OFFICE OF THE CHIEF ECONOMIST
IP DATA HIGHLIGHTS

Number 5, October 2020

Inventing AI

Tracing the diffusion of artificial intelligence with U.S. patents

UNITED STATES
PATENT AND TRADEMARK OFFICE
uspto

U.S. Patent and Trademark Office ▪ Office of the Chief Economist

IP DATA HIGHLIGHTS ▪ Number 5, October 2020

Inventing AI

Tracing the diffusion of artificial intelligence with U.S. patents

Project team

This report was prepared by the Office of the Chief Economist at the United States Patent and Trademark Office (USPTO) in collaboration with the USPTO's Patent Organization and the Office of Policy and International Affairs.

Andrew A. Toole
Nicholas A. Pairolero
Alexander V. Giczy
James Q. Forman
Christyann Pulliam
Matthew Such
Kakali Chaki
David B. Orange
Anne Thomas Homescu
Jesse Frumkin
Ying Yu Chen
Vincent M. Gonzales
Christian Hannon
Steve Melnick
Eric Nilsson
Ben M. Rifkin

Acknowledgments: The project team thanks Teresa Verigan and Melissa Harvey for the layout and visualization work for this report. The team also thanks several others who provided insightful comments on earlier drafts, including Shira Perlmutter, Nicholas Rada, and John Ward.

KEY FINDINGS

- Artificial intelligence (AI) is increasingly important for invention, diffusing broadly across technologies, inventor-patentees, organizations, and geography.
- In the 16 years from 2002 to 2018, annual AI patent applications increased by more than 100%, rising from 30,000 to more than 60,000 annually. Over the same period, the share of all patent applications that contain AI grew from 9% to nearly 16%.
- Patents containing AI appeared in about 9% of all technology subclasses used by the USPTO in 1976 and spread to more than 42% by 2018.
- The percentage of inventor-patentees who are active in AI started at 1% in 1976 and increased to 25% by 2018. Growth in the percentage of organizations patenting in AI has been similar.
- Most of the top 30 AI companies are in the information and communications technology sector, with some notable exceptions such as Bank of America, Boeing, and General Electric.
- AI diffusion is occurring widely across the United States. For example, inventor-patentees in Oregon are using AI in fitness training and equipment, and in North Dakota, AI is used in agriculture.

Introduction

In a seminal paper on artificial intelligence (AI) published in 1950, Alan Turing considered the question "Can machines think?" and focused on how machines might imitate humans.[1] Today, progress in AI has advanced in ways that Turing could appreciate. Adults and children can call out questions in the comfort of their homes, and digital assistants will recognize their voices, interpret the questions, and respond with answers.[2] Meanwhile, robotic vacuums navigate the complicated terrain of their living rooms. On the streets, automobiles scan and interpret their surrounding environments and are beginning to navigate with increased autonomy.[3] Decision-making throughout the economy—such as in commerce, transportation logistics, health care, and finance—is increasingly improved by the incorporation of predictions made by machines.[4]

The broad scope of new products and services that build on AI technologies suggests that AI has the potential to fundamentally change how people perceive the world around them and live their daily lives. This is the essence of technological progress, and realizing these changes happens through innovation. AI is poised to revolutionize the world on the scale of the steam engine and electricity.[5]

The question is how to gauge the potential impact of AI. One indicator is the nature and diffusion of AI technologies through patents. As the primary form of legal protection for inventions, patents can reveal whether AI technologies are growing in volume and, importantly, whether they are diffusing across a broad spectrum of technical areas, inventors, companies, and geographies.

In this report, we use AI to discover AI. That is, we use a machine learning AI algorithm to determine the volume, nature, and evolution of AI and its component

1 See Turing (1950), 433, in which Turing introduces the "imitation game."

2 These AI systems cannot answer every question, but they are increasingly able to assist with routine tasks, improving their understanding over time with machine learning. Additional improvements are potentially possible by incorporating aspects of developmental psychology, cognitive science, and neuroscience. See Knight (2019).

3 Many more advances are necessary before automobiles become fully autonomous, although the state of the art has recently improved rapidly. See Mallozzi et al. (2019); and Yurtsever et al. (2020).

4 See Agrawal, Gans, and Goldfarb (2018).

5 See Bresnahan and Trajtenberg (1995); Brynjolfsson and McAfee (2014); and Brynjolfsson, Rock, and Syverson (2019).

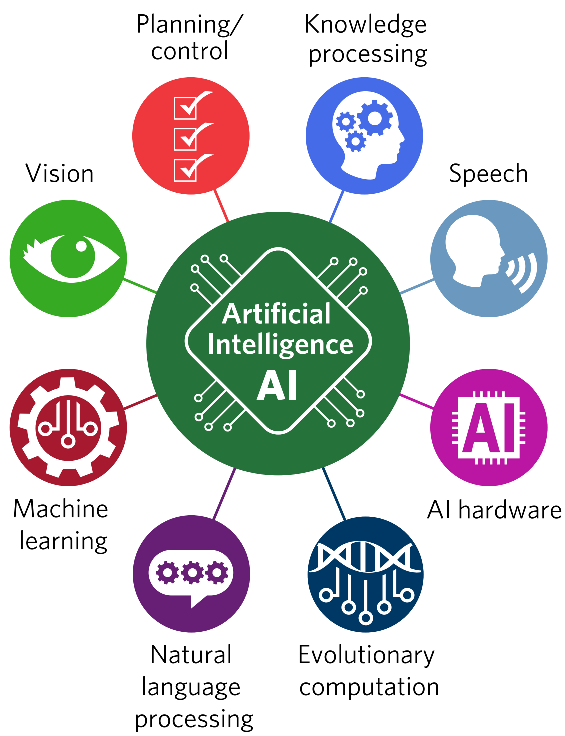
technologies as contained in U.S. patents from 1976 through 2018 (called a patent landscape). The report builds on recent AI landscaping efforts by the European Patent Office (EPO), the World Intellectual Property Organization (WIPO), and others.[6] Our primary advancement over those landscapes involves using an AI method that flexibly learns from the text of patent documents without being overly constrained by specific classifications and keywords.[7] This approach improves the accuracy of identifying AI patents.[8]

What is AI?

The U.S. National Institute of Standards and Technology (NIST) define AI technologies and systems to "comprise software and/or hardware that can learn to solve complex problems, make predictions or undertake tasks that require human-like sensing (such as vision, speech, and touch), perception, cognition, planning, learning, communication, or physical action."[9] Although carefully constructed, this definition is not specific enough for a patent level analysis. For patent applications and grants, we define AI as comprising one or more of eight component technologies (as illustrated in Figure 1). These components span software, hardware, and applications, and a single patent document may contain multiple AI component technologies.

The following brief definitions and examples help to explain the meaning of each AI component technology.

Figure 1: AI component technologies used in the patent landscape



Knowledge processing

The field of knowledge processing involves representing and deriving facts about the world and using this information in automated systems. For example, U.S. Patent No. 7,685,082, issued to the financial software company Intuit Inc., describes an algorithm that uses a pre-defined "knowledge base" to automatically detect accounting errors. One application is real-time error detection for online income tax preparation.

