Chicago-Kent Journal of Intellectual Property

6-25-2020

AI Patents: A Data Driven Approach

Brian S. Haney
Notre Dame Law School, bhaney3@alumni.nd.edu

Follow this and additional works at: https://scholarship.kentlaw.iit.edu/ckjip

Part of the Intellectual Property Law Commons

Recommended Citation

AI Patents: A Data Driven Approach

AI PATENTS: A DATA DRIVEN APPROACH

BRIAN S. HANEY*

Contents

2020   *AI PATENTS: A DATA DRIVEN APPROACH*   409

I. INTRODUCTION

While artificial intelligence (AI) research brings challenges,[1] the resulting systems are no accident. In fact, academics, researchers, and industry professionals have been developing AI systems since the early 1900s.[2] AI is a field uniquely positioned at the intersection of several scientific disciplines including computer science, applied mathematics, and neuroscience.[3] The AI design process is meticulous, deliberate, and time-consuming – involving intensive mathematical theory, data processing, and computer programming.[4] All the while, AI's economic value is accelerating.[5] As such, protecting the intellectual property (IP) springing from this work is a keystone for technology firms acting in competitive markets.[6]

A. Definition

The term AI has been discussed at length by various scholars and industry leaders. Google's Ray Kurzweil describes AI as "the art of creating machines that perform functions that require intelligence when performed by people."[7] Stanford Professor Nils Nilsson states, AI is "concerned with intelligent behavior in artifacts."[8] Carnegie Mellon University's Center for AI and Patent analysis develops machine learning[9] algorithms to define AI within patents.[10] But, perhaps the most important element is defining intelligence.

---

1. For example, de-bugging software beneath an API, re-writing bad code, or fixing problems related to new software versions.

2. Gely P. Basharin, et. al, *The Life and Work of A.A. Markov*, 386 LINEAR ALGEBRA AND ITS APPLICATIONS 3, 15 (2004); *see also* C. E. Shannon, *A Mathematical Theory of Communication*, Bell Systems Technical J. 1, 8 (1948).

3. PETER J. DENNING & MATTI TEDRE, COMPUTATIONAL THINKING 90-91 (2019) (Dissemination of computer science across fields including physics, biology, and economics lead to AI's growth as field of study and practice).

4. MAXIM LAPAN, DEEP REINFORCEMENT LEARNING HANDS-ON (2018)(textbook on deep reinforcement learning programming in Python); *see also* SEBASTIAN RASCHKA & VAHID MIRJALILI, PYTHON MACHINE LEARNING 18, 21-22 (2017) (textbook on machine learning in Python).

5. *See* Neha Soni, *Impact of Artificial Intelligence on Businesses: from Research, Innovation, Market, Deployment to Future Shifts in Business Models*, CORNELL U. 1, 7 (2019), https://arxiv.org/abs/1905.02092 (worldwide spending on cognitive and AI systems will increase prominently from $12 billion in 2017 to $58 billion in 2021).

6. Mark A. Lemley & Mark. P. McKenna, *Unfair Disruption*, 100 B.U. L. REV. 104 (2020) (forthcoming 2020) (manuscript at 104) (on file with author) (discussing the competition between incumbents and new market entrants).

7. RAY KURZWEIL, THE AGE OF INTELLIGENT MACHINES 14 (1992).

8. NILS J. NILSSON, ARTIFICIAL INTELLIGENCE: A NEW SYNTHESIS 1 (1998).

9. A sub-field of AI focused on neural networks, deep learning, and reinforcement learning models.

An early article defining machine intelligence argued, "[i]ntelligence measures an agent's ability to achieve goals in a wide range of environments."[11] The definition has garnered acceptance within the field, having major influence over AI model design.[12] MIT Professor Max Tegmark adopted the definition in 2017,[13] adding intelligence requires three elements: memory, computation, and the ability to learn.[14] Machine learning is a sub-field of AI, including deep learning, reinforcement learning, supervised learning, unsupervised learning, and other techniques designed to allow machines to derive knowledge from information.[15] Generally, and for the purposes of this Article, AI refers to any machine replicating the human mind's thoughtful processes. Now, AI technology is affecting industries across the economy including law, healthcare, and defense.[16]

B. Applications

In the legal industry, technology assisted review is changing the discovery process.[17] In the context of corporate litigation, millions of documents may require searching and examination for relevance.[18] As such, clients now commonly call on litigators to establish e-discovery relevancy hypotheses and to implement predictive coding models for discovering electronic information.[19] In this process, litigators first identify keywords to search and select an initial set of documents to be

---

10.   Dean Alderucci, et al., *Mapping the Movement of AI into the Marketplace with Patent Data*, CARNEGIE MELLON U. (2019)   https://www.cmu.edu/block-center/images/center-images/AI-patent-project-media-summary.pdf.

11.   Shane Legg & Marcus Hutter, *Universal Intelligence: A Definition of Machine Intelligence*, CORNELL U. 1, 12 (2007), https://arxiv.org/abs/0712.3329.

12.   This is particularly with respect to Markovian models for reinforcement learning. *See* U.S. Patent No. 10,346,741 (July 9, 2019) (assigned to DeepMind Technologies – a Google subsidiary).

13.   MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 38 (2017).

14.   Learning is particularly important because machine learning is the predominant area of AI research. *Id.* at 71; *see also* Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1278 (2018).

15.   JOHN D. KELLEHER, DEEP LEARNING 26-28, 123 (2019).

16.   HEMANT TANEJA, UNSCALED: HOW AI AND NEW GENERATION OF UPSTARTS ARE CREATING THE ECONOMY OF THE FUTURE 1 (2018).

17.   *See* FED. R. CIV. P. 26. Rule 26(a) requires the parties produce all "documents, electronically stored information, and tangible things" to be used in the course of litigation.

18.   Michael Simon, et. al., *Lola v. Skadden and the Automation of the Legal Profession*, 20 YALE J.L. & TECH. 234, 254 (2018); *see also* Chris D. Birkel, *The Growth and Importance of Outsourced E-Discovery: Implications for Big Law and Legal Education*, 38 J. LEGAL PROF. 231 (2014).

19.   KEVIN D. ASHLEY, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS 240–42 (2017).

reviewed.[20] Then, document review attorneys review, code, and score the initial set of documents based on the occurrence of certain keywords in relation to a document's relevance.[21] As this review takes place, e-discovery attorneys train machine learning algorithms to classify documents based upon the document review attorneys' decisions in classifying documents in the initial set of documents.[22] In other words, the algorithm learns what documents are relevant by analyzing and replicating the decisions of real attorneys.[23]

Healthcare is another industry being impacted by AI.[24] Data driven AI technologies are disseminating into the practice of medicine.[25] Medical professionals practicing in modern hospitals now store patient data in electronic databases with electronic healthcare records.[26] This allows machine-learning algorithms to analyze patient healthcare data and improve patient care.[27] These resources allow a doctor to know much about a patient's medical history without ever meeting the patient.[28] Further, data-driven analytics and automated patient diagnostics drastically reduce costs associated with healthcare because machines are now capable of doing medical work.[29] However, despite

20.   Nicholas Barry, *Man Versus Machine Review: The Showdown Between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology*, 15 VAND. J. ENT. & TECH. L. 343, 351 (2013)*.*

21.   GORDON V. CORMACK & MAURA R. GROSSMAN, EVALUATION OF MACHINE-LEARNING PROTOCOLS FOR TECHNOLOGY-ASSISTED REVIEW IN ELECTRONIC DISCOVERY 154 (2014), http://plg2.cs.uwaterloo.ca/~gvcormac/calstudy/study/sigir2014-cormackgrossman.pdf.

22.   Barry, *supra* note 20, at 354.

23.   *Id.*

24.   TEGMARK, *supra* note 13, at 102..

25.   *Id.*

26.   Kate Monica, *Apple EHR Patient Data Viewer Now in Use at 39 Health Systems*, EHR INTELLIGENCE (Apr. 2, 2018), https://ehrintelligence.com/news/apple-ehr-patient-data-viewer-now-in-use-at-39-health-systems.

27.   Xiaoqian Jiang, et. al., *A Patient-driven Adaptive Predication Technique to Improve Personalized Risk Estimation for Clinical Decision Support*, J. AM. MED. INFORMATICS ASS'N, 137, 137 (2012).

28.   *Id.*

29.   Alvin Rajkomar, et. al., *Scalable and Accurate Deep Learning with Electronic Health Records*, NATURE PARTNER JOURNALS (May 8, 2018), https://www.nature.com/articles/s41746-018-0029-1.pdf.; *see also* Lloyd Minor, *Crunching the Image Data Using Artificial Intelligence to Look at Biopsies*, STAN. MED. (2017), https://stanmed.stanford.edu/2017summer/artificial-intelligence-could-help-diagnose-cancer-predict-survival.html.;   Brian   S.   Haney, *Quantum_Machine_Learning_Cancer_Diagnostics*,   GITHUB, https://github.com/Bhaney44/Leap/blob/master/Quantum_Machine_Learning_Cancer_Diagnostics.py. (Another example is D-Wave's Adiabatic Quantum Computer, which is capable of running machine learning algorithms for cancer diagnostics.).

the reduced costs and improved efficiency, it is unlikely AI will make an impact on healthcare at a societal scale.[30]

The defense industry is also being impacted by developments in AI technology. Northwestern Law Professor, John McGinnis argues, "The way to think about the effects of AI on war is to think of the consequences of substituting technologically advanced robots for humans on the battlefield."[31] However, McGinnis' mode of thought completely fails to communicate AI security threats. Indeed, today the battlefield is everywhere, and the United States is bombarded with cyber-attacks every day.[32] McGinnis further argues "The existential dread of machines that become uncontrollable by humans and the political anxiety about machines' destructive power on a revolutionized battlefield are overblown."[33] Yet, China has developed and made publicly available state-of-the-art AI guided missile technology and computer programs.[34] And, Russia routinely uses AI to manipulate United States voters on social media for the purposes of influencing political elections.[35] In short, AI is the most important weapon in modern warfare, defense, and national security.[36]

---

30.   Access problems plague the healthcare industry due to excessive government regulation and corruption. On a societal scale, the problem with the healthcare industry is not limitations in diagnostic functions, or even information management. Instead the problem is that insurance companies profit from public funds by intentionally restricting access to care for patients – to drive up demand and profit. *See Restoring Fairness in Western Pennsylvania*, OFF. ATTORNEY GENERAL COMMONWEALTH PA., https://www.attorneygeneral.gov/upmc/.

31.   John O. McGinnis, *Accelerating AI*, 104 NW. U. L. REV. 1253, 1265-66 (2010).

32.   John P. Carlin, *Detect, Disrupt, Deter: A Whole-of-Government Approach to National Security Cyber Threats*, 7 HARV. NAT'L SEC. J. 391, 398 (2016); *see also Significant Cyber Incidents*, CENTER FOR STRATEGIC AND INTERNATIONAL STUDIES (Aug. 2019), https://www.csis.org/programs/technology-policy-program/significant-cyber-incidents. (For example, in May 2019, hackers affiliated with the Chinese intelligence service reportedly had been using NSA hacking tools since 2016, more than a year before those tools were publicly leaked).

33.   McGinnis, *supra* note 31, at 1254.

34.   Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare*, 7 IEEE Access , 37432, 37447 (2019); *see also* youshixun, vCEW New model of cognitive electronic warfare with countermeasures, GITHUB (2019), https://github.com/youshixun/vCEW .

35.   U.S. DEP'T OF JUSTICE, REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION, VOL I, 4 (2019), https://www.justice.gov/storage/report_volume1.pdf.

36.   *See* Hyrum S. Anderson, et.al., *Learning to Evade Static PE Machine Learning Malware Models via Reinforcement Learning*, CORNELL U. LIBR. (2018), https://arxiv.org/abs/1801.08917. (Specifically, detailing reinforcement learning malware models and open-sourced the code on GitHub); You, *supra* note 34, at 37438.. youshixun, *supra* note 34. (open source code for deep reinforcement learning missile control systems sponsored by China).

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

C. Dataset

The dataset gathered for this article consists of 2,459 patents. The patents were collected by searching the claims of all patents in the USPTO database for keywords.[37] The keywords searched are natural language processing, deep learning, reinforcement learning, and deep reinforcement learning.[38] The dataset is tailored to provide a window into four narrow portions of the AI patent market, and is not meant to be comprehensive in scope.[39] Figure 1 depicts the breakdown of this Article's AI patent dataset by subject matter.

AI Patent Dataset

■ Natural Language Processing   ■ Deep Learning
■ Reinforcement Learning   ■ Deep Reinforcement Learning
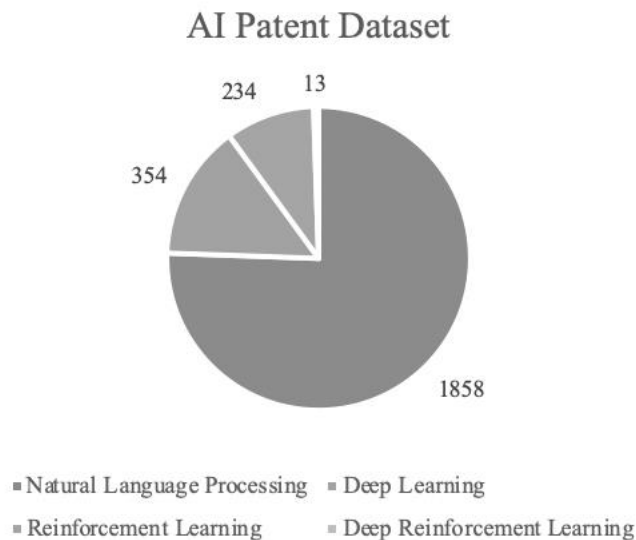
Figure 1[40]

The search results returned a majority of patents for natural language processing (1,858). Deep learning returned (354), reinforcement learning returned (234), and deep reinforcement learning[41] returned (13). Data on

---

37. *Search for Patents*, U.S. PATENT AND TRADEMARK OFFICE (2020), https://www.uspto.gov/patents-application-process/search-patents.

38. These words were selected to reflect sub-fields of machine learning.

39. Throughout this paper the term market is used referring to the total number of patents returned from keyword searches.

40. Brian S. Haney, AI Patents (2019) (A copy of the data is on file with the author).

41. Both "deep reinforcement learning" and "deep learning AND reinforcement learning" were used as search terms deriving thirteen results. The term "deep reinforcement learning" returned six patents, while the terms "deep learning AND reinforcement learning" returned ten patents.

each of the four types of patents are analyzed individually throughout this Article to provide insights for the AI patent market.

The dataset measures year as the year a particular patent was granted. In the aggregate, the data reflects an increasing number of AI patents granted each year. Further, the dataset shows accelerating five-year growth. In the year 1999, 7 patents were granted; in the year 2004, 8 patents were granted; in the year 2009, 20 patents were granted; in the year 2014, 79 patents were granted; and in the year 2019, 947 patents were granted.[42]
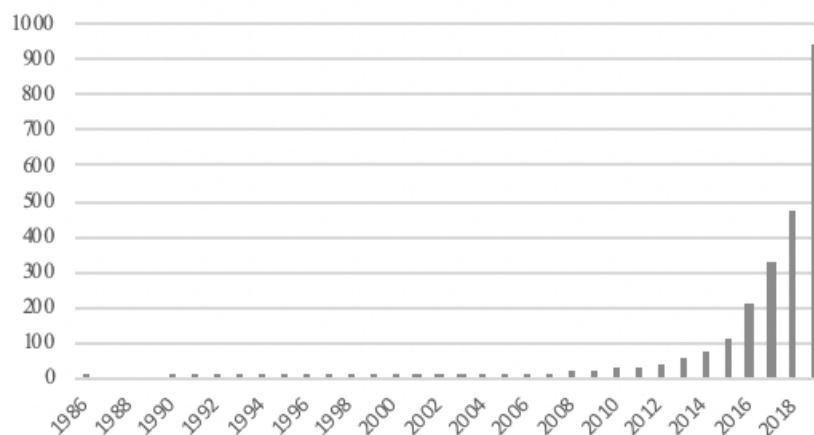
Figure 2[43]

However, one limitation is this dataset does not provide a complete picture of the AI patent market, only a snapshot of a smaller niche market.

Research for this Article revealed one other AI patent dataset. The second dataset consists of graphs published online in an unpublished paper[44] by a team of researchers at Carnegie Mellon University, headed by Dean Alderucci.[45] Figure 3 represents the CMU AI Patent dataset, measuring year, as the year a patent's application was filed.[46]

---

42. Haney, *supra* note 40. (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office).

43. *Id.*

44. Alderucci, *supra* note 11, at Fig. 2.

45. Carnegie Mellon University's Center for AI and Patent Analysis is a research center in Pittsburgh, PA, whose mission includes the ambitious tasks of extracting knowledge and data used for legal, technical, policy, and business decision making. (https://www.cmu.edu/epp/patents/about/index.html)

46. Alderucci, *supra* note 11, at Fig. 3.

CMU AI Patents by Filing Date
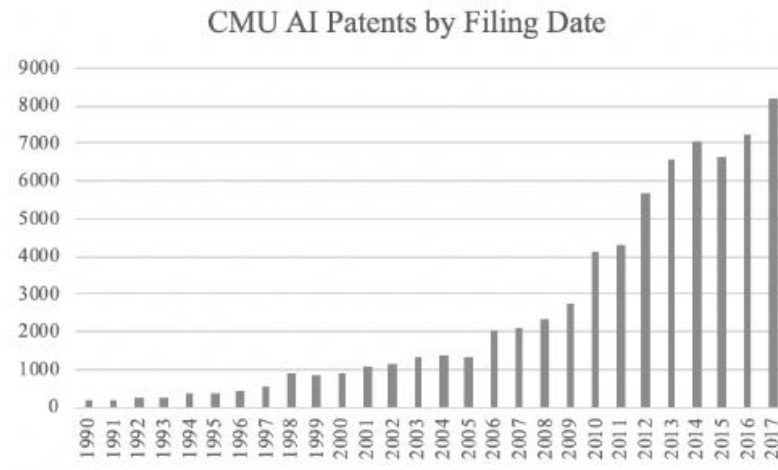
Figure 3[47]

The two datasets are different in a variety of ways, each contributing its own insights, while together creating new questions to be answered. The CMU dataset is much more robust in the scope of patents it includes (70,412).[48] However, the dataset for this Article is much narrower in scope (2,459) – focusing analysis on patents for four specific types of machine learning under AI's broader umbrella. Further, the dataset developed for this Article includes information up to January 1, 2020 – while the CMU dataset is updated through the early part of 2018. Throughout this Article, comparative analysis of the two datasets provides novel observations of the AI patent landscape. But first, each of the four types of technology patents in this Article's dataset are analyzed in depth.

II.   DEEP LEARNING

A. Technology

Deep learning is a sub-field of machine learning concerned with the acquisition of knowledge from large amounts of data.[49] The roots of deep learning date back to the mid-twentieth century.[50] Deep learning involves

---

47*. Id.* (The information contained in this chart was prepared by the author with information from the preceding citation).

48.   *Id.*

49.   ETHEM ALPAYDIN, MACHINE LEARNING 3 (2016); *see also* MICHAEL BUCKLAND, INFO. AND SOC.Y 21-22 (2017) (discussing definitions of information).

50.   RASCHKA & MIRJALILI, *supra* note 5, at 18, 21-22.

modeling the human brain with machines to process information.[51] Both artificial and biological neurons receive input from various sources, mapping information to a single output value.[52] Each neuron in the brain is connected to other neurons through structures called synapses.[53] A biological neuron consists of dendrites—receivers of various electrical impulses from other neurons—that are gathered in the neuron's cell body.[54] Once the neuron's cell body collects enough electrical energy to exceed a threshold amount, the neuron transmits an electrical charge to other neurons in the brain through synapses.[55] This transfer of information in the biological brain provides the foundation for the way in which modern neural networks operate.[56]

i. Data

Deep learning is a process by which neural networks learn from large amounts of data.[57] The internet is the driving force behind modern deep learning strategies because the internet has enabled humanity to organize and aggregate massive amounts of data.[58] Indeed, the explosion in data collection since the inception of the internet continues to result in increasingly available data, as well as improved deep learning applications and models.[59] Critically, every day humans create five exabytes of data,[60] as much data as civilization created from the dawn of time until 1999.[61] This is particularly important because the data – not human programmers – drive progress in deep learning applications.[62] Generally, deep learning systems are developed in four parts: data pre-processing, model design, training, and testing.

The majority of the time spent with deep learning system development is during the pre-processing stage.[63] During this initial phase, machine learning researchers gather, organize, and aggregate data to be analyzed by neural networks.[64] The types of data neural networks process

---

51.   Simon, *supra* note 19 at 254; *see also* ALPAYDIN, *supra* note 49, at 88-90.
52.   U.S. Patent No. 9471884 (assigned to IBM).
53.   MOHEB COSTANDI, NEUROPLASTICITY 6 (2016).
54.   *Id.* at 9.
55.   *Id.* at 7.
56.   RASCHKA & VAHID MIRJALILI, *supra* note 5, at 18.
57.   Brian S. Haney, *The Perils & Promises of Artificial General Intelligence*, 45 J. LEGIS. 151, 157 (2018) (Data are a digital representation of information about the world).
58.   RICHARD SUSSKIND, TOMORROW'S LAWYERS 11 (2017).
59.   DENNING & TEDRE, *supra* note 4, at 93.
60.   An exabyte is 10^{18} or one quintillion byte.
61.   SUSSKIND, *supra* note 58, at 11.
62.   *Id*
63.   JOHN D. KELLEHER & BRENDEN TIERNEY, DATA SCIENCE 97 (2018).
64.   *Id*

vary.[65] In the context of autonomous warfare systems, one example may be images stored as pixel values to be associated with object classification for targeting.[66] The data's organization is in large part dependent on the goal for a deep learning system. If a system is being developed for predictive purposes, the data may be labeled with positive and negative instances of an occurrence.[67] Or, if the system is being learned to gain insight, the data may remain unstructured, allowing the model to complete the organization task.[68]

ii. Model

A deep learning system's model is the part of the system which analyzes the information.[69] The most common deep learning model is the artificial neural network.[70] An artificial neural network is an organized structure of interconnected neurons.[71] Every neural network has an input layer and an output layer.[72] The depth of the model is defined by the number of layers between the input and output layer.[73] Figure 4 is a shallow neural network with one hidden layer.

---

65. *Id.* at 100.

66. Richard Wu, et al., *A Framework Using Machine Vision and Deep Reinforcement Learning for Self-Learning Moving Objects in a Virtual Environment*, AAAI 2017 FALL SYMP. SERIES (2017), https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16003/15319.

