NBER WORKING PAPER SERIES

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON INNOVATION

Iain M. Cockburn
Rebecca Henderson
Scott Stern

Working Paper 24449
http://www.nber.org/papers/w24449

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018

The authors would like to thank the organizers and participants at the first NBER conference on the Economics of Artificial Intelligence, and in particular our discussant Matthew Mitchell for many helpful suggestions and ideas. Michael Kearney provided extraordinary research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Funding for this paper was provided by the MIT Sloan School of Management, by the HBS Division of Research and by the Questrom School of Management.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24449.ack

The Impact of Artificial Intelligence on Innovation
Iain M. Cockburn, Rebecca Henderson, and Scott Stern
NBER Working Paper No. 24449
March 2018
JEL No. L1

ABSTRACT

Artificial intelligence may greatly increase the efficiency of the existing economy. But it may have an even larger impact by serving as a new general-purpose "method of invention" that can reshape the nature of the innovation process and the organization of R&D. We distinguish between automation-oriented applications such as robotics and the potential for recent developments in "deep learning" to serve as a general-purpose method of invention, finding strong evidence of a "shift" in the importance of application-oriented learning research since 2009. We suggest that this is likely to lead to a significant substitution away from more routinized labor-intensive research towards research that takes advantage of the interplay between passively generated large datasets and enhanced prediction algorithms. At the same time, the potential commercial rewards from mastering this mode of research are likely to usher in a period of racing, driven by powerful incentives for individual companies to acquire and control critical large datasets and application-specific algorithms. We suggest that policies which encourage transparency and sharing of core datasets across both public and private actors may be critical tools for stimulating research productivity and innovation-oriented competition going forward.

Iain M. Cockburn
School of Management
Boston University
595 Commonwealth Ave
Boston, MA 02215
and NBER
cockburn@bu.edu

Scott Stern
MIT Sloan School of Management
100 Main Street, E62-476
Cambridge, MA 02142
and NBER
sstern@mit.edu

Rebecca Henderson
Heinz Professor of Environmental Management
Harvard Business School
Morgan 445
Soldiers Field
Boston, MA 02163
and NBER
rhenderson@hbs.edu

I. Introduction

Rapid advances in the field of artificial intelligence have profound implications for the economy as well as society at large. These innovations have the potential to directly influence both the production and the characteristics of a wide range of products and services, with important implications for productivity, employment, and competition. But, as important as these effects are likely to be, artificial intelligence also has the potential to change the innovation process itself, with consequences that may be equally profound, and which may, over time, come to dominate the direct effect.

Consider the case of Atomwise, a startup firm which is developing novel technology for identifying potential drug candidates (and insecticides) by using neural networks to predict the bioactivity of candidate molecules. The company reports that its deep convolutional neural networks "far surpass" the performance of conventional "docking" algorithms. After appropriate training on vast quantities of data, the company's AtomNet product is described as being able to "recognize" foundational building blocks of organic chemistry, and is capable of generating highly accurate predictions of the outcomes of real-world physical experiments (Wallach et al., 2015). Such breakthroughs hold out the prospect of substantial improvements in the productivity of early stage drug screening. Of course, Atomwise's technology (and that of other companies leveraging artificial intelligence to advance drug discovery or medical diagnosis) is still at an early stage: though their initial results seem to be promising, no new drugs have actually come to market using these new approaches. But whether or not Atomwise delivers fully on its promise, its technology is representative of the ongoing attempt to develop a new innovation "playbook", one that leverages large datasets and learning algorithms to engage in precise prediction of biological phenomena in order to guide design effective interventions. Atomwise, for example, is now deploying this approach to the discovery and development of new pesticides and agents for controlling crop diseases.

Atomwise's example illustrates two of the ways in which advances in artificial intelligence have the potential to impact innovation. First, though the origins of artificial intelligence are broadly in the field of computer science, and its early commercial applications have been in relatively narrow domains such as robotics, the learning algorithms that are now being developed

suggest that artificial intelligence may ultimately have applications across a very wide range. From the perspective of the economics of innovation (among others, Bresnahan and Trajtenberg (1995)), there is an important distinction between the problem of providing innovation incentives to develop technologies with a relatively narrow domain of application, such robots purpose-built for narrow tasks, versus technologies with a wide—advocates might say almost limitless—domain of application, as may be true of the advances in neural networks and machine learning often referred to as "deep learning." As such, a first question to be asked is the degree to which developments in artificial intelligence are not simply examples of new technologies, but rather may be the kinds of "general purpose technologies" (hereafter GPTs) that have historically been such influential drivers of long-term technological progress.

Second, while some applications of artificial intelligence will surely constitute lower-cost or higher-quality inputs into many existing production processes (spurring concerns about the potential for large job displacements), others, such as deep learning, hold out the prospect of not only productivity gains across a wide variety of sectors but also changes in the very nature of the innovation process within those domains. As articulated famously by Griliches (1957), by enabling innovation across many applications, the "invention of a method of invention" has the potential to have much larger economic impact than development of any single new product. Here we argue that recent advances in machine learning and neural networks, through their ability to improve both the performance of end use technologies and the nature of the innovation process, are likely to have a particularly large impact on innovation and growth. Thus the incentives and obstacles that may shape the development and diffusion of these technologies are an important topic for economic research, and building an understanding of the conditions under which different potential innovators are able to gain access to these tools and to use them in a pro-competitive way is a central concern for policy.

This essay begins to unpack the potential impact of advances in artificial intelligence on innovation, and to identify the role that policy and institutions might play in providing effective incentives for innovation, diffusion, and competition in this area. We begin in Section II by highlighting the distinctive economics of research tools, of which deep learning applied to R&D problems is such an intriguing example. We focus on the interplay between the degree of generality of application of a new research tool and the role of research tools not simply in

enhancing the efficiency of research activity but in creating a new "playbook" for innovation itself. We then turn in Section III to briefly contrasting three key technological trajectories within AI—robotics, symbolic systems, and deep learning. We propose that these often conflated fields will likely play very different roles in the future of innovation and technical change. Work in symbolic systems appears to have stalled and is likely to have relatively little impact going forwards. And while developments in robotics have the potential to further displace human labor in the production of many goods and services, innovation in robotics technologies per se has relatively low potential to change the nature of innovation itself. By contrast, deep learning seems to be an area of research that is highly general-purpose and that has the potential to change the innovation process itself.

We explore whether this might indeed be the case through an examination of some quantitative empirical evidence on the evolution of different areas artificial intelligence in terms of scientific and technical outputs of AI researchers as measured (imperfectly) by the publication of papers and patents from 1990 through 2015. In particular, we develop what we believe is the first systematic database that captures the corpus of scientific paper and patenting activity in artificial intelligence, broadly defined, and divides these outputs into those associated with robotics, symbolic systems, and deep learning. Though preliminary in nature (and inherently imperfect given that key elements of research activity in artificial intelligence may not be observable using these traditional innovation metrics), we find striking evidence for a rapid and meaningful shift in the application orientation of learning-oriented publications, particularly after 2009. The timing of this shift is informative, since it accords with qualitative evidence about the surprisingly strong performance of so-called "deep learning" multi-layered neural networks in a range of tasks including computer vision and other prediction tasks. Supplementary evidence (not reported here) based on the citation patterns to authors such as Geoffrey Hinton who are leading figures in deep learning suggests a striking acceleration of work in just the last few years that builds on a small number of algorithmic breakthroughs related to multi-layered neural networks.

Though not a central aspect of the analysis for this paper, we further find that, whereas research on learning-oriented algorithms has had a slow and steady upward swing outside of the

United States, US researchers have had a less sustained commitment to learning-oriented research prior to 2009, and have been in a "catch up" mode ever since.

Finally, we begin to explore some of the organizational, institutional and policy consequences of our analysis. We see machine learning as the "invention of a method of invention" whose application depends, in each case, on having access not just to the underlying algorithms but also to large, granular datasets on physical and social behavior. Developments in neural networks and machine learning thus raise the question of, even if the underlying scientific approaches (i.e., the basic multi-layered neural networks algorithms) are open, prospects for continued progress in this field—and commercial applications thereof—are likely to be significantly impacted by terms of access to complementary data. Specifically, if there are increasing returns to scale or scope in data acquisition (there is more learning to be had from the "larger" dataset), it is possible that early or aggressive entrants into a particular application area may be able to create a substantial and long-lasting competitive advantage over potential rivals merely through the control over data rather than through formal intellectual property or demand-side network effects. Strong incentives to maintain data privately has the additional potential downside that data is not being shared across researchers, thus reducing the ability of all researchers to access an even larger set of data that would arise from public aggregation. As the competitive advantage of incumbents is reinforced, the power of new entrants to drive technological change may be weakened. Though this is an important possibility, it is also the case that, at least so far, there seems to be a significant amount of entry and experimentation across most key application sectors.