Speech

Speech recognition includes techniques to understand a sequence of words given an acoustic signal. U.S. Patent No. 10,043,516, issued to Apple Inc., and titled "Intelligent automated assistant," describes an invention like Apple's Siri, Amazon's Alexa, or Microsoft's Cortana, that answers articulated questions and responds to spoken commands.

6 See EPO (2017); WIPO (2019); CISTP (2018); IP Australia (2019); JPO (2019); OECD (2019); UKIPO (2019); and CIPO (2020).

7 See Trippe (2015); Abood and Feltenberger (2018); and Toole et al. (2020).

8 To learn more about the structure and performance of our AI algorithm, see the overview provided in the Appendix. For additional details and discussion, please refer to the online supplement.

9 NIST (2019), 7-8. In a leading textbook, Russell and Norvig (2016) define AI broadly as the development of machines capable of undertaking human activities in four areas: thinking humanly, acting humanly, thinking rationally, and acting rationally.

AI hardware

Modern AI algorithms require considerable computing power. AI hardware includes physical computer components designed to meet this requirement through increased processing efficiency and/or speed. For instance, U.S. Patent No. 8,892,487, issued to IBM Corp., describes a device for efficient information processing that mimics synapses between biological neurons analogous to a biological brain.

Evolutionary computation

Evolutionary computation contains a set of computational routines using aspects of nature and, specifically, evolution. U.S. Patent No. 7,657,494, issued to the oil and gas company Chevron USA Inc., describes an evolutionary approach to predicting available petroleum reserves. The invention's computerized method evaluates a large number of competing models and selects the model with the highest performance by using a genetically inspired algorithm that "mutates" through different options.

Natural language processing

Understanding and using data encoded in written language is the domain of natural language processing. U.S. Patent No. 8,930,178, issued to the Cincinnati Children's Hospital Medical Center, uses text to build an ontology by simulating various human memory approaches. The resulting ontology can be used to increase the efficiency of various healthcare administrative tasks such as assigning billing codes to clinical records.

Machine learning

The field of machine learning contains a broad class of computational models that learn from data. U.S. Patent 9,390,378, issued to retailer Wal-Mart Stores, Inc., develops an algorithm to optimize an e-commerce platform by classifying product descriptions, reviews, and other product features using crowdsourcing to resolve ambiguous results.

Vision

Computer vision extracts and understands information from images and videos. U.S. Patent No. 10,055,843, issued to the Mayo Foundation for Medical Education and Research and to Arizona State University, automates the detection of abnormalities in images taken during colonoscopies.

Planning and control

Planning and control contains processes to identify, create, and execute activities to achieve specified goals. For example, U.S. Patent No. 10,031,490, issued to Fisher-Rosemount Systems Inc., may help to reduce costly workflow analyses when abnormal conditions occur in processing plants. The invention describes a method for detecting potential problems through visual, sound or other environmental conditions and uses an expert system to identify and address those problems.

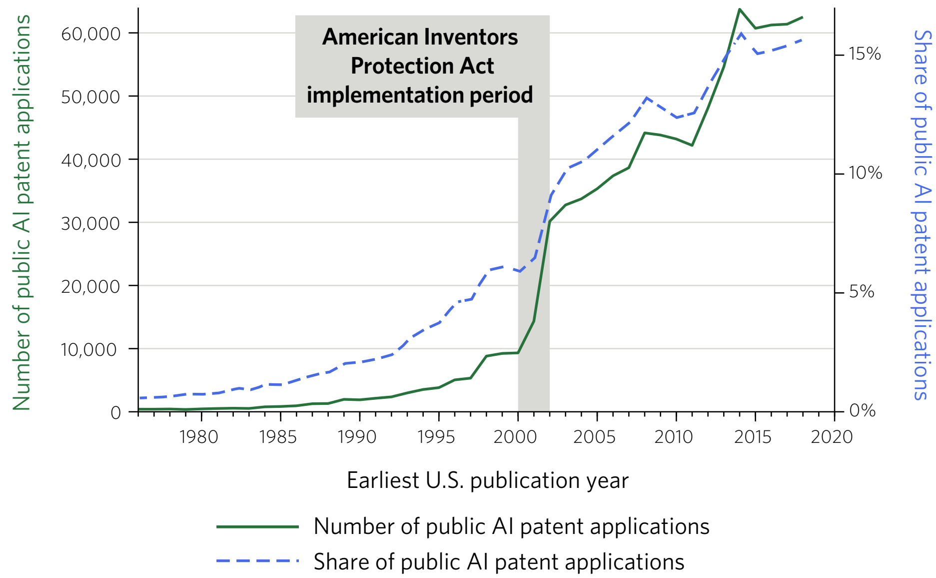
AI is increasingly important for invention

One hallmark of valuable new technologies is an increase in patent applications. These applications reflect the expectations and decisions of investors and innovators who seek to use or to build on the new technologies for innovation. AI technologies exhibit this increase. Figure 2 illustrates the long-term trends from 1976 through 2018 in the volume of public AI patent applications and their share among all public patent applications.[10] Because of changes made by the American Inventors Protection Act (AIPA) at the end of 1999 and its implementation period (the gray area in Figure 2), the trends are most informative after 2002.[11] From 2002 through 2018, both the volume and share of AI patent applications generally increased. In that 16-year period, annual AI

10 Public patent applications are patent applications that have been published before being granted (called pre-grant publications) and, in applications without pre-grant publications, the granted patents.

11 The AIPA, subtitle E, provides for publication of patent applications 18 months after filing. These pre-grant publications increased the volume of publicly available patent applications, which had previously been restricted to only granted patents. This increase is apparent in Figures 2 and 3.

Figure 2: The volume and share of public AI patent applications, 1976–2018



Note: The earliest U.S. publication year is either the year of the first pre-grant publication for a granted or pending application or the year a granted patent was published.

patent applications increased by more than 100%, rising from 30,000 to more than 60,000. Although all patent applications at the USPTO increased during that time, the share of AI applications, which adjusts for this overall trend, also shows notable growth—from 9% in 2002 to nearly 16% by 2018.

Although the overall trend in AI patent applications shows substantial growth, it does not reveal the nature of the AI involved. As mentioned earlier, a patent may fall into one or more of the eight component technologies. For instance, U.S. Patent No. 7,392,230, titled "Physical neural network liquid state machine utilizing nanotechnology," is classified by our methodology as both machine learning and AI hardware component technologies.

