High-level summary and review

The GATE model is an impressive achievement, offering a highly detailed and practical framework connecting developments in artificial intelligence to macroeconomic growth theory. It explicitly integrates AI progress—most importantly, improvements in hardware and software compute efficiency—into an economic framework designed to model endogenous investment in these durable technologies.

The paper's realistic representation of technological progress will become a benchmark in detailed economic treatments of AI. For example, by distinguishing between hardware upgrades, which enhance only newly installed computational capital, and software advancements, which benefit the entire stock, the model captures asymmetries that are very real for the industry and which are also of central importance to the macroeconomic dynamics of investment. To my knowledge, this has not been done before, and is critical to do.

Equally important to the discussion of the AI future is the paper's modeling of limits on AI: physical limits and complementarity-driven bottlenecks. Both AI's capacity to do work with currently existing capacities, as well as AI technological progress, is to some extent constrained by human inputs. The model takes this seriously, very much unlike many discussions by technology leaders. This makes the model capable of structuring discussions about this dimension, which many economists, going back to Baumol, consider decisive.

These reasons suffice to put the paper on any reading list for economists interested in the pragmatic macroeconomics of AI. I enjoyed studying the paper, and found it both illuminating and provocative.

One provocative dimension is a novel description of the mechanics of AI work, treating labor as a factor that can effectively be accumulated through the creation of "digital workers." The paper puts this conceptual perspective front and center. On the one hand, this is a natural way to think literally about AI labor—software engineers certainly feel nowadays that they are competing and collaborating with digital workers. But I was not sure whether it is the most natural way in terms of macroeconomics. To my mind, the underlying accumulable items are more accurately interpreted as ideas (algorithms) and software, as well as specialized physical capital. Digital workers can be activated or deactivated as needed using these inputs and some others, but I'm not sure I want to think of an actual "worker"—an instance of Claude solving my problem through the API—as accumulable.

That's my only real disagreement with the authors. Everything else I have to say is a series of next steps that I expect to be addressed in follow-up work to this important contribution.

A significant limitation of the model is the fixed representation of complementarities between human labor and AI, captured by a constant technological parameter rho. Recent work, such as Ide and Talamàs (2025), emphasizes that such complementarities evolve dynamically as AI capabilities improve, potentially reshaping the qualitative outcomes of the model. For example, AI workers who performed routine tasks and amplified the value of experienced knowledge workers may become managers and replace or supervise those knowledge workers. In any case, the nature of human-AI interaction is likely to change substantially over the time frame considered by the model rather than remaining static. This seemed to me like one of the most important changes that could significantly alter the model's predictions about automation trajectories and economic impacts.

The paper works only with a dynamic social planner's problem. It explicitly does not consider this as a proxy for a market solution in the spirit of the First Welfare Theorem, because it does not envisage R&D markets as existing. While the planner's solution is a totally reasonable place to start, it does restrict somewhat the analysis of relevant policy questions. Introducing economic wedges—such as investment taxes or subsidies—seems very important.

AI R&D can generate positive externalities, such as knowledge spillovers that boost productivity and economic growth. However, it also carries negative externalities including the potential for the loss of control of weapons and biohazards, as well as political instability. The actors actually controlling R&D investment internalize neither the positive nor negative externalities. It seems valuable to study how distortions relative to the planner's solution affect the growth paths driven by AI investments. For researchers seeking to follow up on the paper, this seems like the lowest-hanging fruit.

Finally, clearer interpretation and economic justification of certain model components, particularly the inference-training tradeoffs, are necessary. The paper makes a variety of assumptions about adjustment costs and the cost of various models that are controversial. Clarifying how inference costs scale with model size would enhance the transparency and persuasiveness of the model, especially for readers trying to connect the formalism to real-world AI economics. I'd like to emphasize, though, that macroeconomic modeling of these things has to start somewhere, and putting down functional forms gives us something very valuable: a starting point to argue about.

Overall, GATE provides a rich analytical framework linking AI and investments at a macro level. I expect it will make a real impact.