II. The Economics of New Research Tools: The Interplay between New Methods of Invention and the Generality of Innovation

At least since Arrow (1962) and Nelson (1959), economists have appreciated the potential for significant underinvestment in research, particularly basic research or domains of invention with low appropriability for the inventor. Considerable insight has been gained into the conditions under which the incentives for innovation may be more or less distorted, both in terms of their overall level and in terms of the direction of that research. As we consider the potential impact of advances in AI on innovation, two ideas from this literature seem particularly

important—the potential for contracting problems associated with the development of a new broadly applicable research tool, and the potential for coordination problems arising from adoption and diffusion of a new "general purpose technology." In contrast to technological progress in relatively narrow domains, such as traditional automation and industrial robots, we argue that those areas of artificial intelligence evolving most rapidly—such as deep learning—are likely to raise serious challenges in both dimensions.

First, consider the challenge in providing appropriate innovation incentives when an innovation has potential to drive technological and organizational change across a wide number of distinct applications. Such "general purpose technologies" (David, 1990; Bresnahan and Trajtenberg, 1995) often take the form of core inventions that have the potential to significantly enhance productivity or quality across a wide number of fields or sectors. David's (1990) foundational study of the electric motor showed that this invention brought about enormous technological and organizational change across sectors as diverse as manufacturing, agriculture, retail, and residential construction. Such "GPTs" are usually understood to meet three criteria that distinguish them from other innovations: they have pervasive application across many sectors; they spawn further innovation in application sectors, and they themselves are rapidly improving.

As emphasized by Bresnahan and Trajtenberg (1995), the presence of a general-purpose technology gives rise to both vertical and horizontal externalities in the innovation process that can lead not just to underinvestment but also to distortions in the direction of investment, depending on the degree to which private and social returns diverge across different application sectors. Most notably, if there are "innovation complementarities" between the general purpose technology and each of the application sectors, lack of incentives in one sector can create an indirect externality that results in a system-wide reduction in innovative investment itself. While the private incentives for innovative investment in each application sector depend on its the market structure and appropriability conditions, that sector's innovation enhances innovation in the GPT itself, which then induces subsequent demand (and further innovation) in other downstream application sectors. These gains can rarely be appropriated within the originating sector. Lack of coordination between the GPT and application sectors, as well as across application sectors, is therefore likely to significantly reduce investment in innovation. Despite

these challenges, a reinforcing cycle of innovation between the GPT and a myriad of application sectors can generate a more systemic economy-wide transformation as the rate of innovation increases across all sectors. A rich empirical literature examining the productivity impacts of information technology point to the role of the microprocessor as a GPT as a way of understanding the impact of IT on the economy as a whole (among many others, Bresnahan and Greenstein (1995); Brynjolfsson and Hitt (1999); and Bresnahan, Brynjolfsson, and Hitt (2001)). Various aspects of artificial intelligence can certainly be understood as a GPT, and learning from examples such as the microprocessor are likely to be a useful foundation for thinking about both the magnitude of their impact on the economy, and associated policy challenges.

A second conceptual framework for thinking about AI is the economics of research tools. Within the research sectors, some innovations open up new avenues of inquiry, or simply improve productivity "within the lab". Some of these advances appear to have great potential across a broad set of domains, beyond their initial application: as highlighted by Griliches (1957) in his classic studies of hybrid corn, some new research tools are inventions that do not just create or improve a specific product—instead they constitute a new way of creating new products, with much broader application. In Griliches' famous construction, the discovery of double-cross hybridization "was the invention of a method of inventing." (Hereinafter, "IMI".) Rather than being a means of creating a single a new corn variety, hybrid corn represented a widely applicable method for breeding many different new varieties. When applied to the challenge of creating new varieties optimized for many different localities (and even more broadly, to other crops) the invention of double-cross hybridization had a huge impact on agricultural productivity.

One of the important insights to be gained from thinking about IMIs, therefore, is that the economic impact of some types of research tools is not limited to their ability to reduce the costs of specific innovation activities—perhaps even more consequentially they enable a new approach to innovation itself, by altering the "playbook" for innovation in the domains where the new tool is applied. For example, prior to the systematic understanding of the power of "hybrid vigor," a primary focus in agriculture had been improved techniques for self-fertilization (i.e., allowing for more and more specialized natural varietals over time). Once the rules governing hybridization (i.e., heterosis) were systematized, and the performance advantages of hybrid vigor

demonstrated, the techniques and conceptual approach for agricultural innovation was shifted, ushering in a long period of systematic innovation using these new tools and knowledge.

Advances in machine learning and neural networks appear to have great potential as a research tool in problems of classification and prediction. These are both important limiting factors in a variety of research tasks, and, as exemplified by the Atomwise example, application of "learning" approaches to AI hold out the prospect of dramatically lower costs and improved performance in R&D projects where these are significant challenges. But as with hybrid corn, AI based learning may be more usefully understood as an IMI than as a narrowly limited solution to a specific problem. One the one hand, AI based learning may be able to substantially "automate discovery" across many domains where classification and prediction tasks play an important role. On the other, they may also "expand the playbook" is the sense of opening up the set of problems that can be feasibly addressed, and radically altering scientific and technical communities' conceptual approaches and framing of problems. The invention of optical lenses in the 17th century had important direct economic impact in applications such as spectacles. But optical lenses in the form of microscopes and telescopes also had enormous and long-lasting indirect effects on the progress of science, technological change, growth, and welfare: by making very small or very distant objects visible for the first time, lenses opened up entirely new domains of inquiry and technological opportunity. Leung et al. (2016), for example, evocatively characterize machine learning as an opportunity to "learn to read the genome" in ways that human cognition and perception cannot.

Of course, many research tools are neither IMIs nor GPTs, and their primary impact is to reduce the cost or enhance the quality of an existing innovation process. For example, in the pharmaceutical industry, new kinds of materials promise to enhance the efficiency of specific research processes. Other research tools can indeed be thought of as IMIs but are nonetheless relatively limited in application. For example, the development of genetically engineered research mice (such as the Oncomouse) is an IMI that has had a profound impact on the conduct and "playbook" of biomedical research, but has no obvious relevance to innovation in areas such as information technology, energy, or aerospace. The challenge presented by advances in AI is that they appear to be research tools that not only have the potential to change the method of innovation itself but also have implications across an extraordinarily wide range of fields.

Historically technologies with these characteristics—think of digital computing—have had large and unanticipated impacts across the economy and society in general. Mokyr (2002) points to the profound impact of IMIs that take the form not of tools per se, but innovations in the way research is organized and conducted, such as the invention of the university. GPTs that are themselves IMIs (or vice versa) are particularly complex phenomena, whose dynamics are as yet poorly understood or characterized.

From a policy perspective, a further important feature of research tools is that it may be particularly difficult to appropriate their benefits. As emphasized by Scotchmer (1990), providing appropriate incentives for an upstream innovator that develops only the first "stage" of an innovation (such as a research tool) can be particularly problematic when contracting is imperfect and the ultimate application of the new products whose development is enabled by the upstream innovation is uncertain. Scotchmer and her co-authors emphasized a key point about a multi-stage research process: when the ultimate innovation that creates value requires multiple steps, providing appropriate innovation incentives are not only a question of whether and how to provide property rights in general, but also of how best to distribute property rights and incentives across the multiple stages of the innovation process. Lack of incentives for early-stage innovation can therefore mean that the tools required for subsequent innovation do not even get invented; strong early-stage property rights without adequate contracting opportunities may result in "hold-up" for later-stage innovators and so reduce the ultimate impact of the tool in terms of commercial application.

The vertical research spillovers created by new research tools (or IMIs) are not just a challenge for designing appropriate intellectual property policy.[1] They are also exemplars of the core innovation externality highlighted by endogenous growth theory (Romer, 1990; Aghion and Howitt, 1992); a central source of underinvestment in innovation is the fact that the intertemporal spillovers from innovators today to innovators tomorrow cannot be easily captured. While tomorrow's innovators benefit from "standing on the shoulders of giants," their gains are not easily shared with their predecessors. This is not simply a theoretical idea: an increasing body of evidence suggests that research tools and the institutions that support their development and

[1] Challenges presented by AI-enabled invention for legal doctrine and the patent process are beyond the scope of this essay.

diffusion play an important role in generating intertemporal spillovers (among others, Furman and Stern, 2011; Williams, 2014). A central insight of this work is that control—both in the form of physical exclusivity as well as in the form of formal intellectual property rights—over tools and data can shape both the level and direction of innovative activity, and that rules and institutions governing control over these areas has a powerful influence on the realized amount and nature of innovation.

Of course, these frameworks cover only a subset of the key informational and competitive distortions that might arise when considering whether and how to provide optimal incentives for the type of technological change represented by some areas of AI. But these two areas in particular seem likely to be important for understanding the implications of the current dramatic advances in AI supported learning. We therefore turn in the next section to a brief outline of the ways in which AI is changing, with an eye towards bringing the framework here to bear on how we might outline a research agenda exploring the innovation policy challenges that they create.

