictive models can identify neighborhoods at risk of foreclosure spikes, detect where predatory lending might proliferate due to economic stress, or flag concentrated patterns of illegal practices. Instead of waiting for annual reports or scheduled examinations, this would allow regulators to act immediately by issuing consumer advisories, investigating problematic lending practices, or collaborating with community groups to offer financial counseling as soon as issues emerge.

The spirit of the Community Reinvestment Act (CRA) (i.e., meeting local credit needs) might be better realized if evaluations use all available information, including alternative data and machine learning, to tailor expectations to each community's context (Meursault et al., 2022). These models could also integrate multiple data streams such as property records, demographic metrics, and local market trends to assess banks' CRA compliance, predict economic stress, identify areas of underinvestment, and help regulators allocate resources effectively for community development (Singh and Patel, 2024).

2.6 Research

Research by Federal Reserve economists underpins policy decisions across all core functions. Fed economists study a range of topics, including inflation dynamics, labor market shifts, financial innovation, and systemic risk; their work supports the Fed's mandate of maximum employment and price stability while adapting U.S. economic policy to rapid technological change (see, e.g., Bordo and Prescott, 2019).

AI is changing how this research is done, affecting both the scale and scope of analysis: models can ingest more information, selection of model features can be more data-driven, and empirical validation can be more exhaustive. Traditional economic methods typically require simplifying assumptions and strictly limit the number of variables due to analytical and data constraints. AI methods relax these limitations and allows researchers to evaluate many more potential predictors (Ludwig and Mullainathan, 2024). Traditionally, an economist might focus on only a handful of time series variables for forecasting, but machine learning methods can simultaneously analyze hundreds or even thousands. For instance, the FRED database has over 527,000 series (an amount that is impossible to track manually), yet ML methods let researchers use them all while keeping models both accurate and manageable (Smalter Hall and Cook, 2017). Once a hypothesis is formulated, AI can also assist in testing it: by rapidly coding up simulations or pulling relevant data, an AI can perform quick preliminary tests before the researcher commits full resources (Manning et al., 2024).

AI also enables richer models like agent-based simulations and reinforcement learning (e.g., Axtell and Farmer, 2025). For example, Lopez-Lira (2025) built an open-source market simulator where LLMs act as diverse trading agents, with a realistic order book, market and limit orders, dividends, and detailed microstructure. These types of new modeling frameworks can help regulators spot and prevent risks like herding or flash crashes, for example.

More broadly, Korinek (2023, 2024c,b) illustrates how generative AI can transform economic research in day-to-day workflows. Tailored specifically for economists, this work provides a detailed taxonomy of AI applications, from idea generation and literature reviews to coding and manuscript drafting. Korinek argues that a carefully structured human-AI partnership has the potential to significantly accelerate economic discovery. Novy-Marx and Velikov (2025) showcase this need for clearly structured partnership by implementing a fully automated approach in which economists define overarching research guidelines and domain-specific evaluation protocols, while AI autonomously performs analysis, goes through relevant literature, and then drafts complete manuscripts. As they show, such large-scale automation highlights critical risks in this process. In particular, automated research increases the potential for Hypothesizing After Results are Known (HARKing), given that AI can effortlessly produce artificially coherent narratives and post-hoc theoretical justifications to fit observed data patterns. Novy-Marx and Velikov (2025) argue that industrial-scale hypothesis generation could potentially flood the literature with plausible yet spurious findings.

This finding underscores the importance of economists treating AI as a collaborative assistant rather than fully autonomous actor, and maintaining human oversight.

It is clear that while generative AI promises to accelerate scientific insight, it also challenges foundational norms of science. In response, an interdisciplinary panel convened by the National Academy of Sciences (Blau et al., 2024) urges the community to uphold five principles of human accountability whenever AI is used in research: (1) fully disclose and attribute AI tools and their outputs; (2) rigorously verify AI-generated content and analyses; (3) clearly document the presence of synthetic data; (4) build ethical considerations into every stage of AI deployment; and (5) ensure ongoing oversight.

3 Preparing Data and Infrastructure for AI

The opportunities discussed in Section 2 depend on both the right technological backbone and human skills. Central to that backbone is a unified data architecture: breaking down data silos (i.e., isolated repositories of information within departments, systems, or regions) that, while sometimes necessary in highly regulated contexts, ultimately limit efficiency, undermine collaboration, and constrain advanced analytics.

Over time, the Federal Reserve's federated structure, in which each Reserve Bank operates as its own legal entity, has led to prevalence of such fragmented data architectures. During the COVID-19 pandemic, the challenges became particularly evident through the Main Street Lending Program, when siloed loan performance data across regional banks complicated oversight and delayed reporting to the Government Accountability Office (U.S. Government Accountability Office, 2024).

At the same time, although silos have clear drawbacks, they often remain necessary in large, tightly regulated organizations where privacy and governance often require stringent controls over sensitive data. A data lake provides a strategic approach to addressing these issues by offering a single, unified repository for storing vast amounts of raw data in its original form—whether structured (e.g., SQL databases), semi-structured (e.g., JSON files), or unstructured (e.g., PDF reports).[16] Unlike traditional data warehouses, a data lake relies on a "schema-on-read" philosophy (e.g., a flexible data storage approach that applies structure only when data is retrieved, not upfront). For all the benefits of consolidation, concentrating sensitive information in a single repository does create additional security concerns.[17] Pairing data lake or AI platform with clear ownership, audit trails and privacy compliance helps prevent misuse.

Of course, technology alone is insufficient; the workforce forms the other essential component. We touch on this topic next.

4 Preparing the Workforce

Evidence of the value in pairing human judgment with AI is widespread. Choudhary et al. (2023), for example, demonstrate how merging human and AI assessments significantly boosts accuracy by capitalizing on distinct yet complementary strengths.[18] Likewise, Ide and Talamas (2024) show that AI can reduce both time and knowledge barriers in "knowledge work." Janssen et al. (2022) add that explainable AI (XAI), coupled with experienced decision-makers, can enhance government decisions, emphasizing the importance of using transparent models, offering training, and maintaining a balanced approach between human oversight and algorithmic reliance.

[16] In essence, a data lake acts as a central library for all types of information, as opposed to separate filing cabinets that don't talk to each other.

[17] In the Federal Reserve's case, these risks could be mitigated by either encrypting data both at rest and in transit (e.g., AES-256 for storage and TLS 1.3 for transmissions) or employing attribute-based access control (ABAC), which grants or restricts access based on user roles and data classification levels.

[18] Ensembles are particularly effective when human and AI systems make different kinds of errors, allowing one to offset the other.