Appendix: Detailed Comments

Write down a summary of the main contributions of the piece

The GATE model is an integrated assessment framework that systematically connects AI development, task automation, and macroeconomic outcomes. The model's principal contribution is combining three critical elements: a compute-based model of AI development that tracks hardware and software efficiency gains, a task automation framework linking effective compute to labor task automation, and a semi-endogenous growth model with endogenous investment decisions and adjustment costs.

GATE captures how AI transforms labor from a non-accumulable to an accumulable factor through the creation of "digital workers," potentially leading to growth acceleration while being constrained by complementarities between tasks, adjustment costs, and physical limits. The paper provides a mathematical framework for understanding how AI investment choices affect automation trajectories and subsequent economic outcomes.

Write down the main strengths of the piece

The paper's greatest strength lies in its interdisciplinary integration of AI engineering and economic growth modeling. It successfully translates AI development processes into tractable economic frameworks, making AI concepts accessible to economists while incorporating economic realities like adjustment costs and investment tradeoffs for AI researchers. The compute-based approach to modeling AI progress is justified with empirical discussions, and the explicit modeling of both hardware and software efficiency improvements captures important asymmetries (e.g., software upgrades benefit all compute while hardware only improves new additions). The model sensibly incorporates physical constraints and recognizes bottlenecks from task complementarities.

Write down any significant weaknesses

The comments below offer a more complete list of areas for improvement. Some highlights in brief: The model lacks a detailed treatment of human-AI complementarities, instead using a fixed production function with static complementarity parameters. This misses how the nature of human-machine relationships is likely to evolve as AI capabilities advance. The paper's terminology around "labor accumulation" requires clarification, as (in my view) it's not labor itself being accumulated but ideas and specialized capital, which produce temporary laborers. Finally, the inference-training tradeoff mechanisms need more explicit economic/functional interpretation.

The social planner framework, while analytically tractable, sidesteps questions about investment wedges (taxes/subsidies) that would be valuable for policy discussions. These can be incorporated at low cost and might make for some striking takeaways.

There are a bunch of expository suggestions. The exposition is often convoluted with redundancies. More prominent and illustrative examples would help readers grasp key insights intuitively.

Write down concrete recommendations to improve the article

See below

Our figures and graphs tend to attract the most attention in our papers. Do you have any suggestions on how to improve them?

See below. Mainly, have more of them in the main text and have some smoothing in the illustrations in the web product.

Would you recommend publishing this article after addressing your feedback?

Yes, definitely. I could see it coming out in a general interest economics journal or a top macroeconomics field journal.

More straightforward comments

- Make sure to define new notation as soon as it is introduced even though it's standard within some literatures. Otherwise, the paper appears confusing to the first-time reader – I expect your paper will have a broader audience than most.
- The language needs to be tightened significantly. The paper has many long and convoluted sentences. These often interfere with the reader's understanding. Relatedly, there are many redundancies. The paper could be significantly shortened without losing in content, and I think that would increase its impact significantly.
- The main body of the paper should include an explanation of how the model is solved. Even a concise explanation would do.
- Expository: The paper makes a series of functional form assumptions–many of them standard–which are fine with me. However, some choices are left unexplained; since your readers will likely not be primarily ones steeped in the macroeconomic modeling assumptions. The macroeconomic model crucially lacks this; even though you are adapting an existing framework, it would still be valuable to explain the rationale

behind the specific formulas you adopt. This is perhaps especially true for the capital adjustment costs assumption.

- The paper states "it would be interesting to analyze market equilibrium counterparts to the planner setting studied in GATE, under alternative market structures." It would be useful at this point to make a note for the broader audience: except for the externalities of R&D investment, given sufficiently complete markets, the planner's solution is the market equilibrium solution. (This could even be literally true in some version of this model if there were sufficiently good IP markets to compensate the R&D.) This could connect to the discussion of wedges that I suggest below in my more significant/substantive comments.
- The curves shown in the web product are too jagged for my taste. A little bit of spline-based smoothing, as long as honestly noted, will help visually.
- The main paper would benefit from more in-text illustrations – showing some scenarios and discussing them using the visuals. (Some economists put these in the back of papers but I think in the middle of papers is the right place.)