III. The Evolution of Artificial Intelligence: Robotics, Symbolic Systems, and Neural Networks

In his omnibus historical account of AI research, Nilsson (2010) defines AI as "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment." His account details the contributions of multiple fields to achievements in AI, including but not limited to biology, linguistics, psychology and cognitive sciences, neuroscience, mathematics, philosophy and logic, engineering and computer science. And, of course, regardless of their particular approach, artificial intelligence research has been united by from the beginning by its engagement with Turing (1950), and his discussion of the possibility of mechanizing intelligence.

Though often grouped together, the intellectual history of AI as a scientific and technical field is usefully informed by distinguishing between three interrelated but separate areas: robotics, neural networks, and symbolic systems. Perhaps the most successful line of research in the early years of AI—dating back to the 1960s—falls under the broad heading of symbolic

systems. Although early pioneers such as Turing had emphasized the importance of teaching a machine as one might a child (i.e., emphasizing AI as a learning process), the "symbol processing hypothesis" (Newell, Shaw, and Simon, 1958; Newell and Simon, 1976) was premised on the attempt to replicate the logical flow of human decision making through processing symbols. Early attempts to instantiate this approach yielded striking success in demonstration projects, such as the ability of a computer to navigate elements of a chess game (or other board games) or engage in relatively simple conversations with humans by following specific heuristics and rules embedded into a program. However, while research based on the concept of a "general problem solver" has continued to be an area of significant academic interest, and there have been periodic explosions of interest in the use of such approaches to assist human decision-making (e.g., in the context of early-stage expert systems to guide medical diagnosis), the symbolic systems approach has been heavily criticized for its inability to meaningfully impact real-world processes in a scalable way. It is of course possible that this field will see breakthroughs in the future, but it is fair to say that, while symbolic systems continues to be an area of academic research, it has not been central to the commercial application of AI. Nor is it at the heart of the recent reported advances in AI that are associated with the area of machine learning and prediction.

A second influential trajectory in AI has been broadly in the area of robotics. While the concepts of "robots" as machines that can perform human tasks dates back at least to the 1940s, the field of robotics began to meaningfully flourish from the 1980s onwards through a combination of the advances in numerically controlled machine tools and the development of more adaptive but still rules-based robotics that rely on the active sensing of a known environment. Perhaps the most economically consequential application of AI to date has been in this area, with large scale deployment of "industrial robots" in manufacturing applications. These machines are precisely programmed to undertake a given task in a highly controlled environment. Often located in "cages" within highly specialized industrial processes (most notably automobile manufacturing), these purpose-built tools are perhaps more aptly described as highly sophisticated numerically controlled machines rather than as robots with significant AI content. Over the past twenty years, innovation in robotics has had an important impact on manufacturing and automation, most notably through the introduction of more responsive robots that rely on programmed response algorithms that can respond to a variety of stimuli. This

approach, famously pioneered by Rod Brooks (1990), focused the commercial and innovation orientation of AI away from the modeling of human-like intelligence towards providing feedback mechanisms that would allow for practical and effective robotics for specified applications. This insight led, among other applications, to the Roomba and to other adaptable industrial robots that could interact with humans such as Rethink Robotics' Baxter). Continued innovation in robotics technologies (particularly in the ability of robotic devices to sense and interact with their environment) may lead to wider application and adoption outside industrial automation.

These advances are important, and the most advanced robots continue to capture public imagination when the term AI is invoked. But innovations in robotics are not, generally speaking, IMIs. The increasing automation of laboratory equipment certainly improves research productivity, but advances in robotics are not (yet) centrally connected to the underlying ways in which researchers themselves might develop approaches to undertake innovation itself across multiple domains. There are of course counterexamples to this proposition: robotic space probes have been a very important research tool in planetary science, and the ability of automated remote sensing devices to collect data at very large scale or in challenging environments may transform some fields of research. But robots continue to be used principally in specialized end-use "production" applications.

Finally, a third stream of research that has been a central element of AI since its founding can be broadly characterized as a "learning" approach. Rather than being focused on symbolic logic, or precise sense-and-react systems, the learning approach attempts to create reliable and accurate methods for the prediction of particular events (either physical or logical) in the presence of particular inputs. The concept of a neural network has been particularly important in this area. A neural network is a program that uses a combination of weights and thresholds to translate a set of inputs into a set of outputs, measures the "closeness" of these outputs to reality, and then adjusts the weights it uses to narrow the distance between outputs and reality. In this way, neural networks can learn as they are fed more inputs (Rosenblatt, 1958; 1963). Over the course of the 1980s, Hinton and his co-authors further advanced the conceptual framework on which neural networks are based through the development of "back-propagating multi-layer" techniques that further enhance their potential for supervised learning.

After being initially heralded as having significant promise, the field of neural networks has come in and out of fashion, particularly within the United States. From the 1980s through the mid-2000s, their challenge seemed to be that there were significant limitations to the technology that could not be easily fixed by using larger training datasets or through the introduction of additional layers of "neurons." However, in the mid-2000s, a small number of new algorithmic approaches demonstrated the potential to enhance prediction through back propagation through multiple layers. These neural networks increased their predictive power as they were applied to larger and larger datasets, and were able to scale to an arbitrary level (among others, a key reference here is Hinton and Salakhutdinov (2006)). These advances exhibited a "surprising" level of performance improvement, notably in the context of the ImageNet visual recognition project competition pioneered by Fei-Fei Li at Stanford (Krizhevsky, Sutskever and Hinton, 2012).

IV. How Might Different Fields within Artificial Intelligence Impact Innovation?

Distinguishing between these three streams of AI is a critical first step towards developing a better understanding of how AI is likely to influence the innovation process going forward, since the three differ significantly in their potential to be either GPTs or IMIs—or both.

First, though a significant amount of public discussion of AI focuses on the potential for AI to achieve super-human performance over a wide range of human cognitive capabilities, it is important to note that, at least so far, the significant advances in AI have not been in the form of the "general problem solver" approaches that were at the core of early work in symbolic systems (and that were the motivation for considerations of human reasoning such as the Turing test). Instead, recent advances in both robotics and in deep learning are by and large innovations that require a significant level of human planning and that apply to a relatively narrow domain of problem-solving (e.g., face recognition, playing Go, picking up a particular object, etc.) While it is of course possible that further breakthroughs will lead to a technology that can meaningfully mimic the nature of human subjective intelligence and emotion, the recent advances that have attracted scientific and commercial attention are well removed from these domains.

Second, though most economic and policy analysis of AI draws out consequences from the last two decades of automation to consider the future economic impact of AI (e.g., in job displacement for an ever-increasing number of tasks), it is important to emphasize that there is a sharp difference between the advances in robotics that were a primary focus of applications of AI research during the 2000s and the potential applications of deep learning which have come to the fore over the last few years.

As we suggested above, current advances in robotics are by and large associated with applications that are highly specialized and that are focused on end-user applications rather than on the innovation process itself and these advances do not seem as of yet to have translated to a more generally applicable IMI. Robotics is therefore an area where we might focus on the impact of innovation (improved performance) and diffusion (more widespread application) in terms of job displacement versus job enhancement. We see limited evidence as yet of widespread applications of robotics outside industrial automation, or of the scale of improvements in the ability to sense, react to, and manipulate the physically environment that the use of robotics outside manufacturing probably requires. But there are exceptions: developments in the capabilities of "pick and place" robots and rapid progress in autonomous vehicles point to the possibility for robotics to escape manufacturing and become much more broadly used. Advances in robotics may well reveal this area of AI be a GPT, as defined by the classic criteria.

Some research tools/IMIs based on algorithms have transformed the nature of research in some fields, but have lacked generality. These types of algorithmic research tools, based on a static set of program instructions, are a valuable IMI, but do not appear to have wide applicability outside a specific domain and do not qualify as GPTs. For example, while far from perfect, powerful algorithms to scan brain images (so-called functional MRI imaging) have transformed our understanding of the human brain, not only through the knowledge they have generated but also by establishing an entirely new paradigm and protocol for brain research. However, despite its role as a powerful IMI, fMRI lacks the type of general-purpose applicability that has been associated with the most important GPTs. In contrast, the latest advances in deep learning have the potential to be both a general-purpose IMI and a classic GPT.

The following table summarizes these ideas:

		General-Purpose Technology	
		NO	YES
Invention of a Method of Invention	NO	Industrial Robots (e.g. Fanuc R2000)	'Sense & React" Robots (e.g. Autonomous vehicles)
	YES	Statically-coded Algorithmic Tools (e.g. fMRI)	**Deep Learning**

How might the promise of deep learning as a general-purpose IMI be realized? Deep learning promises to be an enormously powerful new tool that allows for the unstructured "prediction" of physical or logical events in contexts where algorithms based on a static set of program instructions (such as classic statistical methods) perform poorly. The development of this new approach to prediction enables a new approach to undertaking scientific and technical research. Rather than focusing on small well-characterized datasets or testing settings, it is now possible to proceed by identifying large pools of unstructured data which can be used to dynamically develop highly accurate predictions of technical and behavioral phenomena. In pioneering an unstructured approach to predictive drug candidate selection that brings together a vast array of previously disparate clinical and biophysical data, for example, Atomwise may fundamentally reshape the "ideas production function" in drug discovery.

