AI governance through AI markets

Joshua Z. Tan

Department of Computer Science
University of Oxford, Oxford, U.K.

Jeffrey Ding

Department of Politics and International Relations
University of Oxford, Oxford, U.K.

August 16, 2019

1 Introduction

Many of our existing templates for AI governance suffer from two critical issues:

1. They lack a clear institutional vehicle to express and enforce the norms and guidelines for safer AI development in the near-term (e.g. proposals for a "CERN" for AI)
2. They either directly conflict with or at least do not benefit the economic interests of the major stakeholders involved (e.g. firms having to disclose trade secrets to keep track of progress toward artificial general intelligence)

Leveraging the emerging market system[1] for AI products and services ("AI market") could address these two issues. Market-based solutions aim to cohere with and shape the economic interests of major stakeholders. Compared to direct regulation, market-based solutions are more adaptable to fast-changing technologies and easier to enforce since the market mechanism enforces them, by design. And even direct regulations should at least be market-aware: the success of many non-market regulatory decisions about the safety and reliability of AI systems, as well as who gets to benefit from the fruits of AI technologies, will be shaped by market forces.

Perhaps most importantly, market integration (for example, via the construction of a global AI market) is one of the few available mechanisms for addressing the dangerous dynamics of an AI race: in other words, a competition over AI between nations.

In Section 2, we consider some plausible examples of future AI markets. In Section 3, we study the problem of predicting how different AI markets will evolve and interact, especially given the possibility of

[1] By *market* or *market system* we mean a system which enables the exchange of goods, services, and information, often but not always via the mechanism of price (e.g. open-source software is also "exchanged" via the market). The total or global AI market includes every transaction involving an AI product, service, technology, or information asset: every robot purchase, every machine learning consultant fee, every IBM Watson maintenance contract, every Deepmind hire. This larger market is organized through a variety of mechanisms: national AI markets, markets for particular goods and services within AI, proprietary platforms (e.g. stock markets, auction houses, or matching services), and AI marketplaces both online and offline. We refer to the total system of funding, production, exchange, and usage of AI products, services, and information as the *AI economy*. The AI economy has certain characteristics—it is expanding, expertise-driven, and extremely well-capitalized—but not all of these are relevant to understanding the AI market.

new market forms enabled by AI technologies, and propose a framework / research program for solving this problem. In Section 4, we discuss the policy implications of AI markets for the AI race, and conclude with some practical recommendations.

2 Three plausible futures

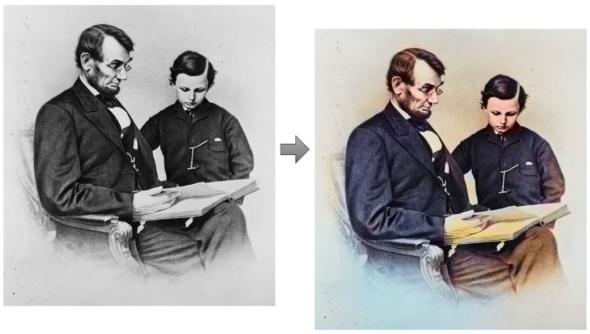
The global AI market in 2018 is a modest extension (around $2 billion) of the overall tech market [6]. There is no specialized market structure for AI products and services, nor is there (yet) a special regulatory framework that divides them out in any country that we know of. And robotics aside, many AI technologies are still so embedded within larger computational systems that it is hard to buy or sell them apart from those systems—witness the proliferation of custom-built "AI layers" within large corporations, or the fact that there is still no major, independent company whose exclusive activity is AI.

What will AI markets (and AI goods and services) look like 10, 20 years from now? Will we be able to govern these future markets, and if so, how? In this section, we set some reference points for discussion: three simple narratives of how AI markets will develop. Each "plausible future" is based on a snippet of cutting-edge AI research or an analogy with an existing technology. These futures will be our first data points.