Figure 3 shows the number of public AI patent applications in each component technology from 1990 to 2018.[12] The largest are planning/control (dashed red line) and knowledge processing (dashed light blue line). These two components include inventions that control systems, develop plans, and process information (see sidebar). They are the most general AI component

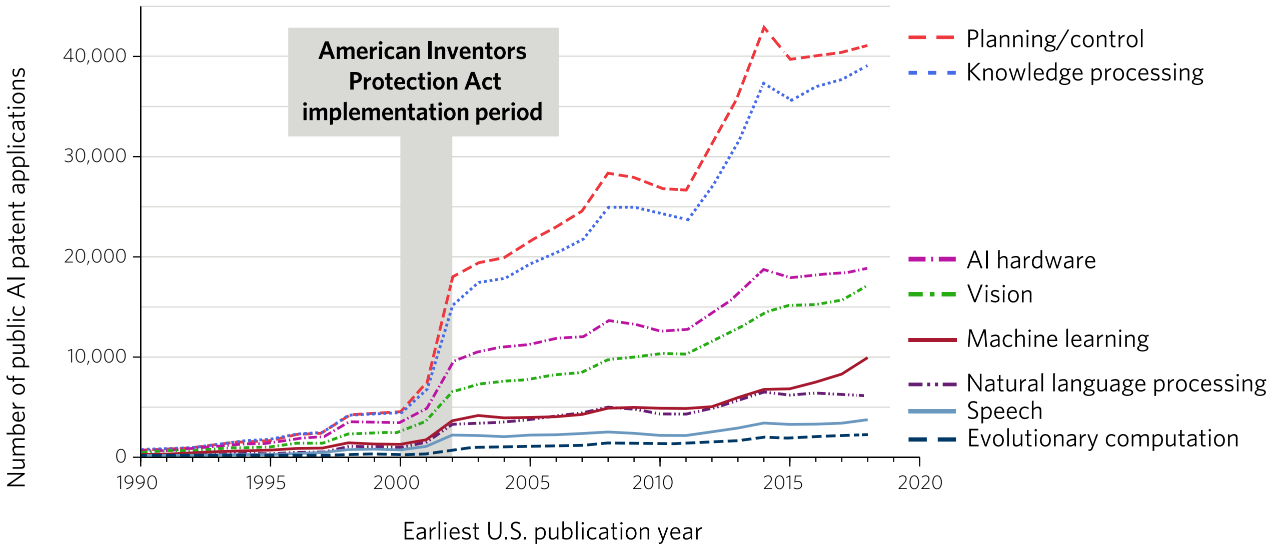
AI patent classified as both planning/control and knowledge processing

U.S. Patent No. 9,378,459 was issued to Avaya Inc. in June 2016. The invention, titled "Cross-domain topic expansion," is used in customer service operations to automatically identify and answer questions. For instance, call center employees often need fast and efficient ways to answer customer questions.

The USPTO's machine learning algorithm identified this patent as containing AI in planning/control and knowledge processing. The invention is an automated method for identifying and filling gaps in a company's knowledge database. The system exercises a degree of planning/control by synthesizing external data, question/answer histories, and user feedback to update the knowledge base for answering queries.

12 The figure starts in 1990 because the volume of patent applications in each AI component technology is low and generally uninformative before that year.

Figure 3: The volume of public AI patent applications by AI component, 1990–2018



Notes: A patent application may be classified in multiple AI component technologies. Before 1990, the lines are indistinguishable at the graph scale.

technologies, and patents in other component technologies such as machine learning often include an element of planning/control or knowledge processing.

Since 2012, patent applications in machine learning and computer vision show pronounced increases. Both of these AI technologies were central to the 2012 success of AlexNet, which was part of the 2010 ImageNet Large Scale Visual Recognition Challenge.[13] AlexNet was a watershed achievement that changed the technological trajectories for image recognition and machine learning, particularly for deep learning.[14]

Notably, patent applications in AI hardware have increased along with those in computer vision. The close association of applications in these two component technologies probably reflects the interplay between advances in image recognition and the need for computational power and performance. Specialized hardware includes accelerators for computer processors and specialized memory. Other applications of AI, such as autonomous vehicles, also involve specialized hardware.[15]

An invention lens on AI diffusion

Technology diffusion is the spread and adoption of a new technology by inventors, companies, and other innovators. When a new technology is developed, it takes time for that technology to be understood and adopted, and even more time before innovators can effectively use the technology in their invention and production processes. Technologies that diffuse broadly have potentially large effects on innovation,

13 The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was created in 2010 by Fei-Fei Li as a competition to improve computer vision, and it uses 1.4 million images from more than 1,000 categories. AlexNet, created by Alex Krizhevsky and Ilya Sutskever, was entered in 2012 as the first deep learning model in the ILSVRC. AlexNet demonstrated a remarkable decrease in error rate and won by a 40% margin. Deep learning models have since garnered the top results in the ILSVRC. See the discussion of AlexNet in Krohn, Beyleveld, and Bassens (2020).

14 See LeCun, Bengio, and Hinton (2015). Traditionally, machine learning practitioners developed informative measures (called features) meant to help the algorithm learn (called data preprocessing). The machine learning model learns on the basis of these precomputed features, rather than feeding in the raw data itself. Deep learning models generally increase performance by limiting the amount of necessary preprocessing, allowing the algorithm to fully learn which aspects of the data are most important. See Batra et al. (2018).

15 See Batra et al. (2018).

productivity, and economic growth. For example, steam power, electricity, and information technology greatly enhanced the volume, as well as the variety, of goods produced within the economy.[16]

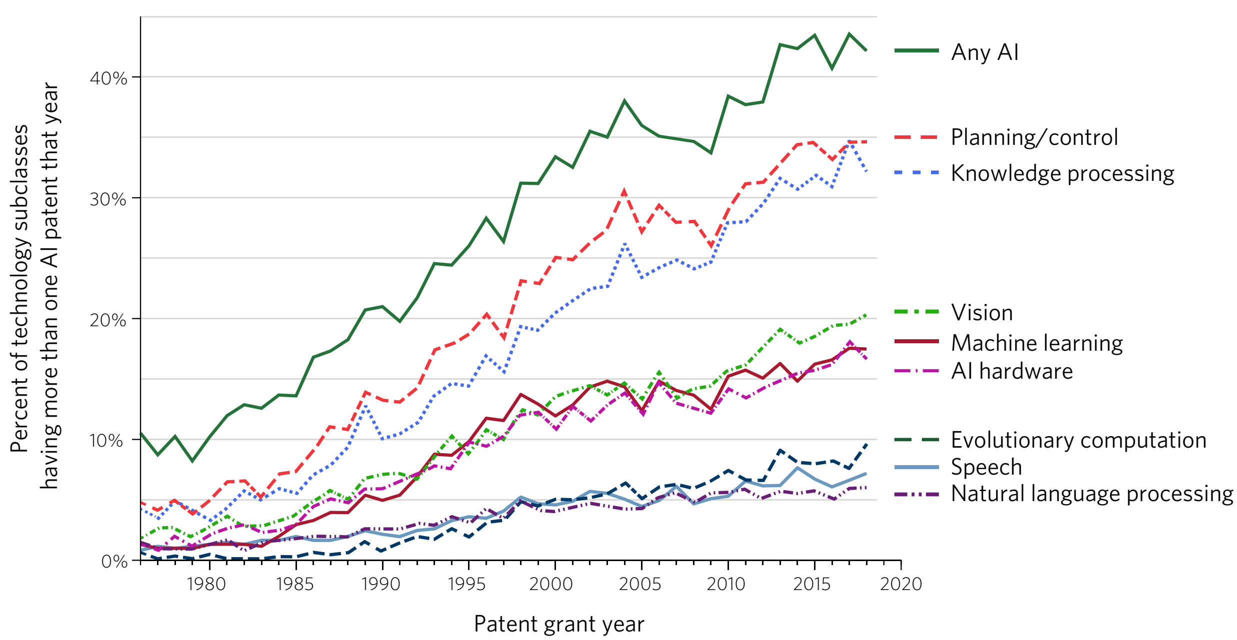
Patent documents offer a unique "invention lens" into diffusion. These documents contain detailed technical descriptions of the inventions as well as other metadata that identify the patents' technological classifications, inventors, assigned owners, locations, and key dates. Our analysis of diffusion relies on granted AI patents linked to identifiers from PatentsView.[17]

Diffusion of AI across technologies

This section explores whether AI technologies are spreading to new areas of invention. For every patent application, the USPTO reviews its technical content and assigns the application to a specific technology grouping on the basis of common subject matter.[18] The current system has more than 600 subclasses that cover a vast array of subject matter, including chemicals, electronics, machinery, and materials.