If advances in deep learning do represent the arrival of a general-purpose IMI, it is clear that there are likely to be very significant long-run economic, social, and technological consequence. First, as this new IMI diffuses across many application sectors, the resulting explosion in technological opportunities and increased productivity of R&D seem likely to generate economic growth that can eclipse any near-term impact of AI on jobs, organizations, and productivity. A more subtle implication of this point is that "past is not prologue": even if automation over the recent past has resulted in job displacement (e.g., Acemoglu and Restrepo, 2017a), AI is likely to have at least as important an impact through its ability to enhance the potential for "new tasks" (as in Acemoglu and Restrepo, 2017b).

Second, the arrival of a general-purpose IMI is a sufficiently uncommon occurrence that its impact could be profound for economic growth and its broader impact on society. There have

been only a handful of previous general-purpose IMIs and each of these has had an enormous impact not primarily through their direct effects (e.g., spectacles, in the case of the invention of optical lenses) but through their ability to reshape the ideas production function itself (e.g. telescopes and microscopes). It would therefore be helpful to understand the extent to which deep learning is, or will, causing researchers to significantly shift or reorient their approach in order to enhance research productivity (in the spirit of Jones (2009)).

Finally, if deep learning does indeed prove to be a general-purpose IMI, it will be important to develop institutions and a policy environment that is conductive to enhancing innovation through this approach, and to do so in a way that promotes competition and social welfare. A central concern here may be the interplay between a key input required for deep learning—large unstructured databases that provide information about physical or logical events—and the nature of competition. While the underlying algorithms for deep learning are in the public domain (and can and are being improved on rapidly), the data pools that are essential to generate predictions may be public or private, and access to them will depend on organizational boundaries, policy and institutions. Because the performance of deep learning algorithms depends critically on the training data that they are created from, it may be possible, in a particular application area, for a specific company (either an incumbent or start-up) gain a significant, persistent innovation advantage through their control over data that is independent of traditional economies of scale or demand-side network effects. This "competition for the market" is likely to have several consequences. First, it creates incentives for duplicative racing to establish a data advantage in particular application sectors (say, search, autonomous driving, or cytology) followed by the establishment of durable barriers to entry that may be of significant concern for competition policy. Perhaps even more importantly, this kind of behavior could result in a balkanization of data within each sector, not only reducing innovative productivity within the sector, but also reducing spillovers back to the deep learning GPT sector, and to other application sectors. This suggests that the proactive development of institutions and policies that encourage competition, data sharing, and openness is likely to be an important determinant of economic gains from the development and application of deep learning.

Our discussion so far has been largely speculative, and it would be useful to know whether our claim that deep learning may be both a general-purpose IMI and a GPT, while

symbolic logic and robotics are probably not, have any empirical basis. We turn in the next section to a preliminary examination of the evolution of AI as revealed by bibliometric data, with an eye towards answering this question.

V. Data

This analysis draws upon two distinct datasets, one that captures a set of AI publications from Thompson Reuters Web of Science, and another that identifies a set of AI patents issued by the U.S. Patent and Trademark Office. In this section, we provide detail on the assembly of these datasets and summary statistics for variables in the sample.

. As previously discussed, peer-reviewed and public-domain literature on AI points to the existence of three distinct fields within AI: robotics, learning systems and symbol systems, each comprised of numerous subfields. To track development of each of these using this data, we began by identifying the publications and patents falling into each of these three fields based on keywords. Appendix 1 lists the terms we used to define each field and identify the papers and patents belonging to it. .[2] In short, the robotics field includes approaches in which a system engages with and responds to environmental conditions; the symbolic systems field attempts to represent complex concepts through logical manipulation of symbolic representations, and the learning systems field processes data through analytical programs modeled on neurologic systems.

Publication Sample and Summary Statistics

Our analysis focuses on journal articles and book publications through the Web of Science from 1955 to 2015. We conducted a keyword search utilizing the keywords described in Appendix A (we tried several variants of these keywords and alternative algorithmic approaches but this did not result in a meaningful difference in the publication set). We are able to gather detailed information about each publication, including publication year, journal information, topical information, as well as author and institutional affiliations.

[2] Ironically enough, we relied upon human intelligence rather than machine learning to develop this classification system and apply it to this data set.

This search yields 98,124 publications. We then code each publication into one of the three main fields of AI, as described above. Overall, relative to an initial dataset of 98,124, we are able to uniquely classify 95,840 publications as symbolic systems, learning systems, robotics, or "general" AI (we drop papers that involve combinations of these three fields). Table 1A reports the summary statistics for this sample.

Of the 95,840 publication in the sample, 11,938 (12.5 percent) are classified as symbolic systems, 58,853 (61.4 percent) as learning and 20,655 (21.6 percent) as robotics, with the remainder being in the general field of "artificial intelligence." To derive a better understanding of the factors that have shaped the evolution of AI, we create indicators for variables of interest including organization type (private versus academic), location type (US domestic versus International), and application type (computer science versus other application area, in addition to individual subject spaces, e.g. biology, materials science, medicine, physics, economics, etc.).

We identify organization type as academic if the organization of one of the authors on the publication is an academic institution. 81,998 publications (85.5 percent) and 13,842 (14.4 percent) are produced by academic and private sector authors, respectively. We identify publication location as US domestic if one of the authors on the publication lists the United States as his or her primary location. 22,436 publications (25 percent of the sample) are produced domestically.

We also differentiate between subject matter. 44 percent of the publications are classified as Computer Science, with 56 percent classified as other applications. Summary statistics on the other applications are provided in Table 2A. The other subjects with the largest number of publications in the sample include Telecommunications (5.5 percent), Mathematics (4.2), Neurology (3.8), Chemistry (3.7), Physics (3.4), Biology (3.4), and Medicine (3.1).

Finally, we create indicator variables to document publication quality, including journal quality (top 10, top 25 and top 50 journals by impact factor[3]) and a count variable for cumulative citation counts. Less than one percent of publications are in a top 10 journal with two percent and 10 percent in top 25 and top 50 journals. The average citation count for a publication in the sample is 4.9.

[3] The rankings are collected from Guide2Research, found here: http://www.guide2research.com/journals/

Patent Sample and Summary Statistics

We undertake a similar approach for gathering a dataset of AI patents. We start with the public-use file of USPTO patents (Marco, Carley et al., 2015; Marco et al., 2015,), and filter the data in two ways. First, we assemble a subset of data by filtering the USPTO Historical Masterfile on the U.S. Patent Classification System (USPC) number.[4] Specifically, USPC numbers 706 and 901 represent "Artificial Intelligence" and "Robots," respectively. Within USPC 706, there are numerous subclasses including "fuzzy logic hardware," "plural processing systems," "machine learning," and "knowledge processing systems," to name a few. We then use the USPC subclass to identify patents in AI fields of symbolic systems, learning systems and robotics. We drop patents prior to 1990, providing a sample of 7,347 patents through 2014.

Second, we assemble another subset of AI patents by conducting a title search on patents, with the search terms being the same keywords used to identify academic publications in AI.[5] This provides an additional 8,640 AI patents. We then allocate each patent into an AI field by associating the relevant search term with one of the overarching fields. For example, a patent that is found through the search term "neural network," is then classified as a "learning" patent. Some patents found through this search method will be duplicative of those identified by USPC search, i.e. the USPC class will be 706 or 901. We drop those duplicates. Together these two subsets create a sample of 13,615 unique AI patents. Summary statistics are provided in Table 1B.

In contrast to the distribution of learning systems, symbolic systems and robotics in the publication data, the three fields are more evenly distributed in the patent data: 3,832 (28 percent) learning system patents, 3,930 (29 percent) symbolic system patents, and 5,524 (40 percent) robotics patents. The remaining patents are broadly classified only as AI.

Using ancillary datasets to the USPTO Historical Masterfile, we are able to integrate variables of interest related to organization type, location, and application space. For example,

[4] We utilized data from the Historical Patent Data Files. The complete (un-filtered) data sets from which we derived our data set are available here: https://www.uspto.gov/learning-and-resources/electronic-data-products/historical-patent-data-files

[5] We utilized data from the Document ID Dataset that is complementary to Patent Assignment Data available on the USPTO website. The complete (un-filtered) data sets from which we derived our data set are available here: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-assignment-dataset

Patent Assignment Data tracks ownership of patents across time. Our interest in this analysis relates to upstream innovative work, and for this reason, we capture the initial patent assignee by organization for each patent in our sample. This data enables the creation of indicator variables for organization type and location. We create an indicator for academic organization type by searching the name of the assignee for words relating to academic institutions, e.g. "University", "College" or "Institution." We do the same for private sector organizations, searching for "corp", "business", "inc", or "co", to name a few. We also search for the same words or abbreviations utilized in other languages, e.g. "S.p.A." Only seven percent of the sample is awarded to academic organizations, while 91 percent is awarded to private entities. The remaining patents are assigned to government entities, e.g. U.S. Department of Defense.

