

The Early Economic Impacts of Transformative AI: A Focus on Temporal Coherence

April 28, 2025

Félix Dorn
Independent

Xavier Ferrés
Universitat Pompeu Fabra

Elsie Jang
Independent

Valmik Nahata
University of California, San Diego

With
BlueDot Impact & Apart Research

Abstract

This paper investigates the economic potential of Transformative AI, focusing on “temporal coherence”—the ability to maintain goal-directed behavior over time—as a critical, yet underexplored, factor in task automation. We argue that temporal coherence represents a significant bottleneck distinct from computational complexity. Using a Large Language Model to estimate the ‘effective time’ (a proxy for temporal coherence) needed for humans to complete remote O*NET tasks, the study reveals a non-linear link between AI coherence and automation potential. A key finding is that an 8-hour coherence capability could potentially automate around 80-84% of the analyzed remote tasks.

1 Introduction

Transformative AI (TAI) refers to AI systems that precipitate a transition comparable to, or more significant than, the agricultural or industrial revolutions (Karnofsky (2016)). We take a broad and inclusive view of transformative AI, including systems that could fundamentally alter economic structures even if they lack many capabilities humans possess. TAI overlaps with, but is distinct from, concepts like “superintelligence” and “artificial general intelligence” (AGI).

Recent arguments suggest that transformative AI is not a distant prospect, but a current reality. Agüera y Arcas and Norvig (2023) argue that frontier models such as GPT-4, Claude, and LLaMA already meet key criteria for AGI: broad task competence, multi-modality, instructability, and in-context learning. Though flawed, these systems represent the first true examples of general intelligence, much as the ENIAC was the first general-purpose computer despite its limitations. Thus, we frame our inquiry around the early economic impacts of transformative AI already in deployment.

If transformative AI is here, the immediate question is how its integration will reshape the economy. Prior work highlights both utopian and dystopian possibilities. Autor (2024) argues that AI could “rebuild the middle class” by extending decision-making capabilities to a broader set of workers, counteracting the hollowing out caused by previous waves of computerization. Meanwhile, Korinek and Suh (2024) model transitions to AGI under different assumptions, emphasizing that outcomes could range from wage collapse under full automation to indefinite wage growth if complex human tasks persist. Their work highlights the critical importance of the structure of human work and the pace of technological progress in determining economic outcomes.

In parallel, empirical work by Barnett (2025) at Epoch AI offers a compelling methodology for tracking the early economic impacts of transformative AI. Barnett investigates how the automation of remote work—tasks performable entirely via digital interfaces—could dramatically expand economic output. By systematically reclassifying tasks in the O*NET database using large language models, Barnett forecasts large, rapid GDP gains even under conservative assumptions.

Building on this work, we propose a framework that extends task-based models by incorporating the concept of temporal coherence: the degree to which completing a task requires maintaining consistent goals, plans, and reasoning over time. We argue that temporal coherence constitutes an independent axis of task difficulty, distinct from traditional measures of computational complexity or routineness.

Our key contribution is to highlight temporal coherence as a critical, underexplored bottleneck for AI-driven automation. By focusing on short-run indicators and early occupational impacts, we aim to complement longer-term scenario work on transformative AI and to provide actionable insights for policymakers, firms, and workers navigating the transition.

2 Framework

2.1 Background: Task-based models and Automation

Labor economists have traditionally approached the question of technological automation through a task-based framework (Autor (2013); Acemoglu and Autor (2011)). In this view, tasks—rather than occupations—are the fundamental units of economic production. Technologies substitute for, complement, or redefine specific tasks, and occupational change is a consequence of underlying task-level shifts.

Korinek and Suh (2024) refine this perspective by modeling human work as composed of atomistic tasks that vary in computational complexity, meaning the amount of compute (e.g., floating point operations) a machine would need to perform them. Technological progress is conceptualized as expanding an “automation frontier,” gradually enabling machines to perform tasks of increasing complexity.

An illustrative example they provide is teaching an economics course. While occupational databases like O*NET list “teach economic theories” as a task, this high-level label can be decomposed into subtasks: planning a syllabus, preparing lectures, delivering explanations, answering questions, recognizing confusion, adjusting content dynamically, and grading assignments. Each subtask differs in computational complexity. Planning a lecture may be relatively simple; dynamically adapting explanations in response to subtle student cues remains much more challenging.