67. ALPAYDIN, *supra* note 49, at 68.

68. Alec Radford, et. al., *Language Models are Unsupervised Multitask Learners*, OPENAI (2019), https://openai.com/blog/better-language-models/ (providing a method by which the structures of human language may be learned through unsupervised machine learning).

69. KELLEHER & TIERNEY, *supra* note 63, at 121;sSee also KERAS:THE PYTHON DEEP LEARNING LIBRARY, https://keras.io/ for code for layered neural networks. Keras is an Application Programming Interface (API) written on top of Google's Tensforflow.

70. TEGMARK, *supra* note 14, at 76.

71. EUGENE CHARNIAK, INTRODUCTION TO DEEP LEARNING 8-9 (2018) (The network's interconnected neurons are modeled with weight coefficients, while learning algorithms adjust the weights between neurons until a model is optimized for performance. Typically, matrix multiplication and partial derivative calculations are the learning algorithm's mathematical core. Importantly, neural networks are universal function approximators, meaning they can approximate any function with desired accuracy given enough perceptrons); *see also* U.S. Patent No. 10,146,286 (Dec. 4, 2018) (assigned to Intel Corporation).

72. KELLEHER, *supra* note 16, at 68.

73. TEGMARK, *supra* note 14, at 76.
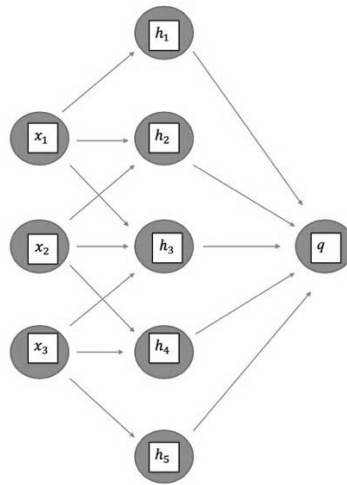
Figure 4[74]

Each layer of hidden neurons[75] acts as a feature extractor by providing analysis of slightly more complicated features.[76] Feature extraction is a method of dimensionality reduction—decreasing input attributes—allowing the observable manifestation of hidden features.[77] The later neurons extract hidden features by combining the previous features of a slightly larger number of neurons.[78] Finally, the output layer observes the whole input to produce a final prediction.[79] In other words, deep neural networks learn more complicated functions of their initial input when each hidden layer combines the values of the preceding layer.[80]

Interestingly, deep neural networks may be used for both supervised[81] and unsupervised learning tasks.[82] In unsupervised learning a deep neural network may be used to recognize patterns in unstructured or

74. In figure 4, the *x* values represent neurons in the input layer, the *h* values represent the neurons in a hidden layer and the *q* value represents the output layer.

75. KELLEHER, *supra* note 16, at 68 (A deep neural network contains multiple hidden layers between the input and output layer).

76. ALPAYDIN, *supra* note 49, at 75.

77. *Id.* at 76.

78. U.S. Patent No. 10,540,588 (Jan. 21, 2020) (assigned to Microsoft).

79. U.S. Patent No. 10,467,495 (Nov. 5, 2019) (assigned to Siemens Healthcare).

80. *Id.*

81. U.S. Patent No. 7,395,251 (July 1, 2008) (assigned to International Business Machines Corporation).

82. U.S. Patent, No. 10,460,215 (Oct. 29, 2019) (assigned to Microsoft).

unlabeled data.[83] Unsupervised learning is critical for AI development because the majority of data on the internet is unlabeled.[84] In other words, unlabeled data is cheaper, more voluminous, and more readily available.[85] One example of an unsupervised learning task is clustering, which are commonly used for document classification and discovery during in law suits.[86]

Alternatively, in supervised learning neural networks make predictions about future occurrences.[87] For example, a supervised learning algorithm may be used for computer vision in an autonomous vehicle.[88] In such a case, the supervised learning algorithm may predict whether an object is a pedestrian or another object.[89] Depending on the algorithm's classification, the car is designed to take different actions to ensure driver, passenger, and bystander safety.[90] Supervised neural networks learn using pre-labeled data to minimize an error function.[91] In the context of driverless cars, the pre-labeled data may be examples of pedestrians and other objects.[92] During training, the neural network makes a prediction of value, which is measured against a pre-labeled true value.[93] Then, an error function calculates the error in a network's prediction, allowing for iterative updates minimizing the error rate.[94] The process of iterative improvement is accomplished with a backpropagation algorithm, perhaps the most critical element of deep learning systems.[95]

---

83.   KELLEHER, *supra* note 16, at 27.

84.   ALPAYDIN, *supra* note 49, at 117.

85*.   Id.*

86.   Document review automation is made possible because neural networks are able to learn patterns in unstructured data without human supervision. *Id.* at 112; *see also* Sergio David Becerra, *The Rise of Artificial Intelligence in the Legal Field: Where we are and Where we are Going*, 11 J. BUS. ENTREPRENEURSHIP & L. 27, 39 (2019); ASHLEY, *supra* note 20, at 239.

87.   Barry, *supra* note 21, at 354.

88.   Brian S. Haney, *The Optimal Agent: The Future of Autonomous Vehicles & Liability Theory*, 29 ALB. L.J. SCI. & TECH. (Forthcoming 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261275; *see also* U.S Patent 10,474,964 (Nov. 12, 2019) (assigned to Ford Global Technologies).

89.   Gary Marcus, Deep Learning: A Critical Appraisal 3 (Jan. 8, 2018) (unpublished research paper), (accessed at https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf).

90.   Damien Matti et al., *Combining LiDAR Space Clustering and Convolutional Neural Networks for Pedestrian Detection*, CORNELL U. 1, 3 (2017), https://arxiv.org/abs/1710.06160; *see also* U.S. Patent No. 10,061,316 (Aug. 28, 2018) (assigned to Toyota).

91.   RASCHKA & MIRJALILI, *supra* note 5, at 35-36.

92.   Matti, supra note 90, at 3.

93.   U.S. Patent No. 10,540,588 \y (Jan. 21, 2020) (assigned to Microsoft).

94.   Melissa Mortazavi, *Rulemaking Ex Machina*, 117 COLUM. L. REV. ONLINE 202, 209 (2017).

95.   Steven M Bellovin et al., *Privacy and Synthetic Datasets*, 22 STAN. TECH. L. REV. 1, 29 (2019).

iii. Backpropagation

In the 1970s and 1980s, researchers developed backpropagation as a way to train neural networks.[96] Backpropagation is an algorithm for updating the weights in a neural network, improving accuracy over time.[97] In other words, backpropagation is how neural networks learn.[98] Technically, backpropagation's central task is to minimize an error function.[99] The error function is minimized through an iterative process, updating the network's weights[100] toward a set of weights capable of generalizing to make accurate predictions for the whole data set.[101] After consistent iteration, the network converges, capturing a general pattern and allowing the network to generalize about new instances, rather than merely memorizing training data.[102]

There are variations of backpropagation algorithms.[103] More generally, a backpropagation algorithm has three steps: (1) an instance enters the network, flowing forward until the network generates a prediction;[104] (2) the network's error for the prediction is calculated by comparison to the correct output;[105] and (3) the error is propagated back through the network,

---

96.   ALPAYDIN, *supra* note 49, at 92.

97.   U.S. Patent No. 10,540,588 (Jan. 21, 2020) (assigned to Microsoft).

98.   Paul John Werbos is considered the first person to explore backpropagation through neural networks in his seminal 1974 Ph.D. thesis, *The Roots of Backpropagation*. One the key contributions of Werbos' work is the idea of backpropagation through time. By applying a temporal element to the process, Werbos showed the utility of neural networks in dynamic control tasks for robotics systems. *See* Paul John Werbos, The Roots of Backpropagation from Ordered Derivatives to Neural Networks and Political Forecasting 279-280 (1994).

99.   An error function is a measure of the difference between the network's output and the actual value associated with the instance. The backpropagation algorithm's objective is to minimize the error function. *See* U.S. Patent No. 10,346974 (July 9, 2019) (assigned to Toshiba Medical Systems Corporation); *see also* U.S. Patent No. 10,112,113 (Oct. 30, 2018) (assigned to Sony Computer Entertainment).

100.   The learning rate determines the pace at which the weights are updated. *See* RASCHKA & MIRJALILI, *supra* note 5, at 22.

101.   KELLEHER, *supra* note 16, 214-215.

102.   U.S. Patent No. 8,595,167 (Nov. 26, 2013) (assigned to Google).

103.   U.S. Patent No. 10,096,107 (Oct. 9, 2018). (assigned to Siemens Healthcare) (Discussing gradient descent methods); *see also 18.01SC Single Variable Calculus: Chain Rule*, MIT OPENCOURSEWARE, (2010) (A commonly used back propagation algorithm in NLP is the Chain Rule).

$$\lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta x} \cdot \frac{\Delta x}{\Delta t}.$$

Here, y is a function of x and x is a function t. The derivative of y with respect to t is $\lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t}$. In other words, the chain rule takes the dot product of the derivative of y with respect to x and the derivative x with respect to t).

104. KELLEHER & TIERNEY, *supra* note 63, 130.

105*. Id.*

updating the weights.[106] In other words, the essential function of the algorithm adjusts the weights of a neural network to reduce error.[107] The algorithm's ultimate goal is convergence to an optimal network, but probabilistic maximization also provides state-of-the-art performance in real world tasks.[108] While the backpropagation algorithm remains a foundational achievement in AI studies, a critical idea in deep learning remains; deep learning is about the data – not algorithms.[109]

B. Patents

i. By Year

Rina Dechter first introduced the term *deep learning* in the year 1986.[110] However, the first patent with the term appearing in a claim was not granted until the year 2014.[111] Since, then there has been a sudden and rapid growth in the number of patents granted each year with a claim to some deep learning application. Figure 5 depicts the number of patents granted each year by the USPTO.

---

106. Mathematically, backpropagation is a method of computing the partial derivatives of error functions in neural networks. The backpropagation algorithm's goal is to learn and optimize weight coefficients, defining the network's parameters. The algorithm iterates the network toward a set of weights producing a desirable result. *See* KELLEHER, *supra* note 16, 130, 214-215.

107. RASCHKA & MIRJALILI, supra note 5, at 35-36.

108. KELLEHER, *supra* note 16, at 215. *See also* Lise Getoor, Selectivity Estimation using Probabilistic Models 461, 462 (2001), https://dl.acm.org/doi/pdf/10.1145/375663.375727 (discussing probabilistic graphical models).

109. ALPAYDIN, *supra* note 49, at 10-11.

110. Rina Dechter, *Learning While Searching in Constraint-Satisfaction Problems*, AAAI-86 PROCEEDINGS (1986), https://www.aaai.org/Papers/AAAI/1986/AAAI86-029.pdf.

111. U.S. Patent No. 8,775,332 (July 8, 2014. (The first patent granted with the term deep learning appearing in a claim).

Deep Learning Patents by Year
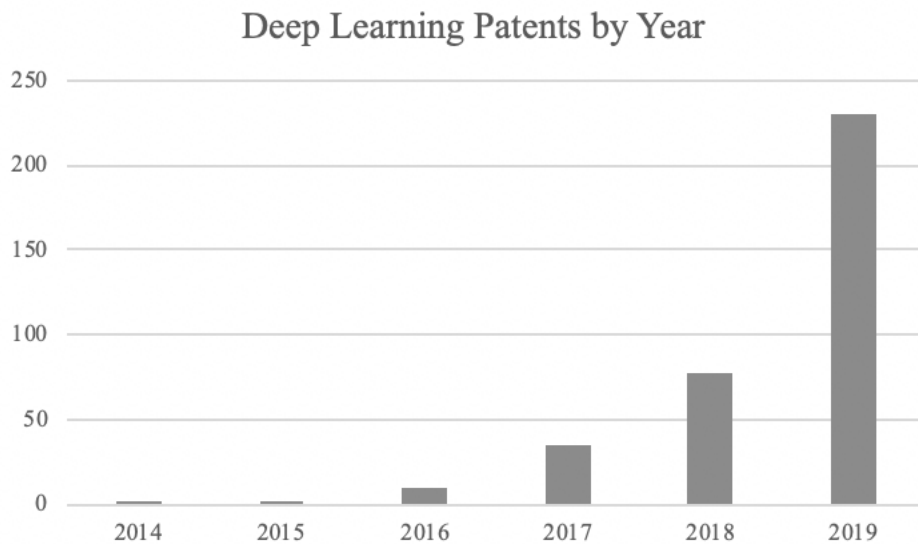
Figure 5[112]

Interestingly, in the year 2014, 2 deep learning patents were granted; in the year 2016, 9 deep learning patents were granted; in the year 2018, 77 deep learning patents were granted; and in the year 2019, 230 deep learning patents were granted. In fact, the number of patents issued have at least doubled each year since 2015.[113] The duration for which this trend will continue depends on a variety of factors. One argument is the deep learning patent marketplace is a rapidly growing element of the knowledge economy.[114]

ii. Market

The deep learning patent market apparently sprang out of nowhere. Consider in 2013 there were zero deep learning patents and by the end of 2019 there were 354.[115] Figure 6 graphs the deep learning patent market's growth since its inception – measured by total patents.

112. Brian S. Haney, Deep Learning Patents (2019) (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office) (A copy of the data is on file with the author).

113. *Id.*

114. JAMES W. CORTADA, INFORMATION AND THE MODERN CORPORATION 3-4 (2011) (discussing knowledge as a vital asset class for corporations).

115. Haney, *supra* note 112.

Deep Learning Patent Market Growth
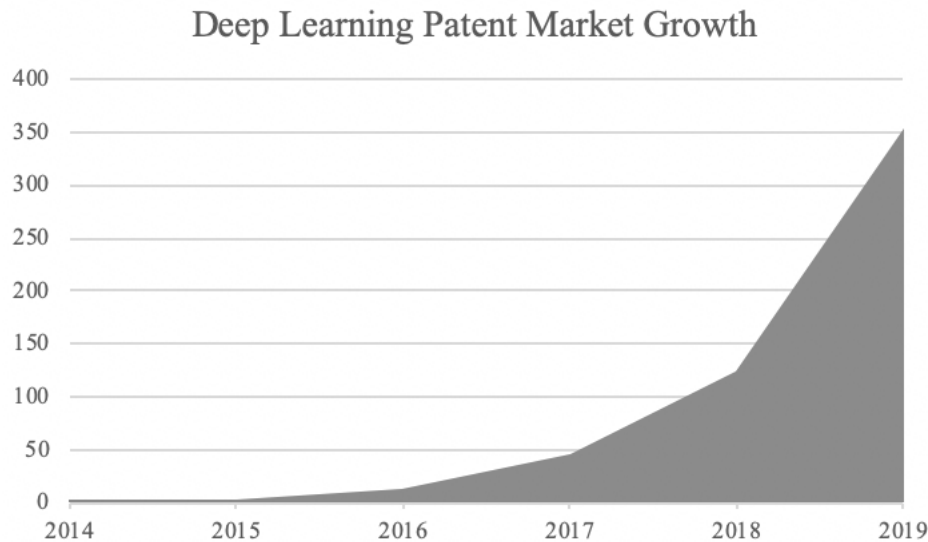
Figure 6[116]

The market grew from 2 patents in the year 2014, to 12 patents in the year 2016, to 124 patents in the year 2018. In considering this market trend, the rate of growth seems symbiotic with the Law of Accelerating Returns (LOAR), which states the price and performance of information technology follows a predictable exponential trajectory.[117] Deep learning is an information technology because it's essential function is data analysis for the derivation of knowledge.[118] As such, one may expect the market for deep learning patents to follow a similar trajectory to that of the information technology more generally.[119]

iii. Firms

The market for deep learning patents is a relatively diverse collection of technology companies. Figure 7 provides a sample of companies with deep learning patents.

---

116. *Id.*

117. RAY KURZWEIL, HOW TO CREATE A MIND 250 (2012).

118. KELLEHER, *supra* note 16, at 79; *see also* CORTADA, *supra* note 116, at 5 (arguing information is the most vital asset for the modern corporation).

119. This is just one of many market growth possibilities.

2020   *AI PATENTS: A DATA DRIVEN APPROACH*   425

Deep Learning Patents by Company
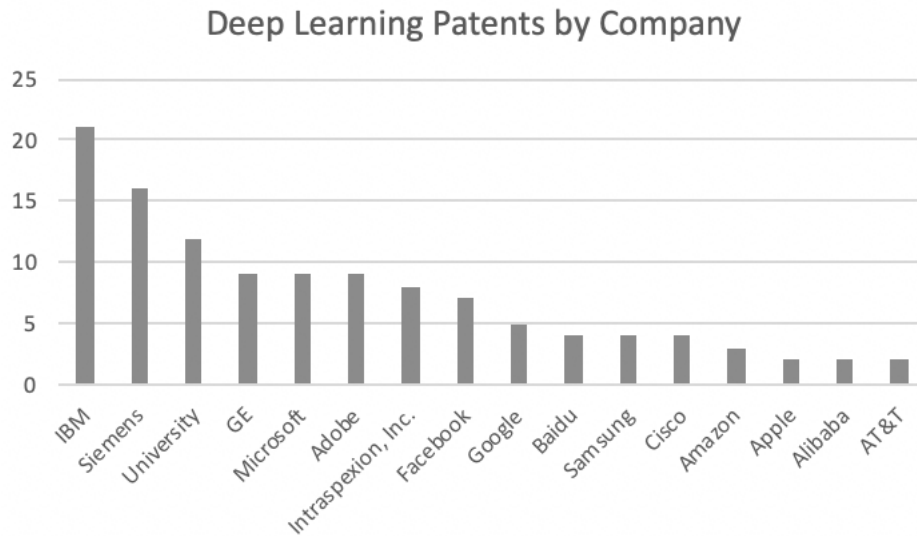
Figure 7[120]

Interestingly, International Business Machines (IBM) has the most deep learning patents to date with 21.[121] Universities own 12   deep learning patents.[122] Further, big technology companies Apple (2), Amazon (3), Google (5), Microsoft (9), and Facebook (7) all have established a modest market share.[123] Surprisingly, the multinational conglomerate Siemens AG (Siemens) holds the second most deep learning patents with 16.[124]

III.   REINFORCEMENT LEARNING

A. Technology

The roots of reinforcement learning date back to the early twentieth century and the work of Russian mathematician, Andrei Markov.[125] Markov's work in probability theory resulted in one of the twentieth century's most important ideas, the Markov Decision Process (MDP).[126] In short, the MDP is a statistical tool for predicting the future. MDPs trace the probabilistic

120. Haney, *supra* note 112.
121. *Id.*
122. *Id.*
123. *Id.*
124. *Id.*
125. Basharin, *supra* note 3, at 15.
126. GEORGE GILDER, LIFE AFTER GOOGLE 75 (2018).

transitions from one state to another through time.[127] Although Markov was a prominent figure in his time, his greatest influence was delayed nearly a century.[128] Today, Markovian techniques pervade the science of modern information theory.[129] Markov's models are used in search algorithms, machine translation, and financial trading.[130] And, the *Markov Decision Process* (MDP) remains the foundation of reinforcement learning.[131]

Reinforcement learning is a type of machine learning concerned with learning how an agent should behave in an environment to maximize a reward.[132] Agents are software programs making intelligent decisions.[133] The purpose of reinforcement learning algorithms is to learn how an agent should makes decisions.[134] Reinforcement learning is particularly important because of its unsupervised nature.[135] In other words, reinforcement learning algorithms learn without human supervisors.[136] Reinforcement learning algorithms contain three elements: (1) model: the description of the agent-environment relationship;[137] (2) reward: the agent's goal;[138] and (3) policy: the way in which the agent makes decisions.[139] In short, the goal of

---

127. Markov's brilliance was realized in his ability to describe the temporal dependencies between events across time. *See also* U.S. Patent No 9,858,171 (Jan. 2, 2018) (assigned to Google).

128. GILDER, *supra* note 126, at 76-77.

129. Basharin, *supra* note 3, at 4 (2004).

130. GILDER, *supra* note 126, at 82-88.

131. Brian S. Haney, Applied Artificial Intelligence in Modern Warfare & National Security Policy, 11 HASTINGS SCI. & TECH. L.J. (forthcoming 2019) (accessed at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3454204); *see also* U.S. Patent No. 10423129 (Sep. 24, 2019) (assigned to Massachusetts Institute of Technology).

132. MYKEL J. KOCHENDERFER, DECISION MAKING UNDER UNCERTAINTY 77 (2015). *See also* Leslie Pack Kaelbling, et al., Reinforcement Learning: A Survey, J. of Artificial Intelligence Research (1996), http://www.cse.msu.edu/~cse841/papers/kaelbling.pdf. (Surveying the field of reinforcement learning.) *See also* Leslie Pack Kaelbling, Learning in Embedded Systems (1990), https://apps.dtic.mil/dtic/tr/fulltext/u2/a323936.pdf.

133. RICHARD S. SUTTON & ANDREW G. BARTO, REINFORCEMENT LEARNING: AN INTRODUCTION 3 (2017).

134. CHARNIAK, *supra* note 71, at 113.

135. *Id.*

136. Alex Kendall, et. al., *Learning to Drive in A Day*, CORNELL U. (2018), https://arxiv.org/abs/1807.00412.

137. Katerina Fragkiadaki, *CMU: 10703:Deep Q Learning*, CARNEGIE MELLON SCH. COMPUTER SCI., (2018), https://www.cs.cmu.edu/~katef/DeepRLFall2018/lecture_DQL_katef2018.pdf.

138. LAPAN, *supra* note 5, at 3.

139. U.S. Patent No. 9,298,172 (Mar. 29, 2016) (assigned to International Business Machines Corporation); *see also* Fragkiadaki, *supra* note 137.

reinforcement learning is to identify and select the policy which maximizes expected reward for an agent acting in an environment.[140]

i. Model

Formally, reinforcement learning is described through an agent-environment interaction, with the MDP.[141] Figure 8 describes the agent-environment interaction in an MDP.
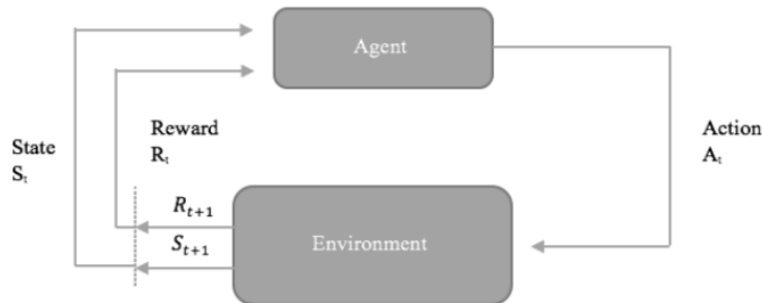
Figure 8[142]

In an MDP, the interaction begins when an agent chooses an action in the environment's initial state.[143] The model continues to the next state, where the agent receives a reward and a set of actions from which to choose, the agent selects an action, the environment returns a reward and the next state.[144] This process continues perpetually until the environment's final state.[145] Ultimately, in reinforcement learning an agent learns to take actions optimizing a reward.[146]

---

140. Jennifer Barry et al., *Quantum Partially Observable Markov Decision Processes*, 90 PHYSICAL REV. A, 032311-1, 032311-2 (2014), https://journals.aps.org/pra/abstract/10.1103/PhysRevA.90.032311.