In what follows, we explore the composition of the FRS workforce by detailing the range of tasks carried out in each occupation and then map these tasks to their potential for AI augmentation.

4.1 Task Augmentation and Generative AI Exposure

As of 2023, the Federal Reserve System employed approximately 22,000 people, including nearly 1,000 economists (FRED, 2023). To capture a holistic view of the Fed's current workforce, we use data from Revelio Labs, a platform that consolidates and standardizes millions of public employment records. Revelio provides data on workforce composition, hiring and attrition, compensation, job postings, employee sentiment, and layoffs, with most records dating back to 2008. Our analysis covers all 12 Federal Reserve Banks as well as the Board of Governors.

Every occupation can be analyzed as a bundle of tasks. Following the approach of Eloundou et al. (2023) and Eisfeldt et al. (2023), we integrate each job posting with the corresponding skills data from the O*NET database to assess a firm's labor exposure to generative AI. This assessment proceeds in three stages: (1) evaluating exposure at the task level, (2) aggregating those results to the occupation level, and (3) deriving an overall measure of exposure at the firm level. The O*NET database catalogs 19,265 tasks that span 923 U.S. occupations, with each occupation comprising a distinct subset of these tasks.[19]

We use a classification system to determine how susceptible each occupational task is to augmentation. Tasks are categorized according to a four-level exposure scheme, based on whether the use of an LLM-based tool can reduce the time required to complete the task by at least half without compromising quality. We start by assigning each task an exposure score $X_T \in \{0, 0.5, 1, 0.75\}$, corresponding to one of the following categories:

[19] Note that in our merged dataset, one position may be linked to multiple skills.

- **No Exposure (E0)** ($X_T = 0$): An LLM-based tool cannot reduce the time to complete this task by at least half without degrading the output, or its use diminishes output quality.

- **Indirect Exposure (E1)** ($X_T = 0.5$): An LLM-based tool alone does not suffice to halve task completion time, but additional software could achieve this reduction without loss of quality.

- **Exposure with Image Capabilities (E2)** ($X_T = 0.75$): An LLM-based tool equipped with image-processing functions (e.g., viewing, captioning, or creating images) substantially reduces the time needed to complete this task without compromising quality.[20]

- **Direct Exposure (E3)** ($X_T = 1$): Employing an LLM-based tool cuts the time needed for this task by at least half while preserving quality.

We then apply this classification to 7,820 tasks across 360 unique occupation codes within the FRS.[21] We aggregate the task-level scores to compute an average exposure measure at the occupation level. Following Eisfeldt et al. (2023), we also distinguish the share of exposure arising from supplementary (rather than core) tasks for each classified occupation, allowing us to separate the contribution of peripheral activities from that of central ones.

4.2 The Exposure of FRS Workforce to AI

Of the 7,820 unique tasks in the sample, 5,290 were classified as E0, 1,197 classified as E1, 14 classified as E2, and 1,319 classified as E3.[22] We next aggregate tasks' expo-

[20] This setup, however, does not support video processing or extract fine-grained details such as precise measurements from images.

[21] The exact prompt used in our classification procedure is provided in the Appendix.

[22] We provide some classification examples in the Appendix. Tasks classified as E2 end up being mostly related to graphic design tasks for this particular sample.

sures to generative AI at the occupation level. For each 8-digit Standard Occupational Classification (SOC) occupation from the O*NET, we calculate the share of the total number of tasks for each occupation that have either a direct or an indirect exposure to generative AI. Our measure of occupation-level exposure X^O is the sum of task-level exposures X^T for $T \in \{0, \ldots, 19{,}265\}$ within each occupation $O \in [1, \ldots, 923]$ divided by the total number of tasks in occupation O:

$$X^O = \frac{\sum_{T=0}^{19{,}265} X^T}{S^O}.$$

Overall, when aggregated to the occupation level, we see that roughly half of the occupations studied (179 out of 360, or approximately 49.7%) exhibit relatively low generative AI exposure. These roles require physical presence, interpersonal interaction or specialized manual skills and are therefore currently resistant to automation. Typical occupations in this category include duties that emphasize direct human interaction and hands-on activities.

The next segment, comprising 26.1% or 94 occupations, falls into the moderate exposure range. These combine administrative or managerial duties with digital tasks, using technology and data while still relying on human oversight. Examples include chief executives, administrative services managers and financial managers, as well as project managers and compliance managers. These occupations benefit from automation in tasks such as document processing but continue to depend on human judgment and contextual knowledge.

Approximately 18.3% (66 occupations) are in the high exposure bracket, characterized by more intensive integration of generative AI tools within their workflows. These roles often require advanced data analysis, familiarity with algorithms, and substantial digital interactions, along with continued human involvement. Examples include technical and analytical specialties, including computer systems analysts, software quality testers, information security analysts, accountants, and auditors. Creative and communication-oriented roles, such as editors, news analysts, and technical writers, similarly benefit from AI-supported editing and research capabilities while still necessitating editorial judgment and nuanced communication.

Finally, a small subset of occupations (5.8%, representing 21 roles) exhibits very high exposure levels (75–100), indicating they have the most to benefit from advanced computational tools and generative AI. These occupations involve extensive use of predictive analytics, algorithms, and digital solutions, and are rapidly transforming due to AI innovations. Examples include data and AI specialists like data scientists, financial quantitative analysts, and business intelligence analysts, whose work heavily depends on sophisticated analytics and machine learning models. Software and system development roles, such as blockchain engineers, database architects, web developers, and economists also fall within this category. We list the top 20 occupations with the highest levels of exposure in Table 1 as a reference.

Finally, to measure GenAI exposure at the Reserve Bank level, we take all of the occupations in that bank's district and average their AI exposure scores. Across all twelve Reserve Bank districts, mean GenAI exposure scores fall in a narrow band (0.52–0.55), and the confidence intervals overlap substantially, indicating no significant cross-district differences. Boston and St. Louis exhibit the highest mean exposure (0.55), whereas Atlanta has the lowest (0.52).

4.3 Potential Productivity Gains

To estimate potential productivity gains from integration of AI across the FRS, we begin by quantifying the total labor effort at each Reserve Bank. Using the 2024 Reserve Bank budgets, we extract the full-time equivalent (FTE) staffing levels for each

Table 1: Top 20 Occupations with Highest AI Exposure within the FRS

Occupation (ONET Title)	**Mean**
Data Warehousing Specialist	1.000
Business Intelligence Analyst	0.926
Financial Quantitative Analyst	0.905
Information Security Engineer	0.899
Database Architect	0.883
Database Administrator	0.877
Blockchain Engineer	0.876
Data Scientist	0.875
Search Marketing Strategist	0.870
Penetration Tester	0.864
Financial Risk Specialist	0.843
Management Analyst	0.818
Proofreader and Copy Marker	0.818
Web Developer	0.817
Business Continuity Planner	0.810
Clinical Data Manager	0.810
Economist	0.807
Chief Sustainability Officer	0.792
Regulatory Affairs Specialist	0.767
Logistics Engineer	0.767

Note: Mean denotes the average generative AI exposure score for each occupation, scaled between zero and one where one corresponds to the highest observed exposure across all 360 occupations in the FRS. Occupations are defined by their ONET Titles. Exposure scores are calculated based on the intensity of AI-applicable tasks.