More substantive

- The integration of AI in the economy seems to be missing a very important element: Humans often complement AI, and the nature and evolution of these complementarities is likely to be a major part of the story of AI's impact on the economy. It seems important to discuss, and some modeling of this would make me way more confident in the paper's predictive dimension. See Ide and Talamàs, 2025 for ideas.
    - To be clear, the CES modeling of task aggregation already incorporates a humans-and-machines complementarity.
    - But the nature of the complementarity is not, in reality, a fixed technological parameter (see the above cited paper).
    - I don't recommend you do it in this model, but it seems like many of the qualitative predictions might be fragile to a dynamic rho, or major heterogeneities in rho across the economy.
    - Separately, the first pages of Ide and Talamàs, 2025 contain some writing that are related to the conceptual contributions of this paper regarding labor being accumulable etc.
- "(…) once a model is large enough to accommodate a complex task, the additional inference resources needed to run that task can be proportionally smaller:" It would help to be a lot more explicit about the "production function" you envision for a large model here.
    - At least at a superficial level, inference marginal prices per token are generally highest for the largest models.

    - Is the idea that fewer tokens are effectively required through some mix of more efficient completion/less prompting? As far as I understand it, the key thing being modeled is that "A larger model might be more compute-efficient per task because it can solve problems more directly, even though each individual operation costs more." But this is where writing things out explicitly/quantitatively would be very valuable.
    - Or is part of the issue with increased inference cost in large models that the firms are recouping training costs?
    - References or an explicit argument would be great.
- Externalities/wedges: Since investment plays a major role in the model, it would be natural from a macroeconomic perspective to discuss wedges (effective taxes and subsidies) relative to the planner's dynamic solution. Government can subsidize investment in labor productivity relative to AI productivity. In winner-take-all(ish) markets, there could conceivably be effective subsidies on R&D investment arising from the fact that some firms overinvest in the short run to have a stronger claim on monopoly rents.

    - The paper mentions these issues, but mostly just to explain/defend the social-planner (equivalently, complete markets and perfect competition) benchmark.
    - Wedges are easily incorporated into exercises such as this.
    - I think a fairly low-cost extension that would be expected and valued by macroeconomists would study how the paths in the illustrations would be affected by various wedges, because policy discussions seem to be one of the main downstream impacts of this project/product. Given the dynamic nature of investment, wedges can potentially have large effects on timelines.
    - I can see the case for leaving it to future work, but starting on it seems like low hanging fruit.
    - As a side point, it's not obvious that R&D is underprovided in reality, since AI risks are socially bad and the firms don't internalize the negative externalities. So slowing down development may be socially good. I may have missed your mention of this—but if it's not there, it's worth mentioning.
- The paper acknowledges that new tasks may emerge as AI systems automate old ones. (These tasks may include things like participating in medical trials for AI-developed drugs and other tasks we can't imagine yet.) This seems very relevant in the context of discussing a full-automation transition. While adding a new-task module may be beyond scope, a more thorough discussion of the implications of ignoring new tasks for the most important qualitative predictions would be helpful. This is related to the dynamics-of-complementarity point discussed above.

- My sense is that macroeconomists will prefer to think of the production function of the AI economy as follows: tasks are done using a stock of ideas (algorithms) in combination with specific types of physical capital. Your notion of accumulable labor is of course a particular way to operationalize this, but I didn't find the terminology that helpful, because the labor itself is not being accumulated. When I think practically, the new "laborers" are individual instances of models running, but these are shut down and turned on as needed. The capital that they run on is the thing that is capable of being accumulated. Maybe I'm misunderstanding something, but clarifying the discussion around this would help.

- Some readers would probably appreciate versions of your assumptions (e.g. software efficiency doubles every ~16 months) that treated the growth curves as "logistic curves masquerading as exponential." How much would those changes affect our conclusions from this model?

A big expository comment

I think part of what makes the papers of Chad Jones so effective is that Section 2 often has a very stripped-down quantitative model highlighting one big implication or insight obtained from the big model. This model doesn't lack for striking qualitative changes to how we think about growth, but they're presented as parts of one big machine. For example, the idea might be that "'Labor accumulation' is the key mechanism through which AI development affects output in GATE." I didn't walk away from the paper with a crystal clear encapsulation of this idea that I could explain to a colleague while walking from our offices to the parking lot, and I think the paper would be a lot more powerful if it had this.