141. Fabian Ruehle, *Data Science Applications to String Theory*, PHYSICS REPORTS 134 (https://doi.org/10.1016/j.physrep.2019.09.005.

142. SUTTON & BARTO, *supra* note 133, at 38 (model created by author based on illustration at the preceding citation); *see also* U.S. Patent No. 8,478,642 (July 2, 2013) (assigned to Carnegie Mellon University).

143. CHARNIAK, *supra* note 71, at 113.

144. *Id.*

145. Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 NATURE INT'L J. SCI. 529, 529 (2015).

146. Barry, *supra* note 140, at 032311-2.

In reinforcement learning, the environment[147] represents the problem.[148] For example, in robotics control systems, the environment is made up of states for moments in time in which the environment exists.[149] In other words, one way to think about states is that each state represents a moment in time.[150] Alternatively, in a trading algorithm the environment may be made up of a portfolio of stocks.[151]

An agent is an algorithm solving the environment or problem.[152] For example, in the case of autonomous vehicles, an agent may control the car's steering.[153] And, a second example is a trading algorithm, where the environment is a portfolio of stocks, an agent would be tasked with buying, selling, or staying at each interval of time.[154] Initially, the agent is presented with a state of the environment, which includes several possible actions.[155] Then, the agent takes an action in the present state advancing to the next state of the environment, where a reward associated with the chosen action is returned.[156] The agent's actions[157] in each state determine the environment's evolution, affecting future states.[158] In turn, the agent's actions affect the opportunities available to the agent at later states.[159] This line of analysis is intuitive. For example, the college one chooses to attend is an action taken in one state and it affects the opportunities available to one in later states.

147. *Id.* (Environments are made up of two types of space, state spaces and action spaces. There are two types of state spaces, observable and partially observable).

148. LAPAN, *supra* note 5, at 8; *see also* U.S. Patent No. 9,298,172 Method and apparatus for improved reward-based learning using adaptive distance metrics, Tesauro , et al. (March 29, 2016) (assigned to International Business Machines Corporation).

149. Kendall *supra* note 136..

150. LAPAN, *supra* note 5 at 20.

151. *Id.* at 217.

152. U. S. Patent No. 10,498,855 (assigned to Cisco Technology, Inc.).

153. Kendall, *supra* note 136.

154. LAPAN, *supra* note 5, at 217.

155. KOCHENDERFER, *supra* note 132, at 77; *see also* U.S. Patent No. 8,060,454 (Nov. (assigned to International Business Machines Corporation).

156. C.E. Shannon, *A Mathematical Theory of Communication*, 27 BELL SYS. TECHNICAL J. 1, 8 (1948).

157. *Part I: Key Concepts in RL*, SPINNING UP (2020). https://spinningup.openai.com/en/latest/spinningup/rl_intro.html (The action space is the set of all actions in a given environment. Generally, there are two types of action spaces, discrete and continuous.).

158. KOCHENDERFER, *supra* note 132, at 79; *see also* U.S. Patent No. 10,346,741 (Jul. 9, 2019) (assigned to DeepMind Technologies – a Google subsidiary).

159. KOCHENDERFER, *supra* note 132, at 79; *see also* U.S. Patent No. 10,346,741, *supra* note 158.

Ultimately, the agent's behavior is defined by two features, a reward and a policy.[160]

ii. Reward

The goal for any agent in an MDP is to maximize its expected rewarded during the episode.[161] In other words, the agent's goal is to maximize its total reward, rather than the reward for its immediate state.[162] The reward is a method of teaching the agent what it should do and is meant to formalize the idea of a goal.[163] For example, the reward for an agent playing a game of chess would be associated with winning the game.[164] The goal would be to allow the agent to make sacrifices for a particular move, reducing immediate reward, at the expense of increasing the probability of winning the overall game, the total reward.

Defining the reward for a reinforcement learning system is often one of the most challenging aspects of algorithmic development.[165] The reward is easier to describe for a task like missile control, where the agent need only take actions to minimize the missile's distance from the target.[166] However, in other tasks like writing, the reward is more difficult to define because good writing is not only subjective, but involves considerable abstraction on the part of the reader.[167] In other words, there isn't a formal list or method for describing what differentiates good writing from bad writing. The mechanics of reinforcement learning are better suited to optimize more objective metrics.[168]

---

160. Fragkiadaki, *supra* note 137.

161. Episode refers to the total experience of an agent progressing through an environment a terminal state. *See* U.S. Patent No. 10,498,855 (Dec. 3, 2019) (assigned to Cisco Technology, Inc.).

162. CHARNIAK, *supra* note 71, at 113.

163. *Id.*

164. LAPAN, *supra* note 5, at 21.

165. NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 239 (Reprt. ed. 2014); *see also* U.S. Patent No. 10,467,274 (Nov. 5, 2019) (assigned to Snap Inc.).

166. Rebecca Crootof, *Autonomous Weapons Systems and the Limits of Analogy*, 9 HARV. NAT'L SEC. J. 51, 59 (2018); *see also* Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare*, 7 IEEE Access 37432, 37438 (2019).

167. Ron Dolin, Measuring Legal Quality: Purposes, Principles, Properties, Procedures, and Problems (June 18, 2017) (unpublished manuscript) (on file with the Harvard Law School, Center on the Legal Profession at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2988647).

168. For example, in the contexts of missiles – minimize distance from target and time. *See* Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare*, 7 IEEE Access 37432, 37438 (2019).

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

The reward acts as a feedback mechanism, allowing the agent to learn independent of human training.[169] The rewards are used to update the agent's knowledge over time, so it learns to take actions returning the highest rewards.[170] For each time step, the reward is a number $r_t \in \mathbb{R}$, which is associated with a corresponding action.[171] The basic idea is to program rational agents that maximize reward in a given environment.[172] However, an important distinction in reinforcement learning is the relationship between reward and value.[173] The reward defines the response from taking an action in a given state, where the value refers to the total amount of reward over an episode.[174] In other words, reward is a measure of short-term gain and value is a measure of long-term reward.[175] The agent's policy determines the value the agent returns over the course of an episode.[176]

iii. Policy

A policy[177] is a mapping from states to probabilities for selecting actions.[178] In other words, a policy is the way in which an agent makes

---

169. CHARNIAK, *supra* note 71, at 10; *see also* U.S. Patent No. 8,595,167 to Grieve, et al., Predicting likelihood of a successful connection between unconnected users within a social network using a learning network (Nov. 26, 2013) (assigned to Google).

170. KOCHENDERFER, *supra* note 132, at 77.

171. *Id.* Formally, the principle of maximum reward is stated:

$$a^* = arg \max_a \mathrm{E}\, \mathrm{R}(s|a).$$

Here, a^* represents to action maximizing reward according to a reward function $\mathrm{R}(s|a)$, which defines the expected reward received from action a given state s. The principle of maximum reward states, a rational agent should choose the action maximizing expected reward and controls the agent's decision-making.

172. U.S. Patent No. 8,429,096 to Soundararajan, et al. Resource isolation through reinforcement learning (Apr. 23, 2013) (assigned to Amazon Technologies, Inc.).

173. CHARNIAK, *supra* note 71, at 113-14.

174. *Id.*

175. *Id.*

176. U.S. Patent Application No. 14/097,862 (filed Dec. 5, 2013).

177. Formally, the policy is represented as π. In general, there are two types of policies, deterministic and stochastic policies. In a deterministic policy, the state determines the action

$$a = \pi(s).$$

In a stochastic policy, the agent randomly decides each action:

$$\pi(a|s) = \mathbb{P}[a|s].$$

The goal for a given environment is to find the optimal policy, π^* which maximizes the agent's reward in an episode. *See* Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 NATURE INT'L J. SCI. 529, 529 (2015); *see also* U.S. Patent No. 8,478,642, (July 2, 2013) (assigned to Carnegie Mellon University).

decisions.[179] For example, a greedy person has a policy routinely guiding their decision making to choose the action returning the highest dollar value.[180] Alternatively, a great athlete has a policy guiding their decision making toward taking actions to excel in their respective sport like weight lifting, practice, or seeking out the best coaches. The goal for reinforcement learning is to develop a policy allowing the agent to maximize the value it returns for a given episode.[181]

One of the main challenges in reinforcement learning is balancing exploration for new rewards and exploitation of learned rewards.[182] In other words, an agent must prefer actions it has found to be effective in producing rewards, but it also must try new actions to discover the environment's best rewards.[183] So, the agent has to exploit its knowledge to gain rewards, but also has to explore to take better actions in the future.[184] Thus, the agent tries a variety of actions, both stochastically and deterministically, progressively favoring those that return the best value.[185]

Generally, an optimal policy is developed to maximize value.[186] A value function[187] is used to compute the value of a given state according to a defined policy.[188] Policy evaluation is the process of computing the expected

---

178. KOCHENDERFER, *supra* note 132, at 80.

179. *Id.*

180. Brian S. Haney, *The Perils and Promises of Artificial General Intelligence*, 45 J. LEGIS. 151, 161 (2018).

181. CHARNIAK, *supra* note 71, at 114-15.

182. MARVIN MINSKY, SOCIETY OF MIND 76 (1986).

183. *Id.*

184. U.S. Patent No. 7,395,252 (July 1, 2008) (assigned to The Trustees of Columbia University in the City of New York).

185. U.S. Patent No. 10,296,004 (May 21, 2019) (assigned to Toyota).

186. WERBOS, *supra* note 98, at 306.

187. A value function is used to compute the value of a given state according to a defined policy. The value function V^π is equal to the expected sum of the discounted rewards for executing policy π:

$V^\pi(s) = \mathrm{E}[R(s_0) + \gamma R(s_1) + \cdots | s_0 = s, \pi(s)].$

The expected future rewards are discounted with a discount factor γ. The discount factor is typically defined:

$0 < \gamma < 1,$

allowing present rewards to have higher value. The discount factor determines the importance of future rewards. *See* Ahmad El Sallab et al., *Deep Reinforcement Learning Framework for Autonomous Driving*, CORNELL U. (2017), https://arxiv.org/pdf/1704.02532.pdf.

188. U.S. Patent No. 8,060,454 (Nov. 15, 2011) (assigned International Business Machines Corporation).

reward from executing a policy in a given environment.[189] Policy evaluation can be used in a general process called policy iteration[190] for computing an optimal policy.[191] Policy iteration is effective because the number of policies for an agent in an MDP are finite.[192] Thus, the iterative process of updating policies must converge to an optimal policy and optimal value function in a finite number of iterations.[193]

B. Patents

i. By Year

As a concept, reinforcement learning is between forty and fifty years older than deep learning's earliest roots.[194] Figure 9 graphs the number of reinforcement learning patents granted by year.
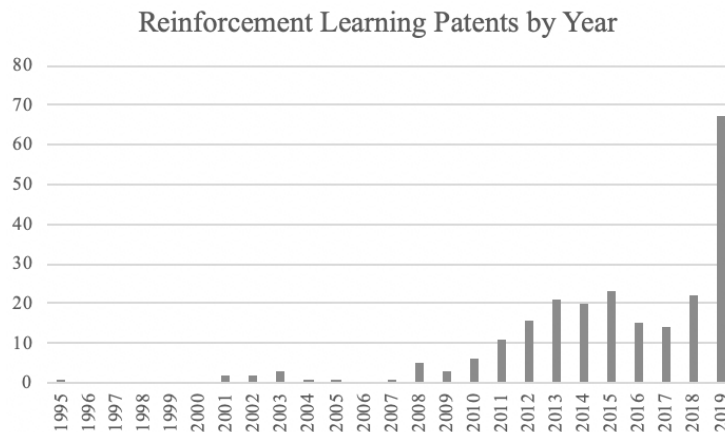
Figure 9[195]

---

189. KOCHENDERFER, *supra* note 132, at 80.

190. Policy iteration is a method of finding the optimal policy by continuously evaluating and improving the policy.

191. U.S. Patent No. 8,468,041 (June 18, 2013) (assigned to Oracle America, Inc.).

192. KOCHENDERFER, *supra* note 132, at 81.

193. U.S. Patent No. 9,661,019 (May 23, 2017) (assigned to Oracle International Corporation).

194. Reinforcement learning was conceived somewhere between 1905 and 1913, where deep learning's origins began somewhere between 1948 and 1957.

195. Brian S. Haney, Reinforcement Learning Patents (2019). (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office) (A copy of the data is on file with the author).

Compared to deep learning patents, the rate at which the USPTO is granting reinforcement learning patents is irregular. In the year 1995, one reinforcement learning patent was granted; in the year 2000, zero reinforcement learning patents were granted; in the year 2005, one reinforcement learning patent was granted; in the year 2010, 6 reinforcement learning patents were granted; in the year 2015, 23 reinforcement learning patents were granted. Yet, In the year 2019, the number of number of patents granted (67) was more than triple the previous year (22) and more than the previous three years combined (51).[196]

ii. Market

The reinforcement learning patent market has seen consistent growth since its inception in the year 1995.[197] Figure 10 graphs the reinforcement learning patent market's growth – measured by total patents.
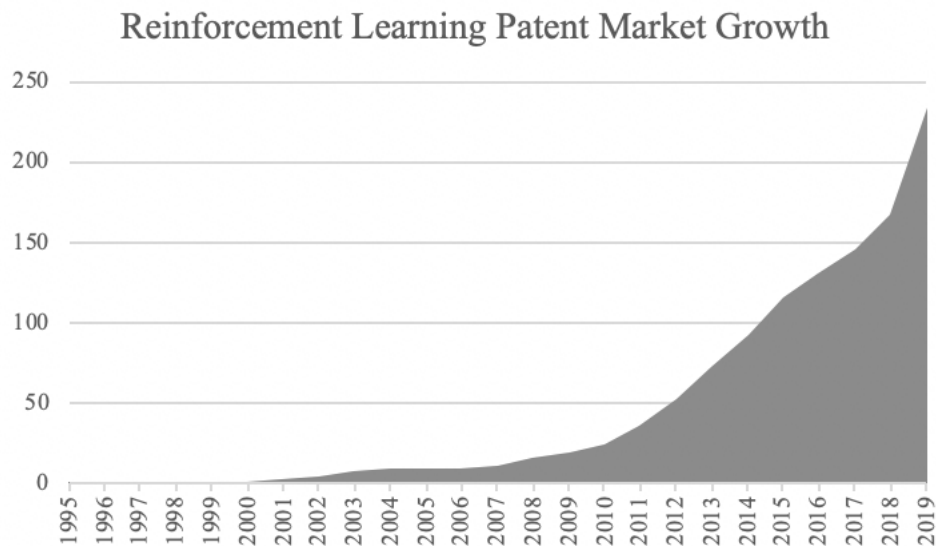
Figure 10[198]

The market's growth until the year 2010 was relatively linear, with the year 2011 providing the first noticeable departure toward a more accelerated growth.[199] In the year 2012 the total market included 52 patents; in the year

196. *Id.*
197. *Id.*
198. *Id.*
199. *Id.*

2014 the total market included 93 patents; in the year 2016 the total market included 131 patents; and in the year 2018 the total market included 167 patents. The year 2019 brought a significant increase in market size, moving from 167 patents in the year 2018 to 234 patents in 2019.[200]

iii. Firms

The reinforcement learning patent market is less diverse than the deep learning patent market. Figure 11 provides a sample of firms with a stake in the reinforcement learning patent market.
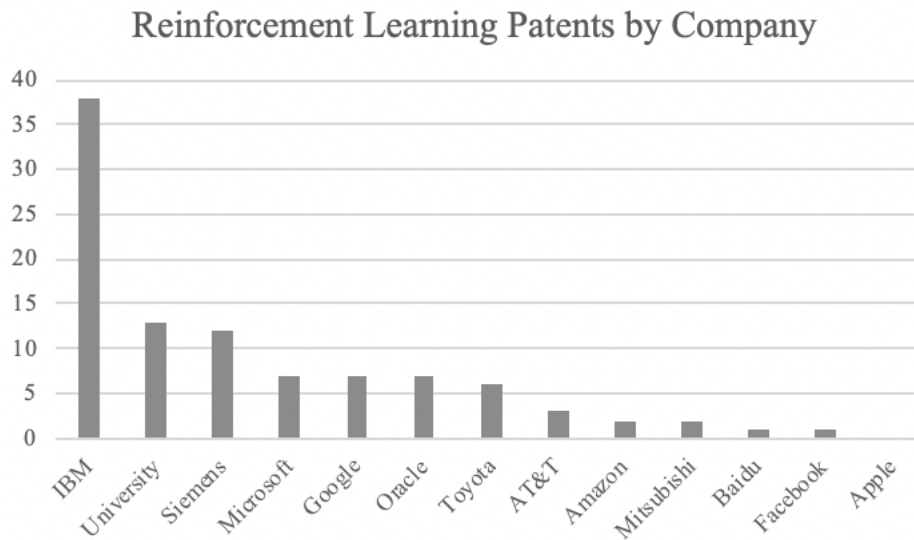
Figure 11[201]

IBM has a stronghold on the reinforcement learning patent market, owning 38 of 234 patents.[202] While universities own 13 thirteen patents, the next closest corporate actor is Siemens (12), followed by Microsoft (7), Google (7) and Oracle (7).[203] Interestingly, Apple has not laid a stake in this market despite being one of the world's leading technology companies.[204]

---

200. *Id.*
201. Haney, *supra* note 195.
202. *Id.*
203. *Id.*
204. *Id.* Apple's value is over $1 trillion.

IV.   Deep Reinforcement Learning

A. Technology

The integration of deep learning and reinforcement learning is the cutting edge in AI research.[205] Deep Reinforcement Learning is an intelligence technique combining deep learning and reinforcement learning.[206] The assimilation of the two systems began in literature during the 1980s with the work of Paul John Werbos at Harvard.[207] Werbos later patented his designs and remains one of the AI's most influential figures.[208] However, Max Tegmark suggests the deep reinforcement learning model was not implemented as computer code until 2013 in Volodymyr Mnih's seminal piece – *Human Level Control Through Reinforcement Learning*.[209] A researcher at Google's Deep Mind, Mnih's work was a major breakthrough for AI.[210]

Arguably, deep reinforcement learning is a method of general intelligence because of its theoretic capability to solve any continuous control task.[211] For example, deep reinforcement learning systems show state-of-the-art performance in tasks such as collision avoidance in driverless cars, automated landing systems for aerial vehicles, and autonomous weapons control.[212] However, deep reinforcement learning algorithms show poorer performance on other types of tasks like writing, because mastery of human language is – for now – not describable as a continuous control problem.[213]

---

205. Tegmark, *supra* note 14, at 39 (2017).

206. Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 Nature Int'l J. Sci. 529, 529 (2015).

207. Werbos, *supra* note 98 at 306.

208. Neural Networks for Intelligent Control, U.S. Patent No. 6,882,992 (Apr. 19, 2005).

209. Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 Nature Int'l J. Sci. 529, 529 (2015). *See also* Volodymyr Mnih, Koray Kavukcuoglu, Methods and Apparatus for Reinforcement Learning, U.S. Patent Application No. 14/097,862 at 5 (filed Dec. 5, 2013), https://patents.google.com/patent/US20150100530A1/en. *See also* Volodymyr Mnih, Koray Kavukcuoglu, Methods and Apparatus for Reinforcement Learning, U.S. Patent No. 9,679,258 B2 (2017) (https://patents.google.com/patent/US9679258B2/en).

210. *Id*. *See also* United States Patent No. 10,346,741 to Mnih, et al. Asynchronous deep reinforcement learning (July 9, 2019) (Assigned to DeepMind Technologies – a Google subsidiary).

211. Tegmark, *Supra* note 14, at 39 (2017).

212. Alex Kendall, et. al., *Learning to Drive in A Day* (2018), https://arxiv.org/abs/1807.00412. *See also* U.S Patent 10,474,964 to Micks, et al. Training algorithm for collision avoidance (Nov. 12, 2019) (assigned to Ford Global Technologies) (Deep neural network models for collision avoidance).

213. Noam Chomsky, Syntactic Structures 17 (1957).

Regardless of its scalable nature toward general intelligence, deep reinforcement learning is a powerful type of AI.[214] Generally, there are three different frameworks for deep reinforcement learning: action-value, policy gradient, and actor-critic.[215]

i. Deep Q-Network

An example of an action-value framework for deep reinforcement learning algorithm is the Deep Q-Network (DQN).[216] The DQN algorithm is a type of model-free-learning.[217] In model-free-learning, the agent randomly explores the environment, gathering information about the environment's states, actions, and rewards.[218] All the while, the agent stores the information in memory, called experience.[219] The DQN is perhaps the most important deep reinforcement learning algorithm in research and is discussed at length in many AI patents.[220]

The DQN algorithm develops an optimal policy[221] for an agent with a Q-learning algorithm.[222] More specifically, the DQN algorithm combines Q-learning[223] with a neural network to maximize an agent's reward.[224] The DQN

---

214. TEGMARK, *supra* note 14, at 39 (2017).

215. Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare,* IEEE Access Vol. 7, 37432, 37438 (2019).

216. *See* Yuval Tassa, et. al., *DeepMind Control Suite*, 12 (January 3, 2018)(https://arxiv.org/abs/1801.00690).(The Deep Mind Control Suit is a set of tasks for benchmarking continuous RL algorithms developed by Google Deep Mind) *See also* U.S. Patent No. 10,296,830, to Cai, et al. Dynamic topic guidance in the context of multi-round conversation (May 21, 2019). (Assigned to International Business Machines Corporation).

217. KOCHENDERFER, *supra* note 132 at 121-122. In model-free-learning, there isn't a formal description of the agent-environment relationship.