2.1 Future 1

Algorithmia (www.algorithmia.com) is a new, dedicated marketplace for AI algorithms. In its marketplace, sellers upload code applets that run on servers in the cloud, and buyers call the code via a web API call to the platform; payment is per call or per runtime on the server. Buyers can find sellers on an open market, supplemented by a rating/reputation system. The current top-rated algorithm is "Colorful Image Colorization" (see Figure 1), which takes as input a black-and-white image and outputs a colorized version of the image.

20 years from now, the AI market could look like an extended version of Algorithmia: private researchers and inventors load everything from motion-planning algorithms to game-playing bots to vast general-purpose neural nets into a vast, transparent marketplace. Everyone, from individuals to small companies to large multinationals and governments, can load to the market and everyone can purchase, much as they can buy stocks from the stock market. Technological improvements and supportive regulations lead to a well-functioning free market in what some some say is an industry ripe for market failure.



```
curl -X POST -d '{
    "image": "data://deeplearning/example_data/lincoln.jpg"
}' -H 'Content-Type: application/json' -H 'Authorization: Simple YOUR_API_KEY'
https://api.algorithmia.com/v1/algo/deeplearning/ColorfulImageColorization/1.1.13
```

Figure 1: An example algorithm, "Colorful Image Colorization", hosted on Algorithmia [1] along with its corresponding API call.

Instead of an AI business landscape dominated by a monopoly or oligopoly, there would be freewheeling competition among hundreds of AI startups and established firms. Competition through free market structures drives up efficiency and rewards innovation, which proved critical in AI research and development. This Algorithmia-style future would be similar to a decentralized electricity grid in which the electricity generated by distributed sources (e.g. one household's solar panels) can be sold to others, in contrast to current models of national electricity grids controlled by powerful utilities. In the AI realm, a decentralized cloud would facilitate a distributed market for AI goods and services.

2.2 Future 2

Or will the future look less egalitarian? Suppose that general-purpose AI systems proved too hard to modularize (or at least it was too hard to monetize the pieces) and too dependent on expensive data sets, so only a few mega-corporations now build them. Now Facebook and Google each own their in-house AI system, to which they parcel access. Network externalities, the expense of R&D, and onerous government compliance rules akin to the ones for pharmaceutical drugs mean that these few corporations provide even the most specialized AI products and services—or that those services are not provided at all. In China, Baidu, Tencent, and the Chinese government have created the Chinese Panopticon.

Under this scenario, AI technologies would further entrench the near-monopoly power of technology giants. "Superapps" like Tencent's Wechat, which can collect data on user searches, navigation patterns, reading interests, and payments in a single platform, could swallow up products that provide even the most specialized AI products and services. Additionally, some governments, such as China's, could be more willing to designate certain large companies as national champions in order to better manage a smaller number of outsized AI giants rather than a larger number of freewheeling firms.

2.3 Future 3

Will there even be humans in the future market? Imagine that the AI market becomes fully automated; the only participants are AIs (representing human individuals and corporations), who purchase new algorithms and hire other AI on behalf of their erstwhile masters. (So an "AI market" in both senses of the word!) This third future is more speculative and opens up many avenues for exploration. It is also not mutually exclusive with either of the first two features, so one could imagine a more egalitarian fully automated AI market as well as a more exclusive fully automated AI market.

Together, the three scenarios above make two points.

First, the traditional policy instruments we use to govern markets today—international trade treaties, corporation regulation through reporting & compliance, national and international fiscal & monetary policy—*may not work on the markets of the future.* On the other hand, some new technologies may offer regulators new tools to regulate and even design markets "for the common good", so it may be wise, from an AI governance perspective, to prioritize the development and funding of these technologies. But which technologies? To predict the development of AI markets, we need to build a model that relates market variables (e.g. the relative power of market participants, the cost structure of producing and selling goods, and so on) to technological variables (e.g. prevalence of machine learning techniques, capabilities of mobile and edge computing, progress toward artificial general intelligence, and so on).