Figure 4 shows the technological diffusion of AI beginning in 1976 by plotting the percentage of technology subclasses containing at least two granted AI patents. Much like the growth in the overall volume of AI patent applications, AI technologies are diffusing across a larger percentage of technology subclasses (solid green line). In 1976, patents containing AI appeared in about 10% of the subclasses. By 2018, they had spread to more than 42% of all patent technology subclasses (see sidebars on page 8 and 9 for examples).

Figure 4: Diffusion of AI across patent technology subclasses, overall and by AI component, 1976–2018



16 See Bresnahan and Trajtenberg (1995); Jovanovic and Rousseau (2005); Gordon (2017); and Brynjolfsson, Rock, and Syverson (2019).

17 PatentsView is a free online platform for visualizing, disseminating, and promoting a better understanding of U.S. patent data. It is supported by the USPTO's Office of the Chief Economist.

18 The USPTO uses a hierarchical classification system called the Cooperative Patent Classification (CPC) system, developed jointly with the European Patent Office.

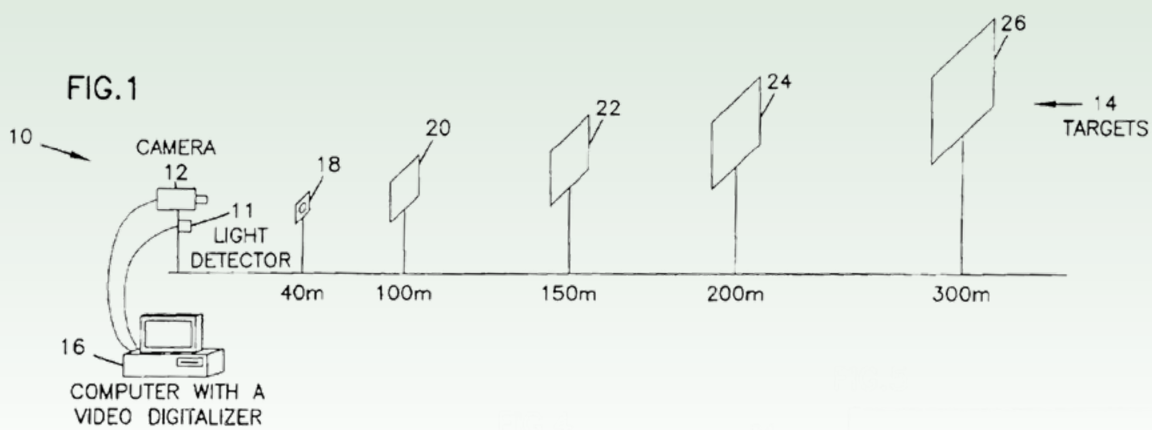
The AI component technologies show three distinct clusters with different diffusion rates. The first cluster, knowledge processing and planning/control, is diffusing the fastest across patent technology classes. This status reflects the general applicability of these AI components to a wide variety of technical areas. For the second cluster (vision, machine learning, and AI hardware), the diffusion rate is slower, but it is still increasing. Diffusion for the third cluster (evolutionary computing, speech, and natural language processing) is the slowest, hovering around 5% through the late 1990s and only recently expanding to near 10% of all technology subclasses. These clusters suggest a form of technological interdependence among the AI component technologies; however, more research is required to understand the factors behind these patterns.

Diffusion of AI across inventors and patent owners

The economic impact of AI is larger when a growing number of inventors, companies, and other organizations use AI in their invention and production processes. Growth in the percentage of inventors and organizations that received AI patents each year is one indicator of diffusion. This metric could be calculated using the names of inventors and organizations as they appear on granted patents. However, using raw patent data would miscount both inventors and organizations because of the multiple variations in how names are recorded. For instance, "International Business Machines" and "IBM" would be counted as two distinct organizations. To overcome this limitation, we once again relied on PatentsView. PatentsView provides unique IDs for inventors (hereinafter inventor-patentees) and organizations named on patents.[19]

Example of AI diffusing to other technology areas

U.S. Patent No. 7,016,045 was issued to the Regents of the University of Minnesota in March 2006. It is part of Cooperative Patent Classification (CPC) subclass G01N, which pertains to analyzing materials by determining chemical or physical properties. The invention, titled "Video camera-based visibility measurement system," is used to measure atmospheric visibility in a manner that is similar to that of the human eye. Using a curve-fitting technique, the invention processes an image from a video camera to account for environmental conditions such as fog, rain, and snow. The USPTO algorithm classified this patent in the vision AI component technology.



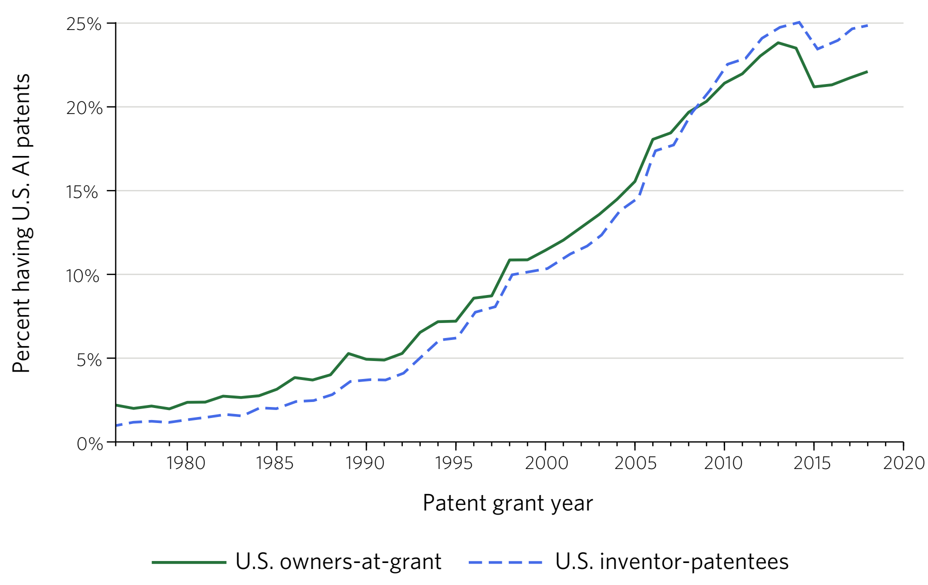
19 PatentsView uses a machine learning algorithm to assign unique IDs to inventor-patentees, to owners-at-grant, and to locations.