218. LAPAN, *supra* note 5 at 127.

219. CHARNIAK, *supra* note 71 at 133.

220. Many patents discuss the DQN algorithm. However, U.S. Patent No. 10,296,830 and its sister U.S. Patent No. 10,296,832, are the only two patents with claims including a DQN. *See also* U.S. Patent No. 10,032,281, Multi-scale deep reinforcement machine learning for N-dimensional segmentation in medical imaging (July 24, 2018), Ghesu, et al. (Assigned to Siemens Healthcare) *See also* U.S. Patent No. 10,296,830, to Cai, et al. Dynamic topic guidance in the context of multi-round conversation (May 21, 2019) (assigned to International Business Machines Corporation).

221. KOCHENDERFER, *supra* note 132 at 81. The optimal policy is the best method of decision making for an agent with the goal of maximizing reward.

222. Volodymyr Mnih et al., Human-Level Control Through Deep Reinforcement Learning, 518 NATURE INT'L J. SCI. 529, 529 (2015).

223. *Q-Learning* is a model-free reinforcement learning technique; it does not require an environment to learn stochastic transitions. *See* Brian S. Haney, *The Perils & Promises of Artificial General Intelligence*, 45 J. LEGIS. 151, 162 (2018). *See also* U.S. Patent No. 8,060,454 to Das, et al.,

algorithm's most important aspect is the Bellman Equation.[225] The Bellman Equation does two things; it defines the optimal policy and allows the agent to consider the reward in its present state as greater relative to similar rewards in future states. In other words, the Bellman Equation is a Q-learning algorithm defining the optimal policy by expressing the relationship between the value of a state and the values of future states.[226] However, the Bellman Equation is a slower algorithm in practice and can be computationally expensive.

Thus, a neural network is used as an approximator for a state-action value function, allowing for more efficient programming and model development.[227] After the optimal policy is defined, the agent engages in the exploitation of its environment.[228] During the exploitation phase, the agent maximizes its reward by making decisions according to the optimal policy.[229] The DQN is an off-policy algorithm, meaning it uses data to optimize

Method and apparatus for improved reward-based learning using nonlinear dimensionality reduction (Nov. 15, 2011) (assigned to International Business Machines Corporation).

224. WERBOS, *supra* note 98 at 306-307.

225. The algorithm continues perpetually until the convergence of the Q-value function. The convergence of the Q-value function represents Q^* and satisfies the Bellman Equation, defined:

$$Q^*(s,a) = E_{s'\sim\varepsilon}\left[r + \gamma \max_{a'} Q^*(s',a')|s,a\right].$$

Here, $E_{s'\sim\varepsilon}$ refers to the expectation for all states, r is the reward, γ is a discount factor. Additionally, the max function describes an action at which the Q-value function takes its maximal value for each state-action pair. An agent's optimal policy π^* corresponds to taking the action in each state defined by Q^*. *See also* U.S. Patent No. 8,060,454 to Das, et al., Method and apparatus for improved reward-based learning using nonlinear dimensionality reduction (November 15, 2011) (assigned International Business Machines Corporation) (Claim 14 and claim 23 both discuss applications of Bellman equations for optimality).

226. Volodymyr Mnih, Koray Kavukcuoglu, Methods and Apparatus for Reinforcement Learning, U.S. Patent Application No. 14/097,862 at 5 (filed Dec. 5, 2013), https://patents.google.com/patent/US20150100530A1/en.

227. However, one issue that arises is that the value of $Q(s,a)$ must be computed for every state-action pair, which may be computationally infeasible. For example, computing the value of every state-action pair, where the raw input is pixels in an Atari game would require tremendous computational power. One solution is to use a function approximator to estimate the Q-value function:

$$Q(s,a;\emptyset) \approx (s,a).$$

Here, $\emptyset$ represents the function parameters. Thus, the Q-value correlates with an optimal policy, telling the agent which actions to take in any given state. *See* Volodymyr Mnih, Koray Kavukcuoglu, Methods and Apparatus for Reinforcement Learning, U.S. Patent Application No. 14/097,862 at 5 (filed Dec. 5, 2013), https://patents.google.com/patent/US20150100530A1/en.

228. LAPAN, *supra* note 5 at 127.

229. *Id.*

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

performance.[230] Indeed, DQN is essentially a reinforcement learning algorithm, where the agent uses a neural network to decide which actions to take.

ii. Proximal Policy Optimization

A second variant of deep reinforcement learning is the Proximal Policy Optimization ("PPO") algorithm, a gradient technique.[231] Similar to the DQN algorithm, the PPO algorithm is a method of model-free learning.[232] In contrast to the DQN algorithm, PPO is an on-policy algorithm, meaning it does not learn from old data and instead directly optimizes policy performance.[233] One advantage of the PPO model is that it can be used for environments with either discrete or continuous action spaces.[234] In general, PPO works by computing policy gradient estimation and iterating with a stochastic gradient optimization algorithm.[235] In other words, the algorithm continuously updates the agent's policy based on the old policy's performance.[236]

---

230. Hado van Hasselt, Arthur Guez, and David Silver, *Deep Reinforcement Learning with Q-Learning*, Google DeepMind, 2098 (2018) https://arxiv.org/abs/1509.06461.

231. John Schulman, et al., High-Dimensional Continuous Control Using Generalized Advantage Estimation (2016), https://arxiv.org/abs/1506.02438. *See also* Brian S. Haney, Open_AI_Lunar_Lander,   GitHub https://github.com/Bhaney44/OpenAI_Lunar_Lander/blob/master/LunerLader_4.py   (Code example for proximal policy optimization algorithm for lunar lander in 2D).

232. U.S. Patent No. 10,146,286, to Lee, et al., Dynamically updating a power management policy of a processor (December 4, 2018) (assigned to Intel Corporation) (describing ". . .a portable, adaptive and model-free RL approach. . .").

233. OpenAI, Proximal Policy Optimization, OpenAI Spinning UP (2018) https://spinningup.openai.com/en/latest/algorithms/ppo.html.

234. *Id.*

235. John Schulman, et al., Proximal Policy Optimization Algorithms, OpenAI at 2 (2017), https://arxiv.org/abs/1707.06347. *See also* U.S. Patent No. 10,467,274, to Ren, et al. Deep reinforcement learning-based captioning with embedding reward (November 5, 2019) (assigned to Snap Inc.).

236. *Id.* The PPO update algorithm may be defined:

$$\theta_{k+1} = \arg\max_{\theta} \mathbb{E}_{s,a\sim\pi_{\theta_k}} [L(s,a,\theta_k,\theta)].$$

Here, $L(s,a,\theta_k,\theta)$ is the objective function, θ are the policy parameters, θ_k are the policy parameters for k experiment. Generally, the PPO update is a method of incremental improvement for a policy's expected return. *See also* U.S. Patent No. 10,467,274, to Ren, et al. Deep reinforcement learning-based captioning with embedding reward (November 5, 2019) (Assigned to Snap Inc.); *see also* U.S. Patent No. 8,478,642, System, method and device for predicting navigational decision-making behavior (July 2, 2013) (assigned to Carnegie Mellon University) (describing Stochastic Exponentiated Gradient Ascent); s*ee also* United States Patent No.

2020 *AI PATENTS: A DATA DRIVEN APPROACH* 439

The PPO algorithm's key to the success is obtaining good estimates of an advantage function.[237] The advantage function describes the advantage of a particular policy relative to another policy.[238] The algorithm's goal is to make the largest possible improvement on a policy, without stepping so far as to cause performance collapse.[239] To achieve this goal, PPO relies on clipping the objective function to remove incentives for the new policy to step far from the old policy.[240] In essence, the clipping serves as a regularizer, minimizing incentives for the policy to change dramatically.[241]

iii. Deep Deterministic Policy Gradient

A third deep reinforcement learning variant and an example of the actor-critic[242] framework is the Deep Deterministic Policy Gradient ("DDPG") algorithm.[243] Like both DQN and PPO, DDPG is a model-free learning

10,346,741 to Mnih, et al. Asynchronous deep reinforcement learning (July 9, 2019) (assigned to DeepMind Technologies – a Google subsidiary).

237. *Id*.

238. For example, if the advantage for the state-action pair is positive, the objective reduces to:

$$L(s, a, \theta_k, \theta) = min\left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}, (1 + \epsilon)\right) A^{\pi_{\theta_k}}(s, a).$$

Here, $A^{\pi_{\theta_k}}$ is the advantage estimate for the policy given parameters $\pi_\theta(a|s)$, and the hyperparameter ϵ corresponds to how far away the new policy can step from the old while still profiting the objective. Where the advantage is positive the objective increases and the *min* function puts a limit to how much the objective can increase. The limitation on the objective increase is called clipping.

239. Brian S. Haney, *Applied Artificial Intelligence in Modern Warfare & National Security Policy*, 11 HASTINGS SCI. & TECH. L.J. 61, 73 (2019).

240. JOHN SCHULMAN, ET AL., HIGH-DIMENSIONAL CONTINUOUS CONTROL USING GENERALIZED ADVANTAGE ESTIMATION (2016), https://arxiv.org/abs/1506.02438.

241. Haney, *supra* note 239 at 73.

242. See U.S. Patent No. 9,134,707 to Vamvoudakis, et al., Optimal online adaptive controller (September 15, 2015) (assigned to Board of Regents, The University of Texas System).

243. U.S. Patent No. 10,061,316 to Nishi, Control policy learning and vehicle control method based on reinforcement learning without active exploration (August 28, 2018) (assigned to Toyota); *see also* U.S. Patent No. 8,060,454 to Das, et al., Method and apparatus for improved reward-based learning using nonlinear dimensionality reduction (November 15, 2011) (assigned International Business Machines Corporation); TensorFLow, GitHub, DDPG (2020). https://github.com/tensorflow/agents/tree/master/tf_agents/agents/ddpg (Code for DDPG from TensorFlow under an Apache license).

method.[244] However, unlike PPO, DDPG is only applicable in continuous action spaces.[245] In form DDPG is relatively similar to DQN.[246] DDPG is an off-policy algorithm, meaning it re-uses old data.[247] Importantly, DDPG learns a deterministic policy.[248] In short, DDPG is a method of deep reinforcement learning using two function approximators,[249] an actor and a critic.[250]

Ultimately, the actor decides which action to take.[251] But, to optimize an agent's reward, after each action, the critic defines the necessary adjustment for performance improvement.[252] The DDPG algorithm shows promise in continuous control tasks for robotics systems.[253] For example,

---

244. TIMOTHY P. LILLICRAP, ET AL., CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING, 1 (2016), https://arxiv.org/abs/1509.02971.

245. Haney, *supra* note 239 at 73-74.

246. TIMOTHY P. LILLICRAP, ET AL., CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING, 1 (2016), https://arxiv.org/abs/1509.02971.

247. APRIT AGARWAL, KATHARINA MUELLING, KATERINA FRAGKIADAKI, MODEL LEARNING FOR LOOK-AHEAD EXPLORATION IN CONTINUOUS CONTROL, CARNEGIE MELLON UNIVERSITY, AAAI (2019), https://arxiv.org/abs/1811.08086.

248. DDPG learns a deterministic policy $\pi_\theta(s)$ which gives the action maximizing:
$$Q_\phi(s,a): \max_\theta \mathbb{E}_{s\sim\mathcal{D}}\big[Q_\phi\big(s,\pi_\theta(s)\big)\big].$$
Here, the Q-function parameters Q_ϕ are constants and $s\sim\mathcal{D}$ is the state sampled from the replay buffer. *See* Brian S. Haney, *Applied Artificial Intelligence in Modern Warfare & National Security Policy*, 11 HASTINGS SCI. & TECH. L.J. 61, 74 (2019).

249. *See* Control system and technique employing reinforcement learning having stability and learning phases, U.S. Patent No. 6,665,651 (filed Dec. 16, 2003) (assigned to Colorado State University Research Foundation) (Neural networks are known as universal function approximators).

250. The actor-critic framework may be thought of as dueling neural networks. The critic estimates the optimal action-value function $a^*(s)$. Generally, the action-value function is tailored to continuous action spaces, defined:
$$a^*(s) = arg \max_a Q^*(s,a).$$

Here, the optimal action $a^*(s)$ is defined as a value of $Q^*(s,a)$ at which a takes it's optimal value according to the Bellman Equation. The critic's role is to minimize loss, typically using a means squared error function, or target network, which gives consistent target values. *See* U.S. Patent No. 8,060,454 to Das, et al., Method and apparatus for improved reward-based learning using nonlinear dimensionality reduction (November 15, 2011) (assigned International Business Machines Corporation) (Claim 14 and claim 23 both discuss applications of Bellman equations for optimality).

251. CHARNIAK, *supra* note 71 at 130.

252. ALEKSANDRA FAUST, OSCAR RAMIREZ, ET AL., PRM-RL: LONG-RANGE ROBOTIC NAVIGATION TASKS BY COMBINING REINFORCEMENT LEARNING AND SAMPLING-BASED PLANNING (2018) https://arxiv.org/abs/1710.03937v2.

253*. Id.*

DDPG has shown state-of-the-art success for self-driving cars.[254] However, the off-policy nature of the algorithm makes it much slower because it takes more computational power to train compared to the PPO and other on-policy algorithms. As computational hardware develops, quantum computers provide a faster method of computing than classical methods and may be able speed up off-policy machine learning algorithms.[255]

In sum, DQN, PPO, and DDPG are foundational algorithms for the state-of-the-art in AI technology.[256] While the mathematical models underlying these systems are not new,[257] their capabilities have shown rapid recent improvement.[258] Most importantly, these AI systems are capable of generalizing about information to make predictions and achieve goals.[259] As a result, deep reinforcement learning is transforming the foundations of the defense industry, national security threats, and global warfare.[260]

B. Patents

i. By Year

Interestingly, despite its conception in the 1980s, the first deep reinforcement learning patent was not granted until the year 2016.[261] Perhaps unsurprisingly, the patent was granted to IBM.[262] Figure 12 graphs the number of patents granted by year.

---

254. ALEX KENDALL, ET. AL., LEARNING TO DRIVE IN A DAY (2018), https://arxiv.org/abs/1807.00412.

255. JACOB BIAMONTE, ET. AL. QUANTUM MACHINE LEARNING 2 (2018) https://arxiv.org/abs/1611.09347.

256. United States Patent No. 10,346,741 to Mnih, et al. Asynchronous deep reinforcement learning (filed July 9, 2019) (Assigned to DeepMind Technologies – a Google subsidiary).

257. C.E. Shannon, *A Mathematical Theory of Communication*, Bell Systems Technical Journal (1948).

258. GILDER, *supra* note 126 at 75.

259. TEGMARK, *supra* note 14, at 85-86 (2017).

260. *See* Gregory C. Allen, *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security, Center for a New American Security* 1 (February 2019), https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy. *See also* Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare,* IEEE Access Vol. 7, 37432, 37438 (2019); *see also* youshixun, vCEW New model of cognitive electronic warfare with countermeasures, GitHub https://github.com/youshixun/vCEW (2019) (Open source code for deep reinforcement learning missile control systems sponsored by China).

261. U.S. Patent No. 9471884 (filed May 30, 2014).

262. *Id*.

Deep Reinforcement Learning Patents by Year
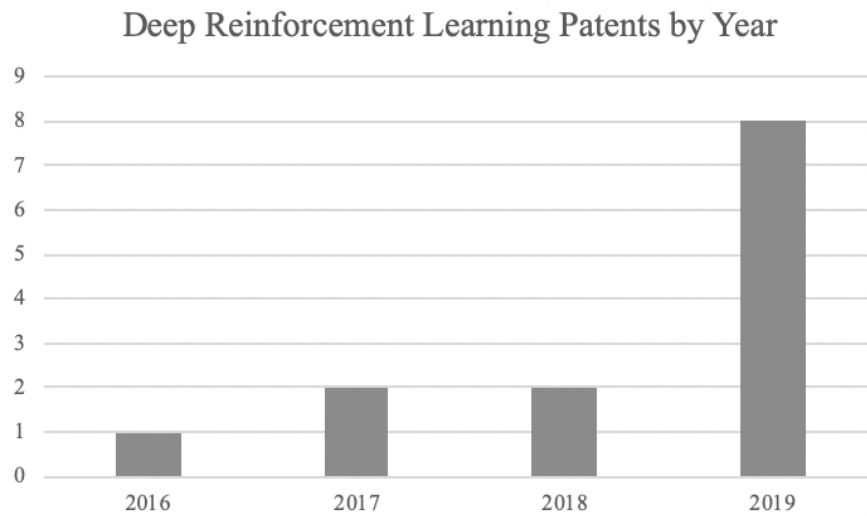
Figure 12[263]

However, the rate at which the USPTO is granting deep reinforcement learning patents appears to be accelerating.[264] Indeed, the number of patents granted in the year 2019 (8) is larger than every preceding year combined (5).[265]

ii. Market

The market for patents on technologies integrating deep learning and reinforcement learning is staunchly smaller than the patent market for the two technologies independently. Figure 13 graphs the reinforcement learning patent market's growth – measured by total patents.

---

263. BRIAN S. HANEY, DEEP REINFORCEMENT LEARNING PATENTS 39 (2019) (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office) (a copy of the data is on file with the author).

264. *Id.*

265. *Id.*

Deep Reinforcement Learning Patent Market Growth
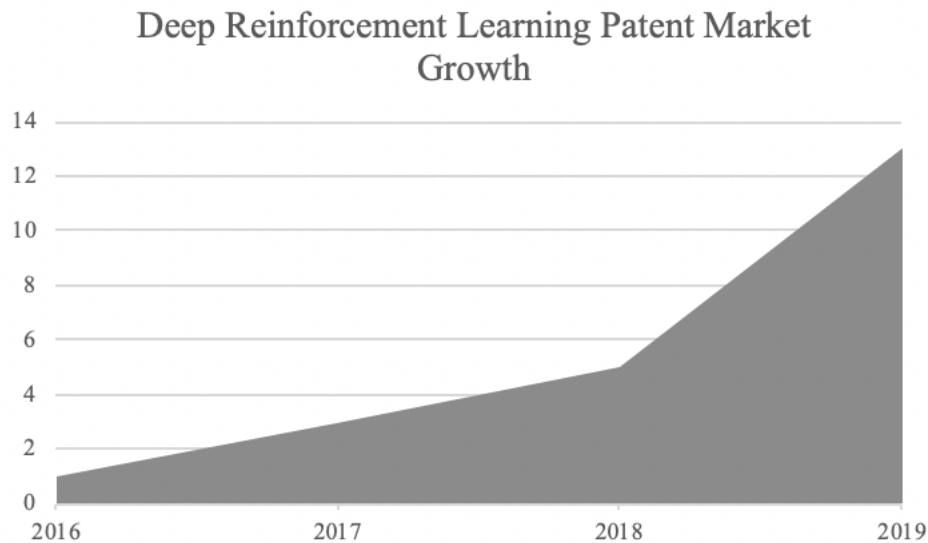
Figure 13[266]

Despite its smaller size, the deep reinforcement learning patent market appears to be following similar growth trends compared to deep learning and reinforcement learning patents.[267] In fact, from the year 2016 to the year 2019 the market grew from nothing to 13 total patents.[268] The 13 patents represent a relatively wide spectrum of industry, including healthcare, telecommunications, and robotics.[269]

iii. Firms

Interestingly, of the four AI patent markets surveyed in this Article, the deep reinforcement learning market is the only market not led by IBM. Figure 14 graphs a sample of firms with a stake in the deep reinforcement learning patent market.

---

266. Haney, *supra* note 263.

267. *Id.*

268. *Id.*

269. U.S. Patent No. 10049301 (filed Aug. 1, 2016); *see* U.S. Patent No. 10498855 (filed June 17, 2016); *see also* U.S. Patent No. 10375585 (filed July 6, 2017)*;* U.S. Patent No. 10416618 (filed July 29, 2016).

Deep Reinforcement Learning Patents by Company
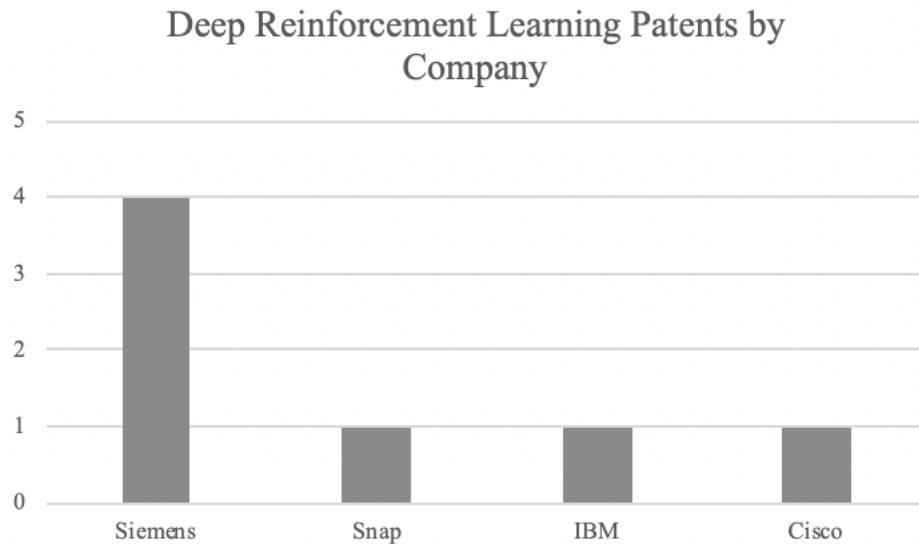
Figure 14[270]

Instead, the market is led by Siemens, the only firm with more than one patent.[271] All four of Siemens deep reinforcement learning patents relate to applications in healthcare and are held by a Siemens healthcare subsidiary.[272] Noticeably absent from the chart are big technology companies: Amazon, Apple, Facebook, Microsoft, and Google. Yet, Google and Microsoft have both developed significant research in deep reinforcement learning.[273]

V.   NATURAL LANGUAGE PROCESSING

A. Technology

Natural language processing (NLP) sits at the intersection of computer science, artificial intelligence, and computational linguistics.[274] NLP

---

270. HANEY 42, *supra* note 263.

271. *Id.*

272. U.S. Patent No. 10339695 (filed July 7, 2017)*; see also* U.S. Patent No. 10049301 (filed Aug 1, 2016)*;* U.S. Patent No. 10032281 (filed July 27, 2018)*;* U.S. Patent No. 9760690B1 (filed June 23, 2016).