Second, this model must go both ways. That is, *market forces help drive (and constrain) technological innovation, but markets also evolve with technological innovation.* Market forces drove the adoption (and to some degree the development) of new AI technologies like deep learning, but new technologies from AI and computer science are already changing the way that traditional markets operate: advances have enabled new kinds of markets (e.g. the Google ad market), code (vs. law) can regulate markets automatically [7], and the price mechanism is being replaced by machine learning and big data [8]. In other words, there is a complicating feedback loop between market mechanisms and technological development. This feedback loop makes it very hard to predict the future of AI markets. But it also exposes new opportunities: if we can understand both directions of the feedback loop, we will not only be able to steer AI markets toward positive technologies, we can also *design* AI markets (and invent new market interventions) by using the right technologies!

Just to refresh, here is the argument so far:

1. AI usage and research is driven by market forces.
2. To govern AI usage and research, we need to explore ways to govern (and even design) AI markets.
3. Traditional policy instruments for market gover-

	Traditional free market	Matching platforms, e.g. dating sites	Data-rich market [8]
Preference information	Price, mainly	Fixed/learned ontology	Price + data flows
Decision-making mechanism	Price mechanism	Matching algorithm	Learned models
Centralized / decentralized	Decentralized	Centralized	Decentralized
Market participants	Anyone	Specialized group	Anyone
Market interventions	Taxation, price ceilings, monetary policy, subsidies and tariffs	Modifying the matching algorithm + higher-level competition between platforms	Dynamic matching across time/individuals, recommendation engines, metadata standards

Table 1: Different market types differ by a number of different market variables.

nance may not work 10, 20 years down the line. We need to predict how the market will evolve.

4. The feedback loop between technology and markets makes prediction really hard... but also really valuable.

How do we make better predictions about AI markets? So far, we have just three data points, created by hand. What we would like is a framework that classifies these three data points, puts them into context, compares them with historical examples from the financial markets or the semiconductor industry, and spits out a quantitative prediction of their relative likelihood. This is a lot to ask for. In the next section, we propose a research program towards building such a framework.

3 Modeling AI markets

For centuries, financial markets have been the prime example of a tech/market feedback loop. For example, [5] reviews how a long series of improvements in communications technology allowed more efficient transactions between markets, thereby increasing market integration across. In turn, more market integration meant even more demand for improvements in such technologies.

What is the relationship between the market variables and the technological variables of the AI economy, and how can we study this relationship within the context of the tech/market loop? Any framework for answering these two questions should also be able to answer at least a subset of the following questions:

1. What is particular about the tech/market feedback loop, compared to other feedback loops in economics [3]?

2. How do different market systems interact, especially at the international level, and under what conditions can we consider them integrated?

3. How do we classify or distinguish different AI markets, e.g. in each of the three plausible futures above in Section 2?

4. How do we compute the relatively probability of a given future AI market?

These questions cannot be answered or even posed with the traditional tools of policy analysis or econometrics; the question is considered qualitatively, when at all. Thus:

Proposal: use open dynamical systems to model the role of technology in the market, and apply this model in order predict the behavior of AI markets.

Background: feedback loops are one of the signature features of a complex system and have been studied extensively in systems theory and complexity science. Often, the more quantitative approaches begin by defining the system in question as some sort of dynamical system. For example, a market can be thought of as a kind of open dynamical system which outputs buy/sell transactions and which takes as input all the economic information available to the market, including previous transactions. Treating it as an *open* dynamical system (with exposed input and output connectors) emphasizes the behavioral and compositional aspects of the system [9]. Compare this to more typical "closed" approaches, common in analyses of financial markets, that emphasize the solution/computation of particular equilibrium states of a dynamical system described by PDEs, for example the Black-Scholes equation for security pricing.

Thinking about the market as a dynamical system (whether open or closed) gives us a convenient way of talking about the *market state*: essentially, a set of parameters that tell you *why* the market outputs so-and-so transactions when fed this-and-that information. Now we can be more precise: by *market variables*, we mean any set of latent variables used to parameterize the market state.

Thinking about the market as an open dynamical system gives us a clean way of distinguishing the market, especially the AI market, from other markets and from the overall economy in which it is embedded. This feature is especially important if we want to consider inter-market interaction and market integration in the context of a hypothetical global AI market.