Korinek and Suh’s model implies that automation will proceed according to the computational complexity of these atomistic tasks: low-complexity tasks will be automated earlier, while high-complexity tasks will resist automation until later technological stages.

2.2 Temporal coherence as a key bottleneck

While computational complexity is crucial, it does not fully capture the challenges AI faces in automating economically meaningful work. Many high-value tasks require not only solving individual problems, but maintaining consistent goals, reasoning, and plans over extended periods of autonomous activity.

Returning to the teaching example: even if an AI system could perform each subtask—preparing slides, answering factual questions—it would still struggle to deliver a coherent semester-long course. Effective teaching requires maintaining thematic and conceptual consistency over time, adapting to cumulative student understanding, and revising instructional strategies based on longitudinal feedback. These demands cannot be easily decomposed into independent tasks solved in isolation.

We refer to the ability to sustain goal-directed consistency and reasoning across time as

temporal coherence. Temporal coherence can be roughly defined as the capacity to plan and act consistently over a given time-horizon t .

A simple example of a task requiring temporal coherence that current AI systems struggle with is booking a flight. Other, more complex tasks include writing an economics paper, creating a PowerPoint presentation, or negotiating with someone over the course of a day. Each of these tasks requires maintaining a stable objective, coordinating subgoals, and adjusting plans dynamically—all over a sustained timeframe.

Current AI systems exhibit remarkable capabilities on isolated subtasks but continue to struggle with long-horizon coherence. Failures of planning, consistency, and state tracking are among the most common errors observed in today’s frontier models. This suggests that temporal coherence is a distinct, independent bottleneck to automation, separate from computational complexity.

Accordingly, we propose that automation potential depends not only on the computational difficulty of individual subtasks, but also on the degree of temporal coherence required to integrate those tasks into a meaningful whole. Advances in temporal coherence capabilities are likely to be key drivers of AI’s expanding economic impact.

2.3 Measuring temporal coherence and Automation potential

Building on the task-based framework and the concept of temporal coherence introduced above, we focus our inquiry on two central questions:

- (1) Which occupations are likely to be most affected by AI-driven automation in the short to medium term?
- (2) What percentage of tasks do we expect to be automatable by AI systems, and by when?

In task-based economic models, completing a task requires applying specific skills. Though skills are often abstractly classified (e.g., low-, medium-, or high-skill), we argue that a key determinant of automability is whether an AI possesses the right bundle of capabilities—including the ability to sustain coherent, goal-directed behavior over extended periods.

Several skills or capabilities have been proposed as bottlenecks (see, for instance, Aschenbrenner (2024)), but we believe temporal coherence is the most important one at this stage. Among missing capabilities, temporal coherence appears to be the largest current bottleneck for unlocking economic value. If AI systems could maintain stable plans and act autonomously across arbitrary time-horizons, they would be able to automate a much wider range of economically valuable tasks—even without major improvements in other areas like multimodality, hallucination reduction, or memory.

Thus, we propose the following measurement approach: We estimate the degree of temporal coherence required to perform different tasks by measuring the time it would take a human to complete them autonomously. Tasks requiring longer autonomous execution are presumed to demand higher levels of temporal coherence.

There are two main reasons why we focus on temporal coherence rather than some other skill that AI systems currently lack. First, among the set of missing capabilities, temporal coherence is arguably the biggest bottleneck in terms of unlocking economic value (EpochAI (2025); Patel (2025)). One way to phrase this is that if temporal coherence were “solved”—if AI systems could act consistently and autonomously across any time-horizon t —this would enable the automation of more tasks than solving any other capability bottleneck (such as multimodality, hallucination reduction, or memory improvements). Second, the temporal coherence of different LLMs has been recently estimated by METR using coding tasks (METR (2025)). These empirical benchmarks allow us to predict when models might become capable of completing longer-horizon tasks autonomously, thus providing a concrete basis for forecasting patterns of automation.

3 Methodology

Our methodology aims to estimate the temporal coherence required for tasks within the US economy, focusing specifically on remote work potentially automatable by AI agents. We leverage the O*NET database, build upon prior work classifying remote tasks, employ Large Language Models (LLMs) for time estimation, and validate our approach using a manually annotated dataset and a defined loss metric.

3.1 Data Sources and Preparation

We utilize the Occupational Information Network (O*NET) database (v29.1) as our primary source for detailed descriptions of occupations and associated tasks in the US economy, following established practices in the literature (Arntz et al. (2016); Brynjolfsson et al. (2018); Duckworth et al. (2019); Eloundou et al. (2023); Felten et al. (2021); Frey and Osborne (2017)).