273. U.S. Patent No. 10,540,588 (filed June 29, 2015); U.S. Patent No. 9,679,258 B2 (filed Dec. 5, 2013).

274. PENG LAI LI, *NATURAL LANGUAGE PROCESSING*, 1 GEO. L. TECH. REV. 98, 98 (2016)*; see also* U.S. Patent No. 10,445,429 (filed Jan. 10, 2018).

is the study of computational linguistics, which includes natural language understanding and natural language generation.[275] In other words, NLP uses formal logic to analyze the informal structures of human language.[276] Pattern recognition is fundamental to this practice.[277] Generally, NLP systems learn patterns from a text corpus, which is a body of natural language.[278] NLP studies strive to develop machines which process, understand, and generate language representations as well as humans.[279] However, language representation is a difficult task because human language interpretation depends on real world presence, common sense, and context.[280] Thus, NLP endeavors to bridge the divide enabling computers to analyze syntax, and process semantics.[281]

Modern theories of NLP began in the 1950s with the seminal work of Noam Chomsky.[282] Chomsky's key insight in *Syntactic Structures*, was the independence of grammar from semantics.[283] According to Chomsky, grammar is a device generating all of the grammatical sequences of a language and none of the ungrammatical devices.[284] And, grammar may be set up to include clear sentences and clear non-sentences.[285] Chomsky presents an example of sentence, which is grammatically correct, but lacks any meaning, "Colorless green ideas sleep furiously."[286] From this sentence, Chomsky

275. *Id.*

276. STEVEN BIRD, ET. AL., NATURAL LANGUAGE PROCESSING WITH PYTHON 39 (2009).

277. *Id.* at 221.

278. ASHLEY, *supra* note 20 at 234 (2017).

279. MILES BRUNDAGE, ET. AL. THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE: FORECASTING, PREVENTION, AND MITIGATION, 12 (2018), https://arxiv.org/pdf/1802.07228.pdf.

280. BIRD, ET. AL., *supra* note 276 at 32.

281. PENG LAI LI, 98 (2016). *See also* U.S. Patent No. 10,445,429 (filed Jan. 10, 2018).

282. NOAM CHOMSKY, SYNTACTIC STRUCTURES 34 (1957). *See also* C.E. Shannon, *A Mathematical Theory of Communication*, Bell Systems Technical Journal (1948). (http://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf) (Shannon's work provided key influence for Chomsky in Syntactic Structures and is an earlier example of NLP. However, Chomsky's work in Syntactic Structures is foundational to modern NLP theory in its discussion of generative grammar. Shannon's key contribution was modeling Markovian techniques for generating text sequences).

283. *Id.* at 17.

284. *Id.* at 13.

285. *Id.* at 14.

286. *Id.* at 15.

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

concluded grammar is independent of meaning.[287] As a result, Chomsky focused his analysis on rule-based language models.[288]

Generally, a language model is a probabilistic system for processing natural language.[289] In other words, a language model is a formalization of a language's sentences.[290] However, other language models have also been developed. For example, Zoltan Torey described language as a method of communicating the mind's percepts.[291] According to Torey, "Since percepts are private, first person experiences, they cannot be accessed, handled, or communicated without a carrier."[292] In Torey's language model, the word is a percept carrier, allowing the brain to generate mental experiences.[293] In the context of NLP, most language learning models can be understood as consisting of three elements: text corpora, vector space representations, and learning models.

i. Text Corpora

Language learning starts with problem definition and data collection.[294] NLP uses data in the form of a text corpus, which is a body of text commonly stored in various formats including SQL, CSV, TXT, or JSON.[295] The majority of time developing a deep learning system is spent on the pre-processing stage, aggregating and organizing the corpus.[296] During this initial

287. *Id.* at 15.

288. *Id.* at 17, 18. Chomsky was deeply opposed to probabilistic based models of language. Instead, he analyzed linguistic description in terms of a system with levels of representations. In large part, Chomsky's preferences for rule-based systems of language may have been due to the lack of data and computing resources available in the 1950s and 60s. Beginning in the 1980s, NLP research and development began to focus on statistics and probability models*; see also* PENG LAI LI, 99 (2016).

289. DEAN ALDERUCCI, *THE AUTOMATION OF LEGAL REASONING: CUSTOMIZED AI TECHNIQUES FOR THE PATENT FIELD*, DUQ. L.R. (2020) (Forthcoming) (on file with author) (Language modeling is a general technique that considers the word order for sentences and is used for in predicting the next word. Neural language models can use all words in a sentence or set of sentences to predict the sequences of words that likely precede or follow a word. Language modeling significantly increases the power of NLP systems to process text).

290. *Id.*

291. ZOLTAN TOREY, THE CONSCIOUS MIND 40 (2014).

292*. Id.*

293*. Id.*

294. David Lehr, Paul Ohm, *Playing with The Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 668 (2017).

295. JOHN D. KELLEHER, BRENDEN TIERNEY, DATA SCIENCE 10 (2018).

296. *Id.* at 65.

phase, machine learning researchers gather, organize, and aggregate data to be analyzed by neural networks.[297] How the data is organized is in large part dependent on the goal for the NLP system.[298] For example, in a system being developed for predictive purposes the data may be labeled with positive and negative instances of an occurrence.[299] The labels allow a supervised learning algorithm to learn how to classify future instances of data – making predictions.[300]

A critical component of corpora development is the normalization process. The normalization process allows the corpora to be consistent, readable, and searchable.[301] In general, normalization refers to the reduction of text toward a more basic or simplistic form.[302] For example, reducing all the text in a corpus to lowercase form is a method of normalization.[303] A second example of normalization is stemming.[304] Stemming refers to the process of stripping affixes from words, typically with regular expressions.[305] A third method of normalizing a raw text corpus is segmentation.[306] Text segmentation is the process of dividing written text into more meaningful units.[307] One way this may be accomplished is by representing characters with Boolean values, indicating word breaks.[308] The normalization process supports further preprocessing activity toward the development of a vector space model.[309] After a text corpus is adequately developed with normalization techniques it may be vectorized.

---

297. *Id*. at 1*; see also* U.S. Patent No. 10,445,429 to Sayed Ibrahim, et al., Natural language understanding using vocabularies with compressed serialized tries (October 15, 2019) (assigned to Apple Inc.).

298. BIRD, ET. AL., *supra* note 276 at 106.

299. ALPAYDIN, *supra* note 49 at 68.

300. *Id*.

301. BIRD, ET. AL., *supra* note 276 at 39.

302. *Id*.

303. *Id*.

304. KARMRAN KOWSARI, ET. AL., TEXT CLASSIFICATION ALGORITHMS: A SURVEY, at 5 (2019) https://arxiv.org/abs/1904.08067.

305. BIRD, ET. AL., *supra* note 276 at 107. (Regular expressions are algorithms defining patterns in text).

306. *Id*. at 112.

307. *Id*.

308. *Id*. at 113*; see also* NOAM CHOMSKY, SYNTACTIC STRUCTURES 32 (1957) (Morphemes are fundamental meaningful units of language data which cannot be further sub-divided).

309. Aashish R. Karkhanis, Jenna L. Parenti, *Toward an Automated First Impression on Patent Claim Validity: Algorithmically Associating Claim Language with Specific Rules of Law*, 19 STAN. TECH. L. REV. 196, 207 (2016).

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

ii. Vectorization

Vector space language models are collections of word vectors, which represent words as vector values and are associated with abstract features.[310] For example, vector values may be associated with information retrieval, document classification, or question and answering.[311] Vector space models represent words in a three-dimensional vector space.[312] Within this three-dimensional space, words are associated via co-occurrences, the rate at which words co-occur within a defined window.[313] The cosine similarity[314] of two vectors is a standard measure of how close the two vectors are to one another.[315] However, vector space models are blind to synonyms, idioms, and antonyms – which is a significant limitation.[316] Yet, vector space models still provide state of the art performance in research and industry.[317]

A critical task for developing vector space models for NLP is creating word embeddings.[318] Word embeddings are mappings of words to vectors,[319] allowing deep learning models to computationally process textual

---

310. THOMAS MIKOLOV, ET. AL., EFFICIENT ESTIMATION OF WORD REPRESENTATIONS IN VECTOR SPACE (2013) https://arxiv.org/abs/1301.3781.

311. JEFFREY PENNINGTON, ET AL., GLOVE: GLOBAL VECTORS FOR WORD REPRESENTATION 1532 (2014).

312. *Id.*

313. *Id.*

314. The computation for arbitrary-dimension cosine similarity is formally expressed:

$$\cos(x, y) = \frac{x \cdot y}{\sqrt{\left(\sum_{i=1}^{i=n} x_i^2\right)\left(\sum_{i=1}^{i=n} y_i^2\right)}}$$

The cosine similarity is computed for each word with respect to all preceding words in the model. *See also* Dean Alderucci, *The Automation of Legal Reasoning: Customized AI Techniques for the Patent Field*, Duq. L.R. (Forthcoming 2020) (On file with author) ("Although the software does not understand any of the words it processes, calculating word co-occurrences permits NLP software to perform feats of apparent text comprehension.").

315. CHARNIAK, *supra* note 71 at 75.

316. Dean Alderucci, *The Automation of Legal Reasoning: Customized AI Techniques for the Patent Field*, DUQ. L.R. (Forthcoming 2020) (On file with author) ("Discussing the relationship between statistical models, knowledge, and reasoning.").

317. Pennington, et al., *supra* note 311.

318. HONGLIANG FEI, ET AL., HIERARCHICAL MULTI-TASK WORD EMBEDDING LEARNING FOR SYNONYM PREDICTION, (2019).

319. CHARNIAK, *supra* note 71 at 73. A floating-point number is a number with an arbitrary, un-restricted number of digits after the decimal. For example, 0.883, 1.45, and 17.989891 are all floating-point numbers.

2020   *AI PATENTS: A DATA DRIVEN APPROACH*   449

information.[320] Conceptually, word embeddings are based on the distributional hypothesis: words with similar meanings tend to occur in similar context.[321] Indeed, word embeddings provide a way to quantify meaning because embedding similarity mirrors meaning similarity.[322] The process of developing word embeddings supports vector space model production.[323] In essence, word embeddings are a way to vectorize text copra for computational processing.[324]

The preprocessing stage accounts for the majority of time spent on NLP projects and is arguably the most important.[325] Indeed, the data define machine learning systems.[326] Thus, it is critical the dataset developed for any particular project is accurate and valid.[327] Once the pre-processing stage is complete, machine learning algorithms analyze the data.[328] There are various machine learning methods and models employable for NLP.[329]

iii. Models

In the last few years, deep learning models have shown the best performance in NLP tasks.[330]   For example, deep learning models are the foundation of document review systems.[331] Indeed, pre-trial discovery in lawsuits involves processing parties' requests for materials to reveal facts and

---

320. LINGPENG KONG, ET AL., A MUTUAL INFORMATION MAXIMIZATION PERSPECTIVE OF LANGUAGE REPRESENTATION LEARNING(2019) https://arxiv.org/abs/1910.08350.

321. TOM YOUNG ET. AL., *RECENT TRENDS IN DEEP LEARNING BASED NATURAL LANGUAGE PROCESSING*, 2 (Computational Intelligence Magazine 2018) https://arxiv.org/abs/1708.02709v5. *See also* Dean Alderucci, The Automation of Legal Reasoning: Customized AI Techniques for the Patent Field Duq. L. Rev. (Forthcoming 2020) (On file with author) ("Although the software does not understand any of the words it processes, calculating word co-occurrences permits NLP software to perform feats of apparent text comprehension.").

322. *Id.*

323. HONGLIANG FEI, ET AL., HIERARCHICAL MULTI-TASK WORD EMBEDDING LEARNING FOR SYNONYM PREDICTION, (2019).

324. Pennington, et al., *supra* note 311.

325. JOHN D. KELLEHER, BRENDEN TIERNEY, DATA SCIENCE 65 (2018).

326. ALPAYDIN, *supra* note 49 at 12.

327. *Id.* at 156.

328. *Id.* at 104.

329. YOUNG et. al., supra note 321 at 2.

330. *Id. See also* U.S. Patent No. 10,504,518 (issued Dec. 10, 2010).

331. Simon, et. al., *supra* note 19 at 254*; see also* Sergio David Becerra, *The Rise of Artificial Intelligence in the Legal Field: Where we are and Where we are Going*, 11 J. BUS. ENTREPRENEURSHIP & L. 27, 39 (2019).

develop evidence for trial.[332] In practice, this type of discovery often requires the processing of millions of documents and is thus automated with NLP.[333] In particular, two types of deep learning models are most commonly used in research and practice, Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs).[334]

A recurrent neural network (RNN), is a neural network tailored for processing sequential series of information in which the output contributes to the input.[335] RNNs improved previous NLP methods by incorporating an artificial memory mechanism.[336] In fact, the term *recurrent* refers to the way in which the network processes information, depending on preceding calculations.[337] RNNs only have one hidden layer, but they also use a replay buffer for memory.[338] The memory mechanism is inspired by a biological counterpart in the human brain.[339] In the brain, memories are formed by the strengthening of synaptic connections.[340] As such, RNNs work by strengthening the relationships between certain nodes in the network through a recurrent feed-forward model.[341] In general, RNNs are appropriate for problems where specific prior nodes influence later nodes in the network[342] because RNNs process sequences of data one element at a time.[343] Thus, RNNs are frequently used for language-modeling in particular because language learning is often defined through a problem framework requiring memory.[344] In addition to RNNs, Convolutional Neural Networks (CNNs) are also commonly used in NLP tasks.[345]

---

332. *Id.*

333. ASHLEY, *supra* note 20 at 239.

334. *Id.*

335. CHARNIAK, *supra* note 71, at 82.

336. *Id.* at 83.

337. Young et. al., supra note 321, at 2.

338. KELLEHER, *supra* note 16, at 170-171 (An RNN's depth arises from the fact that the memory vector is propagated forward and iteratively improved).

339. MOHEB COSTANDI, NEUROPLASTICITY 55 (2016).

340. *Id.*

341. CHARNIAK, *supra* note 71, at 83.

342. Nal Kalchbrenner, et. al., A Convolutional Neural Network for Modeling Sentences, University of Oxford (2014) https://arxiv.org/abs/1404.2188.

343. KELLEHER, *supra* note 16, at 172.

344. CHARNIAK, *supra* note 71, at 83.

345. *See* Yoon Kim, *Convolutional Neural Networks for Sentence Classification* (2014), https://arxiv.org/abs/1408.5882; *see also* U.S. Patent No. 10,460,215 (issued Oct. 29, 2019).

Similar to RNNs, CNNs draw inspiration in design from the biological brain. Indeed, CNNs are modeled based upon the biological visual cortex.[346] The biological visual cortex is composed of receptive fields made up of cells that are sensitive to small sub-regions of the visual field.[347] In a CNN, these small sub-regions are modeled with a kernel, as described by figure 15.
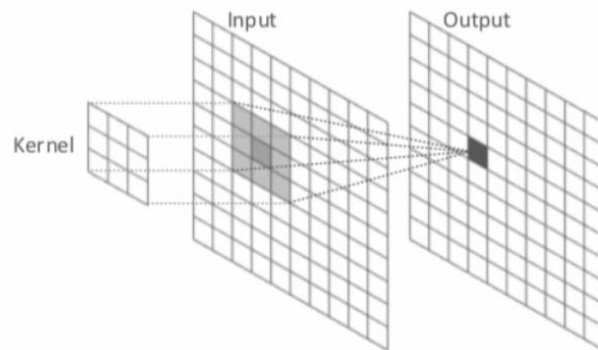

Figure 15[348]

A kernel is a small square matrix that is applied to each element of the input matrix.[349] Each kernel is convolved across an input matrix and the resulting output is called a feature map.[350]

Further, in a CNN, a neuron's response to a stimulus in its receptive field is modeled with a mathematical convolutional operation, similar to the way in which light is convoluted by the eye as it passes through the lens to the retina.[351] Convolution is a mathematical operation for classification, relying on matrix multiplication between certain kernels and the network's later

---

346. Manon Legrand, *Deep Reinforcement Learning for Autonomous Vehicle Control among Human Drivers*, Universite Libre de Bruxelles, 23 (2017).

347. Brian S. Haney, *The Future of Autonomous Vehicles & Liability Theory*, 29 ALB. L.J. SCI. & TECH. (2019) (Forthcoming) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261275.

348. Manon Legrand, *Deep Reinforcement Learning for Autonomous Vehicle Control among Human Drivers*, Universite Libre de Bruxelles, 23 (2017) (Model based upon preceding citation); *see also* Brian S Haney, CNN, GITHUB, https://github.com/Bhaney44/CNN (providing various CNN coding examples).

349. CHARNIAK, *supra* note 71, at 52.

350. Legrand, *supra* note 346, at 24.

351. *Id.* at 22-23 (The retina transfers electrical signals across the optic nerve to the occipital lobe, where the image is transposed in the visual cortex, the visual processing center of the human brain).

layers.[352] The convolutional operation allows CNNs to classify objects based upon their similarity.[353] The process of learning to optimize functions is the core of both RNNs and CNNs and is achieved by learning the appropriate set of weights for the connections in the network.[354]

B. Patents

i. By Year

The first NLP patent, titled *Method and Apparatus for Analyzing the Syntactic Structure of a Sentence*, was awarded to Tokyo Shibaura Denki Kabushiki Kaisha,[355] in the year 1986.[356] Figure 16 graphs the number of patents granted by year.
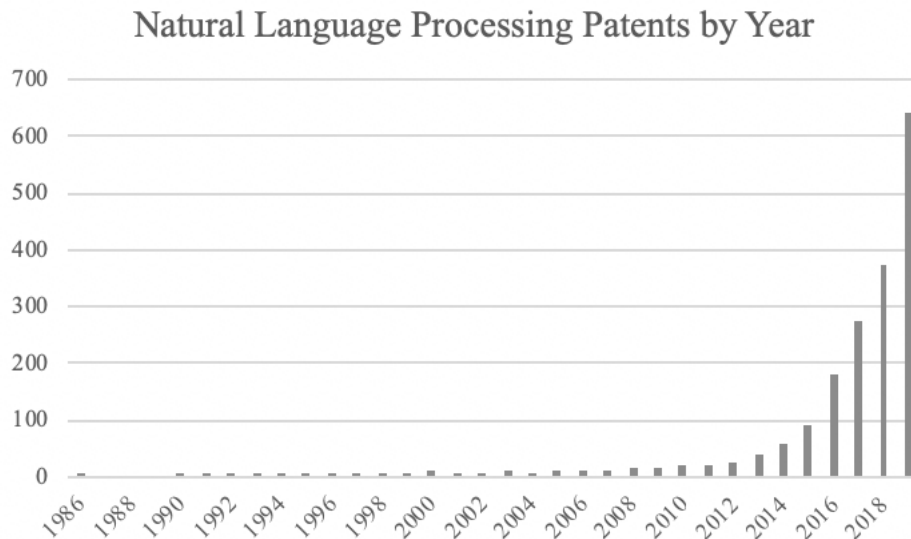
Figure 16[357]

---

352. ALPAYDIN, *supra* note 49, at 101-02.

353. Kabita Thaoroijam, *A Study on Document Classification using Machine Learning Techniques*, IJCSI INTERNATIONAL JOURNAL OF COMPUTER SCIENCE ISSUES Vol. 11 Issue 2 (March 2014).

354. KELLEHER, *supra* note 16, at 161.

355. A subsidiary of Tokyo Shibaura Denki, a multinational conglomerate that evolved into what is now Toshiba – headquartered in Tokyo, Japan.

356. U.S. Patent No. 4,586,160, (issued Apr. 29, 1986).

357. Brian S. Haney, NLP Patents (2019). (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office) (A copy of the data is on file with the author).

More NLP patents were granted than any other sample in this Article's dataset. From the year 1986 to 2004, less than ten patents were granted each year.[358] However, from 2016 to 2019 no less than 182 NLP patents were granted in a single year.[359] And, the number of NLP patents granted has increased each and every year since 2012.[360]

ii. Market

Figure 17 graphs the NLP patent market's growth – measured by total patents.[361] From the year 1986 to 2019 the NLP patent market has grown from one to 1,858 patents.[362]

Natural Language Processing Patent Market
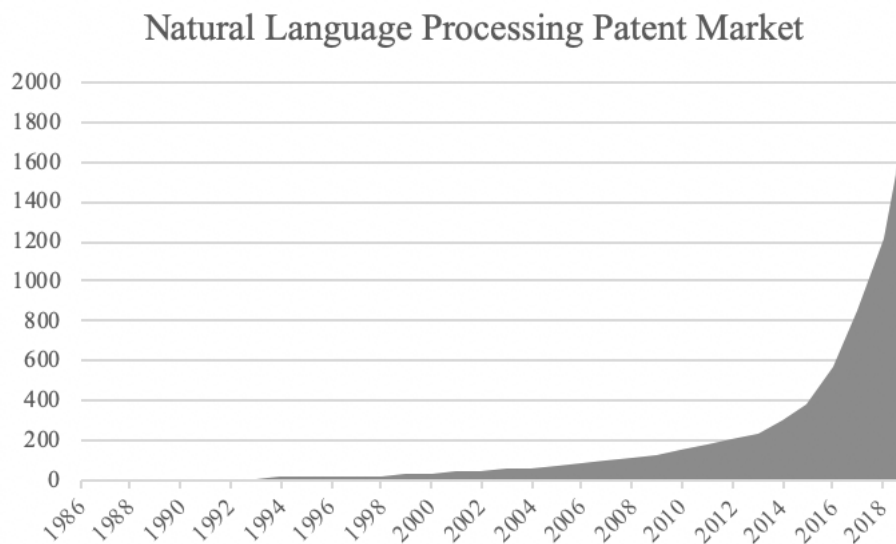
Figure 17[363]

The market's growth rate accelerated significantly from the year 2014 to 2019 in particular. In 2014 the total market size was 297 patents and in 2019

358. *Id.*
359. *Id.*
360. *Id.*
361. *Id.*
362. *Id.*
363. *Id.*

the total market size grew to 1,858, an increase of over a 600%.[364] Further, the growth rate by total patents has increased each year since 2012.[365]

iii. Firms

The NLP patent market is unique due to the extreme concentration of patents with one firm. Indeed, IBM owns 681 of 1,858 total NLP patents.[366] Figure 18 graphs a sample of firms with a stake in the NLP patent market.