Bank (Division of Reserve Bank Operations and Payment Systems, 2023). We assume each FTE represents 2,080 working hours per year, a standard measure of annual labor time.[23]

To estimate the share of labor that could be augmented by generative AI tools, we combine these labor totals with Reserve Bank–level exposure scores. These scores represent the average fraction of task time within each Bank that could plausibly be reduced or assisted by large language models. We multiply each Bank's total staff-hours by its exposure score to compute estimated "AI-augmentable hours", or the share of labor time most likely to benefit from AI integration, summarized in Table 2.

Table 2: Estimated Labor Inputs and AI-Augmentable Hours by Reserve Bank

Bank	**FTEs (2024)**	**ELH**	**AI Exposure**	**AI-Augmentable Hours**
Boston	1,296	2,695,680	0.55	1,482,624
St. Louis	1,508	3,136,640	0.55	1,725,152
Cleveland	1,114	2,317,120	0.54	1,251,245
Kansas City	2,072	4,307,360	0.54	2,325,974
Minneapolis	1,147	2,389,760	0.54	1,290,470
Philadelphia	884	1,838,720	0.54	992,909
Richmond	1,617	3,363,360	0.54	1,816,214
San Francisco	1,910	3,976,800	0.54	2,147,472
Chicago	1,726	3,588,080	0.53	1,901,683
Dallas	1,343	2,794,240	0.53	1,481,947
New York	3,073	6,391,840	0.53	3,387,675
Atlanta	1,780	3,702,400	0.52	1,925,248

Note: **FTEs (2024)** denote full-time-equivalent staffing levels reported in the 2024 Reserve Bank budgets. **(ELH)**, or Estimated Labor Hours equal FTEs × 2,080 hours per year. **AI Exposure** is the average share of tasks potentially augmentable by generative-AI tools (see Section 4.2). **AI-Augmentable Hours** are ELH × AI Exposure. Board of Governors FTE data not available.

Next, to estimate the total labor capacity devoted to each of the Federal Reserve's *core functions*, we begin with the official 2024 Reserve Bank budget Division of Reserve Bank Operations and Payment Systems (2023). The budget allocates $6.05 billion in

[23]This number (2,080 = 40 hours/week x 52 weeks/year) should be treated as an upper bound, given that it doesn't take into account holidays, paid time off, or sick time.

operating expenses across major operational areas, including monetary policy, supervision, cash operations, and Treasury services. Again, assuming that labor costs account for a proportional share of each functional budget, we assign full-time equivalents and corresponding staff hours to each function in proportion to its share of total expenses. Table 3 presents the resulting estimates, and provides a functional decomposition of total Reserve Bank labor inputs. However, these figures alone do not tell us where (geographically or institutionally) those hours are concentrated.

Table 3: 2024 Estimated Labor Inputs by Federal Reserve Function

Function	**2024 Budget ($M)**	**% of Expense**	**FTEs**	**Labor Hours**
Supervision	1,784.5	29.5%	6,263	13,025,040
Treasury Services	819.8	13.5%	2,867	5,960,160
Cash Operations	897.9	14.8%	3,142	6,537,360
Monetary Policy	629.1	10.4%	2,209	4,595,120
Fee-Based Services	775.2	12.8%	2,719	5,655,520
Open Market Ops	299.8	5.0%	1,062	2,209,040
All Other Services	566.4	9.4%	1,979	4,120,320
Total	**6,053.2**	**100%**	**21,238**	**44,102,560**

Notes: Budget figures are from the 2024 Federal Reserve Bank operating expense plan. Estimated FTEs are computed assuming labor shares proportional to budget shares. Estimated labor hours are based on 2,080 hours per FTE per year.

With this exercise, we are now able to connect the top-down and bottom-up perspectives developed in this paper. By estimating the labor devoted to each function and using what is known about which Reserve Banks lead which activities, we begin to see where AI-driven productivity gains are most likely to happen. While precise allocations are not available, this mapping enables a back-of-the-envelope exercise: we can distribute the total labor hours associated with each function (as derived above) across the Reserve Banks based on their known specializations.

We turn to the known functional specializations of individual Reserve Banks. While all Reserve Banks contribute to multiple areas of the Fed's mandate, certain Banks have historically assumed lead roles in specific functions. These institutional responsibilities are evident in program-level assignments, published budget narratives, and long-standing operational structures. While each Reserve Bank contributes broadly to the Fed's mandate, several have formal System-wide specialties. New York runs all open-market operations, monitors market stability, and leads LISCC supervision of the largest banks. Atlanta houses the Retail Payments Office and processes certain savings-bond deposits. Boston and Kansas City co-lead the development and enterprise services for the FedNow Service. Richmond operates National IT, providing cloud, cybersecurity, and application hosting across the System. The Board of Governors, whose staff costs are outside Reserve Bank budgets, is responsible for monetary policy design, macroeconomic research, supervision policy, and rule-making.[24]

For instance, most of the 2.2 million hours assigned to Open Market Operations (OMO) can reasonably be attributed to the New York Fed ($\alpha_{\text{OMO,NY}} \simeq 1$), $H_{\text{OMO}} = 2.21$ million hours (Table 3) combined with its AI exposure score, $X_{\text{NY}} = 0.53$ to give

$$\widehat{H}^{\text{AI}}_{\text{OMO}} \;=\; 0.53 \times 2.21\,\text{million} \;\approx\; 1.17\,\text{million},$$

so that **about 53 % of all OMO staff hours are plausibly AI-augmentable**.

With more data, one could extend this to other FRS core functions, e.g.:

Table 4: AI-augmentable Staff Hours for Key Function by Bank, 2024

Function	**Lead Bank(s) & X_b**	**Labor hours H_f (m)**	$\widehat{H}^{\mathbf{AI}}_f$ **(m)** [%]
Open-Market Operations	NY (0.53)	2.21	1.17 (53)
Treasury Services[a]	Atlanta (0.52), St. Louis (0.55)	5.96	3.18 (53)
Cash Operations[b]	SF (0.54), KC (0.54), Dallas (0.53)	6.54	3.51 (54)

Notes: [a] Hours split $\alpha_{\text{ATL}} = 0.54$, $\alpha_{\text{STL}} = 0.46$ by 2024 FTEs. [b] Hours split $\alpha_{\text{SF}} = 0.36$, $\alpha_{\text{KC}} = 0.39$, $\alpha_{\text{DAL}} = 0.25$. Percentages in brackets are $\widehat{H}^{\text{AI}}_f / H_f$.