Our analysis focuses on tasks amenable to automation by current and near-term AI agents, which primarily operate in digital environments. Therefore, we build upon the work of Barnett (2025), using their annotated dataset identifying O*NET tasks that can be performed remotely. Barnett defines remote tasks as those accomplishable entirely via digital tools, a computer, and an internet connection, without requiring physical presence. This excludes tasks demanding physical manipulation or location-specific presence.

3.2 LLM-based Estimation of Effective Task Duration

We employ an LLM to estimate the 'effective time' required for a skilled human to complete each remote task. This duration serves as our proxy for the temporal coherence demanded by the task. The underlying heuristic is that tasks requiring longer periods of sustained, autonomous human effort necessitate a higher degree of temporal coherence for an AI agent to successfully automate.

Definition of Effective Time: We define 'effective time' as the active, focused work duration needed to complete the task, crucially excluding any waiting periods, communication delays, or time spent on unrelated activities. This represents the continuous, productive time investment required if the worker could operate without interruption.

LLM Prompting Strategy: We utilized the OpenAI's GPT-4.1-mini model. For each task, the model received:

- A system prompt defining the objective: estimate the lower and upper bounds of 'effective time' within which 80% of instances are completed by a skilled human, excluding wait times, and selecting only from predefined duration categories.
- A user message containing the specific task details: Occupation Category, Occupation Description, Task Description, and associated Detail Work Activities from the O*NET database.

Discrete Time Categories: To standardize the LLM's output and simplify the estimation task, we restricted the possible lower and upper bound estimates to the following discrete, exponentially scaled durations: ['10 minutes', '30 minutes', '1 hour', '2 hours', '4 hours', '8 hours', '16 hours', '3 days', '1 week', '3 weeks', '6 weeks', '3 months', '6 months', '1 year', '3 years', '10 years']

Conversion to Quantitative Measures: For quantitative analysis (e.g., calculating loss, aggregations), these categorical estimates were converted into minutes. We used the conversion factors detailed in Table 1.

3.3 Validation and Optimization

To assess the reliability of the LLM-generated estimates and refine our prompting strategy, we implemented a validation process.

Validation Dataset Creation: We randomly sampled 45 remote tasks from the Barnett-annotated O*NET dataset. These tasks were then manually annotated by the authors to establish 'golden' lower and upper bound effective time estimates, using the same discrete duration categories provided to the LLM.

Table 1: Conversion Factors for Time Categories

Category	Minutes Equivalent
10 minutes	10
30 minutes	30
1 hour	60
2 hours	120
4 hours	240
8 hours	480
16 hours	960
3 days	1440
1 week	2400
3 weeks	7200
6 weeks	14400
3 months	28800
6 months	57600
1 year	115200
3 years	345600
6 years	691200
10 years	1152000

Loss Function: We defined a loss metric to quantify the discrepancy between the LLM’s predictions (*pred*) and the manually annotated golden values (*golden*) for both lower and upper bounds on the validation set. We used the Mean Absolute Error (MAE) on the log scale, calculated as follows:

$$\text{Loss}_{\text{lower}} = \left| \ln \left(\frac{\text{pred}_{\text{lower}}}{\text{golden}_{\text{lower}}} \right) \right|$$

$$\text{Loss}_{\text{upper}} = \left| \ln \left(\frac{\text{pred}_{\text{upper}}}{\text{golden}_{\text{upper}}} \right) \right|$$

$$\text{Total Loss} = \exp \left(\frac{\text{mean}(\text{Loss}_{\text{lower}}) + \text{mean}(\text{Loss}_{\text{upper}})}{2} \right)$$

This loss function penalizes estimates proportionally to their deviation from the golden standard on a logarithmic scale, providing an interpretable measure of average error magnitude (e.g., a loss of 1.5 means predictions are off by a factor of 1.5 on average).

Hyperparameter Tuning: We iteratively refined the system prompt, user message structure, and experimented with different LLM models to minimize the validation loss. The final model and prompts reported above yielded a validation loss of 2.1.

3.4 Assumptions and Limitations

This methodology relies on several key assumptions:

- O*NET task descriptions, augmented with DWAs, provide sufficient detail for meaningful time estimation by LLMs.
- The chosen LLM can reasonably interpret the concept of 'effective time' and exclude waiting periods based on the provided instructions.
- The selected discrete time categories adequately capture the relevant range of task durations.
- The manual annotations on the validation set (N=45) provide a representative benchmark for the broader set of remote tasks.