Natural Language Processing Patents by Company
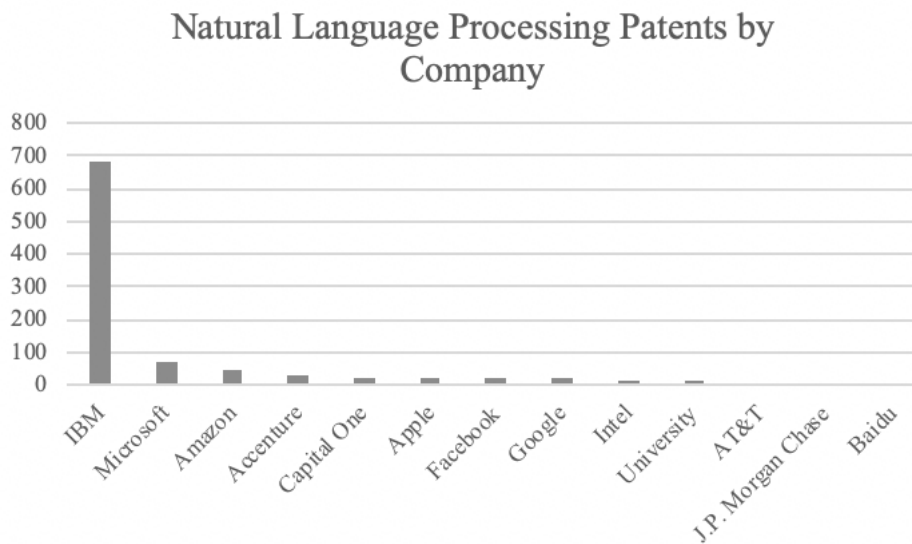
Figure 18[367]

IBM owns a significant portion of the market with over a 36% market share.[368] Microsoft and Amazon own the second and third largest portions of the market with 70 and 49 patents respectively.[369] In fact, Microsoft, Amazon, Apple, Facebook, and Google have a combined 174 NLP patents, about 9% of the total market.[370] Thus, IBM owns more than three times as many NLP patents as the five companies combined.[371]

---

364. *Id.* (Technical increase 625.5892%).

365. *Id.*

366. *Id.*

367. *Id*

368. IBM owns 36.65 % of the total market.

369. Haney *supra* note 357.

370. *Id.* (Technically 9.3649% of the market).

371. *Id.* (Technically 3.91379 times more).

VI.   Intellectual Property Strategy

Intellectual Property (IP) is information springing from the human mind.[372] Broadly, IP's umbrella covers any intelligence, skills, code, writing, or data.[373] IP plays many roles within the firm and defines a firm's structures and strategies in knowledge and information management.[374] Further, IP describes the knowledge and capabilities of a firm and its employees, providing freedom of action in innovation and growth strategy.[375] Moreover, IP is a flexible asset class providing access to new markets, the ability to improve existing products, and opportunities to develop new revenue streams.[376]

In short, IP is a vital asset for any firm competing in a global knowledge economy.[377] As a result, a firm's ability to safeguard and protect its IP is crucial to firm success because proprietary technology is the most substantive advantage a company can have.[378] As such, top firms are increasingly developing IP strategies.[379] Conventional wisdom teaches a theory of IP rights based on the sword and the shield.[380] Yet, modern firms must challenge this conventional wisdom to remain relevant in today's viciously competitive economy.[381] Every firm needs to innovate in terms of how they develop products and services.[382] Similarly, firms need to innovate in terms of how they choose to protect or disclose information about those products and services to the outside world.[383] This Part explores three considerations firms take into account during IP strategic planning and development for AI technologies: protection, litigation, and valuation.

---

372. Andrew Beckerman-Rodau, *The Problem with Intellectual Property Rights: Subject Matter Expansion*, 13 Yale J. L. & Tech. 36, 37 (2010-2011).

373. *Id*.

374. James W. Cortada, Information and the Modern Corporation 4 (MIT Press 2011).

375. John Palfrey, Intellectual Property Strategy 3 (MIT Press 2012).

376. Ted Hagelin, *A New Method to Value Intellectual Property*, 30 AIPLA Q.J. 353, 363 (2002).

377. *Id.*

378. Peter Thiel, Zero to One 48 (2014). (Proprietary IP in the form of technologies are the most valuable assets any business can possess because it makes a product difficult to replicate, increases the firm's substantive rights, and improves company prestige).

379. Palfrey, *supra* note 375 at 35.

380. *Id*. at 2 (As a sword, IP rights are used to attack competitors infringing on rights. As a shield, IP rights defends against attacks and accusations of infringement).

381. Anne Kelley, *Practicing in the Patent Marketplace*, 78 U. Chi. L. Rev. 115, 115 (2011).

382. *Id.*

383. Palfrey, *supra* note 375 at 141.

A. Protection

i. Patents

The most traditional form of IP protection for new technologies is a patent.[384] A patent provides the holder the legal right to prohibit others from using, making, or selling an invention without permission.[385] Indeed, in conferring the exclusive right to discoveries to its inventors, a patent confers an essential temporary monopoly to the holder.[386] This concept is foundational to our modern economy. In short, a patent confers the exclusive rights to use and profit from an invention to the holder, backed by the Government.

The United States Patent and Trademark Office ("USPTO") reviews applications to determine whether a claimed invention:

1. Is statutory subject matter;[387]
2. Is useful;
3. Is novel;
4. Would not be considered obvious by a hypothetical person of ordinary skill in the field; and
5. Is described well enough that those in the field can make and use the invention.[388]

The USPTO's granting of patent rights provides typical property rights,[389] including the right of the patent owner to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States.[390] Notre Dame Law Professor Stephen Yelderman argues the U.S. patent system's fundamental goal is to provide an adequate incentive to motivate innovators to publish their invention in exchange for rights.[391] Thus, the system Congress created

---

384. PALFREY, *supra* note 375 at 55.

385. Stephen Yelderman, *The Value of Accuracy in The Patent System*, 84 U. CHI. L. REV. 1217, 1270 (2017)*; see* U.S. Const. art. I, § 8, cl. 8.

386. Bryce C. Pilz, Student Intellectual Property Issues on the Entrepreneurial Campus, 2 MICH. J. PRIVATE EQUITY & VENTURE CAP. L. 1, 16 (2012).

387. 35 U.S.C. § 101. (The first element of the statutory requirements, statutory subject matter, includes any new process, machine, manufacture, or composition of matter, or any new and useful improvement thereof).

388. 35 U.S.C. § 112.

389. Andrew Beckerman-Rodau, The Problem with Intellectual Property Rights: Subject Matter Expansion, 13 Yale J. L. & Tech. 36, 55 (2010-2011).

390. *Id.*

391. Yelderman, *supra* note 385 at 1262-63.

provides a delicate balance.[392] In exchange for monopoly rights, the innovator must provide a description of how to make and use the invention.[393] However, some firms have begun taking a different approach indicative of a changing paradigm in IP protection.[394] While firms traditionally used patents as the sole means to protect inventions and innovations, there is a recent trend for firms to utilize trade secrets as a protective measure in addition to patents.[395]

ii. Trade Secrets

In contrast to filing a patent application, inventors may be able to profit from their work while keeping the invention confidential and relying on trade secret protection, rather than making the invention public.[396] In theory, trade secret disclosure benefits society more broadly than does maintaining a trade secret, since it permits more people to make use of the information as a starting point for further innovation.[397] However, the unique nature of the technology industry calls this theory to question. For example, according to SpaceX Founder & CEO Elon Musk, "our primary long-term competition is China – if we published patents, it would be farcical because the Chinese would just use them as a recipe book."[398] Professor Yelderman argues, trade secret law has evolved as an alternative mode of protection for firms not willing to disclose their inventions or other proprietary technologies.[399]

Trade secret law confers an exclusive right on the possessor of valuable information not generally known to competitors.[400] Generally, trade secrets include formulas, patterns, programs, devices, methods, techniques,

---

392. Max Stul Oppenheimer, Patents 101: Patentable Subject Matter and Separation of Powers, 15 Vand. J. Ent. & Tech. L. 1, 8 (2012).

393. *Id.* at 9.

394. Yelderman, *supra* note 385 at 1264.

395. *Id.*

396. Oppenheimer, *supra* note 392 at 9.

397. Yelderman, *supra* note 385 at 1262.

398. Chris Anderson, *Elon Musk's Mission to Mars*, Wired Magazine (October 21, 2012),https://www.wired.com/2012/10/ff-elon-musk-qa/. *See also* Gregory C. Allen, *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security, Center for a New American Security* 1 (February 2019). In July 2017, China's State Council, released an AI plan and strategy calling for China to pass the United States by 2020 and become the world's leader in AI by 2030, committing $150 billion to the goal. By the end of 2018, Chinese leadership assessed the program's development as surpassing the United States, achieving its objective earlier than expected).

399. Yelderman, *supra* note 385 at 1263.

400. Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 Stan. L. Rev. 311, 329 (2008).

and processes.[401] The traditional conception of the trade-off between patents and trade secrets views the patent system's disclosure function as a principal drawback.[402] All the while, trade secrets have advantages of their own. For example, trade secret protections are immediate, while it takes years to get a patent.[403] Further, trade secret law confers an exclusive right on the possessor of valuable information not generally known to competitors.[404] In other words, trade secret law allows firms to protect their proprietary technologies and without publicly disclosing sensitive firm information.[405]

Traditionally, trade secrets are protected by state law.[406] The Uniform Trade Secrets Act (UTSA) was published in 1979 by the National Conference of Commissioners on Uniform State Laws and has been adopted by 47 states and the District of Columbia.[407] The UTSA defines "trade secret" as information, including a formula, pattern, compilation, program, device, method, technique, or process, that:

1. derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and

2. is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.[408]

The crux of the UTSA provides a remedy for claimants in the event of trade secret misappropriation.[409] Generally, misappropriation includes the malicious or unauthorized disclosure of firm trade secrets.[410] Damages for

---

401. Trade Secret, Black's Law Dictionary (10th ed. 2014).

402. Lemley, *supra* note 400 at 314.

403. *Id*. at 326.

404. *Id.* at 329.

405. Robert G. Bone, *A New Look at Trade Secret Law: Doctrine in Search of Justification*, 86 Calif. L. Rev. 241, 249 (1998).

406. Unif. Trade Secrets Act Refs. & Annos (2019).

407. Reid, et al., *supra* note, at 137, at 122.

408. Uniform Law Commission Annual Conference, *The Uniform Trade Secrets Act With 1985 Amendments*, Uniform Law Commission, (Aug. 2-9, 1985), http://www.uniformlaws.org/shared/docs/trade%20secrets/utsa_final_85.pdf.

409. Unif.Trade Secrets Act § 3 (2019). *See also* 18 U.S.C. § 1836 (2016). *See also* Patrick J. Manion*, Two Steps Forward, One Step Back: The Defend Trade Secrets Act of 2016 and Why The Computer Fraud and Abuse Act of 1984 Still Matters for Trade Secret Misappropriation*, 43 J. Legis. 289, 294 (2017).

410. Brittany S. Bruns, *Criticism of the Defend Trade Secrets Act of 2016: Failure to Preempt*, 32 Berkeley Tech. L.J. 469, 484 (2017).

misappropriation include both the actual loss caused by misappropriation and the unjust enrichment caused by misappropriation.[411]

Yelderman argues, the added protections at the federal level make firms more likely to pursue trade secret protections as opposed to traditional patent protections.[412] Yet, despite the DTSA's added protections, the protection of competing interests in confidential information remains a difficult and complex task.[413] For example, in the AI technology industry, much work stems from government contracts which carry a plethora of compliance issues.[414] Now, firms are adopting more dynamic and complex strategies for protecting IP.[415] According to John Palfrey, the former Harvard Law Professor and current President of the MacArthur Foundation, the most innovative organizations in any given market have the most innovative IP strategies.[416]

iii. Open Source

The open-source strategy is unique because companies give IP resources away for free.[417] In the modern world, this strategy makes sense because the decentralized nature of information across the internet has dismantled notions of truly proprietary or classified information.[418] Further, open-source strategies allow firms to profit from their IP in non-traditional ways. For example, Google open sources search engine and machine learning

411. Unif.Trade Secrets Act § 3 (2019).

412. Yelderman, *supra* note 385 at 1264.

413. Suellen Lowry, *Inevitable Disclosure Trade Secret Disputes: Dissolutions of Concurrent Property Interests*, 40 STAN. L. REV. 519, 519 (1998).

414. Veronica Root, *Coordinating Compliance Incentives*, 102 CORNELL L. REV. 1003, 1029 (2017) (discussing regulatory agencies deficiencies in information and coordination). *See also* 35 U.S.C. §207. *See also* Gregory N. Mandel, *Leveraging the International Economy of Intellectual Property*, 75 OHIO ST. L.J. 733, 736 (2014).

415. For example, Google has a diverse intellectual property AI IP portfolio including copyrights, patents, trade secrets, and open source software.

416. PALFREY, *supra* note 375 at 88.

417. *Id.* at 105.

418. Shixun You, et al., *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare*, IEEE Access Vol. 7, 37432, 37438 (2019) (for example, in a recent study funded by the National Natural Science Foundation of China, *Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare* (China AI Missile Study), researchers demonstrate Chinese capabilities in deep reinforcement learning control systems for missile control).

tools, and users flock.[419] In turn, Google continues to make no less than 95% of its profits from advertisements.[420]

Indeed, there are sometimes strong reasons to let others use IP with fewer restrictions than the law establishes on a firm's behalf automatically.[421] The idea behind open innovation is the creators of new ideas don't have to be within your organization to be helpful.[422] One possibility for firms building their business, is to build on the IP of others by using open source innovations.[423] One example of open source development in the technology industry is OpenAI's Lunar Lander.[424] The OpenAI Lunar Lander allows anyone with a computer to access a simulated lunar environment, where a deep reinforcement learning control system can be trained to land a lunar module.[425] The benefits for a company like OpenAI are users contribute to, train, and develop OpenAI's software free of charge to the company.[426] Importantly, in open source models the creator does not give away all the rights free and clear to their creations.[427] Instead, the open-access strategy allows a company to give away certain rights, retaining those deemed more valuable.[428]

419. TensorFlow,   TF   AGENTS, https://github.com/tensorflow/agents/tree/master/tf_agents/agents/dqn (TensorFlow is a Google software package for machine learning;GitHub is a website and repository where programmers post code). According to the website the code is Licensed under the "Apache License, Version 2.0", available at https://www.apache.org/licenses/LICENSE-2.0 (An Apache License is a type of patent license).

420. GILDER, *supra* note 126 at 37.

421. PALFREY, *supra* note 375 at 107.

422. *Id.* at 66.

423. *Id.* at 59.

424. OpenAI, LunarLander-v2 (2019), https://gym.openai.com/envs/LunarLander-v2/.

425. *Id.* A second example of an open source strategy is the Bitcoin Network, which is available on GitHub and available under an open source license. *See* Bitcoin, *Bitcoin Core integration/staging tree*, GITHUB (2019), https://github.com/bitcoin/bitcoin. *See also* Brian Seamus Haney, *Blockchain: Post-Quantum Security & Legal Economics*, 24 N.C. BANKING INST. (2019) (Forthcoming)https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3444695.

426. OpenAI GitHub (2020), https://github.com/openai.

427. PALFREY, *supra* note 375 at 106.

428. *Id*. at 105.

2020   *AI PATENTS: A DATA DRIVEN APPROACH*   461

B. Litigation

i. Patent Claims

Patent Claims mark the invention's boundaries, defining the particular thing invented and making the public aware of the invention.[429] Patent claims generally define devices, structures, or methods.[430] The USPTO will issue a patent only for claims it determines satisfy the statutory requirements, and a challenge to an issued patent will succeed if the challenger can show that any of these requirements have not been met.[431] Further, courts construe patent claims by starting with the plain meaning of their terms as they would be understood by a person having ordinary skill in the art.[432] Claims are the most important part of a patent[433] because claims are the only part of the patent that can be infringed.[434] As such, aggressively asserting patent claims has a place in IP strategy, but can lead to destructive consequences if allowed to take control.[435]

Patent claims directed to AI have tended to focus on machine learning, which inverts the programming paradigm.[436] AI patent claims tend to utilize functional claiming[437] and emphasize algorithmic structures and the functional elements of software such as data structures.[438] This form of patent claiming in a digital technology represents another instance of a divided infringement possibility, where separate actors can divide the performance of the patented method among themselves.

There are varying opinions on AI patentability. Thus, an AI's nature effects the patent claims.[439] Experts suggest many AI patents tend to

---

429. KEVIN F. O'MALLEY, ET. AL., 3A FED. JURY PRAC. & INSTR. § 158:41 (6th ed. 2019).

430. Mark A. Lemley, *The Changing Meaning of Patent Claim Terms,* 104 MICH. L. REV. 101, 107 (2005).

431. Oppenheimer, *supra* note 392 at 4.

432. Lemley, *supra* note 430 at 102.

433. *Id.* at 101.

434. KEVIN F. O'MALLEY, ET. AL., 3A FED. JURY PRAC. & INSTR. § 158:41 (6th ed. 2019).

435. HOWARD C. ANAWALT AND EVE J. BROWN, IP STRATEGY: COMPLETE INTELL. PROP. PLANNING § 5:1 (2019).

436. Tabrez Y. Ebrahim, *Dynamicism, Deep Learning, & Patent Theory* (2020) (on file with author) (draft available at https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/Ebrahim_Patent-Infringement.pdf).

437. 35 U.S.C. § 112(f) (2011).

438. Ebrahim, *supra* note 436.

439. *Id.* (AI patent trend analysis demonstrates three categories of AI patents: (1) algorithms, (2) platforms, and (3) applications).

implement at least one method patent claim and at least one system patent claim.[440] However, the complexity of the systems creates difficulties in defining claim scope and application.[441] For example, consider the first claim of Google's 258' patent:

1.   A method of reinforcement learning, the method comprising:

     a.   inputting training data relating to a subject system, the subject system having a plurality of states and, for each state, a set of actions to move from one of said states to the next said state; wherein said training data is generated by operating on said system with a succession of said actions and comprises starting state data, action data and next state data defining, respectively for a plurality of said actions, a starting state, an action, and a next said state resulting from the action; and training a second neural network using said training data and target values for said second neural network derived from a first neural network; the method further comprising generating or updating said first neural network from said second neural network.[442]

One issue is whether this claim covers all applications of DQN methods, another is whether the claim covers applications of this particular method in different contexts.[443] Thus, from an IP strategy perspective, one difficulty is interpreting the boundaries of Google's ownership rights.

One of the biggest challenges in drafting patent claims may be the syntax of the industry. Consider the complex relationships between the terms: neural network, reinforcement learning, supervised learning, unsupervised learning, machine learning, states, and actions.[444] A neural network is a learning algorithm modeling associative properties which may be supervised or

---

440. *Id.*

441. Mark A Lemley, *Software Patents and The Return of Functional Claiming* 2013 Wis. L. Rev. 905, 906 (2013) (Arguing software patents create "thickets" of overlapping inventions). *See also* Mark A. Lemley & Mark P. McKenna, *Scope*, 57 Wm. & Mary L. Rev 2197, 2240 (2015).

442. U.S.   Patent   No.   9,679,258   B2   (issued   Jun.   13,   2017) (https://patents.google.com/patent/US9679258B2/en) (The Google 258' patent includes both system and methods claims).

443. U.S. Patent No. 10,346,741 (issued July 9, 2019) (to Mnih, et al. Asynchronous deep reinforcement learning—defining more advances in deep reinforcement learning techniques; assigned to DeepMind Technologies – a Google subsidiary).

444. In the 258' patent, states and actions refer to a Markov model.

unsupervised in nature, but it is not necessarily a deep learning algorithm.[445] Reinforcement learning often incorporates both supervised learning and unsupervised learning techniques.[446] In the 258' patent, a neural network is used to optimize the way in which a reinforcement learning agent chooses actions to navigate the states of an environment[447]—all of which falls under the umbrella of machine learning.

For example, U.S. Patent No. 7,395,251 was awarded to IBM in 2008,[448] but was not cited as prior art in Google's 258' patent, which was awarded in 2017. Consider the similarity between Google's 258' patent claim 1 and IBM's 251' patent claim 26:[449]

> 26. In a method for estimation, control, system identification, reinforcement learning, supervised learning, unsupervised learning, and/or classification, comprising a step of iteratively transforming a first matrix into a second matrix, the improvement comprising the steps of: (a) specifying a functional relationship between said first matrix and a first set of vectors, (b) specifying a transformation of each vector in said first set of vectors into a vector of a second set of vectors, (c) implementing said first set of vectors as a first set of activity vectors in a neural network equivalent system (NNES), (d) implementing an approximation of said first matrix as a first set of connection strength values in said NNES, (e) determining, by means of neural computations, a second set of connection strength values as a function of said first set of activity vectors, and (f) determining, by means of neural computations, a second set of activity vectors as a function of said first set of activity vectors and of said first set of connection strength values, wherein said second set of connection strength values approximates said second matrix.[450]

---

445. One example would be a perceptron algorithm, which is not layered. However, there are few applications of modern neural networks that don't involve deep learning.

446. KELLEHER, *supra* note 16 at 26-28 (discussing the relationship between unsupervised learning, supervised learning, and reinforcement learning).

447. U.S. Patent No. 9,679,258 B2, *supra* note 442.

448. U.S. Patent No. 7,395,251 (issued July 1, 2008) (assigned to International Business Machines Corporation).

449..S. Patent No. 9,679,258 B2, *supra* note 442 (the earliest cited prior art in the 258' Patent is 2010. *See* U.S. Patent No. 2010/0094788 A1 (issued Apr. 15, 2010) (assigned to Siemens Corporation).

450. U.S. Patent No. 7,395,251, *supra* note 448 (assigned to International Business Machines Corporation).

CHICAGO-KENT J. INTELL. PROP.   Vol 19:3

Both patents are describing a process by which a reinforcement learning system interacts with a neural network to optimize a reward.[451] Indeed, what the 258' patent refers to as *states*–the 251' patent refers to as *matrices*.[452] However, both the states and the matrices are fed through a neural network for approximation. Moreover, the states referred to in the 258' patent are composed of matrices.[453]

There are challenges regarding how best to protect IP rights for any new technology.[454] But, for some firms, these challenges should be considered as opportunities.[455] Further, the growth rates in AI technology lead some to claim existing patent protection mechanisms will not satisfy the new industry.[456] As such, understanding who owns the rights to what in this domain may turn out to be whoever can explain it better to a judge or jury.[457] Deciding whether an AI patent is infringed will be a difficult task for courts to grapple with in the near future.

ii. Infringement

John Palfrey argues, having a clearer certainty in IP rights helps to lead to faster and less expensive settlements.[458] And, having control of IP rights from the outset generally decreases the risk of litigation.[459] Yet litigation is an unavoidable part of the patent system's private enforcement

---

451. U.S. Patent No. 9,679,258 B2, *supra* note 442.

452. U.S. Patent No. 7,395,251, *supra* note 448 (assigned to International Business Machines Corporation).

453. Google uses the 258' patent in the TensorFlow Python library. TensorFlow uses tensors to represent states in reinforcement learning problems. According to TensorFlow, "A tensor is a generalization of vectors and matrices to potentially higher dimensions." *See* TensorFlow, TensorFlow Tensors (2020), *https://www.tensorflow.org/guide/tensor*. *See also* U.S. Patent Application No. 14/097,862 at 5 (filed Dec. 5, 2013) (describing the process by which a neural network performs convolutional operations on an 84x84x4 pixel image).