Figure 5 shows the annual percentage of U.S. inventor-patentees and U.S. patent owners-at-grant who received at least one granted AI patent from 1976 through 2018.[20] As this percentage grows, it indicates that a larger fraction of inventor-patentees and patent owners used AI technologies in their invention processes. The trends for inventor-patentees and patent owners show substantial diffusion and have generally increased together. The diffusion trend for inventor-patentees (dashed blue line) starts at just 1% and increases to 25% by 2018. That is, 25% of all unique inventor-patentees in 2018 used AI technologies in their granted patents. Moreover, starting in 2009, the share of inventors using AI is even greater than the share of organizations using AI (the blue dashed line crosses the solid green line). This means that diffusion is not just taking place across organizations, but is happening within organizations—more and more inventor-patentees within organizations are adopting AI in their work.

Example of AI diffusing to other technology areas

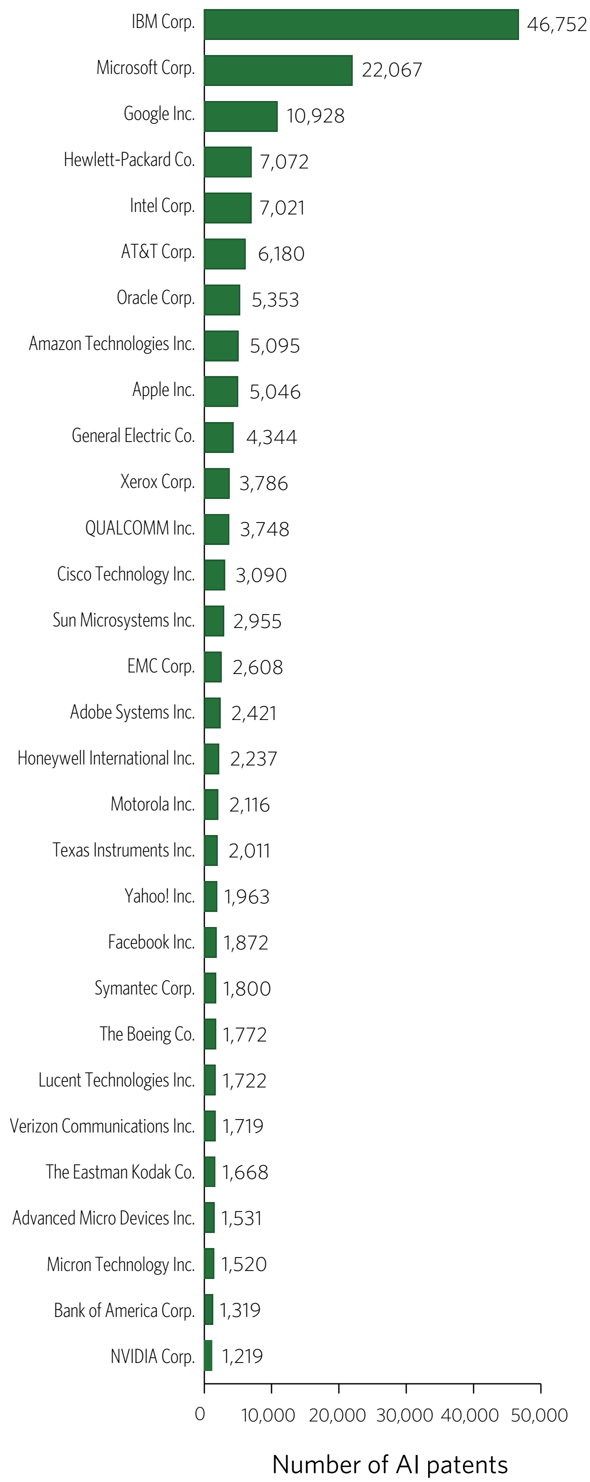
U.S. Patent No. 10,093,277 was issued to Hyundai Motor Company in October 2018. It is part of CPC subclass B60R, which pertains to vehicle fittings and parts. The use of AI in this subclass of inventions may improve, for example, how automobile components interact with drivers. The invention, titled "Method of controlling operation standby time of driver convenience system," uses a neural network algorithm to determine standby times for the operation of a driver convenience system, such as opening a trunk or folding mirrors of an automobile, to provide a user-customized service. The USPTO algorithm classified this patent in several component technologies: machine learning, evolutionary computation, vision, knowledge processing, planning/control, and AI hardware.

Figure 5: Annual percentage of U.S. inventor-patentees and patent owners with AI patents, 1976–2018



20 Patent owners-at-grant (or patent assignees) include not only organizations and individuals who have been assigned patent rights and are listed as such on the published patent, but also inventors, if they did not assign their rights to another entity, and non-inventor applicants who also did not assign their rights. Our data does not include the latter two groups. Since we do not have comprehensive data on patents bought and sold by owners after the patent grant date, we focus on owners as listed on the patent at grant.

Figure 6: Top 30 U.S. AI patent owners-at-grant, 1976–2018



Notes: Some entries may share corporate ownership.

Like most technologies, AI requires specialized knowledge to understand and implement it. When skilled labor and technical information are hard to obtain, diffusion is generally slower, and adoption tends to be restricted to a narrow set of organizations. To explore this phenomenon for AI technologies, Figure 6 lists the top 30 U.S. companies holding AI patents. These companies held 29% of all AI patents granted from 1976 to 2018, as recorded at the time the patents were granted. Most of the top 30 are in the information and communications technology (ICT) sector, but there are some notable exceptions, such as General Electric, Boeing, and Bank of America. For instance, in recent years, General Electric has emphasized adding AI to its mechanical products and processes, such as by creating "digital twins" of jet engines to monitor and forecast maintenance, in addition to building AI into the inspection of parts.[21]

Diffusion of AI across geography

Looking beyond industries and companies, wide geographic diffusion is also associated with a larger economic impact for new technologies. For AI diffusion, Figure 7 shows two maps of U.S. counties covering different time periods. Darker shades indicate higher concentrations of AI inventor-patentees. The top map (Figure 7a), which covers 25 years of AI patenting (1976–2000), shows that AI inventor-patentees tend to be concentrated in larger cities and established technology hubs, such as Silicon Valley, California. These locations have resource advantages that make early adoption easier. For instance, technology hubs are already home to successful companies with employees who have the specialized knowledge required to understand and implement AI technologies. This advantage also extends to regions with major research universities.

Despite these advantages, the location of AI inventor-patentees since 2001 shows that AI technologies are diffusing widely across U.S. states and counties (Figure 7b). For instance, Maine and South Carolina are active in digital data processing and data processing adapted for business. Inventor-patentees in Oregon are using AI in fitness training and equipment. In Montana, AI is incorporated into inventions for analyzing the chemical and physical properties of materials.