Similarly, we create indicator variables for patents assigned to U.S. firms and international firms, based on the country of the assignee. The international firm data can also be more narrowly identified by specific country (e.g. Canada) or region (e.g. European Union). 59 percent of our patent sample is assigned to U.S. domestic firms, while 41 percent is assigned to international firms. Next to the United States, firms from non-Chinese, Asian nations account for 28 percent of patents in the sample. Firms from Canada are assigned 1.2 percent of the patents, and firms from China, 0.4 percent.

Additionally, the USPTO data includes NBER classification and sub-classification for each patent (Hall, Jaffe and Trajtenberg (2001); Marco, Carley, et al., (2015)). These sub-classifications provide some granular detail about the application sector for which the patent is intended. We create indicator variables for NBER sub-classifications related to chemicals (NBER sub-class 11, 12, 13, 14, 15, 19), communications (21), computer hardware and software (22), computer science peripherals (23), data and storage (24), business software (25), medical fields (31, 32, 33, and 39), electronics fields (41, 42, 43, 44, 45, 46, and 49), automotive fields (53, 54, 55), mechanical fields (51, 52, 59), and other fields (remaining). The vast majority of these patents (71 percent) are in NBER subclass 22, Computer Hardware and Software. Summary Statistics of the distribution of patents across application sectors are provided in Table 2B.

VI. Deep Learning as a GPT: An Exploratory Empirical Analysis

These data allow us to begin examining the claim that the technologies of deep learning may be the nucleus of a general-purpose invention for the method of invention.

We begin in Figures 1A and 1B with a simple description of the evolution over time of the three main fields identified in the corpus of patents and papers. The first insight is that the overall field of AI has experienced sharp growth since 1990. While there are only a small handful of papers (less than a hundred per year) at the beginning of the period, each of the three fields now generates more than a thousand papers per year. At the same time, there is a striking divergence in activity across fields: each start from a similar base, but there is a steady increase in the deep learning publications relative to robotics and symbolic systems, particularly after 2009. Interestingly, at least through the end of 2014, there is more similarity in the patterns for all three fields in terms of patenting, with robotics patenting continuing to hold a lead over learning and symbolic systems. However, there does seem to be an acceleration of learning-oriented patents in the last few years of the sample, and so there may be a relative shift towards learning over the last few years which will manifest itself over time as publication and examination lags work their way through.

Within the publication data, there are striking variations across geographies. Figure 2A shows the overall growth in learning publications for the US versus rest-of-world, and Figure 2B maps the fraction of publications within each geography that are learning related. In the US on learning is far more variable. Prior to 2000 the US has a roughly equivalent share of learning related publications, but the US then falls significantly behind, only catching up again around 2013. This is consistent with the suggestion in qualitative histories of AI that that learning research has had a "faddish" quality in the US, with the additional insight that the rest of the world (notably Canada) seems to have taken advantage of this inconsistent focus in the United States to develop capabilities and comparative advantage in this field.

With these broad patterns in mind, we turn to our key empirical exercise: whether in the late 2000s deep learning shifted more towards "application-oriented" research than either robotics or symbolic systems. We begin in Figure 3 with a simple graph that examines the number of publications over time (across all three fields) in computer science journals versus application-oriented outlets. While there has actually been a stagnation (even a small decline) in

the overall number of AI publications in computer science journals, there has been a dramatic increase in the number of AI-related publications in application-oriented outlets. By the end of 2015, we estimate that nearly 2/3 of all publications in AI were in fields beyond computer science.

In Figure 4 we then look at this division by field. Several patterns are worthy of note. First, as earlier, we can see the relative growth through 2009 of publications in learning versus the two other fields. Also, consistent with more qualitative accounts of the fields, we see the relative stagnation of symbolic systems research relative to robotics and learning. But, after 2009, there is a significant increase in application publications in both robotics and learning, but that the learning boost is both steeper and more long-lived. Over the course of just seven years, learning-oriented application publications more than double in number, and now represent just under 50% of *all* AI publications.[6]

These patterns are if anything even more striking if one disaggregates them by the geographic origin of the publication. In Figure 5, we at rates of publication in computer science versus applications for the US versus rest-of-world. The striking upward swing in AI application papers that begins in 2009 turns out to be overwhelmingly driven by publications ex US, though US researchers begin a period of catch-up at an accelerating pace towards the final few years of the sample.

Finally, we look at how publications have varied across application sectors over time. In Table 3, we examine the number of publications by application field in each of the three areas of AI across two three-year cohorts (2004-2006 and 2013-2015). There are a number of patterns of interest. First, and most importantly, in a range of application fields including medicine, radiology and economics, there is a large relative increase in learning-oriented publications relative to robotics and symbolic systems. A number of other sectors, including neuroscience and biology, realize a large increase in both learning-oriented research as well as other AI fields. There are also some more basic fields such as mathematics that have experienced a relative decline in publications (indeed, learning-oriented publications in mathematics experienced a

[6] The precise number of publications for 2015 are estimated from the experience of the first nine months (the Web of Science data run through September 30, 2015). We apply a linear multiplier for the remaining three months (i.e., estimating each category by 4/3).

small absolute decline, a striking different relative to most other fields in the sample). Overall, though it would be useful to identify more precisely the type of research that is being conducted and what is happening at the level of particular subfields, these results are consistent with our broader hypothesis that, alongside the overall growth of AI, learning-oriented research may represent a general-purpose technology that is now beginning to be exploited far more systematically across a wide range of application sectors.

Together, these preliminary findings provide some direct empirical evidence for at least one of our hypotheses: learning-oriented AI seems to have some of the signature hallmarks of a general-purpose technology. Bibliometric indicators of innovation show that it is rapidly developing, and is being applied in many sectors—and these application sectors themselves include some of the most technologically dynamic parts of the economy. This preliminary analysis does not trace out the important knowledge spillovers between innovation in the GPT and innovation and application sectors, but it is probably far too early to look for evidence of this.

VII. Deep Learning as a General-Purpose Invention in the Method of Invention: Considerations for Organizations, Institutions and Policy

With these results in mind, we now consider the potential implications for innovation and innovation policy if deep learning is indeed a general-purpose technology (GPT) and/or a general-purpose invention in the method of invention (IMI). If deep learning is merely a GPT, it is likely to generate innovation across a range of applications (with potential for spillovers both back to the learning GPT and also to other application sectors) but will not itself change the nature of the innovation production function. If it is also a general purpose IMI, we would expect it to have an even larger impact on economy-wide innovation, growth, and productivity as dynamics play out—and to trigger even more severe short run disruptions of labor markets and the internal structure of organizations.

Widespread use of deep learning as a research tool implies a shift towards investigative approaches that use large datasets to generate predictions for physical and logical events that have previously resisted systematic empirical scrutiny. These data are likely to have three

sources: prior knowledge (as in the case of "learning" of prior literatures by IBM's Watson), online transactions (e.g., search or online purchasing behavior) and physical events (e.g., the output from various types of sensors or geolocation data) What would this imply for the appropriate organization of innovation, the institutions we have for training and conducting research over time, and for policy, particularly as we think about private incentives to maintain proprietary datasets and application-specific algorithms?

The Management and Organization of Innovation

Perhaps most immediately, the rise of general-purpose predictive analytics using large datasets seems likely to result in a substitution towards capital and away from labor in the research production process. Many types of R&D and innovation more generally are effectively problems of labor-intensive search with high marginal cost per search (Evenson and Kislev, 1975, among others). The development of deep learning holds out the promise of sharply reduced marginal search costs, inducing R&D organizations to substitute away from highly-skilled labor towards fixed cost investments in AI. These investments are likely to improve performance in existing "search intensive" research projects, as well as to open up new opportunities to investigate social and physical phenomena that have previously been considered intractable or even as beyond the domain of systematic scientific and empirical research.

It is possible that the ability to substitute away from specialized labor and towards capital (that in principle could be rented or shared) may lower the "barriers to entry" in certain scientific or research fields—particularly those in which the necessary data and algorithms are freely available—while erecting new barriers to entry in other areas (e.g. by restricting access to data and algorithms). As of yet, there are few if any organized markets for "trained" research tools or services based on deep learning, and few standards to evaluate alternatives. Our analysis suggests that the development of markets for shared AI services and the widespread availability of relevant data may be a necessary precursor to the broad adoption and dissemination of deep learning.