454. *See* Mark P. McKenna & Christopher Jon Springman, *What's In, and What's Out: How IP's Boundary Rules Shape Innovation*, 30 Harv. J. L. & T. 491, 494 (2017) (arguing utility patent claims are undermined by the law's lack of clarity and inconsistency).

455. Mark A. Lemley & Mark P. McKenna, *Scope*, 57 Wm. & Mary L. Rev 2197 (2015) (arguing patent owners can and do exploit gaps in patent law for financial fain with regularity).

456. Ebrahim, *supra* note 436.

457. See Mark A Lemley, *Software Patents and The Return of Functional Claiming*, 2013 Wis. L. Rev. 905, 930 (2013) (discussing problems relating to the uncertainty associated with the meaning and scope of software patent claims).

458. Palfrey, *supra* note 375 at 32.

459. *Id.*

scheme.[460] As AI technology becomes more commonplace in products and services, AI patentees will file patent infringement actions against their competitors.[461] In fact, there has been a rapid raise in AI patents despite doctrinal claim drafting issues.[462] According to Professor Tabrez Ebrahim, AI technology will trigger expensive patent wars, similar to other high technology industry patents.[463]

Unclear statutory language creates opportunities for asserting AI patent infringement.[464] The words in the patent infringement statute and the steps in utilizing it have been applied to a variety of technologies over many years.[465] Direct infringement is the broadest clause conferring infringement liability in the Patent Act. The Patent Act defines direct infringement under 35 U.S.C. § 271(a):

> Except as otherwise provided in this title, whoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefor, infringes the patent.[466]

Further, 35 U.S.C. § 271(a) has been recognized as requiring no more than the unauthorized use of a patented invention by performing one of the enumerated activities—either making, using, offering for sale, selling, or importing the invention.[467] Therefore, any firm that makes a patented AI technology and goes on to use, offer for sale, sell, or import the technology plainly is a direct infringer.[468] In fact, the mere act of making a patented AI technology is a direct infringement, and distinct from any subsequent use, sale, offer for sale, or importation.[469]

Thus, AI patent disputes are making their way to court.[470] For example, a recent patent infringement case in federal court centers on a

---

460. R. Polk Wagner, *Understanding Patent-Quality*, 157 U. PA. L. REV. 2135, 2143 (2009).

461. Ebrahim, *supra* note 436.

462. Brian S. Haney, AI Patents (2019) a copy of the data is on file with the author).

463. Ebrahim, *supra* note 436.

464. *See* Christopher J. White & Hamid R. Piroozi, *Drafting Patent Applications Covering Artificial Intelligence Systems*, 11 Landslide No. 3 at 10 (2019).

465. Ebrahim, *supra* note 436.

466. 35 U.S.C. § 271(a) (2010).

467. Ebrahim, *supra* note 436.

468. *Id.*

469. *Id.*

470. *Id.*

dispute concerning predictive analytics.[471] Generally, patent infringement assessment is based on first determining the meaning in each patent claim and second showing the accused infringement meets each claim term.[472] However, determining the meaning in each claim is a difficult problem for AI patents. For example, if another technology company uses a model incorporating a DQN model for reinforcement learning or another deep reinforcement learning variant – Google may have grounds for an infringement claim.[473] At the same time, it will be incredibly difficult to know what competitors are making and using in terms of an AI system's technical detail.[474] Yet, the code for the DQN algorithm is available as open source code on the Mnih's website at the University of Toronto and on the TensorFlow GitHub.[475]

Generally, patent law aims to provide patentees with payment for lost profits or other competitive harm suffered through infringement.[476] Further, patent damages are a make-whole remedy, intended to restore the patentee to

---

471. *PurePredictive, Inc. v. H2O.AI, Inc.*, Case No. 17-cv-03049-WHO, N.D. Cal. (Aug. 29, 2017). (Patent infringement case involving AI for predictive modeling).

472. Christopher J. White & Hamid R. Piroozi, *Drafting Patent Applications Covering Artificial Intelligence Systems*, 11 Landslide No. 3 at 10, 14 (2019) (the U.S. Code generally divides infringement in to two categories: direct infringement and indirect infringement).

473. Consider IBM may also have strong claim to the DQN algorithm and its implementations. *See* U.S. Patent No. 10,296,830 (issued May 21, 2019) (assigned to International Business Machines Corporation). This patent along with its sister U.S. Patent No. 10,296,832, are the only two patents with claims including a DQN. The DQN is claimed together by claim 1 and 6: "1. A computer-implemented topic guidance method for a call between an agent and a customer, the method comprising: in a training phase of a conversation model: creating, via a processor on a computer, the conversation model by learning a conversation pattern from a conversation topic segment based on successful and unsuccessful recorded dialog for all agents and customers in a history database; and in a nun-time phase of the conversation model: suggesting, via the processor on the computer, a conversation topic for the agent to engage the customer based on the learned conversation model and via a multi-round conversation to assist the agent to make a successful selling, the conversation model is unchanged during the run-time phase. 6. The computer-implemented method of claim 1, wherein one of a Q-learning technique and a deep Q-network is used in the creating to create the conversation model").

474. TensorFlow,   TF   Agents, https://github.com/tensorflow/agents/tree/master/tf_agents/agents/dqn. (TensorFlow is a Google software package for machine learning) (GitHub is a website and repository where programmers post code).

475. Volodymyr Mnih, https://www.cs.toronto.edu/~vmnih/. *See also* TensorFlow, TF Agents, https://github.com/tensorflow/agents/tree/master/tf_agents/agents/dqn.

476. Mark A. Lemley, *Distinguishing Lost Profits from Reasonable Royalties*, 51 WM. & MARY L. REV. 655, 669 (2009). (Or under a reasonable-royalty model the rate that would have both compensated patentees and allowed users of the technology to make a reasonable profit).

the same position as before the infringement.[477] Yet, the patent infringement statute is relatively silent as to definitions and courts have struggled with associating consistent semantics to the statute's syntax.[478] One potentially lucrative theory of AI patent infringement is direct infringement by firms selling machine learning models or offering AI as Service (AIaS). [479] Consequently, a patentee may improve the probability of victory by asserting a sufficiently large number of patents.[480] For example, IBM may have an advantage in litigation due to the robust nature of its machine learning patent portfolio.[481]

Ebrahim argues a doctrinal assessment of the patent infringement statute provides little likelihood for success in AI patent infringement lawsuits.[482] According to Ebrahim, a liability loophole results from multi-actor, divided infringement scenarios.[483] Indeed, Ebrahim argues "artificial intelligence technology creates a patent litigation liability loophole."[484] The liability loophole is in large part the product of AI supply chain development creating divided infringement scenarios.[485] Another recent article argues that

---

477. Amy L. Landers, *Patent Valuation Theory and the Economics of Improvement*, 88 Tex. L. Rev. 163, 166 (2010).

478. Ebrahim, *supra* note 436.

479. 35 U.S.C. § 271(a) (2010). For example, Google uses AI on the back-end of its search engine and offers AIaS through Google Cloud. *See also* Mark A Lemley, *Software Patents and The Return of Functional Claiming* 2013 WIS. L. REV. 905, 934 (2013) (arguing if a software product is successful, its maker can expect to be hit with dozens of suits and hundreds of threat letters from patent owners seeking a royalty from that product).

480. Sinan Utku, The Near Certainty of Patent Assertion Entity Victory in Portfolio Patent Litigation, 19 J. TECH. L. & POL'Y 1, 26 (2014).

481. Brian S. Haney, AI Patents (2019) (a copy of the data is on file with the author). *See also* U.S. Patent No. 9,298,172 (issued March 29, 2016) (assigned to International Business Machines Corporation). *See also* U.S. Patent No. 7,395,251 (issued Jul. 1, 2008) (assigned to International Business Machines Corporation).

482. Ebrahim, *supra* note 436.

483. *Id*.

484.*Id*. (Further asserting clever claim-drafting by patent prosecution will not avoid the multiple actor scenarios since artificial intelligence necessitates that parties divide the performance of machine learning. "The need for some connection between the parties in machine learning presents problems for patent holders of artificial intelligence method patents."

485. Divided infringement occurs when the actions of multiple entities are combined to perform every step of a claimed method, but no single party acting alone has completed the entire patented method. Multi-actor patent claims arise from infringement in a multi-party value chain and accompanying multi-party actions. Thus, the AI supply chain may make firms liable even though their innocent activities were combined with those of another party to violate another party's patent right. Ebrahim further argues, since machine learning requires access to a dynamic,

the unique attributes of AI—autonomous ability to function without humans, to modify and evolve over time in response to new data—causes doctrinal uncertainties in patent infringement analysis.[486] This is especially true considering the open availability of potentially patented software on GitHub.[487]

iii. Patent Assertion Entities

In the last decade, the landscape of patent litigation has radically shifted.[488] Entities that do not manufacture products have become important players in the patent litigation system.[489] Non-practicing entities (NPEs) provide ways for patentees to monetize their patents, often when there is not an alternative.[490] In fact, some small companies have been able to sell or monetize their patent portfolios to support ongoing or new business ventures.[491] However, few patents are economically valuable.[492] Thus, most companies cannot necessarily rely on their patents for an exit or revenue strategy.[493]

Interestingly, a recent study suggests NPEs represent slightly more than a quarter[494] of patent litigation cases.[495] As such, complaints that trolls are interfering with innovation are common.[496] The pejorative term "troll" is used by some to refer to any party that doesn't actually produce goods or

---

trainable data set as a data source and since other parties a need to work in concert, then no single party can perform all of the steps alone.

486. Mark Lemley & Mark McKenna, *Unfair Disruption*, 100 B.U. L. REV. 71 (2020) (drawing from antitrust injury doctrine to recognize that for disruptive technologies, cases of infringement are sometimes challenges to market disruption).

487. *See* TensorFlow, *TF Agents*, https://github.com/tensorflow/agents/tree/master/tf_agents/agents/dqn. (TensorFlow is a Google software package for machine learning); *see also* U.S. Patent No. 9,679,258 B2 (issued June 13, 2017) (https://patents.google.com/patent/US9679258B2/en).

488. Christopher A. Cortopia, et al., *Unpacking Patent Assertion Entities*, 99 MINN. L. REV. 649, 649 (2014).

489. *Id.*

490. Colleen Chien, *Startups and Patent Trolls*, 17 STAN. TECH. L. REV. 461, 479 (2014).

491. *Id.*

492. *Id.* at 481.

493. *Id.*

494. 264 out of 945 decisions.

495. John R. Allison, et al., *How Often Do Non-Practicing Entities Win Patent Suits?*, 32 BERKELEY TECH. L. J. 237, 257 (2017) (the study covers all patent lawsuits filed in federal district courts between January 1, 2008, and December 31, 2009).

496. Allison, et al., *supra* note 495 at 238.

services.[497] As Texas Law Professor John Allison explains, the debate over patent trolls has occupied policy makers in the patent system for the last several years.[498] For example, former U.S. President Barack Obama stated, "They don't actually produce anything themselves . . . They are essentially trying to leverage and hijack somebody else's idea and see if they can extort some money out of them."[499]

Generally, a patent troll is a person or entity who acquires ownership of a patent without the intention of actually using it to produce a product.[500] Yet, some arguments suggest patent trolls actually benefit society.[501] The argument follows: trolls act as a market intermediary for patents.[502] Not to mention, many well-known and highly respected companies have been accused of troll-like behavior, for example giants such as Apple Inc. and Microsoft Corp.[503] NPE proponents claim these entities provide liquidity in the marketplace for patents by permitting inventors who are otherwise excluded from the marketplace.[504] For instance, individuals who are capable of inventing new products, but cannot raise the capital to manufacture products may be admitted to the market.[505] Indeed, small inventors are the ones least likely to be able to commercialize their inventions, and therefore the ones most dependent on patent law to create a market for licensing.[506] As

497. *Id.* at 242. Indeed, some use troll to refer to anyone who is suing them, even practicing entities.

498. *Id.* at 296.

499. Gene Sperling, *Taking on Patent Trolls To Protect American Innovation*, THE WHITE HOUSE BLOG (June 4, 2013, 1:55 PM), http:// www.whitehouse.gov/blog/2013/06/04/taking-patent-trolls-protect-american-innovation.

500. James F. McDonough, *The Myth of The Patent Troll: An Alternative View of the Function of Patent Dealers in an Idea Economy*, 56 EMORY L.J. 189, 189 (2006).

501. *Id*. at 190.

502. *Id.* (stating the value of corporations used to be grounded in land, natural resources, and human capital, but the driving force in the U.S. economy today is intellectual property).

503. Christopher Hu, *Some Observations on The Patent Troll Litigation problem*, 26 NO. 8 INTELL. PROP. & TECH. L.J. 10, 10 (2014).

504. Christopher A. Cortopia, et al., *Unpacking Patent Assertion Entities*, 99 MINN. L. REV. 649, 653 (2014).

505. In fact, Thomas Edison has been branded by some as the king of patent trolls – as the awardee of 1,093 patents. Edison was an inventor and despite never practicing many of his inventions, they were incorporated in other products. *See* McDonough, *supra* note 500 at 198. *See also* U.S. Patent No. 265,786 Apparatus for The Transmission of Electrical Power, to Edison (1882). *See also* U.S. 219,268 Electric-Light, to Edison (1879).

506. John R. Allison, et al. *Valuable Patents*, 92 GEO. L.J. 435, 468 (2004). *See also* Tabrez Y. Ebrahim, *Automation & Predictive Analytics in Patent Prosecution: USPTO Implications & Policy*, 35 GA. ST. U. L. REV. 1185, 1214 (2019).

such, a recent paper argues a more suitable, market-contextual term for nonpracticing patent owners who license or enforce their patents is "patent dealers."[507]

Regardless of how they are defined, NPEs exist because the ownership and assertion of patents is a way to make money. For example, in 2009 Nokia and Samsung paid a small semiconductor[508] firm in King of Prussia, Pennsylvania called InterDigital a combined $653 million over a portfolio of patents for smart phone technology.[509] One advantage for NPEs is they are generally immune from the effects of defensive patenting because they do not manufacture products, and therefore a basis for a potential countersuit is often lacking.[510] Thus, given the cost of litigation, cases are cheaper to settle because there are few consistent methods of obtaining early dismissal and no realistic chance of recovering attorney fees and costs.[511]

Interestingly, recent studies reveal significant forum selection advantages in NPE cases.[512] For example, the Eastern District of Texas decided a disproportionate number of NPE cases.[513] Further, the percentage of all patent lawsuits and accused infringers attributable to NPE–instituted litigation is even higher in the technology industry.[514] Yet, often times, if a technology's potential licensee reads the patent documentation or is presented with the technology by an inventor with ambitions of licensing the technology, the corporation can simply use the patented technology without permission.[515] However, due to the vast syntactic overlap and complexities in AI patents claims, the war chest strategy will likely be successful for NPEs.[516]

---

507. McDonough*, supra* note 500 at 201.

508. A semiconductor is a solid substance that has a conductivity between that of an insulator and most other metals. Silicon semiconductors are essential components of most electronic circuits.

509. PALFREY, *supra* note 375 at 18. *See also In Matter of Arbitration Between InterDigital Communications Corp. and Samsung. . .,* 528 F.Supp.2d 340 (2007). *See also* InterDigital Communications Corp. v. Nokia Corp., 407 F.Supp.2d 522 (2005).

510. W. Michael Shuster, *Artificial Intelligence and Patent Ownership*, 75 WASH. & LEE L. REV. 1945, 1986 (2018).

511. Christopher Hu, *Some Observations on The Patent Troll Litigation problem*, 26 NO. 8 INTELL. PROP. & TECH. L.J. 10, 12 (2014).

512*.* Mark A. Lemley, *Where to File Your Patent Case*, 38 AIPLA Q. J. 4, 1 (2010).

513. Allison, et al., *supra* note 495 at 260-261.

514. *Id.* at 239.

515. *See* McDonough*, supra* note 500 at 209.

516. Allison, et al., *supra* note 495 at 285. (Discussing the reasons NPEs employ the war chest strategy); *see also* Mark A Lemley, *Software Patents and The Return of Functional Claiming*, 2013 WIS. L. REV. 905 (2013).

The war chest strategy involves asserting the patent against numerous parties, settling with weaker parties to finance the ongoing litigation, and then litigating more aggressively and longer against parties with more capital.[517] Interestingly, although large companies tend to dominate patent headlines, most unique defendants to NPE suits are small.[518] Thus, aggressive litigation against the final defendants is possible because the patent's value is captured during early settlements with smaller companies.[519] In turn, this allows NPEs the opportunity to play with house money. In such instances, the strategy relies significantly on the defendant's risk exposure, rather than the claim's merits.[520] Perhaps, the most critical aspect for AI patent development and IP strategy is developing a valuable portfolio.

C. Valuation

The way in which IP is valued is a crucial consideration for a firm's strategic planning, growth strategy, and bottom line. As a whole, the IP system is designed to encourage innovation by offering a temporary monopoly over inventions or works of authorship.[521] Some investors and firms have come to view patents as economic assets, *per se*.[522] Yet, many patents turn out to be worthless.[523] The truth is patent valuation is more art as science, often relying on an array of factors, without bright-line rules.[524]

Interestingly, Professor Allison argues valuable patents can be identified, at least in the aggregate.[525] According to Allison litigated patents tend to be more valuable.[526] Substantively, Allison argues valuable patents

---

517. Allison, et al., *supra* note 495 at 285-286.

518. Colleen Chien, *Startups and Patent Trolls*, 17 Stan. Tech. L. Rev. 461, 464 (2014).

519. *Id*.

520. Allison, et al., *supra* note 495 at 286.

521. Benjamin N. Roin, *Intellectual Property Versus Prizes: Reframing the Debate*, 81 U. Chi. L. Rev. 999, 1001 (2014).

522. Malcom T. Meeks, Charles A. Eldering, PhD, *Patent Valuation: Aren't we forgetting something? Making the case for claims analysis in patent valuation by Proposing a Patent Valuation Method and a Patent-Specific Discount Rating Using the CAPM*, 9 Nw. J. Tech. Tech. & Intell. Prop. 194, 194 (2010).

523. John R. Allison, et al. *Valuable Patents*, 92 Geo. L.J. 435, 437 (2004).

524. Pablo Fernandez, *Company Valuation of Brands and Intellectual Capital* (2019) (accessed at https://papers.ssrn.com/sol3/papers.cfm?abstractid=274973).

525. Allison et al. *supra* note 523 at 438.

526. *Id.*

CHICAGO-KENT J. INTELL. PROP. Vol 19:3

cite more prior art and contain more claims.[527] Indeed, litigated patents tend to have more claims, prior art citations, and citations received.[528] Allison's work provides strong support for general correlations between valuable and non-valuable patents.[529] However, in the context of specific patent valuation, three valuation methods are most commonly used: income models, cost models, and market models.

i. Income Models

Income models value assets based on the economic benefit expected to be received over the asset's life.[530] As Peter Thiel argues, "[s]imply stated, the value of a business today is the sum of all the money it will make in the future."[531] The theory is the extent to which patents affect a technologies ability to generate income influences valuation.[532] Factors included in income models include unjust enrichment, lost profits, reasonable royalty, and cash flow analysis.[533] Income models may be the strongest valuation for patents involved in infringement litigation. Indeed, patent law aims to provide patentees with payment for lost profits and other competitive harm suffered through infringement.[534]

Particularly among income models, the reasonable royalty model is appealing as it can be implemented regardless of the alleged misappropriator's actions.[535] Under a reasonable-royalty model, patent law aims to provide patentees with payment for the "rate that would have both compensated patentees and allowed users of the technology to make a

527. *Id.* Allison argues six key characteristics of litigated patents are: (1) They tend to be young— litigated soon after they are obtained. (2) They tend to be owned by domestic rather than foreign firms. (3) They tend to be issued to inventors or small companies, not to large companies. (4) They cite more prior art than non-litigated patents, and in turn are more likely to be cited by others. (5) They spend longer in prosecution than ordinary patents. (6) They contain more claims than ordinary patents.

528. Allison et al. *supra* note 523 at 451.

529. *Id.* at 438.

530. Ted Hagelin, *supra* note 376 at 363.

531. PETER THIEL, ZERO TO ONE 44 (2014). (Thiel qualifies to properly value a business you have to discount future cash flows to their present worth).

532. Ted Hagelin, *supra* note 376 at 363.

533. Gavin C. Reid, et al., *What's it Worth to Keep a Secret?*, 13 DUKE L. & TECH. REV. 116, 137 (2015).

534. Mark A. Lemley, *Distinguishing Lost Profits from Reasonable Royalties*, 51 WM. & MARY L. REV. 655, 669 (2009).

535. Gavin C. Reid, et al., *What's it Worth to Keep a Secret?*, 13 DUKE L. & TECH. REV. 116, 138 (2015).

reasonable profit."[536] For example, the twenty-five percent rule may be taken into account in income models.[537] The twenty-five percent rule suggests that a licensee pay a royalty rate equivalent to twenty-five percent of its expected profits for the patent or the product that incorporates the patent.[538] The rule has been historically used as a bedrock technique in patent license valuation.[539]

In the context of AI, income models may be difficult to develop. AI requires extensive R&D costs directed at dataset development before a technology may be commercialized.[540] For this reason, it may be months or even years before a company derives income from AI technology. Additionally, marketing companies like Google and Facebook use AI to target ads at consumers – making the total amount of income derived from the models a hazy number to calculate.

ii. Cost Models

Cost models consider factors including time, labor, replacement costs, actual damages, and research and development costs.[541] The assumption underlying cost models is the expense of developing a new asset is commensurate with the economic value the asset can provide during its life.[542] In other words, cost models are based on the idea that the technology is worth the amount it cost its owner to develop and protect.[543] Cost models incentivize firms to keep good accounts of R&D costs, making the model appealing for its ease of application.[544]

However, one concern with cost models is the lack of theoretical robustness, which may result in damages associated with the misappropriation independent of the technology's underlying value.[545] One

---

536. Mark A. Lemley, *Distinguishing Lost Profits from Reasonable Royalties*, 51 WM. & MARY L. REV. 655, 669 (2009).

537. Heather Hamel, *Valuing the Intangible: Mission Impossible? An Analysis of The Intellectual Property Valuation Process*, 5 CYBARIS AN INTELL. PROP. L. REV. 183, 188 (2014).