[24]We are also unable to include National IT in this analysis. National IT, although not a Bank in the traditional sense, supports the entire System with technical infrastructure, cybersecurity, and data modernization efforts that cut across all functional domains.

4.4 Literacy, Up-skilling, and Re-skilling

To address the varying degrees of potential LLM-based task augmentation, we propose the following, in alignment with our exposure classification. Professionals whose roles fall under *Direct Exposure*, such as examiners, financial analysts, research assistants, economists, and other knowledge-based, data intensive occupations, might benefit substantially by integrating AI into their day-to-day responsibilities. This could involve using AI-driven automation tools, developing prompt engineering skills, and interpreting results from software "co-pilot" solutions, especially in coding and complex data tasks.

Individuals seeking a more substantial shift into AI-centric careers, especially those in roles with *No Exposure*, as well as those with partial or indirect exposure aiming to become data scientists, ML engineers, or automation specialists, might benefit from comprehensive re-skilling. This path involves learning coding and model development tools, along with principles of AI explainability, bias mitigation, and model deployment.

In sum, depending on the level of augmentation made possible by LLM-based technologies, job roles could be mapped to one of three pathways. *Literacy* is ideal for positions with *Indirect* or *No Exposure* that nonetheless require a foundational awareness to interact effectively with AI systems. *Upskilling* is tailored to roles already benefiting from low *Direct Exposure* or partial reliance on LLM tools, where deepening AI-related competencies significantly enhances productivity. *Re-skilling* suits those whose current roles face obsolescence (*No Exposure* or declining relevance) or very high *Direct Exposure*, and who seek new career paths with substantial AI components.

This pathway closely aligns with the BIS's recent strategic guidance on workforce adaptation. BIS (2024) lay out a dual-scenario framework: "AI copilots" that enhance human productivity and "AI agents" that perform narrowly defined tasks autonomously. The BIS report stresses that even under the less disruptive copilot scenario, central banks must overcome a few core challenges, retraining among them. Continuous, task-specific learning programs can help existing staff work effectively with new tools.

In addition, rather than depending primarily on external hires, the BIS argues that central banks must cultivate an organizational culture of continuous learning and innovation. This includes establishing dedicated AI training tracks, appointing AI "champions", and creating cross-functional AI governance bodies that oversee model risk frameworks and audit. Most importantly, the Fed already employs individuals with significant AI research expertise or practical AI-related skills, giving it a strong base on which it could build these internal initiatives.

5 Conclusion

Central banking is approaching a major transformation. Imagine a system that rapidly and continuously collects diverse data, from traditional economic metrics and real-time spending patterns to environmental indicators, delivering immediate insights on growth, inflation, and employment. It instantly evaluates various scenarios to predict policy impacts and conducts ongoing stress tests to identify threats such as cyberattacks, natural disasters, and geopolitical crises, proactively addressing manageable risks before they escalate.

Translating this potential into reality hinges on overcoming significant misunderstandings about AI. Nontechnical observers frequently underestimate the rapid pace of AI development, focusing solely on its present capabilities. Technical experts, on the other hand, while deeply familiar with AI's potential, often overlook organizational and societal barriers to widespread adoption. As Dario Amodei, CEO of Anthropic, highlights, "*most people are underestimating just how radical the upside of AI could be,* *just as* [...] *most people are underestimating how bad the risks could be*" Amodei (2024). The realistic and effective adoption of AI requires recognizing that human systems—our rules, behaviors, norms—fundamentally shape technological systems, determining their ultimate form and function (e.g., Orlikowski and Gash, 1994; Lin and Silva, 2005).

What does transformative AI mean for monetary policy and central banking? More than a century after Keynes imagined a future of abundance and shorter workweeks (Keynes, 1930), researchers now unpack potential impact of this new technology on growth trajectories (Trammell and Korinek, 2023), labor scenarios (Korinek and Suh, 2024), market stability (Chow et al., 2024), existential risk (Jones, 2024), and policy design (Korinek, 2024a). Policymakers will play a pivotal role in managing these transitions.

In the short term, disruptions such as rapid labor-market shifts, productivity surges, and widening inequality are likely to amplify central banks' (stabilizing) role. In the longer term, transformative AI might fundamentally alter traditional economic frameworks. In such a scenario, the conventional dynamics of supply, demand, and price mechanisms may diminish in importance, challenging the very essence of monetary policy and central banking. Addressing this transition calls for realistic assessments of AI's impacts and adjustments to our rules, institutions, and norms to preserve economic stability.

References

Agarwal, N., Moehring, A., Rajpurkar, P., and Salz, T. (2023). Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology. Technical report, National Bureau of Economic Research.

Aghion, P., Jones, B. F., and Jones, C. I. (2017). *Artificial Intelligence and Economic Growth*, volume 23928. National Bureau of Economic Research Cambridge, MA.

Agrawal, A. K., Gans, J. S., and Goldfarb, A. (2023). Similarities and Differences in the Adoption of General Purpose Technologies. *National Bureau of Economic Research*.

Aldasoro, I., Doerr, S., Gambacorta, L., Notra, S., Oliviero, T., and Whyte, D. (2024a). Generative Artificial Intelligence and Cyber Security in Central Banking. *Journal of Financial Regulation*.

Aldasoro, I., Doerr, S., Gambacorta, L., and Rees, D. (2024b). The Impact of Artificial Intelligence on Output and Inflation. Technical report, Bank for International Settlements.

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., and Stein, M. (2024c). *Intelligent Financial System: How AI is Transforming Finance*. Bank for International Settlements, Monetary and Economic Department.

Amodei, D. (2024). Machines of Loving Grace. Available from Anthropic.

Angelico, C., Marcucci, J., Miccoli, M., and Quarta, F. (2022). Can We Measure Inflation Expectations Using Twitter? *Journal of Econometrics*, 228(2):259–277.

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., and Shapiro, M. D. (2014). Using Social Media to Measure Labor Market Flows. Technical report, National Bureau of Economic Research.

Association for the Advancement of Artificial Intelligence (2025). AAAI 2025 Presidential Panel on the Future of AI Research. Technical report, Association for the Advancement of Artificial Intelligence.