4 Results

Our analysis focuses on quantifying the relationship between AI temporal coherence and task automability, directly addressing our research questions regarding the scope and timeline of potential automation. We estimated the 'effective time'—our proxy for temporal coherence—required for humans to complete remote tasks identified in the O*NET database.

The first key finding illustrates the cumulative percentage of tasks that become automatable as AI agent coherence time increases, as shown in Figure 1. A striking result is the non-linear relationship observed. Initially, increasing coherence yields modest gains in automation; agents with only 10 minutes of coherence can automate very few tasks. However, the potential for automation accelerates dramatically between coherence times of 1 hour (automating 10% of tasks) and 8 hours. This suggests that achieving coherence levels equivalent to a standard workday represents a critical threshold. Indeed, as highlighted in the figure's title, an agent coherence time of 8 hours corresponds to the potential automation of approximately 83-84% of the analyzed remote tasks. This finding is particularly interesting because it implies that improvements in AI coherence within this specific range (1-8 hours) could unlock a disproportionately large share of economic tasks, supporting our hypothesis that temporal coherence is a major bottleneck. Beyond 8 hours, the rate of gain diminishes significantly, indicating that while longer coherence times enable further automation, the most substantial impact occurs earlier. Reaching near-total automation (approaching 100%) requires coherence times extending to weeks or months.

Building upon this relationship and incorporating forecasts of AI capability development (specifically, temporal coherence improvements based on benchmarks like METR (2025)), we project the timeline for task automation. Figure 2 presents the estimated percentage of tasks that remain *non-automatable* over the period 2022-2030. The projection reveals

Over 80% of the tasks in the economy can be automated with 8 hours of coherence

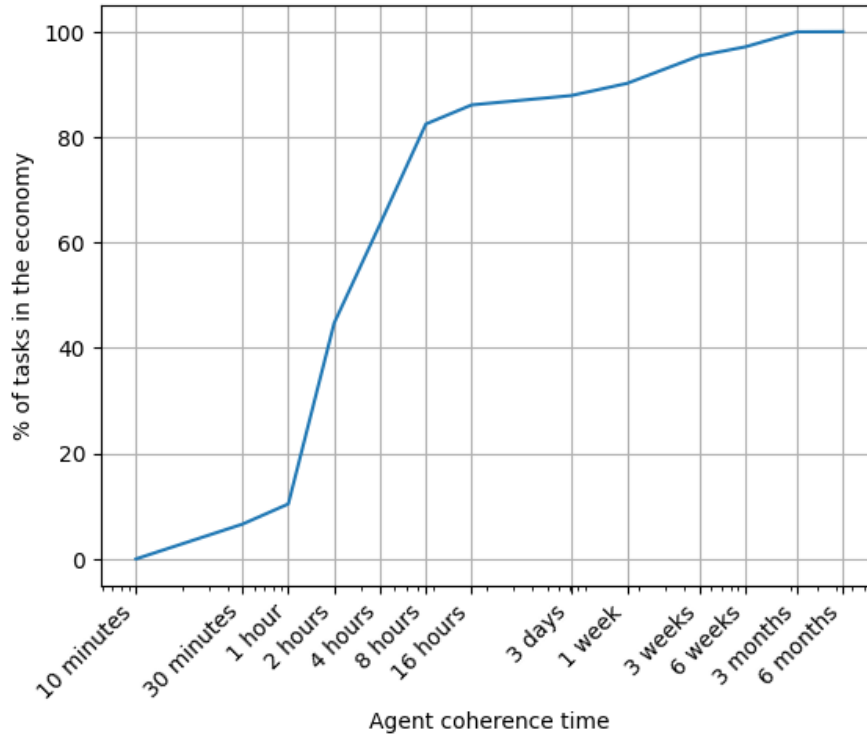


Figure 1: Percentage of tasks automatable as a function of AI agent coherence time. Note the logarithmic scale on the x-axis and the step increase between 1 and 8 hours.

a period of stability until 2024, where essentially all tasks considered require coherence beyond current AI capabilities. However, the model forecasts a remarkably rapid shift shortly thereafter. Between 2024 and 2026, the percentage of non-automatable tasks is projected to plummet from 100% to approximately 36%. This suggests a potential wave of automation impact concentrated in a relatively short timeframe, driven by anticipated breakthroughs in AI’s ability to handle tasks requiring moderate temporal coherence (likely corresponding to the steep part of the curve in Figure 1). Following this rapid transition, the pace of automation slows, with the percentage of non-automatable tasks decreasing further to 14% by 2028 and 3% by 2030. This projected S-curve pattern—slow initial progress, followed by rapid acceleration, and finally diminishing returns—is a crucial insight for policymakers and businesses preparing for AI’s economic integration. It highlights that the economic disruption driven by improving temporal coherence may not be gradual but could manifest significantly over a few critical years.