21 See Woyke (2017).

Figure 7a: Granted AI patents by inventor-patentee location, 1976–2000

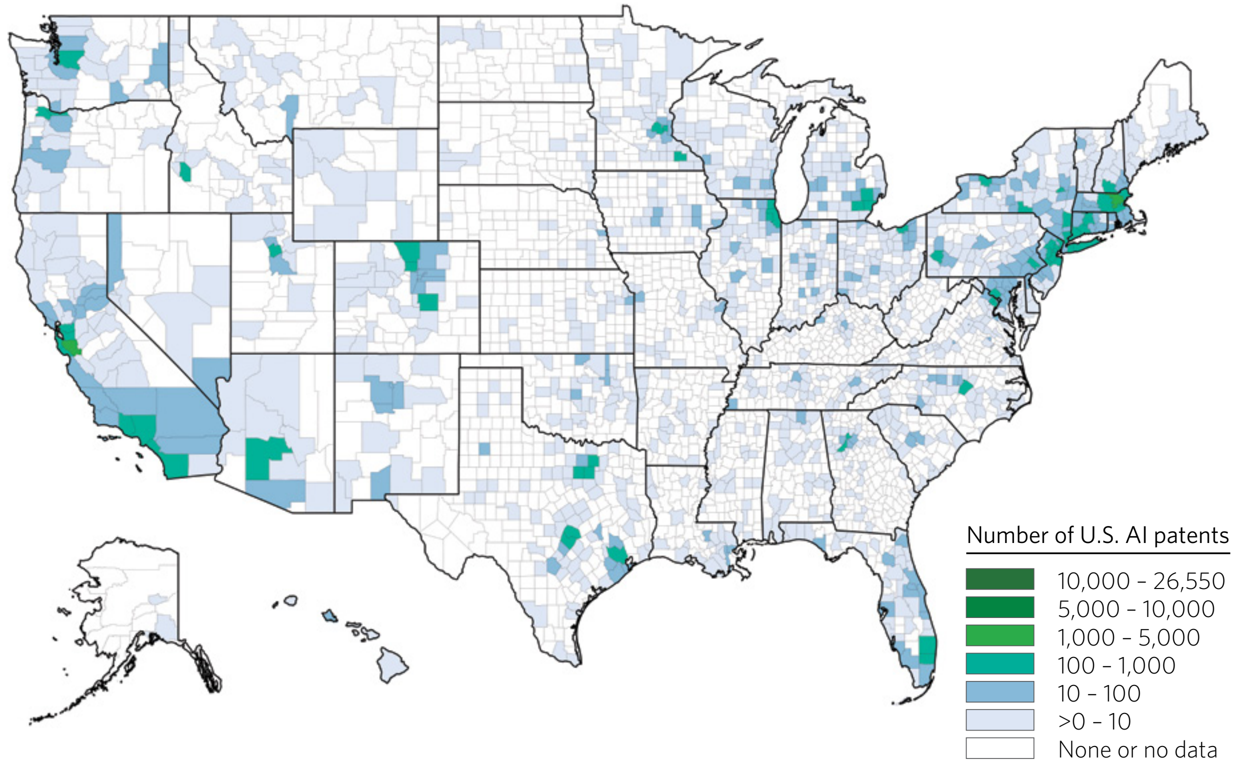
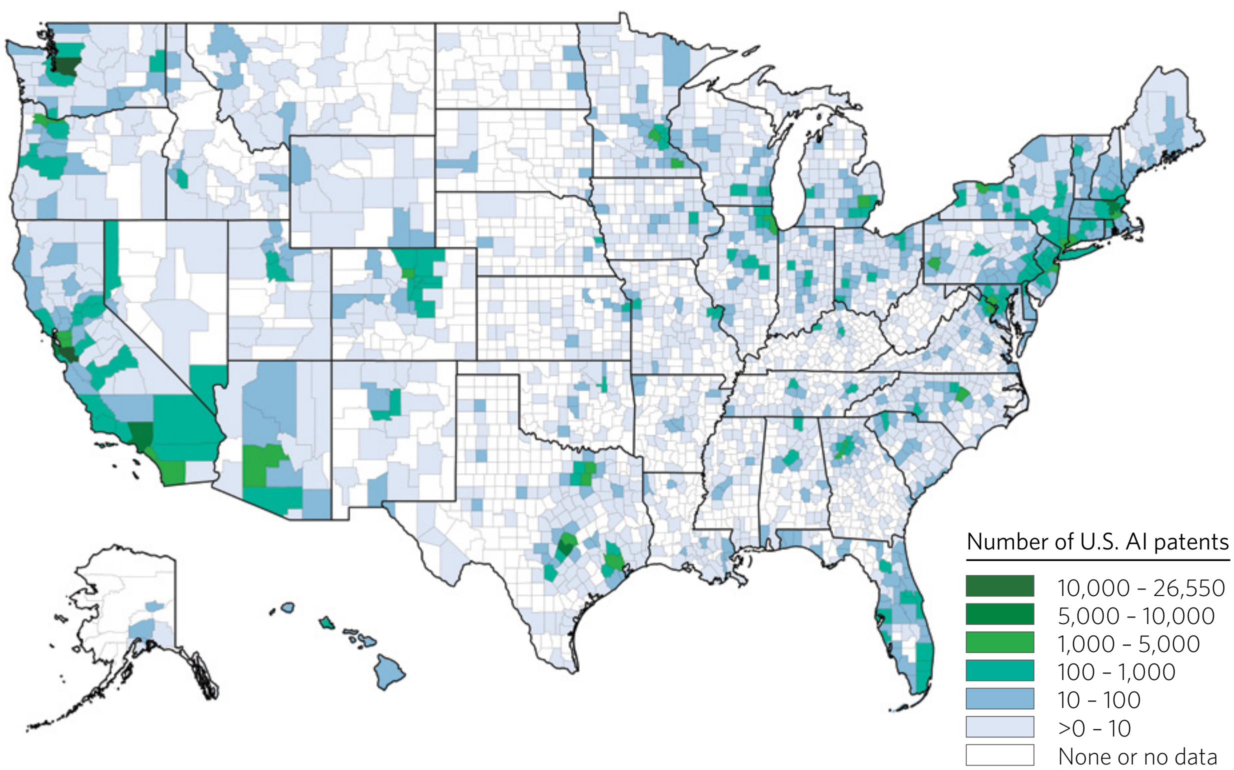


Figure 7b: Granted AI patents by inventor-patentee location, 2001–2018



The American Midwest is also adopting AI technologies, although in significantly fewer numbers. Inventor-patentees are using AI in digital information transmission, image processing, and data recognition and presentation. Wisconsin leads in medical instruments and processes for diagnosis, surgery, and identification, followed by Ohio and Kansas. For example, U.S. Patent No. 9,687,199, titled "Medical imaging system providing disease prognosis," was issued to the Wisconsin Alumni Research Foundation in June 2017. The invention incorporates multiple machine learning models to analyze different patient characteristics that are combined into a complete model for disease prognosis.

In Iowa, Kansas, Missouri, Nebraska, and Ohio, AI technologies are contributing to inventions in telephonic communications. For example, in Ohio, U.S. Patent No. 9,756,185 details an automated call analysis system for assessing the quality of phone conversations and monitoring employee performance. In another example, U.S. Patent No. 8,140,069 was issued to Sprint Spectrum L.P. in March 2012. The invention describes a method to analyze cell phone data using machine learning to assess signal quality. It may improve cell phone service by helping maintenance personnel identify faulty cell towers.