At the same time, the arrival of this new research paradigm is likely to require a significant shift in the management of innovation itself. For example, it is possible that the democratization of innovation will also be accompanied by a lack of investment by individual researchers in specialized research skills and specialized expertise in any given area, reducing the

level of theoretical or technical depth in the work force. This shift away from career-oriented research trajectories towards the ability to derive new findings based on deep learning may undermine long-term incentives for breakthrough research that can only be conducted by people who are at the research frontier. There is also the possibility that the large scale replacement of skilled technical labor in the research sector by AI will "break science" in some fields by disrupting the career ladders and labor markets that support the relatively long periods of training and education required in many scientific and technical occupations.

Finally, it is possible that deep learning will change the nature of scientific and technical advance itself. Many fields of science and engineering are driven by a mode of inquiry that focuses on identifying a relatively small number of causal drivers of underlying phenomena built upon an underlying theory (the parsimony principle as restated by Einstein states that theory should be "as simple as possible but no simpler.") However, deep learning offers an alternative paradigm based on the ability to predict complex multi-causal phenomena using a "black box" approach that abstracts away from underlying causes but that does allow for a singular prediction index that can yield sharp insight. De-emphasizing the understanding of causal mechanisms and abstract relationships may come at a cost: many major steps forward in science involve the ability to leverage an understanding of "big picture" theoretical structure to make sense of, of recognize the implications of, smaller discoveries. For example, it is easy to imagine a deep learning system trained on a large amount of x-ray diffraction data quickly "discovering" the double helix structure of DNA at very low marginal cost, but it would likely require human judgment and insight about a much broader biological context to notice that the proposed structure suggests a direct mechanism for heredity.

Innovation and Competition Policy and Institutions

A second area of impact, beyond the organization of individual research projects or the nature of what counts as "science" in a particular field, will be on the appropriate design and governance of institutions governing the innovation process. Three implications stand out.

First, as discussed above, research over the past two decades has emphasized the important role played by institutions that encourage cumulative knowledge production through low-cost independent access to research tools, materials and data (Furman and Stern, 2012; Murray, et al, 2015). However to date there has only been a modest level of attention to the

questions of transparency and replicability within the deep learning community. Grassroots initiatives to encourage openness organized through online hubs and communities are to be welcomed. But it is useful to emphasize that there is likely to be a significant gap between the private and social incentives to share and aggregate data—even among academic researchers or private sector research communities. One implication of this divergence may be that to the degree any single research result depends on the aggregation of data from many sources, it will be important to develop rules of credit and attribution, as well as to develop mechanisms to replicate the results.

This implies that it will be particularly important to pay attention to the design and enforcement of formal intellectual property rights. On the one hand it will be important to think carefully about the laws that currently surround the ownership of data. Should the data about e.g. my shopping and travel behavior belong to me or to the search engine or ride sharing company that I use? Might consumers have a strong collective interest in ensuring that these data (suitably blinded, of course) are in the public domain, so that many companies can use them in the pursuit of innovation?

On the other, the advent of deep learning has significant implications for the patent system. Though there has so far been relatively little patenting of deep learning innovations, historical episodes such as the discovery and attempted wholesale patenting of express sequence tags and other kinds of genetic data suggests that breakthroughs in research tools—often combined with a lack of capacity at patent offices and conflicting court decisions—can result in long periods of uncertainty that has hampered the issuing of new patents, and this in turn has led to lower research productivity and less competition. Deep learning also presents difficult questions of legal doctrine for patent systems that have been built around the idea of creative authors and inventors. For example, "inventorship" has a specific meaning in patent law, with very important implications for ownership and control of the claimed invention. Can an AI system be an inventor in the sense envisaged by the drafters of the US Constitution? Similarly, standards for determining the size of the inventive step required to obtain a patent are driven by a determination of whether the claimed invention would or would not be obvious to a "person having ordinary skill in the art." Who this "person" might be, and what constitutes "ordinary

skill" in an age of deep learning systems trained on proprietary data, are questions well beyond the scope of this essay.

In addition to these traditional innovation policy questions, the prospect for deep learning raises a wide variety of other issues, including issues relating to privacy, the potential for bias (deep learning has been found to reinforce stereotypes already present in society), and consumer protection (related to areas such as search, advertising, and consumer targeting and monitoring). The key is that, to the extent that deep learning is general-purpose, the issues that arise across each of these domains (and more) will play out across a wide variety of sectors and contexts and at a global rather than local level. Little analysis has been conducted that can help design institutions that will be responsive at the level of application sectors that also internalize the potential issues that may arise with the fact that deep learning is likely to be a GPT.

Finally, the broad applicability of deep learning (and possibly robotics) across many sectors is likely to engender a race within each sector to establish a proprietary advantage that leverages these new approaches. As such, the arrival of deep learning raises issues for competition policy. In each application sector, there is the possibility that firms that are able to establish an advantage at an early stage, and in doing so position themselves to be able to generate more data (about their technology, about customer behavior, about their organizational processes) will be able to erect a deep-learning-driven barrier to entry that will ensure market dominance over at least the medium term. This suggests that rules ensuring data accessibility are not only a matter of research productivity or aggregation, but also speak to the potential to guard against lock-in and anticompetitive conduct. At the present moment there seem to be a large number of individual companies attempting to take advantage of AI across a wide variety of domains (e.g., there are probably more than 20 firms engaging in significant levels of research in autonomous vehicles, and no firm has yet to show a decisive advantage), but this high level of activity likely reflects an expectation for the prospects for significant market power in the future. Ensuring that deep learning does not enhance monopolization and increase barriers to entry across a range of sectors will be a key topic going forward.

VIII. Concluding Thoughts

The purpose of this exploratory essay has not been to provide a systematic account or prediction of the likely impact of AI on innovation, nor clear guidance for policy or the management of innovation. Instead, our goal has been to raise a specific possibility—that deep learning represents a new general-purpose invention of a method of invention—and to draw out some preliminary implications of that hypothesis for management, institutions, and policy.

Our preliminary analysis highlights a few key ideas that have not been central to the economics and policy discussion so far. First, at least from the perspective of innovation, it is useful to distinguish between the significant and important advances in fields such as robotics from the potential of a general-purpose method of invention based on application of multi-layered neural networks to large amounts of digital data to be an "invention in the method of invention". Both the existing qualitative evidence and our preliminary empirical analysis documents a striking shift since 2009 towards deep learning based application-oriented research that is consistent with this possibility. Second, and relatedly, the prospect of a change in the innovation process raises key issues for a range of policy and management areas, ranging from how to evaluate this new type of science to the potential for prediction methods to induce new barriers to entry across a wide range of industries. Proactive analysis of the appropriate private and public policy responses towards these breakthroughs seems like an extremely promising area for future research.

REFERENCES

Aghion, P. and P. Howitt (1992) "A Model of Growth Through Creative Destruction," *Econometrica*, 60(2), 323-251.

Bresnahan, T., E. Brynjolfsson, and L. Hitt (2002) "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 117(1), 339-376.

Bresnahan, T. and S. Greenstein (1999) "Technological Competition and the Structure of the Computer Industry," *Journal of Industrial Economics*, 47(1), 1-40.

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

Brooks, R. (1990) "Elephants Don't Play Chess," *Robotics and Autonomous Systems*, 6, 3-15.

Brooks, R. (1991) "Intelligence Without Representation," *Artificial Intelligence*, 47, 139-159.

Brynjolfsson, E. and K. McElheran (2017) "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review*, 106(5), 133-139

Griliches, Z. (1957) "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25(4), 501-522.

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

Krizhevsky, A., I. Sutskever, G. Hinton (2012) "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing*, 25, *MIT Press*.

Leung, M.K.K., A. Delong, B. Alipanahi, and B.J. Frey (2016) "Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets," *Proceedings of the IEEE*, 104(1): 176-197.

Marco, A., A. Myers, S. Graham, P. D'Agostino, and K. Apple (2015) "The USPTO Patent Assignment Dataset: Descriptions and Analysis," *USPTO Working Paper No. 2015-02*, 1-53.

Marco, A., M. Carley, S. Jackson and A. Myers (2015) "The USPTO Historical Patent Data Files," *USPTO Working Paper No. 2015-01*, 1-57.

Minsky, M. (1961) "Steps Toward Artificial Intelligence," *Proceedings of the IRE*, 8-30.

Mokyr, J (2002) *Gifts of Athena*, Princeton University Press.

Nilsson, N. (2010), *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press.

Romer, P. (1990) "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71-S102.

Rosenberg, N. and M. Trajtenberg (2004) "A General Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States," *Journal of Economic History*, 61(1), 61-99.

Rumelhart, D., G. Hinton, and R. Williams (1986) "Learning Internal Representations by Error Propagation," in J. McClelland and D. Rumelhart (editors), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*, MIT Press, 7-57.

Scotchmer, S. (1991) "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5(1), 29-41.