The main challenge / technical contribution: modeling the behavior of technology. Technology is usually something that shows up directly as properties of goods exchanged on the market—thus part of the information given as input to the market. Consumers use these properties all the time when making purchases (should I buy the iPad, which has optical scan recognition, or the Surface Pro, which doesn't?), as do investors both private and public. We can organize these properties—"dependence on big data", "cloud-based", "required compute", "makes decisions autonomously"—into *technological variables*. Of course, AI governance is also concerned with the technological variables, especially with their social and political externalities.

In a dynamical systems setting, the tech/market feedback loop is modeled by the fact that technological variables are shared between the overall economy and the market state. The precise nature of this "sharing", and its implications for how technology affects the market, is an open question. It is not (just) the usual kind of input-output composition / variable-sharing in open dynamical systems, since major changes in technology often lead not only to discrete, measurable changes in the market state (like improvements in market efficiency, lower transaction costs) but *entirely different ways of parameterizing the market state*, i.e. new market forms.

4 Macroeconomics and the AI race

How will the future AI market affect the dynamics of an AI race between nations? The particular characteristics of this future AI marketplace could significantly influence whether the key stakeholders sacrifice safety procedures in the pursuit of transformative AI as well as the degree to which benefits of AI are shared by the majority of people to benefit humanity.

The previous section outlined a proposal to investigate the feedback loop between technological variables and the evolution of an AI market. This section will outline some possible solutions and interventions in the form of market mechanisms that could address the pitfalls of an AI race.

AI race questions are often understood along a range of typologies, ranging from narrow to broad. A narrow AI race could occur in specific AI-enabled applications, such as autonomous weapons. A broad AI race could develop in pursuit of Artificial General Intelligence (AGI). Most framings of these AI race questions assume that the central actors are governments seeking a decisive strategic advantage.[2] While governments may exercise control over the develop-

[2] Another popular framing is that of labs racing for AI dominance in a vacuum. See [2].

ment of specific AI-enabled applications in the security/defense domain, it is less clear that they will be the principal actors when it comes to developing AI as a general-purpose technology, one that will permeate through global markets. Across a variety of indicators—including amount of AI investment, incentives for top AI talent, quantity and quality of AI research—private firms are racing ahead of governments in developing AI. Structured around these major firms as the dominant actors in an economic race for AI, we analyze AI race problems through the framework of how private firms interact in a market.

In the context of this future AI market, three variable features have significant implications for AI race dynamics: the effect of modularization on the degree of interdependence, the level of inequality in market shares, and the extent to which technical standards are harmonized. It is important to note that these three variables interact with each other (e.g. more compatibility in technical standards adhered to by firms would increase the degree of interdependence). Throughout this section, the analogy of the effect of Internet technologies on global financial markets is employed. Economides [4] articulated an account for how the Internet revolution affected interdependence, inequality, and standardization in global financial markets. The Internet ushered in a more financially integrated globe. By enabling digital processes, goods, and interaction, the Internet technologies lowered transaction costs to trading goods and services across borders. Additionally, the Internet played a significant role in reducing price discrimination based on national borders, resulting in product standardization and an increase in liquidity. Lastly, perhaps most obvious to the everyday user, the Internet cultivated a "network effects" economy with "winner-take-all" dynamics, leading to the rise of technology giants like Facebook.

Will the development of an ecosystem of AI products and services lead to similar market dynamics? Recall the three plausible futures outlined early: 1) a freewheeling, decentralized marketplace, 2) an inegalitarian AI market dominated by a few large firms divided by national lines, 3) an automated AI market. There are potential pitfalls with each of these three scenarios. The first future could introduce too many players into the AI race, making it hard to regulate. In the case of the second scenario, the existence of "too-big-to-fail" AI firms, which govern all the crucial levers of an AI market (data, hardware, key algorithms, etc.) would reduce the incentive of those firms to invest in AI safety public interest, as well as give malicious actors a single, or limited range, of targets. Moreover, relatively closed AI markets, shaped by industrial policy, would nationalize AI competition and increase racing risks. Finally, it is necessary to acknowledge the third scenario as a possibility given the rapid advances in AI, and under this scenario, it is very difficult for markets to catch up to the technology. In contrast, there is a way to shape these three scenarios toward an AI market that is imbued with social values oriented toward making sure transformative AI is beneficial and safe for humanity.