Auer, R. (2024). Project Spectrum: Using Generative AI to Enhance Inflation Nowcasting. In *Proceedings of the Bank for International Settlements*. Bank for International Settlements. Accessed: 2024-12-29.

Auer, R., Köpfer, D., and Švéda, J. (2024). The Rise of Generative AI: Modelling Exposure, Substitution, and Inequality Effects on the US Labour Market. Technical report, CESifo Working Paper.

Axtell, R. L. and Farmer, J. D. (2025). Agent-based Modeling in Economics and Finance: Past, Present, and Future. *Journal of Economic Literature*, 63(1):197–287.

Bank for International Settlements (2025). Governance of AI Adoption in Central Banks. Report.

Barr, M. S. (2025). Artificial Intelligence: Hypothetical Scenarios for the Future. Accessed: 2025-02-18.

Bartel, M., Hanke, M., and Petric, S. (2024). Crisis Identification and Prediction using Machine Learning: The Case of U.S. Regional Banks. *SSRN Working Paper*.

Benchimol, J. and Palumbo, L. (2024). Sanctions and Russian Online Prices. *Journal of Economic Behavior Organization*, 225:483–521.

Berger, A. N., Curti, F., Mihov, A., and Sedunov, J. (2022). Operational Risk Is More Systemic Than You Think: Evidence from US Bank Holding Companies. *Journal of Banking & Finance*, 143:106619.

BIS (2024). Artificial Intelligence and Human Capital: Challenges for Central Banks. BIS Bulletin No. 100.

Blau, W., Cerf, V. G., Enriquez, J., Francisco, J. S., Gasser, U., Gray, M. L., Greaves, M., Grosz, B. J., Jamieson, K. H., Haug, G. H., et al. (2024). Protecting scientific integrity in an age of generative ai.

Bordo, M. D. and Prescott, E. S. (2019). Federal Reserve Structure, Economic Ideas, and Monetary and Financial Policy. NBER Working Paper 26098, National Bureau of Economic Research.

Bowman, M. W. (2024). Artificial Intelligence in the Financial System. Speech at the 27th Annual Symposium on Building the Financial System of the 21st Century: An Agenda for Japan and the United States, Washington, D.C. Available at: https://www.federalreserve.gov/newsevents/speech/bowman20241122a.htm.

Bresnahan, T. F. and Trajtenberg, M. (1995). General Purpose Technologies 'Engines of Growth'? *Journal of Econometrics*, 65(1):83–108.

Bricongne, J.-C., Caldeira, R., and Meunier, B. (2024). Should Central Banks Care About Text Mining? A Literature Review. Working Paper 950, Banque de France.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., and Fox, K. J. (2019). GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy. *National Bureau of Economic Research*.

Brynjolfsson, E. and Mitchell, T. (2017). What Can Machine Learning Do? Workforce Implications. *Science*, 358(6370):1530–1534.

Brynjolfsson, E., Rock, D., and Syverson, C. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372.

Bybee, L., Kelly, B., Manela, A., and Xiu, D. (2024). Business News and Business Cycles. *Journal of Finance*, 79(5):3105–3147.

Cafarella, M., Ehrlich, G., Gao, T., Haltiwanger, J. C., Shapiro, M. D., and Zhao, L. (2023). Using Machine Learning to Construct Hedonic Price Indices. Technical report, National Bureau of Economic Research.

Cajner, T., Crane, L. D., Decker, R. A., Hamins-Puertolas, A., and Kurz, C. (2023). Payroll Employment at the Weekly Frequency. In *AEA Papers and Proceedings*, volume 113, pages 145–150. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Cajner, T., Crane, L. D., Kurz, C. J., Morin, N. J., Soto, P. E., and Vrankovich, B. (2024). Manufacturing Sentiment: Forecasting Industrial Production with Text Analysis. *FEDS Working Paper*.

Cavallo, A. (2017). Are Online and Offline Prices Similar? Evidence From Large Multi-Channel Retailers. *American Economic Review*, 107(1):283–303.

Chen, M., DeHaven, M., Kitschelt, I., Lee, S. J., and Sicilian, M. J. (2023). Identifying Financial Crises Using Machine Learning on Textual Data. *Journal of Risk and Financial Management*, 16(3).

Choudhary, V., Marchetti, A., Shrestha, Y. R., and Puranam, P. (2023). Human-AI Ensembles: When Can They Work? *Journal of Management*, page 01492063231194968.

Chow, T., Halperin, B., and Mazlish, J. Z. (2024). Transformative AI, Existential Risk, and Real Interest Rates. *Working Paper*.

Cipollone, P. (2024). Artificial Intelligence: A Central Bank's View. Keynote speech at the National Conference of Statistics on Official Statistics at the Time of Artificial Intelligence, Rome. Piero Cipollone, Member of the Executive Board of the European Central Bank.

Cook, L. D. (2024). Artificial Intelligence, Big Data, and the Path Ahead for Productivity. Speech at "Technology-Enabled Disruption: Implications of AI, Big Data, and Remote Work," a conference organized by the Federal Reserve Banks of Atlanta, Boston, and Richmond, Atlanta, Georgia. Available at: https://www.federalreserve.gov/newsevents/speech/cook20241001a.htm.

Cook, T. R., Hansen, A. L., Kazinnik, S., and McAdam, P. (2025). Under Pressure: Strategic Signaling in Bank Earnings Calls. Working Paper.

Cook, T. R. and Kazinnik, S. (2024). Social Bias in Financial Applications of Large Language Models. Working paper.

Croushore, D. (2010). An Evaluation of Inflation Forecasts From Surveys Using Real-Time Data. *The BE Journal of Macroeconomics*, 10(1).

Curti, F. and Kazinnik, S. (2023). Let's Face It: Quantifying the Impact of Nonverbal Communication in FOMC Press Conferences. *Journal of Monetary Economics*, 139:110–126.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology*

& Operations Mgt. Unit Working Paper, (24-013).

Desai, A. (2023). Machine Learning for Economics Research: When, What and How. Staff Analytical Note 2023-16, Bank of Canada. Bank of Canada Staff Analytical Note 2023-16.

Desai, A., Kosse, A., and Sharples, J. (2024). Finding a Needle in a Haystack: A Machine Learning Framework for Anomaly Detection in Payment Systems. In *Bank of Canada*.

Deutsche Bundesbank (2025). Using Artificial Intelligence to Decode the Language of Central Banks.

Division of Reserve Bank Operations and Payment Systems (2023). 2024 Federal Reserve Bank Budgets Addendum. Technical report, Board of Governors of the Federal Reserve System, Washington, DC. Approved December 15, 2023.

Doerr, J. F., Kosse, A., Khan, A., Lewrick, U., Mojon, B., Nolens, B., and Rice, T. (2021). DeFi Risks and the Decentralisation Illusion. *BIS Quarterly Review*, 21.