In summary, our results quantify the critical role of temporal coherence. Figure 1 demonstrates that achieving human-like workday coherence (8 hours) could automate a vast majority (80%) of remote tasks, while Figure 2 projects that this capability threshold might be crossed rapidly between 2024 and 2026, leading to a significant potential shift in the labor market landscape within this decade. These findings directly address our research questions by identifying tasks requiring moderate coherence (roughly 1-8 hours) as most susceptible in

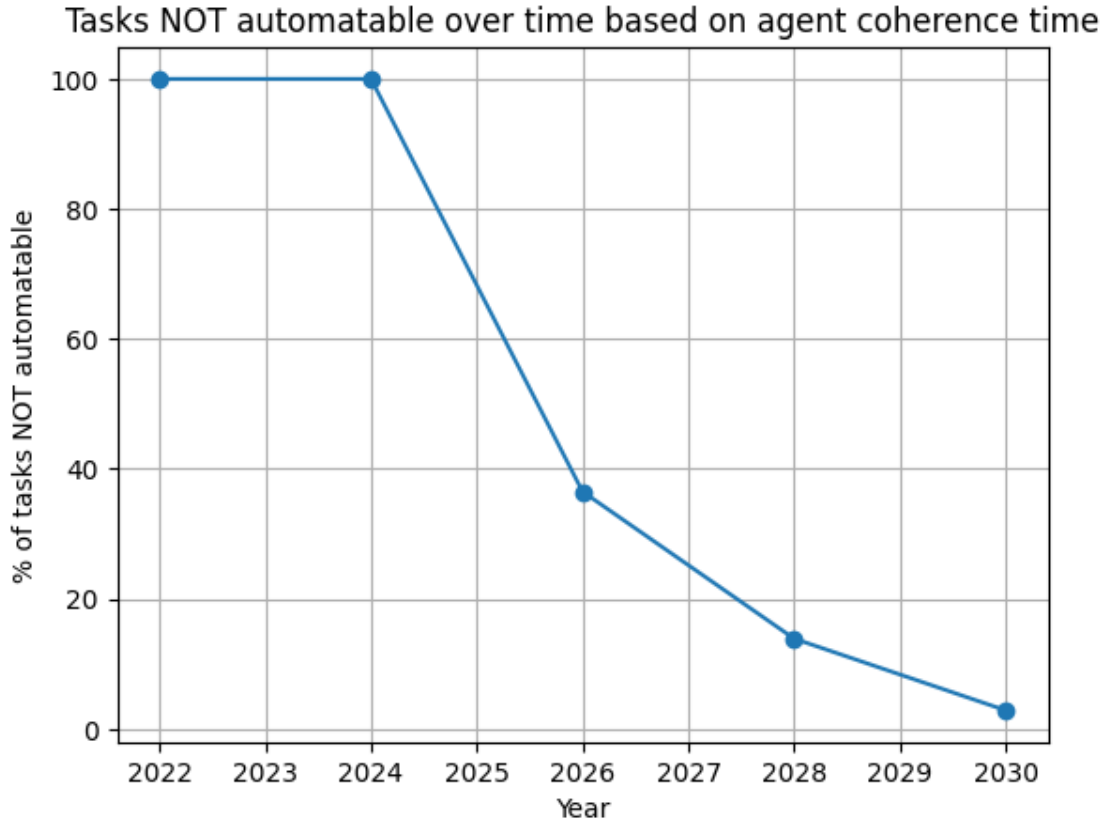


Figure 2: Projected percentage of tasks NOT automatable over time (2022-2030), based on forecasted improvements in AI agent coherence time.

the near-to-medium term and projecting a timeline heavily influenced by anticipated gains in this specific AI capability.

5 Discussion

5.1 Contribution

Our key contribution is to introduce temporal coherence as an independent axis of task difficulty for AI-driven automation. While prior task-based models have emphasized computational complexity or the physical remoteness of tasks, we argue that the requirement to maintain consistent, goal-directed behavior over extended time frames constitutes a distinct and critical bottleneck.