Turing, A. (1950) "Computing Machinery and Intelligence," *Mind*, 59, 433-460.

Wallach, I. Dzamba, M. and Heifels, A. "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery." arXiv:1510.02855 [cs.LG]

Williams, H. .(2013) "Intellectual Property Rights and Innovation: Evidence from the Human Genome,", *Journal of Political Economy,* 121(1): 1-27

Table 1A: Publication Data Summary Statistics

	Mean	Std. Dev.	Min	Max
Publication Year	2007	6.15	1990	2015
Symbolic Systems	.12	.33	0	1
Learning Systems	.61	.48	0	1
Robotics	.21	.41	0	1
Artificial Intelligence	.06	.23	0	1
Computer Science	.44	.50	0	1
Other Applications	.56	.50	0	1
US Domestic	.25	.43	0	1
International	.75	.43	0	1
Observations	95840			

Table 1B: Patent Data Summary Statistics

	Mean	Std. Dev.	Min	Max
Application Year	2003	6.68	1982	2014
Patent Year	2007	6.98	1990	2014
Symbolic Systems	.29	.45	0	1
Learning Systems	.28	.45	0	1
Robotics	.41	.49	0	1
Artificial Intelligence	.04	.19	0	1
Computer Science	.77	.42	0	1
Other Applications	.23	.42	0	1
US Domestic Firms	.59	.49	0	1
International Firms	.41	.49	0	1
Org Type Academic	.07	.26	0	1
Org Type Private	.91	.29	0	1
Observations	13615			

Table 2A: Distribution of Publications across Subjects

	Mean	Std. Dev.
Biology	.034	.18
Economics	.028	.16
Physics	.034	.18
Medicine	.032	.18
Chemistry	.038	.19
Mathematics	.042	.20
Materials Science	.029	.17
Neurology	.038	.19
Energy	.015	.12
Radiology	.015	.12
Telecommunications	.055	.23
Computer Science	.44	.50
Observations	95840	

Table 2B: Distribution of Patents across Application Sectors

	Mean	Std. Dev.
Chemicals	.007	.08
Communications	.044	.20
Computer Hardware and Software	.710	.45
Computer Peripherals	.004	.06
Data and Storage	.008	.09
Business software	.007	.09
All Computer Science	.773	.42
Medical	.020	.14
Electronics	.073	.26
Automotive	.023	.15
Mechanical	.075	.26
Other	.029	.16
Observations	13615	

Table 3: Publications Across Sectors, by AI Field, 2004-2006 versus 2013-2015

		Biology	*Economics*	*Physics*	*Medicine*	*Chemistry*	*Math*	*Materials*	*Neuro.*	*Energy*	*Radiology*	*Telecom.*	*CompSci*
Learning Systems	2004-2006	258	292	343	231	325	417	209	271	172	94	291	3889
	2013-2015	600	423	388	516	490	414	429	970	272	186	404	4582
	% growth	133%	45%	13%	123%	51%	-1%	105%	258%	58%	98%	39%	18%
Robotics	2004-2006	33	10	52	69	24	45	36	31	6	47	653	1431
	2013-2015	65	12	122	83	92	80	225	139	18	25	401	1322
	% growth	97%	20%	135%	20%	283%	78%	525%	348%	200%	-47%	-39%	-8%
Symbol Systems	2004-2006	93	8	68	96	139	54	32	35	15	82	51	827
	2013-2015	105	10	125	84	149	60	101	73	22	56	88	1125
	% growth	13%	25%	84%	-13%	7%	11%	216%	109%	47%	-32%	73%	36%

Table 4: Herfindahl-Hirschman Index for Application Sectors

Application	**H=** $\sum PatShare^2$
Chemical Applications	153.09
Communications	140.87
Hardware and Software	86.99
Computer Science Peripherals	296
Data and Storage	366.71
Computer Science Business Models	222
Medical Applications	290.51
Electronic Applications	114.64
Automotive Applications	197.03
Mechanical Applications	77.51
Other	129.20