Doerr, S., Gambacorta, L., Leach, T., Legros, B., and Whyte, D. (2022). *Cyber Risk in Central Banking*. Bank for International Settlements, Monetary and Economic Department.

Dreyfus, H. L. and Dreyfus, S. E. (2005). Peripheral Vision: Expertise in Real World Contexts. *Organization studies*, 26(5):779–792.

Eisfeldt, A. L., Schubert, G., and Zhang, M. B. (2023). Generative AI and Firm Values. Technical report, National Bureau of Economic Research.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). GPTs Are GPTs: An Early Look At the Labor Market Impact Potential of Large Language Models. *arXiv preprint arXiv:2303.10130*.

Faria-e Castro, M., Louis, F. S., and Leibovici, F. (2023). Artificial Intelligence and Inflation Forecasts. Technical report.

Faust, J. and Wright, J. H. (2009). Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset. *Journal of Business & Economic Statistics*, 27(4):468–479.

Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., and You, J. (2024). How Far Are We From AGI: Are LLMs All We Need? *Transactions on Machine Learning Research*.

Financial Stability Oversight Council (2023). Annual Report.

FRED (2023). Federal Reserve System Employment Since 1915: New Insights from the Research Division of the St. Louis Fed. https://fredblog.stlouisfed.org/2023/10/federal-reserve-system-employment-since-1915/. Posted on October 12, 2023.

Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024). CB-LMs: Language Models for Central Banking. Technical report, Bank for International Settlements.

Gibson, J., Olivia, S., and Boe-Gibson, G. (2020). Night Lights in Economics: Sources and Uses. *Journal of Economic Surveys*, 34(5):955–980.

Goldsmith-Pinkham, P. S., Hirtle, B., and Lucca, D. O. (2016). Parsing the Content of Bank Supervision. *FRB of NY Staff Report*.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2023). The Voice of Monetary Policy. *American Economic Review*, 113(2):548–584.

Gosselin, M.-A. and Taskin, T. (2023). What Can Earnings Calls Tell Us About the Output Gap and Inflation in Canada? Technical report, Bank of Canada Staff Discussion Paper.

Grigsby, J., Hurst, E., and Yildirmaz, A. (2021). Aggregate Nominal Wage Adjustments: New Evidence From Administrative Payroll Data. *American Economic Review*, 111(2):428–471.

Gupta, R., Pandey, G., and Pal, S. K. (2025). Automating Government Report Generation: A Generative AI Approach for Efficient Data Extraction, Analysis, and Visualization. *Digital Government: Research and Practice*, 6(1):1–10.

Hansen, A. L., Horton, J. J., Kazinnik, S., Puzzello, D., and Zarifhonarvar, A. (2024). Simulating the Survey of Professional Forecasters. *Available at SSRN*.

Hansen, A. L. and Kazinnik, S. (2023). Can ChatGPT Decipher Fedspeak. *Available at SSRN*.

Harker, P. T. (2024). Fintech, AI, and the Changing Financial Landscape. Speech delivered at the Carnegie Mellon University Lecture Series, Pittsburgh, PA.

He, S., Lv, L., Manela, A., and Wu, J. (2025). Chronologically Consistent Large Language Models. Working paper.

Hewitt, L., Ashokkumar, A., Ghezae, I., and Willer, R. (2024). Predicting Results of Social Science Experiments Using Large Language Models. *Preprint*.

Hirtle, B. and Kovner, A. (2022). Bank Supervision. *Annual Review of Financial Economics*, 14(1):39–56.

Hornstein, A. (2024). Aggregate Effects of the Adoption of AI. *Federal Reserve Bank of Richmond Economic Brief*, (24-19). This article discusses the potential productivity impacts of AI adoption on the U.S. economy, including implications for monetary policy and labor markets.

Ide, E. and Talamas, E. (2024). Artificial Intelligence in the Knowledge Economy. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 834–836.

International Monetary Fund (2024). Global Financial Stability Report - Steadying the

Course: Uncertainty, Artificial Intelligence, and Financial Stability.

Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., and Kuk, G. (2022). Will Algorithms Blind People? The Effect of Explainable AI and Decision-makers' Experience on AI-supported Decision-making in Government. *Social Science Computer Review*, 40(2):478–493.

Jones, C. I. (2024). The AI Dilemma: Growth Versus Existential Risk. *American Economic Review: Insights*, 6(4):575–590.

Katona, Z., Painter, M. O., Patatoukas, P. N., and Zeng, J. (2024). On the Capital Market Consequences of Big Data: Evidence From Outer Space. *Journal of Financial and Quantitative Analysis*, pages 1–29.

Kazinnik, S., Scida, D., Killen, C., and Wu, J. (2022). News and Networks: Using Financial News Coverage to Measure Bank Interconnectedness. *SSRN Working Paper*.

Keynes, J. M. (1930). Economic Possibilities for our Grandchildren. Essays in Persuasiaon. In *Essays in Persuasion*, Collected Writings, vol. IX, 321-32. London: Macmillan.

Kinder, M., de Souza Briggs, X., Muro, M., and Liu, S. (2024). Generative AI, the American Worker, and the Future of Work. *Brookings Institution*. Retrieved from https://www.brookings.edu/.

Korinek, A. (2023). Generative AI for Economic Research: Use Cases and Implications for Economists. *Journal of Economic Literature*, 61(4):1281–1317.

Korinek, A. (2024a). Economic Policy Challenges for the Age of AI. Working Paper 32980, National Bureau of Economic Research.

Korinek, A. (2024b). LLMs Learn to Collaborate and Reason: December 2024 Update to 'Generative AI for Economic Research: Use Cases and Implications for Economists' published in the Journal of Economic Literature 61(4). *Journal of Economic Literature*.

Korinek, A. (2024c). LLMs Level Up–Better, Faster, Cheaper: June 2024 Update to Section 3 of 'Generative AI for Economic Research: Use Cases and Implications for Economists' published in the Journal of Economic Literature 61(4). *Journal of Economic Literature*.

Korinek, A. and Suh, D. (2024). Scenarios for the Transition to AGI. Working Paper 32255, National Bureau of Economic Research.

Lin, A. and Silva, L. (2005). The Social and Political Construction of Technological Frames. *European Journal of Information Systems*, 14(1):49–59.

Loaiza, I. and Rigobon, R. (2024). The EPOCH of AI: Human-Machine Complementarities at Work. *Available at SSRN 5028371*.

Lopez-Lira, A. (2025). Can Large Language Models Trade? Testing Financial Theories with LLM Agents in Market Simulations. *arXiv preprint arXiv:2504.10789*.