Applying AI technologies relevant to agriculture is a focus in North Dakota. For instance, U.S. Patent No. 9,723,784, titled "Crop quality sensor based on specular reflectance," was issued to Appareo Systems LLC in August 2017. The invention images a crop sample, identifies individual kernels, and determines which kernels are whole and unbroken. The sensor allows for adjustments to the harvesting combine to reduce the percentage of cracked grain.

Looking forward

The volume and diffusion of AI across technologies, inventor-patentees, patent owners, and geography show that AI is increasingly important to U.S. invention. Whether AI turns out to be as revolutionary as electricity or the semiconductor depends, in part, on the ability of innovators and firms to successfully incorporate AI inventions into existing and new products, processes, and services. Our results suggest that AI has this potential.

APPENDIX: METHODOLOGY OVERVIEW

This appendix provides an overview of our methodology to identify AI patents. The [supplementary materials](#) provides further detail.[22]

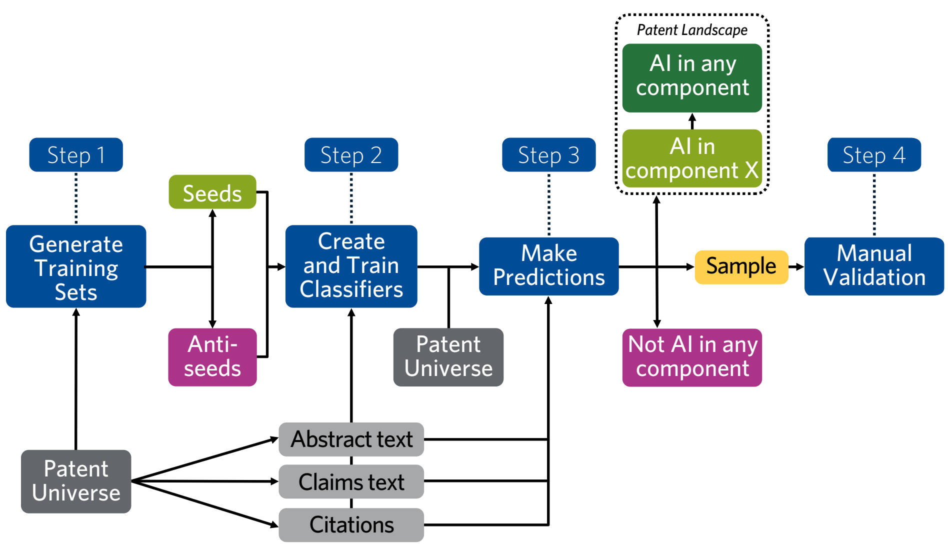
Use of machine learning to identify the AI patent landscape

Traditionally, patent landscapes relied on queries—such as keywords, patent classifications, and citations—to identify the relevant patents.[23] Abood and Feltenberger (2018) have developed an automated approach for patent landscaping that uses machine learning to reduce costs and increase accuracy. We adapted their approach and added a manual validation step. Since we have eight AI categories, we created and trained eight machine learning models, one for each aspect of AI.

From the perspective of machine learning, patent landscaping is a binary classification problem. The model predicts whether a given patent document contains each component technology of AI. Figure A1 provides an overview. The process begins by identifying two sets of patent documents (Step 1)—those representing the technology comprise a "seed" set, and those that do not comprise an "anti-seed" set. Identification of the seed set is accomplished using narrow search queries (like relevant keywords or classifications), ensuring that the results are relevant to the technology of interest (but whose numbers fall short of a robust landscape). Abood and Feltenberger (2018) created the anti-seed set using an "expansion" procedure to identify patent documents unrelated to the seed set and then randomly sampling from those documents.

The next step trains the machine learning classification models on the seed and anti-seed sets (Step 2) with the abstract and claims text, along with forward and backward patent citations (no non-patent references were used). We used the optimal neural network model, as in Abood and Feltenberger (2018). After training, the models generated predictions on the universe of patents to identify those relevant to each AI component technology (Step 3). We also consolidated the categories to generate an indicator (called "any AI") that flags a patent document if it belongs to at least one component technology of AI.

Figure A1: AI landscape methodology process



Source: USPTO, as derived from Abood and Feltenberger (2018).

22 Supplemental materials can be found at www.uspto.gov/sites/default/files/documents/OCE-ai-supplementary-materials.pdf.

23 See Trippe (2015); and Abood and Feltenberger (2018); Toole et al. (2020).

Manual validation

After developing the model, we validated the results through a manual review process (Step 4). From a random sample of 800 patent documents, two experienced examiners reviewed each document and determined which belonged to each component technology of AI.[24] The examiners were provided training materials that defined each AI component technology, along with several examples. If the two examiners disagreed, a third experienced patent examiner adjudicated. This manual validation provided a gold standard by which we assessed our model against other AI landscapes using a number of metrics (Table A1).

We evaluated our model at the level of any AI (that is, whether a given document contained at least one AI component technology). For this reason, we consolidated the seed and anti-seed sets for the eight AI component technologies in our evaluation into one aggregate seed and anti-seed set. We also used examiner agreements, disagreements, and adjudications to create scores for the evaluators (called manual scoring).[25] Finally, we recreated the AI Landscapes in Cockburn et al. (2019) and WIPO (2019) to provide benchmarks for our model. We also evaluated our approach against a "naive" comparison that predicts every document as not AI.

We provide four metrics to benchmark our results. First, "precision" is the percentage of documents predicted to contain some aspect of AI that actually contain some aspect of AI. "Recall" is the percentage of actual AI documents that were predicted to contain at least one aspect of AI. Since there is a tradeoff between precision and recall, the "F1 score" weights the two measures using the harmonic mean.[26]

Table A1: AI patent landscape model validation and comparison

	USPTO Model Seed/ Anti-seed Generation		Comparison of Scoring and AI Model Predictions				
	Seed	Anti-seed	Manual Scoring	USPTO Model	Cockburn (recreated)	WIPO (recreated)	Naive (all not AI)
Precision	0.9213	0.9259	0.3478	0.4054	0	0.6667	0
Recall	1.0000	1.0000	0.8163	0.3750	0	0.1000	0
Accuracy	0.9213	0.9259	0.8142	0.8723	0.8913	0.8967	0.8913
F1 score	0.9590	0.9615	0.4878	0.3896	0	0.1739	0

Source: USPTO analysis.

Notes: USPTO model seed and anti-seed generation compare examiner scoring to the assumption that seed and anti-seed documents are all AI and all not-AI, respectively. Manual scoring results include adjudication. When comparing across methods (manual scoring, USPTO model, Cockburn (recreated), WIPO (recreated) and Naive (all not AI)), we only considered patent documents from our random sample that were not in the seed or anti-seed sets. Cockburn and WIPO results were recreated and limited to the documents reviewed by the patent examiners; naive results are based on the assumption that all documents are predicted as being not-AI. "Precision" is the number of true positives (documents predicted to be AI that are actually AI per a "gold standard") divided by the number of predicted positives. "Recall" is the number of true positives divided by the number of actual positives. "Accuracy" is the number of true positives and true negatives divided by the total. The "F1 score" is a combination of precision and recall metrics using the harmonic mean.