This shift requires three key aspects of the AI market to be realized: a) a level of integration that breeds interdependence among key actors, b) an increase in the number of key actors, and c) harmonization in the technical standards that govern the market. Importantly, it could be the case that there are levers embedded within an AI market that could play an important role in the solution. For instance, increased modularization of AI services could improve transparency in larger AI systems. When system architectures are composed of modular parts, instead of opaque and undifferentiated unitary systems, they can be more easily tested and assessed along a variety of safety measures, including reliability. At the same time, as narrow AI products and services specialize in certain domains, enhancing interoperability among these narrow applications will enable AI services to interact with each other. A smart home robot, for instance, will call upon AI services in the voice recognition, navigation, and task prioritization domains. As was the case with growing modularization, improved interoperability requires each narrow AI service to meet a certain standard of safety and reliability.

If self-governance does not succeed, policy levers are available in each of the three crucial areas. Pushing the AI market towards global integration, a trend that is already emerging, is essential. One such lever that could facilitate this trend is increasing the flow

of data across borders. Additionally, there are many options to increase the number of key actors. Antitrust regulation is the most blunt of instruments, but other vehicles abound. Lastly, in other domains, international standardization has increased the degree of interoperability between products, thereby cultivating greater interdependency and integration between firms. Since standards combine aspects of both public and private goods, they could play an important role in guiding the market away from destructive race dynamics and toward serving the public good.

An integrated, standardized, and egalitarian AI market is not a panacea and also brings risks. For example, while the global financial market promotes interdependence and openness between competing states, it also introduced significant structural instability into the global economy: witness the 1997 Asian financial crisis, the global financial crisis of 2007, and the European debt crisis. A similar danger risk could arise from our model of an AI market designed in the public interest, as automation and the speed of decision-making could spiral out of control.

5 Conclusions

It appears that research on AI risks is fairly divided between short-term and long-term risks of AI development. Researchers who emphasize short-term concerns may miss the forest for the trees, but those who only look at the far-off future may miss out on how that future is rapidly being shaped in the near term. A focus on governing AI through the emerging AI market connects the two sides of the debate. It is important to understand the key microeconomic and macroeconomic features of this AI market, including the degree of interdependence, the number of key actors, and the level of standardization. It is possible to channel competition between racing AI firms toward a globally interdependent, more egalitarian, and standardized AI market, which mitigates safety risks and AI arms races. Now that would be an amazing race.

References

[1] Colorful image colorization. Last accessed on April 23, 2018. https://algorithmia.com/algorithms/deeplearning/ColorfulImageColorization.

[2] S. Armstrong, N. Bostrom, and C. Shulman. Racing to the precipice: a model of artificial intelligence development. Technical report, Future of Humanity Institute, University of Oxford, 2013.

[3] W. B. Arthur. *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press, Oct. 1994. Google-Books-ID: nkc_DwAAQBAJ.

[4] N. Economides. The impact of the internet on financial markets. *Journal of Financial Transformation*, 1:8–13, 2001.

[5] K. D. Garbade and W. L. Silber. Technology, communication and the performance of financial markets: 1840-1975. *The Journal of Finance*, 33(3), 1978.

[6] C. W. Kaul, Aditya. Artificial Intelligence Market Forecasts. Technical report, Tractica, 2016.

[7] L. Lessig. *Code: Version 2.0*. Basic Books, second edition, 2006.

[8] V. Mayer-Schonberger and T. Ramge. *Reinventing Capitalism in the Age of Big Data*. John Murray, 2018.

[9] J. Willems. The Behavioral Approach to Open and Interconnected Systems. *IEEE Control Systems Magazine*, 27(6):46–99, Dec. 2007.