Lopez-Lira, A., Tang, Y., and Zhu, M. (2025). The Memorization Problem: Can We Trust LLMs' Economic Forecasts? *arXiv preprint arXiv:2504.14765*.

Ludwig, J. and Mullainathan, S. (2024). Machine Learning as a Tool for Hypothesis Generation. *The Quarterly Journal of Economics*, 139(2):751–827.

Macias, P., Stelmasiak, D., and Szafranek, K. (2023). Nowcasting Food Inflation With a Massive Amount of Online Prices. *International Journal of Forecasting*, 39(2):809–826.

Manela, A. and Moreira, A. (2017). News Implied Volatility and Disaster Concerns. *Journal of Financial Economics*, 123(1):137–162.

Manning, B. S., Zhu, K., and Horton, J. J. (2024). Automated Social Science: Language Models as Scientist and Subjects. Technical report, National Bureau of Economic Research.

McLemore, P. and Mihov, A. (2025). AI and Operational Losses: Evidence from U.S. Bank Holding Companies. Working paper.

McMahon, C., McGillivray, D., Desai, A., Rivadeneyra, F., Lam, J.-P., Lo, T., Marsden, D., and Skavysh, V. (2024). Improving the Efficiency of Payments Systems Using Quantum Computing. *Management Science*, 70(10):7325–7341.

Metaculus (2025). When Will the First General AI System Be Devised, Tested, and Publicly Announced? Accessed: 2025-03-22.

Meursault, V., Moulton, D., Santucci, L., and Schor, N. (2022). One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-income Areas. *Journal of Policy Analysis and Management*.

Michl, A., Švéda, T., and Josef, A. (2025). First Use of AI in Inflation Forecasting at the CNB. Blog post, Czech National Bank.

Mullin, J. (2023). Artificial Intelligence and Bank Supervision: Regulators are Gathering Information about How Banks Use AI. *Federal Reserve*, 2(2). Second Quarter of 2023.

Novy-Marx, R. and Velikov, M. Z. (2025). AI-Powered (Finance) Scholarship. Technical report, National Bureau of Economic Research.

Orlikowski, W. J. and Gash, D. C. (1994). Technological Frames: Making Sense of Information Technology in Organizations. *ACM Transactions on Information Systems (TOIS)*, 12(2):174–207.

Parks, P., Cruz, R., and Ahn, S. J. G. (2014). Don't Hurt My Avatar: The Use and Potential of Digital Self-Representation in Risk Communication. *International Journal of Robots, Education and Art*, 4(2):10.

Pfeifer, M. and Marohl, V. P. (2023). CentralBankRoBERTa: A Fine-Tuned Large

Language Model for Central Bank Communications. *The Journal of Finance and Data Science*, 9:100114.

Prenio, J. (2024). *Peering through the hype-assessing SupTech tools' transition from experimentation to supervision.* Bank for International Settlements, Financial Stability Institute.

SAS Institute Inc. (2024). Your Journey to a GenAI Future: A Strategic Path to Success for Government. Technical report, SAS Institute Inc. Global research study reveals obstacles and opportunities when integrating GenAI technology.

Singh, R. and Patel, M. (2024). AI Applications in Consumer Protection and Community Reinvestment: Enhancing Compliance Through Data Analytics. *Journal of Financial Regulation*.

Smalter Hall, A. and Cook, T. R. (2017). Macroeconomic Indicator Forecasting with Deep Neural Networks. *Federal Reserve Bank of Kansas City Working Paper*, (17-11).

Trammell, P. and Korinek, A. (2023). Economic Growth under Transformative AI. Technical report, National Bureau of Economic Research.

U.S. Government Accountability Office (2024). Federal Reserve Lending Programs: Nearly Half of Main Street Program Loans Are Fully Repaid, but Losses Have Increased. Report to congressional committees, United States Government Accountability Office.

Vagen, T. (2024). European Central Bank sees supervisory boost from generative AI. Thomson Reuters Regulatory Intelligence. Accessed: 2025-04-25.

Videgaray, L., Aghion, P., Caputo, B., Forrest, T., Korinek, A., Langenbucher, K., Miyamoto, H., and Wooldridge, M. (2024). Artificial Intelligence and Economic and Financial Policy Making. *A High-Level Panel of Experts' Report to the G7.* Prepared for the G7.

Wang, W., Gao, G., and Agarwal, R. (2024). Friend or Foe? Teaming between Artificial Intelligence and Workers with Variation in Experience. *Management Science*, 70(9):5753–5775.

Wong, A., Ho, K., and Chan, I. (2025). Utilising Generative AI Tools to Monitor Systemic Risks in Global Banking: An Analysis of Earnings Call Transcript Data. Research Memorandum 07/2025, Hong Kong Monetary Authority, Market Research Division.

Woodhouse, D. and Charlesworth, A. (2023). Can ChatGPT Predict Future Interest Rate Decisions? *Available at SSRN 4572831.*

Wu, L. and Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. In *Economic Analysis and the Digital Economy.* NBER.

Zarifhonarvar, A. (2024). Experimental Evidence on Large Language Models. *Available at SSRN 4825076*.

Appendix

Impact of LLMs on Various Economist Tasks

Occupation	Task	Class	Explanation
Economist (7538)	Compile, analyze, and report data to explain economic phenomena and forecast market trends, applying mathematical models and statistical techniques.	E1	While LLMs may not directly reduce the time for complex data analysis, they can support the development of applications that assist in compiling and analyzing data, thus potentially reducing time significantly.
Economist (21106)	Explain economic impact of policies to the public.	E3	Explaining economic impacts can be streamlined using an LLM to generate clear, concise explanations and summaries, thus reducing the time needed to prepare and communicate this information effectively.
Economist (7537)	Provide advice and consultation on economic relationships to businesses, public and private agencies, and other employers.	E0	Providing advice and consultation requires a high degree of human interaction and understanding of specific contexts, which cannot be effectively replaced or significantly aided by an LLM.
Economist (7543)	Forecast production and consumption of renewable resources and supply, consumption, and depletion of non-renewable resources.	E1	While LLMs may not directly reduce the time to complete this task by half, they could support the development of applications that analyze data and provide insights, thus potentially streamlining the forecasting process.
Economist (20053)	Conduct research on economic issues, and disseminate research findings through technical reports or scientific articles in journals.	E3	LLMs can assist in writing, editing, and summarizing research findings, significantly reducing the time needed to produce technical reports or articles.
Economist (7541)	Testify at regulatory or legislative hearings concerning the estimated effects of changes in legislation or public policy, and present recommendations based on cost-benefit analyses.	E0	This task requires direct human interaction and the ability to respond to questions and engage in discussions, which cannot be effectively supported by an LLM.