By framing temporal coherence as a measurable capability constraint, we offer a new lens for analyzing early economic impacts of transformative AI. Our approach complements longer-term scenario work by focusing on near-term occupational disruptions and provides ac-

tionable insights for policymakers, firms, and workers seeking to anticipate and navigate AI-driven labor market changes.

Finally, by integrating recent empirical benchmarks on temporal coherence capabilities, we ground our analysis in measurable model performance, enabling more concrete forecasts of when and where automation pressures are likely to emerge.

5.2 Caveats

While our framework focuses on the potential for AI to automate tasks, it is important to note that automation does not immediately or necessarily translate into realized automation. Multiple factors can influence this transition.

Regulations may limit the replacement of human workers with AI systems in some sectors in which there are strict regulations that govern how services are delivered. Automation might be restricted or made less attractive, with high compliance costs. Similarly, there may be a persistent “human premium” – situations where employers or customers prefer human-performed tasks due to perceived higher quality, trust, or reputation. This is also applicable to tasks that are not repetitive or standardized.

Operational costs are another consideration. Even if a model can technically automate a task, its true value depends on whether it is cost-effective compared to human labor. Though in the long term, it has the potential to reduce labor costs, the initial investment in developing, deploying, and maintaining AI models can be expensive.

6 Conclusion

This paper investigated the incipient economic ramifications of Transformative AI, centering on the critical role of temporal coherence—an AI’s capacity to sustain goal-directed action over time. We argue that temporal coherence presents a distinct and significant bottleneck to task automation, complementing existing frameworks focused primarily on computational complexity.

Our analysis, leveraging LLM-based estimations of the ‘effective time’ required for remote O*NET tasks, reveals a stark, non-linear relationship between AI coherence and automation potential. A key threshold appears around the 8-hour coherence mark, potentially unlocking the automation of approximately 80-84% of the analyzed remote tasks. Projections based on anticipated advancements suggest that AI systems may cross this critical threshold rapidly between 2024 and 2026, potentially triggering a significant wave of automation impacting a large fraction of the remote workforce in a compressed timeframe.

While we demonstrate the potential for automation based on technical capabilities, the actual

pace and extent of AI adoption will inevitably be modulated by real-world factors, including regulatory environments, implementation costs, and the persistence of a premium on human interaction or non-standardized tasks.

Nonetheless, our findings underscore the paramount importance of tracking temporal coherence as a leading indicator of AI's transformative economic impact. Monitoring advancements in this specific capability dimension will be crucial for policymakers, firms, and workers seeking to anticipate and navigate the substantial labor market adjustments likely to unfold in the coming years.

References

- Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *NBER Working Paper No. 16082*.
- Agüera y Arcas, B. and Norvig, P. (2023). Artificial general intelligence is already here. Noema Magazine.
- Arntz, M., Gregory, T., and Zierahn, U. (2016). The risk of automation for jobs in oecd countries: A comparative analysis. *Oecd social, employment and migration working papers no. 189*, OECD Publishing.
- Aschenbrenner, L. (2024). Situational awareness: The decade ahead. *Situational Awareness*.
- Autor, D. (2024). Ai could actually help rebuild the middle class. Noema Magazine.
- Autor, D. H. (2013). The “task approach” to labor markets: An overview. *Journal for Labour Market Research*, 46(3):185–199.
- Barnett, M. (2025). The economic consequences of automating remote work. Epoch AI.
- Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108:43–47.
- Duckworth, P., Graham, L., and Osborne, M. (2019). Inferring work task automatability from ai expert evidence. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 485–491.
- 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.
- EpochAI (2025). Is it 3 years, or 3 decades away? disagreements on agi timelines. Transcript of podcast episode.
- Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.
- Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280.
- Karnofsky, H. (2016). Some background on our views regarding advanced artificial intelligence. Open Philanthropy.
- Korinek, A. and Suh, D. (2024). Scenarios for the transition to agi. Centre for the Governance of AI.
- METR (2025). Measuring ai ability to complete long tasks. METR Blog.
- Patel, D. (2025). Agi is still 30 years away — ege erdil & tamay besiroglu. Dwarkesh Podcast.

A Appendix

A.1 Code Repository

The Python code implementing the data processing, LLM calls , time conversion, and loss calculation described in Section 3 is available at: <https://git.xfe.li/dorn/sprint-econtai>