24 Specifically, we used four patent examiners, and each evaluated 400 documents in a way that assured each document was reviewed by exactly two examiners.

25 More details on these procedures are provided in the Supplementary Materials www.link.xxx.

26 Higher recall (successfully predicting more actual AI documents as AI) can be achieved simply by predicting more documents as AI (which sacrifices precision or the percentage of predictions as AI that are actually AI).

For example, if two models have the same recall, but one has higher precision, then the model with higher precision will have a higher F1 score. Finally, "accuracy" is the percentage of predictions that are correct.

The results show that our method for generating seed and anti-seed sets was highly accurate. Specifically, 92% of the documents in the seed set contained at least one component technology of AI, and 93% of the documents in the anti-seed set did not contain any AI component technology. This suggests that our model was trained on high-quality data.

When comparing the performance of alternative AI landscaping methods (manual scoring, USPTO model, Cockburn, WIPO, and Naive), we used the patent documents from our random sample that were not in the seed or anti-seed sets. We made this restriction since evaluating our model against others on the training set (seed and anti-seed) would bias the results favorably toward our model.

The precision, recall, and F1 scores for the manual scoring by patent examiners (with adjudication) give an indication of how difficult it is to classify patent documents as AI (due, in part, to various definitions of AI and differences in opinion). Manual scoring wins out over the other studies in terms of recall and the F1 score. On the other side, the "naive" model illustrates a scenario in which a model can achieve high accuracy by predicting everything as not AI; the other metrics, however, show poor results for this naive classifier.

Cockburn et al. (2019) uses a limited query consisting of U.S. Patent Classification classes 706, artificial intelligence, and 901, robots, along with a keyword search of patent titles. The metrics show that none of the AI patent documents outside of the seed and anti-seed sets in our random sample were in Cockburn's landscape, illustrating the limitations of a narrow, traditional query-based approach to patent landscaping. WIPO (2019) also used a query-based approach, using a complex set of keywords and patent classifications. The WIPO landscaping approach retrieves more AI patent documents than Cockburn et al. (2019), but the WIPO results are still overly narrow (as evidenced by low recall).

In comparison, our machine learning model achieves higher recall, indicating that we successfully retrieved more AI documents than the other approaches, and a higher F1 score, indicating better balance between precision and recall. While not as high as the examiners (manual scoring), our machine learning approach is comparable to the manual evaluators.

REFERENCES

Abood, A., and D. Feltenberger. 2018. "Automated Patent Landscaping." *Artificial Intelligence and Law* 26, no. 2: 103–25.

Agrawal, A., J. S. Gans, and A. Goldfarb. 2018. "Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 89–114. Chicago: University of Chicago Press.

Batra, G., Z. Jacobson, S. Madhav, A. Queirolo, and N. Santhanam. 2018. "Artificial-Intelligence Hardware: New Opportunities for Semiconductor Companies," *Tech Rep*, New York: McKinsey and Company, December.

Bresnahan, T. F., and M. Trajtenberg. 1995. "General Purpose Technologies 'Engines of Growth'?" *Journal of Econometrics* 65, no. 1: 83–108.

Brynjolfsson, E., and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.

Brynjolfsson, E., D. Rock, and C. Syverson. 2019. "Artificial Intelligence and the Modern Productivity Paradox." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 23–57. Chicago: University of Chicago Press.

CIPO (Canadian Intellectual Property Office). 2020. *Processing Artificial Intelligence: Highlighting the Canadian Patent Landscape*. Gatineau: Canadian Intellectual Property Office.

CISTP (China Institute for Science and Technology Policy at Tsinghua University). 2018. *China AI Development Report 2018.* Beijing: China Institute for Science and Technology Policy at Tsinghua University.

Cockburn, I. M., R. Henderson, and S. Stern. 2019. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 115–46. Chicago: University of Chicago Press.

EPO (European Patent Office). 2017. *Patents in the Fourth Industrial Revolution: The Inventions behind Digital Transformation.* Munich: European Patent Office.

Gordon, R. J. 2017. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.* Vol. 70 of *The Princeton Economic History of the Western World*. Princeton, NJ: Princeton University Press.

IP Australia. 2019. *Machine Learning Innovation: A Patent Analytics Report*. Sydney: Australian Computer Society.

Jovanovic, B., and P. L. Rousseau. 2005. "General Purpose Technologies." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven Durlauf, vol. 1, 1181–224. Amsterdam: Elsevier B. V.

JPO (Japan Patent Office). 2019. *Recent Trends in AI-related Inventions—Report.* Tokyo: Japan Patent Office.

Knight, W. 2019. "The Man Who Helped Invent Virtual Assistants Thinks They're Doomed without a New AI Approach." *MIT Technology Review* (March 13). www.technologyreview.com/2019/03/13/136661/virtual-assistants-thinks-theyre-doomed-without-a-new-ai-approach/; accessed May 21, 2020.

Krohn, J., G. Beyleveld, and A. Bassens. 2020. *Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence*. Boston: Addison-Wesley Professional.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521: 436–44. doi.org/10.1038/nature14539.

Mallozzi, P., P. Pelliccione, A. Knauss, C. Berger, and N. Mohammadiha. 2019. "Autonomous Vehicles: State of the Art, Future Trends, and Challenges." In *Automotive Systems and Software Engineering*, edited by Y. Dajsuren and M. van den Brand, 347–67. Cham, Switz.: Springer.

NIST (National Institute of Standards and Technology). 2019. *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools.* Gaithersburg, MD: National Institute of Standards and Technology.

OECD (Organisation for Economic Co-operation and Development). 2019. *Artificial Intelligence in Society.* Paris: OECD Publishing.

Russell, S. J,. and P. Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Essex, UK: Pearson Education Limited.

Toole, A., N. Pairolero, J. Forman, and A. Giczy. 2020. "The Promise of Machine Learning for Patent Landscaping." *Santa Clara High Technology Law Journal* 36, no. 4: 433–38.

Trippe, A. 2015. *Guidelines for Preparing Patent Landscape Reports.* Geneva, Switz.: World Intellectual Property Organization.

Turing, A. M. 1950. "Computing Machinery and Intelligence." *Mind* 49, no. 236: 433–60.

UKIPO (United Kingdom Intellectual Property Office). 2019. *Artificial Intelligence: A Worldwide Overview of AI Patents and Patenting by the UK AI Sector.* Newport, UK: Intellectual Property Office.

WIPO (World Intellectual Property Organization). 2019. *WIPO Technology Trends 2019: Artificial Intelligence.* Geneva, Switz.: World Intellectual Property Organization.

Woyke, E. 2017. "General Electric Builds an AI Workforce." *MIT Technology Review* 120, no. 4. www.technologyreview.com/s/607962/general-electric-builds-an-ai-workforce/; accessed March 16, 2020.

Yurtsever, E., J. Lambert, A. Carballo, and K. Takeda. 2020. "A Survey of Autonomous Driving: Common Practices and Emerging Technologies." *IEEE Access* 8: 58443–69.

U.S. Patent and Trademark Office

600 Dulany Street • Alexandria, Virginia, 22314 • www.uspto.gov