Notes: Classification levels (E0–E2) reflect the degree to which LLMs are expected to reduce task completion time.

Prompts

Classification

System Instructions

The assistant receives the following instructions:

System Instructions

You are a helpful research assistant who classifies tasks according to their exposure to LLMs. For each task, you must produce exactly one valid JSON object with the following three keys:

- `"exposure category"`: one of `["E0", "E1", "E2", "E3"]`
- `"explanation"`: your step-by-step reasoning
- `"confidence"`: one of `["high", "moderate", "low"]`

Your entire response must consist of only this JSON object without any additional commentary or text.
Here's how to determine the exposure category:

E3 - Direct Exposure

Label tasks E3 if direct access to the LLM through an interface like ChatGPT or the OpenAI playground alone can reduce the time it takes to complete the task with equivalent quality by at least half. This includes tasks that can be reduced to:

- Writing and transforming text and code according to complex instructions,
- Providing edits to existing text or code following specifications,
- Writing code that can help perform a task that used to be done by hand,
- Translating text between languages,
- Summarizing medium-length documents,
- Providing feedback on documents,
- Answering questions about a document,
- Generating questions a user might want to ask about a document,
- Writing questions for an interview or assessment,

- Writing and responding to emails, including ones that involve refuting information or engaging in a negotiation (but only if the negotiation is via written correspondence),
- Maintain records of written data,
- Prepare training materials based on general knowledge, or
- Inform anyone of any information via any written or spoken medium.

E1 - Exposure via LLM-Powered Applications

Label tasks E1 if having access to the LLM alone may not reduce the time it takes to complete the task by at least half, but it is easy to imagine additional software that could be developed on top of the LLM that would reduce the time it takes to complete the task by half. This software may include capabilities such as:

- Summarizing documents longer than 2000 words and answering questions about those documents,
- Retrieving up-to-date facts from the Internet and using those facts in combination with the LLM capabilities,
- Searching over an organization's existing knowledge, data, or documents and retrieving information,
- Retrieving highly specialized domain knowledge,
- Make recommendations given data or written input,
- Analyze written information to inform decisions,
- Prepare training materials based on highly specialized knowledge,
- Provide counsel on issues, and
- Maintain complex databases.

E2 - Exposure With Image Capabilities

Suppose you had access to both the LLM and a system that could view, caption, and create images as well as any systems powered by the LLM (those in E1 above). This system cannot take video as an input and it cannot produce video as an output. This system cannot accurately retrieve very detailed information from image inputs, such as measurements of dimensions within an image. Label tasks as E3 if there is a significant reduction in the time it takes to complete the task given access to a LLM and these image capabilities:

- Reading text from PDFs,

- Scanning images, or
- Creating or editing digital images according to instructions. The images can be realistic but they should not be detailed. The model can identify objects in the image but not relationships between those options.

E0 - No Exposure

Label tasks E0 if none of the above clearly decrease the time it takes for an experienced worker to complete the task with high quality by at least half. Some examples:

- If a task requires a high degree of human interaction (for example, in-person demonstrations) then it should be classified as E0.
- If a task requires precise measurements then it should be classified as E0.
- If a task requires reviewing visuals in detail then it should be classified as E0.
- If a task requires any use of a hand or walking then it should be classified as E0.
- Tools built on top of the LLM cannot make any decisions that might impact human livelihood (e.g. hiring, grading, etc.). If any part of the task involves collecting inputs to make a final decision (as opposed to analyzing data to inform a decision or make a recommendation) then it should be classified as E0. The LLM can make recommendations.
- Even if tools built on top of the LLM can do a task, if using those tools would not save an experienced worker significant time completing the task, then it should be classified as E0.
- The LLM and systems built on top of it cannot do anything that legally requires a human to perform the task.
- If there is existing technology not powered by an LLM that is commonly used and can complete the task then you should mark the task E0 if using an LLM or LLM-powered tool will not further reduce the time to complete the task. When in doubt, you should default to E0.

Few-Shot Examples

Here are instructions followed by assistant output examples, strictly in JSON format, for reference during classification:

Few-shot Examples

User prompt:

> *"Occupation: Inspectors, Testers, Sorters, Samplers, and Weighers*
> *Task: Adjust, clean, or repair products or processing equipment to correct defects found during inspections."*

Assistant response:

```
{
  "exposure_category": "E0",
  "explanation": "This task requires physical manipulation
     and repair of equipment which cannot be performed by
     an LLM",
  "confidence": "high"
}
```

—

User prompt:

> *"Occupation: Computer and Information Research Scientists*
> *Task: Apply theoretical expertise and innovation to create or apply new technology."*

Assistant response:

```
{
  "exposure_category": "E1",
  "explanation": "This task involves theoretical knowledge
     and application which can be significantly aided by
     LLMs",
  "confidence": "moderate"
}
```

Revelio Labs

Revelio Labs has developed a job classification system that organizes millions of distinct job titles into standardized occupational categories. At the highest level, Revelio's taxonomy groups roles into seven broad job categories.

Revelio Labs Broad Job Categories

Broad Category	Description	Example Job Titles
Admin	Administrative support, corporate services, and human resources roles.	Human Resources Manager, Office Administrator, Executive Assistant, Administrative Coordinator, Recruiting Specialist
Engineer	Technical and engineering roles involved in product development, IT, and technology.	Software Engineer, Mechanical Engineer, DevOps Engineer, Systems Analyst, IT Network Engineer
Finance	Financial planning, accounting, and related roles managing budgets and capital.	Accountant, Financial Analyst, Investment Analyst, Auditor, Credit Risk Manager
Marketing	Roles focused on market research, advertising, communications, and product marketing.	Marketing Manager, Digital Marketing Specialist, Brand Strategist, Public Relations Coordinator, Content Marketing Analyst
Operations	Operational management, logistics, and general business operations roles.	Operations Manager, Supply Chain Analyst, Project Manager, Logistics Coordinator, Plant Manager
Sales	Customer-facing commercial roles that drive revenue and client relationships.	Sales Representative, Account Executive, Business Development Manager, Sales Associate, Client Success Manager
Scientist	Includes research, scientific R&D, and data-centric roles.	Research Scientist, Data Scientist, Laboratory Researcher, R&D Chemist, Clinical Research Associate

It also provides multiple levels of granularity below these broad groups, allowing users to drill down into more specific job functions.[25]

[25] Beneath the seven broad groups, Revelio Labs provides multiple layers of more granular job classifications. The job taxonomy is structured hierarchically, meaning each broad category breaks down into subcategories and specific job clusters, with the most granular level with roughly 1,500 distinct job groupings.