Figure 1A: Publications by AI field over Time

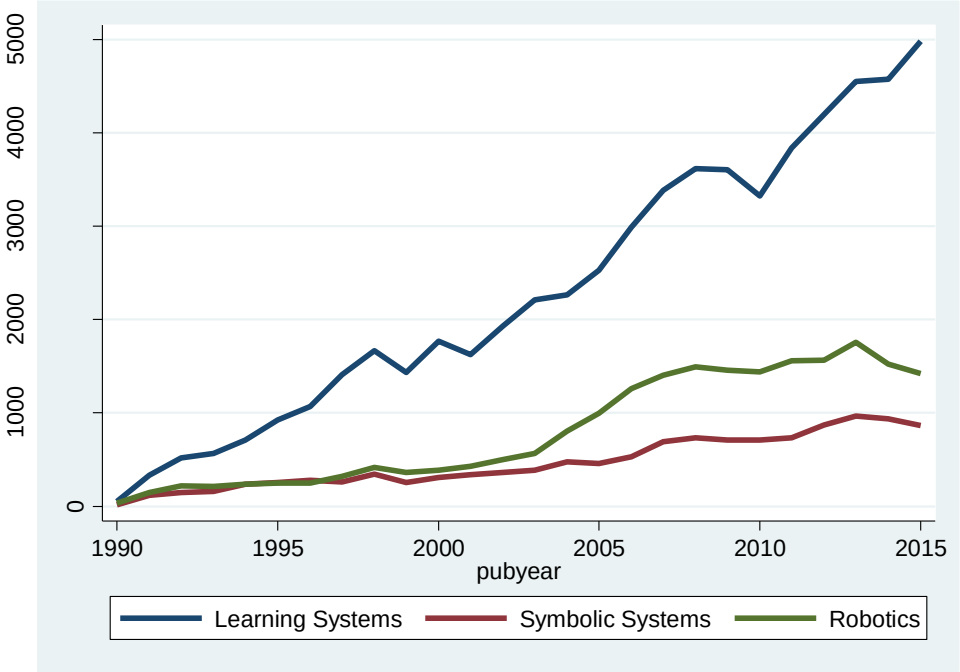


Figure 1B: Patents by AI field over Time

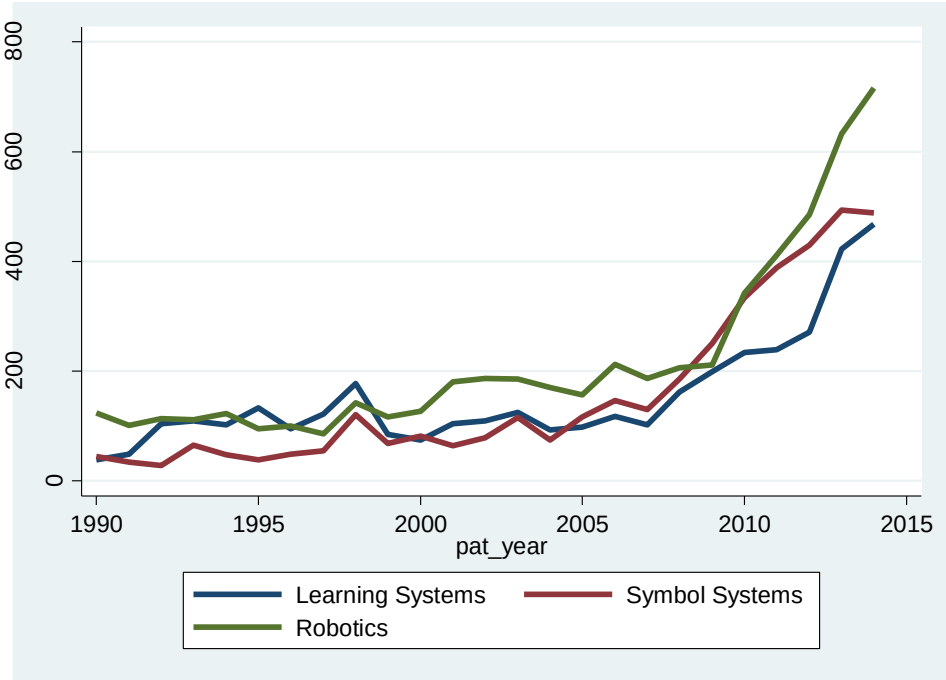


Figure 2A: Academic Institution Publication Fraction by AI Field

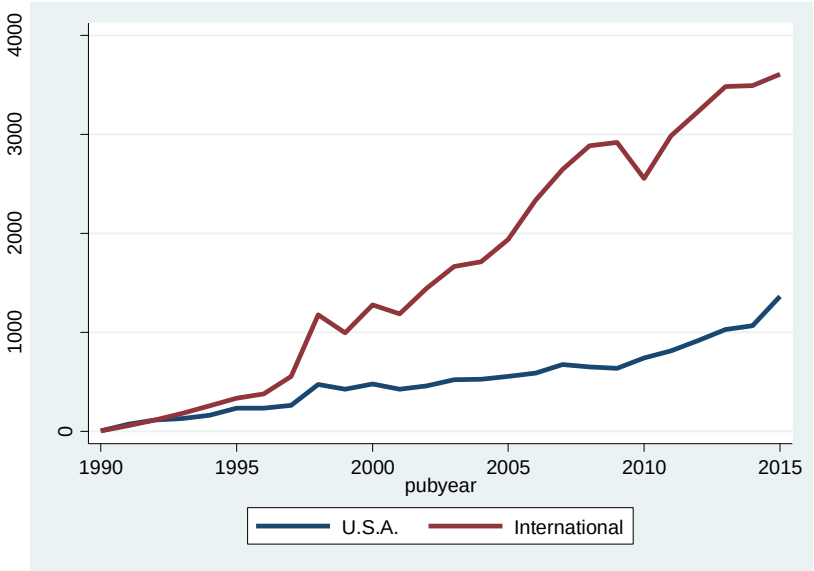


Figure 2B: Fraction of Learning Publications by US versus World

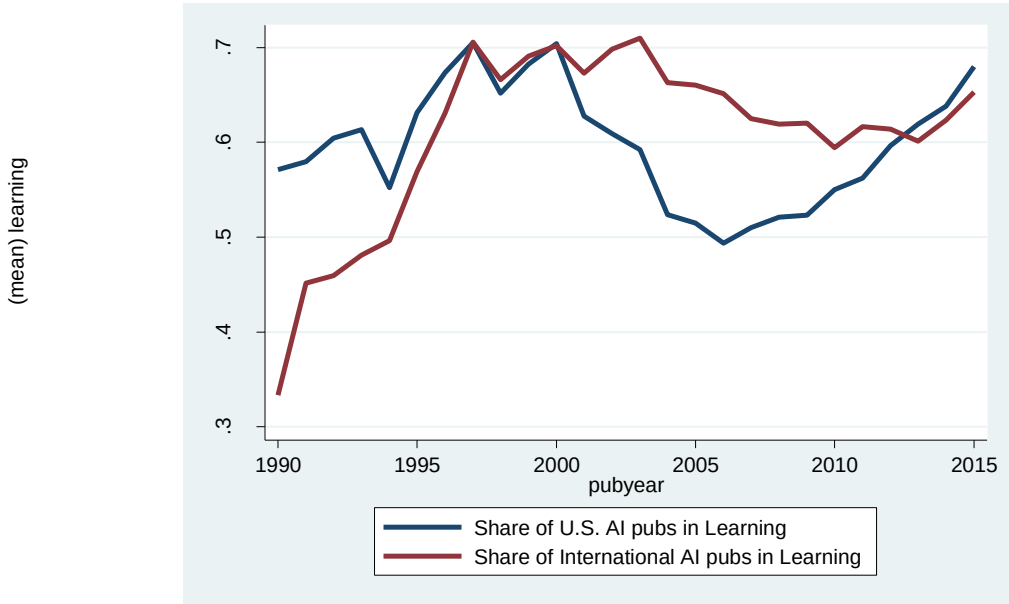


Figure 3: Publications in Computer Science versus Application Journals

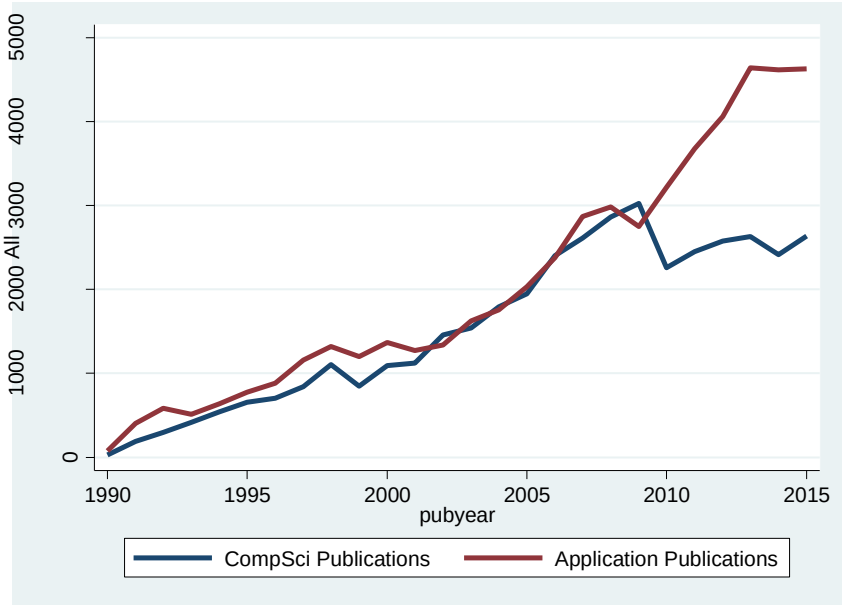


Figure 4: Publications in Computer Science versus Application Journals, by AI Field

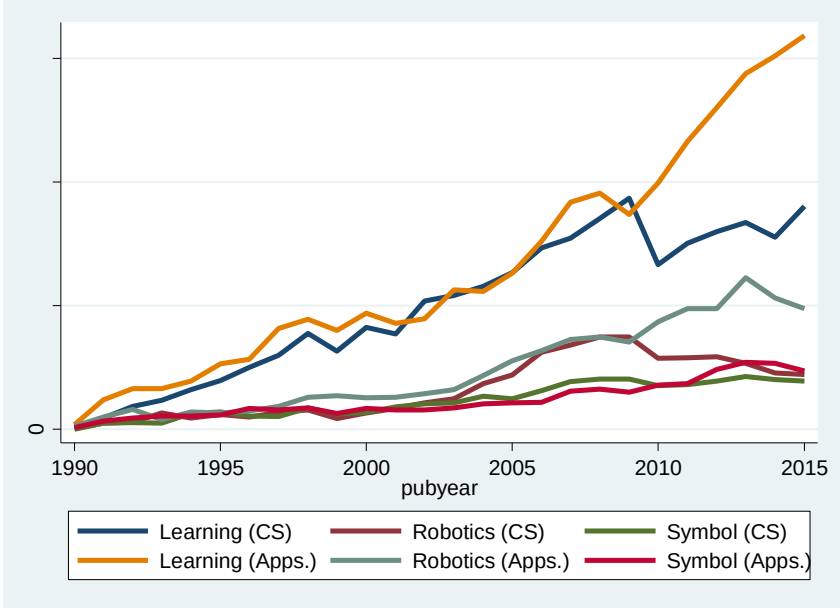
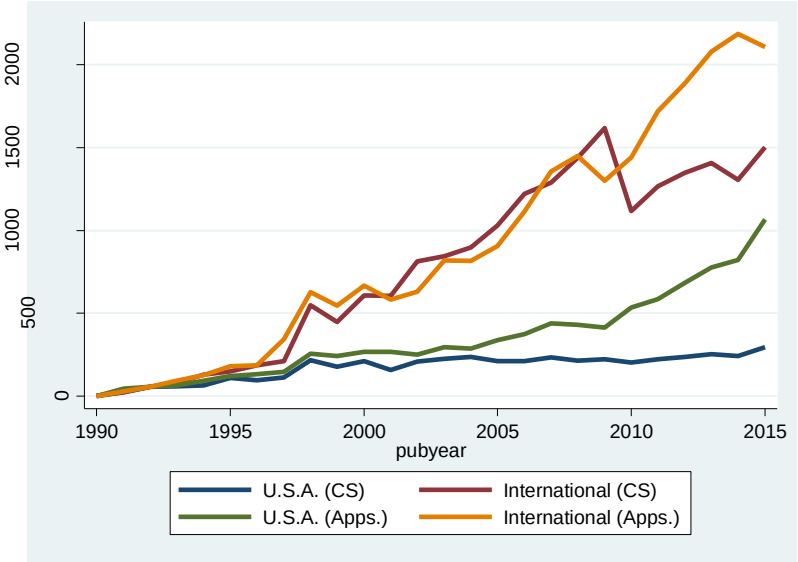


Figure 5: Learning Publications in Computer Science versus Applications, By US versus ROW



Appendix A

Appendix Table 1: Artificial Intelligence Keyword Allocation

Symbols	Learning	Robotics
natural language processing	machine learning	computer vision
image grammars	neural networks	robot
pattern recognition	reinforcement learning	robots
image matching	logic theorist	robot systems
symbolic reasoning	bayesian belief networks	robotics
symbolic error analysis	unsupervised learning	robotic
pattern analysis	deep learning	collaborative systems
symbol processing	knowledge representation and reasoning	humanoid robotics
physical symbol system	crowdsourcing and human computation	sensor network
natural languages	neuromorphic computing	sensor networks
pattern analysis	decision making	sensor data fusion
image alignment	machine intelligence	systems and control theory
optimal search	neural network	layered control systems
symbolic reasoning		
symbolic error analysis		