Agentic AI as a Social Phenomenon: Comment on Hadfield and Koh

Kevin A. Bryan

University of Toronto, Rotman School of Management

kevin.bryan@rotman.utoronto.ca

October 2025

For many in Silicon Valley, 2025 is "the year of the agent". An agent is AI that autonomously uses tools to plan and perform tasks over an extended period. Hadfield and Koh (2025) describe an AI that can take an instruction like set up a web store with $100,000 starting capital, and earn $1,000,000 and then execute adaptive plans autonomously to achieve that goal.

The potential economic implications of AI agents are immense. Traditional machine learning models required highly specialized skills, data, and training to predict outcomes, limiting their diffusion. Large language models with chatbot interfaces spread incredibly quickly (Hartley et al., 2025), but need frequent human intervention. Requiring neither task-specific training nor human involvement as often, the range of economic activity that AI agents can affect is substantially broader. As agents interact with other agents and humans in their autonomous work, equilibrium effects will also become important.

To be clear, AI agents are not science fiction. A well-known benchmark of autonomous work has seen a doubling every seven months of the time horizon over which frontier models can operate, reaching two hours by mid-2025 (Kwa et al., 2025). This exact figure is open

to quibbles, but the rate of improvement is not. In fields where the most advanced AI labs have concentrated effort, such as computer programming, agents are already being used in production environments via tools like OpenAI's Codex and Anthropic's Claude Code.

Hadfield and Koh (2025) describe wonderfully how AI agents could affect the economy. Let me instead turn to why they will not affect the economy nearly as much as they in principle could, and how we might speed things up.

The highest value AI agents are *architectural disruptions* (Henderson and Clark, 1990). Here, architecture refers to the relatively fixed aspects of an organization and society: reputation, communication norms, the skills of upper management, the capital-labor mix, relational contracts, existing laws, licensing rules, and so on. Architectural features tend to exist because they were at some point a good fit for the incumbent state of technology and society.

Why does this matter? To use a chatbot LLM, a worker can simply experiment on their own; this is not architectural, even if perhaps the IT security folks send stern warnings. But for an organization to use an AI agent most productively, changes are required in organizational and societal architecture. Consider the agent use cases suggested by Hadfield and Koh (2025): negotiating purchases, bargaining, reducing coordination costs across firm boundaries and internal divisions, speeding up R&D iteration, monitoring performance in real time, and so on. To actually perform most of these tasks requires changes in sticky institutions, hence the benefit of AI agents will lag their *technical* capability, perhaps by many years.

Let us rank the challenges in deploying an AI agent in order of difficulty. First, we must be able to explain to the agent what task we want done. Second, we need to ensure it actually tries to do what we ask it to do. Third, we must consider whether, in equilibrium, it can do what we want it to do when the agent interacts with others, including other agents. Fourth, we must ensure that the agent gets required information or assistance from other humans or AIs. And fifth, we must ensure that partners, vendors, customers, and regulators actually

permit our agent to do what we want.

The first two challenges are effectively technical. Technical, but not trivial. The AI alignment problem, in a sense, is similar to standard principal-agent economics with transaction costs in "explaining what we want," albeit subject to the now-famous jagged frontier problem (Dell'Acqua et al., 2025). However, for the sake of argument in line with this volume's focus on Transformative AI, let us assume that we are able to express to the agent what we want it to do, and able to guarantee the AI will in fact try to do it.

The last three challenges prove more interesting, since they are not purely technical. Tomasev et al. (2025) point out that the broadest AI agent use cases are likely to be *emergent* in their effects and *permeable* between humans and AI. Their equilibrium effects and even their primary uses will be difficult to predict and will frequently interact with existing human institutions. Why do we say, then, that steps 3, 4, and 5 are the most difficult?

Let us imagine it is 1885, not 2025. The Wizard of Menlo Park has been operating the Pearl Street Station in New York City for three years, providing electricity to homes and industry. Daimler and Benz are about to announce their internal combustion vehicles. I work in agriculture. What would we predict? How quickly, that is, will the dynamo and the petrol engine affect productivity in my sector?

Some use cases are easy. I can take my gas lamp and replace it with an electric light as soon as Edison draws the wires to my village. Horses are slow, but the rutted track between my farm and the market would be easier to reach with a car. There are technical challenges, to be sure - early lightbulbs were inefficient, and cars were expensive and prone to breaking down. But one might foresee that these problems could be solved - "Transformative Energy" and "Transformative Transportation", shall we say?

With hindsight, we know that driving to market and lighting up the farmer's kitchen were not the important impacts of these technologies. From the vehicle, we got the tractor, which changed how crops were planted, the size distribution of farms, and the economics of complementary goods like grain silos and pesticides (Gross, 2017). From electricity came the

water pump, and even the dam (Duflo and Pande, 2007), bringing power and related water control to farms. These innovations mattered much more for productivity. But they took time. If you were a farmer in 1885, thirty years later, you would not have realized many of these benefits. Why?

Consider the tractor. The motor vehicle must be modified for farms. Then complementary technologies like plows must be developed. Then developments like adjustable-width treads make tractors useful for cotton and corn fields. Vendors in your region need to sell these implements. Limits of farm consolidation make economies of scale challenging. Larger economies of scale require storage via grain silos, and even railroad spurs to be able to move this amount of grain to market. You need to be able to hire a skilled mechanic instead of a skilled scytheman - where will they be found?

Electricity tells a similar story. Paul David's example of electricity having small initial effects on industry when it merely replaces belt and turbine technology at slightly lower cost, but eventually permitting "sideways" factories and the assembly line, has parallels in agriculture (David, 1990). The financial and organizational bodies that developed rural electric grids needed to be created, the complementary electrified farm tools needed to be invented, national standards for electrification were needed to coordinate this new technology and speed adoption, and so on.

Coming back to 2025, AI agents will not be any different. In principle, AI agents can search, negotiate, and measure much more quickly than humans can. These are of course some of the famous Coasean transaction costs (Coase, 1960). Consider an automobile manufacturer that wants to source some input, perhaps requiring modifications over time. One may think AI agents will permit much deeper contracts and hence more arm's-length transactions. To do so, the AI agent must have the proper context of what the rest of the automobile manufacturer is attempting. And it must be able to gather that information both from other AIs and from humans who will attempt to get information rents - consider a worker trying to goldbrick the agent by obfuscating the productivity benefit of the new

input (Roy, 1952). The AI agent's choice will affect the career trajectory of many inside the organization. Those humans (or even agents with other goals) will try to influence our agent (Milgrom and Roberts, 1992; Henderson and Gibbons, 2012).

So much for those internal problems. What of external ones? Who is going to sign the 20,000-page contract that covers more contingencies than any human could understand? If a human must sign, ought they be liable for something they are unable to fully read and understand? How are we to know a given agent actually represents a given party? Since these agents will work autonomously and with emergent capabilities, conditional on relational contracts ensuring any remaining incompleteness does not cause the relationship to collapse, what ensures the AI plays the "good" equilibrium (Macaulay, 1963)?

These issues go beyond what our AI can do as an algorithm. Instead, they depend on managers' desires, other AIs' schemes, and regulators' demands. Since the use cases for AI agents are emergent and path-dependent, our humility does not allow us to predict precisely how they will operate. But we can predict that regulatory, organizational, and social choices taken in the very near future will constrain experimentation and diffusion.

What therefore ought we to do to speed up the social benefit of AI agents given these frictions? For firms, the primary goal has to be "loosening" architecture. For instance, if coding agents can generate software for one-off use on projects cost-effectively, then the CIO's process for evaluating software as if it were all procured in large chunks from vendors must change. If monitoring workers in real time is easier, then the HR department rules for how workers are punished, or more positively for how useful information on productivity is passed to other workers, cannot be held constant.

At the societal level, if AI agents can place orders, then the legal infrastructure to ensure that agent is attached to an owner must develop, laws around implicit algorithmic collusion must be created, and so on. We must do this in an environment of uncertainty about what exactly we want AI agents to do, and where externalities or frictions lie. Challenging! But not an impossible - Hadfield and Koh (2025) offer many great ideas. The analogy to the

corporation is useful: when workers combined into large enterprises, we needed new models of careers and new legal institutions. We got them. The question for AI is whether our institutions can both adapt quickly enough and remain flexible enough to allow the technical potential of agentic AI to be fulfilled.

References

Coase, Ronald H. 1960. The Problem of Social Cost. *Journal of Law and Economics*, 3: 1-44.

David, Paul A. 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *American Economic Review*, 80(2): 355-361.

Dell'Acqua, Fabrizio, et al. 2025. Knowledge Worker Productivity and Quality in Generative AI's Jagged Technological Frontier: Field Experimental Evidence. Working paper.

Duflo, Esther, and Rohini Pande. 2007. Dams. *Quarterly Journal of Economics*, 122(2): 601-646.

Gross, Daniel P. 2017. Scale Versus Scope in the Diffusion of New Technology: Evidence from the Farm Tractor. NBER Working Paper No. 24125.

Hadfield, Gillian K., and Andrew Koh. 2025. An Economy of AI Agents. Working paper.

Hartley, Jonathan, Filip Jolevski, Vitor Melo, and Brendan Moore. 2025. The Labor Market Effects of Generative Artificial Intelligence. Working Paper.

Henderson, Rebecca, and Kim B. Clark. 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, 35(1): 9-30.

Henderson, Rebecca, and Robert Gibbons. 2012. Relational Contracts and Organizational Capabilities. *Organization Science*, 23(5): 1350-1364.

Kwa, Thomas, et al. 2025. Measuring AI Ability to Complete Long Tasks. Working Paper.

Macaulay, Stewart. 1963. Non-Contractual Relations in Business: A Preliminary Study. *American Sociological Review*, 28(1): 55-67.

Milgrom, Paul, and John Roberts. 1992. *Economics, Organization and Management*.

Roy, Donald. 1952. Quota Restriction and Goldbricking in a Machine Shop. *American Journal of Sociology*, 57(5): 427-442.

Tomasev, Nenad, et al. 2025. Emergent Agentic Capabilities in AI Systems. Working Paper.