538. *Id.*

539. *Id.*

540. Ebrahim, *supra* note 436.

541. Hamel, *supra* note 537 at 187.

542. Ted Hagelin, *supra* note 376 at 360.

543. Gavin C. Reid, et al., *What's it Worth to Keep a Secret?*, 13 DUKE L. & TECH. REV. 116, 139 (2015).

544. *Id.*

545. *Id.*

factor which may be considered in a cost model is a patent's inventorship.[546] A common argument is the greater the number and prestige of the inventors on a patent, the higher the patent quality because more intelligence and time was dedicated to the patent.[547] It follows, the inventor's prestige and time spent developing a patent may be considered correlational with patent quality.[548] For example, Google's 258' and 741' patents, which both relate to methods and systems for reinforcement learning were invented by Volodymyr Mnih, perhaps the world's most prominent AI researcher.[549] As such, the 258' and 741' patents are likely two of the most valuable AI patents. However, a counterargument against this theory is that such estimations may overlook inventions by a single previously unknown inventor which took substantial time and effort.[550]

Cost models are most favorable to AI technology – which has most of its value in the future. Costs models could include R&D cost for developing AI technology, patent prosecution fees, and engineering fees for the technology. However, cost models are difficult to assert in litigation because the firm claiming infringement must value its own costs. This can be difficult, especially for small startups, who may otherwise have no revenue in early stages. The task requires figuring out exactly how much time was spent developing the technology and what the hourly rates were for each person working on the technology.

iii. Market Models

Market models define fair market value for a technology.[551] Generally, market models value assets based upon comparable transactions

---

546. Hamel, *supra* note 537 at 187.

547. Malcom T. Meeks, Charles A. Eldering, PhD, *Patent Valuation: Aren't we forgetting something? Making the case for claims analysis in patent valuation by Proposing a Patent Valuation Method and a Patent-Specific Discount Rating Using the CAPM*, 9 Nw. J. Tech. Tech. & Intell. Prop. 194, 199 (2010).

548. R. Polk Wagner, *Understanding Patent-Quality*, 157 U. Pa. L. Rev. 2135, 2138 (2009) (prestige and time may also correlate with the capacity of a granted patent to meet the statutory standards of patentability – most importantly, to be novel, nonobvious, and clearly and sufficiently described).

549. Both the 258' and 741' patents stem from Mnih's seminal work published in Nature. *See* Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 Nature Int'l J. Sci. 529, (2015). *See also* U.S. Patent No. 9,679,258 B2, *supra* note 442.. *See also* United States Patent No. 10,346,741 (issued Jul. 9, 2019) (assigned to DeepMind Technologies – a Google subsidiary).

550. Hamel, *supra* note 537 at 188.

551. Gavin C. Reid, *supra* note 543..

between unrelated parties.[552] In essence, the fair market value is determined by assessing the price a buyer would pay a seller for the technology.[553] Other factors included in market valuations are sales and industry surveys.[554] Market models generate the widest range of valuations.[555] One reason for market model's higher variance is the subjectivity in measuring market value compared to other models.[556] A second reason for the higher variance is dependent upon whether market analysis is conducted prospectively or retroactively.[557] Indeed, prospective market valuations tend to be more grounded with the support of financial data as opposed to retroactive valuations.

Intimately intertwined with a technology's market value is the technology's commercialization.[558] In addition to the revenue from licensing, a patent's ability to trigger sales is also relevant in technology valuation.[559] Indeed a patent's ability to influence consumers to buy a product or a newer version of an existing product correlates with increase in value.[560] For example, ownership rights in the latest technology for a computer or cell phone increases firm value.[561] Another example is a patent's ability to trigger sales in an entirely new market – like Edison's electricity empire in the late 19th century.[562]

---

552. Hagelin, *supra* note 376 at 362; *see also* Elona Marku, et al., Disentangling the Intellectual Structure of Innovation and M&A Literature (2017).

553. Hamel, *supra* note 537 at 204.

554. *Id.*

555. Reid, *supra* note 543.

556. *Id.*

557. W. Michael Shuster, *Artificial Intelligence and Patent Ownership*, 75 WASH. & LEE L. REV. 1945, 1987 (2018).

558. *Id.*

559. Malcom T. Meeks, *Charles A. Eldering, PhD, Patent Valuation: Aren't we forgetting something? Making the case for claims analysis in patent valuation by Proposing a Patent Valuation Method and a Patent-Specific Discount Rating Using the CAPM*, 9 NW. J. TECH. TECH. & INTELL. PROP. 194, 199 (2010).

560. W. Michael Shuster, *Artificial Intelligence and Patent Ownership*, 75 WASH. & LEE L. REV. 1945, 1987 (2018).

561. *Id.*

562. Shubha Ghosh, *Decoding and Recoding Natural Monopoly, Deregulation, and Intellectual Property*, 2008 U. ILL. L. REV. 1125, 1170 (2008). *See also* U.S. Patent No. 265,786 (1882). *See also* U.S. 219,268 (1879).

476   *CHICAGO-KENT J. INTELL. PROP.*   Vol 19:3

Market models are likely to be least favorable for valuing AI IP. One reason is because so much AI software is open source,[563] in many instances the technology may be worthless in the market. However, for a larger firm like IBM or Amazon, market models may be favorable because of niche market monopolies in industries like defense and retail.[564] But, proving a patent for AI technology or AI software more generally is what induces the market to act is a difficult task because AI software is often similar in its fundamental structures.[565] Yet, much of a patent's value is in its ability to exclude competitors from the market.[566] In the AI market, virtually no exclusion rights have been exercised.

In sum, there is no established market for intellectual property.[567] Thus, A wide array of factors are considered during technology valuations, which account for more than $12 trillion in annual economic activity.[568] In sum, existing IP valuation regimes are widely understood to exist to promote invention, dissemination, and commercialization of intellectual works.[569] However still, no bright-line rule exists for technology valuation. Instead, valuation factors include the business context of the products relating to the invention, the state of technological progress, and anticipated commercialization opportunities.[570]

563. *See* TensorFlow, GITHUB, https://github.com/tensorflow. *See also* OpenAI, GitHub, https://github.com/openai.

564. One particular interest for the technology industry is the heavy exploitation of the Federal Government as a customer. Federal ownership of IP, whether in whole or in part effects the ownership rights of firms using such IP. *See* 35 U.S.C. §207. *See also* Alexander Rogosa, *Shifting Spaces: The Success of The SpaceX Lawsuit and The Danger of Single-Source Contracts in America's Space Program*, 25 FED. CIRCUIT B.J. 101, 103 (2015).

565. United States Patent No. 10,346,741 (issued Jul. 9, 2019) (assigned to DeepMind Technologies – a Google subsidiary). *See also* U.S. Patent No. 10,032,281, Multi-scale deep reinforcement machine learning for N-dimensional segmentation in medical imaging (July 24, 2018), Ghesu , et al. (assigned to Siemens Healthcare). *See also* U.S. Patent No. 10,296,830, to Cai, et al. Dynamic topic guidance in the context of multi-round conversation (May 21, 2019). (Assigned to International Business Machines Corporation).

566. W. Michael Shuster, *supra* note 557.

567. PALFREY, *supra* note 375 at 126 (IP is worth what someone is willing to pay for it).

568. *See* Digital Spillover, *Measuring the true impact of the Digital Economy*, HUAWEI & OXFORD ECONOMICS (2017), (accessed at https://www.huawei.com/minisite/gci/en/digital-spillover/files/gci_digital_spillover.pdf) (measuring market in 2016 as $11.5 trillion, growing at 2.5x the rate of global GDP). *See also* PALFREY, INTELLECTUAL PROPERTY STRATEGY 126 (MIT Press 2012).

569. Gregory N. Mandel, *Leveraging the International Economy of Intellectual Property*, 75 OHIO ST. L.J. 733, 734 (2014).

570. Landers, *supra* note 477 at 165*.*

VII.   FUTURE CONSIDERATIONS

A. AI Patent Trends

The positive trend in AI patent market growth is accelerating. According to the patents in this Article's dataset, in the year 1999 there were 30 patents; in the year 2004 there were 71 patents; in the year 2009 there were 150 patents; in the year 2014 there were 392 patents; and in the year 2019 there were 2,459 patents in the AI patent market.[571] Figure 19 shows the AI patent market's growth according this Article's dataset.
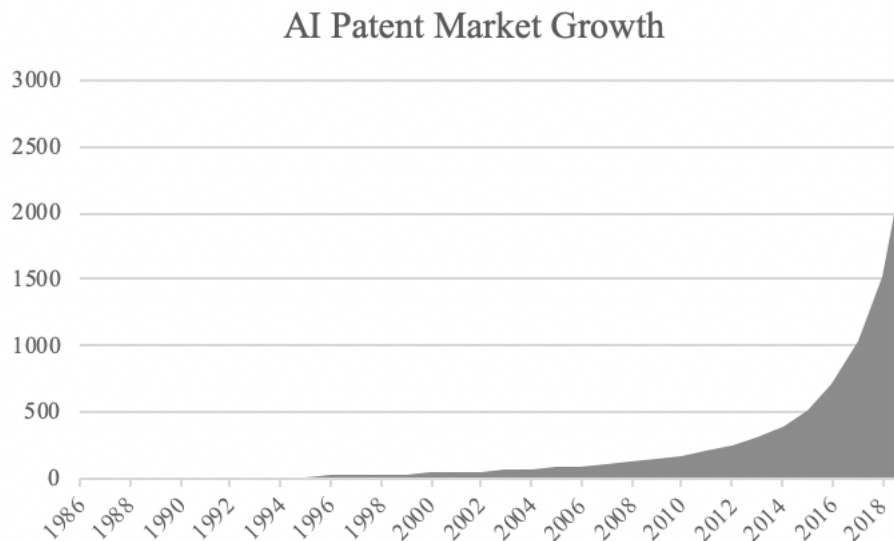
Figure 19[572]

While this Article's dataset captures a small fraction of the total AI patent market, the growth rate in the four markets this Article explores reflect more rapid expansion than AI patents more generally.

According to the patents in the CMU dataset, in the year 1997 there were 2,529 AI patents; in the year 2002 there were 7,329 AI patents; in the year 2007 there were 15,481 AI patents; in the year 2012 there were 34,700 AI patents; and in the year 2017 there were 70,412 AI patents in the

---

571. Haney, *supra* note 40.

572. *Id.* (The information contained in this chart was prepared by the author with information from the United States Patent and Trademark Office).

market.[573] Figure 20 represents the AI patent market's growth according to the CMU dataset.
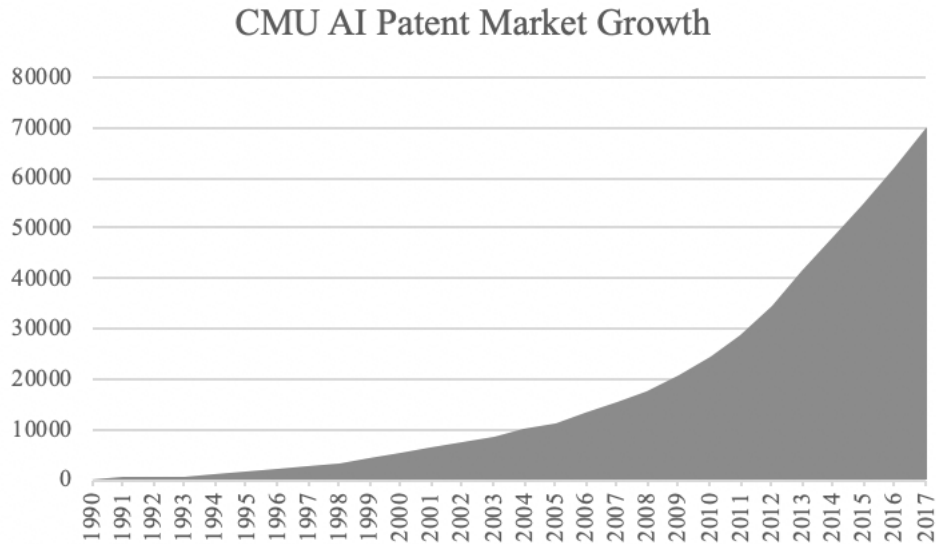
Figure 20[574]

One key difference between this Article's dataset and the CMU dataset is the CMU dataset calculated year by filing date – whereas this Article calculated year by the date a patent was granted. The growth rate presented in the CMU dataset is also accelerating—albeit at a slower rate of change.

The extent to which firms have captured market share in the AI patent market remains less clear. In this Article's dataset, IBM owns a significant portion of the total market with 741 of 2,459 total patents, or just over thirty percent.[575] Figure 21 shows the number AI patents held by each firm according to this Article's dataset.

573. Alderucci, et al., *supra* note 11.
574. Alderucci, et al., *supra* note 11.
575. Technically 30.134%.
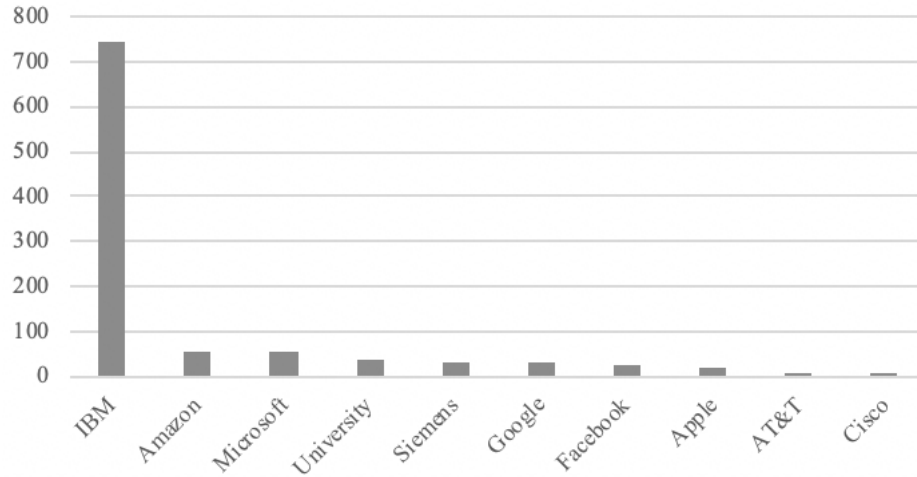
AI Patents by Company

Figure 21[576]

Further, in this dataset Amazon and Microsoft are tied with the second largest market shares at 54 patents each.[577] IBM appears to have a decisive advantage in terms of patents in the four particular types of machine learning in this Article's dataset. Yet on a broader scale, the data reflect a slightly different picture.

By contrast, Microsoft has the largest market share with 3,822 of 70,412 total patents – roughly 5.4% of the total market.[578] Figure 22 represents AI patents by firm according to the CMU dataset.

---

576. Haney, *supra* note 40.
577. *Id.*
578. Technically 5.428%.
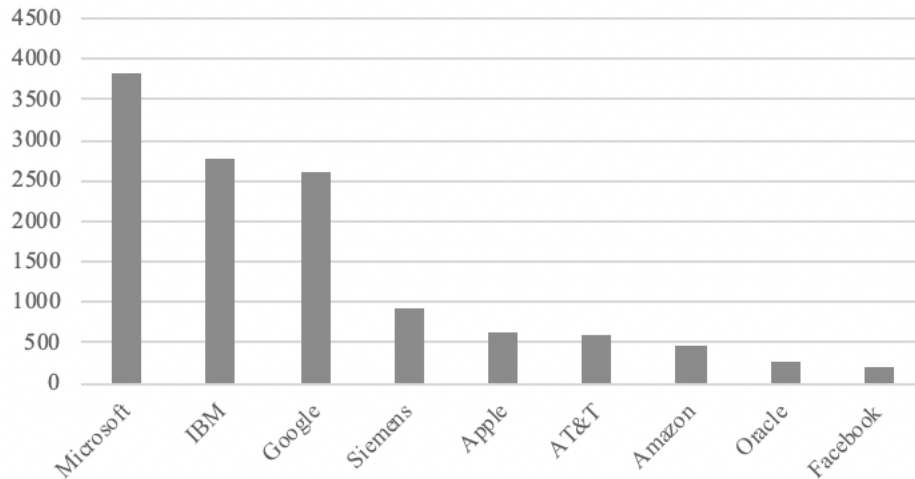
CMU AI Patents by Company

Figure 22[579]

IBM comes in second with 2,761 patents and Google in third with 2,595.[580] One explanation for the different relative positions is IBM focuses its patents on more research focused terms such as natural language processing and less on applied terms like artificial intelligence or machine learning. Further, regarding Facebook and Apple, who have a demonstrably smaller presence in both datasets compared to other big technology companies, one perspective is these companies have less of target for NPEs. At the same time, Microsoft and IBM have a larger sword and shield.

B. Patent Generation

Natural language generation (NLG) is a process of synthesizing language to form sequences with syntactic accuracy and semantic coherence.[581] While some argue this a uniquely human activity,[582] these processes are capable of logical representation.[583] In 2017, a team of researchers from Google and the University of Toronto published the paper,

---

579. Alderucci, et al., *supra* note 11.

580. *Id.*

581. ALPAYDIN, *supra* note 49 at 109.

582. John McGinnis, *Accelerating AI,* 104 NW. U. L. REV. COLLOQUY 366, 368 (2010).

583. ALPAYDIN, *supra* note 49 at 2 (arguing the driving force of computing technology is the realization that every piece of information can be represented as numbers).

2020   *AI PATENTS: A DATA DRIVEN APPROACH*   481

Attention Is All You Need.[584] The paper introduced a novel AI model architecture, the Transformer.[585] Rather than using RNNs or CNNs, the Transformer utilizes an autoencoder[586] with an attention mechanism.[587] The attention mechanism encodes and stores a series of hidden vectors, which are decoded to generate new text.[588] Thus, one approach to developing NLP applications for patent generation is using an attention model.

Indeed, a recent study used GPT-2[589] for patent claim generation.[590] The researchers created a dataset of 555,890 patent claims which were preprocessed for training a GPT-2 model.[591] The study used cloud computing resources from Google to conduct their experiments.[592] The researchers hoped the attention model would show performance improvement compared to ANN models.[593] Unfortunately, a significant portion of the model's generated text was senseless.[594] Yet, the study's authors suggest using a deep learning model in conjunction with the attention mechanism may significantly improve future results.[595]

C. Singularity v. Stagnation

Conventional wisdom teaches technological progress is driven by the LOAR.[596] The LOAR's application to information technology, Moore's Law, projects exponential trends in technological progress toward an ultimate

---

584. Vaswani, et al., *Attention is All You Need 1*, GOOGLE (2017) (https://arxiv.org/abs/1706.03762).

585. *Id*. at 2.

586. An autoencoder is a type of neural network trained to reconstruct its input at its output.

587. *Id*. at 1.

588. Vaswani, et al., *supra* note 584 an attention function is a vectorized mapping a query and a set of key-value pairs to an output).

589. Generative Pre-Training Model (GPT-2) is a large-scale unsupervised language model that generates paragraphs of text, first announced by OpenAI in February 2019.

590. Jieh-Sheng Lee & Jieh Hsiang, *Patent Claim Generation by Fine-Tuning OpenAI GPT-2* (2019) https://arxiv.org/abs/1907.02052.

591. *Id*.

592. *Id*.

593. *Id*. *See also* Ebrahim, *supra* note 436.

594. *Id*. Law of Accelerating Returns ("LOAR").

595. *Id*.

596. Brian S. Haney, *The Perils and Promises of Artificial General Intelligence*, 45 J. LEGIS. 151, 155 (2019).

technological singularity.[597] This notion has developed into a school of thought called Technological Utopianism. Technological Utopianism refers to the idea that digital life is the natural and desirable next step in the cosmic evolution of humanity, which will certainly be good.[598] As a result of Technological Utopianism, a majority of literature on the subject of technology is inherently optimistic, both in terms of outcomes and rates of progress.[599] As a result, utopians argue society as a whole should embrace technology because innovation leads to equality among a society.[600]

However, the utopian perspective is inherently misguided – ignoring the realities of the human condition.[601] Consider, the world's richest men – Bill Gates and Jeff Bezos – both made their fortunes in technology.[602] And, new technologies undoubtedly create winners and losers in the labor market.[603] However, the degree to which winners reap rewards comes at an expense to the losers. It is no surprise Northern California's Bay Area is the center of the world's technological innovation, while simultaneously having the highest percentages of homelessness in the United States.[604]

Peter Thiel tells the story, that our ancestors lived in static, zero-sum societies where success meant seizing things from others.[605] Then, after 10,000 years everything changed in 1600s and progress began to occur due to the development of technology.[606] Society moved from primitive agriculture to medieval windmills, then steam engines in the 1760s, with accelerating technological progress through the industrial revolution until the 1970s.[607]

---

597. RAY KURZWEIL, HOW TO CREATE A MIND 250 (2012).

598. TEGMARK, *Supra* note 14, at 32.

599. NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 34 (2017). *See* KURZWEIL, *supra* note 597 at 261.

600. Eleanor Lumsden, *The Future is Mobile: Financial Inclusion and Technological Innovation in The Emerging World*, 23 STAN. J.L. BUS. & FIN. 1, 5-6 (2017) (arguing the best hope for eradicating poverty is technological innovation).

601. Peter Thiel, *The Education of a Libertarian*, CATO UNBOUND (May 1, 2009), https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian.

602. *The Richest People in The World*, FORBES (March 5, 2019), https://www.forbes.com/billionaires/#77480d02251c.

603. Michael Webb, *The Impact of Artificial Intelligence on the Labor Market*, STANFORD (2019), https://web.stanford.edu/~mww/webb_jmp.pdf.

604. MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION 7 (1971). *See also* The U.S. Dep't of Hous. & Urban Dev., The 2018 Annual Homeless Assessment Report (AHAR) to Congress, 33 (2018), https://www.novoco.com/sites/default/files/atoms/files/hud_ahar_2018_121718.pdf.

605. PETER THIEL, ZERO TO ONE 8-9 (2014).

606. *Id*.

607. *Id*.

But, Thiel's argument technology liberates society from the zero-sum world is mistaken.

Progress happens slower than most suspect. Modern society remains a zero-sum game and perhaps the greatest delusion of modern society is that we ever left the state of nature – or that technology is separate from it.[608] It is unlikely mankind is on the verge of a technological singularity. Looking to the past – we should expect more of the same for the future. Great technology is simple, easy to use, and intuitive. Indeed, the Latin maxim *simplex sigillum veri* stands for the principle – simplicity is the sign of truth.[609] Or, in the words of Richard Branson, Founder of Virgin Group: "If something can't be explained on the back of an envelope, it's rubbish."[610]

608. Thomas Hobbes, Leviathan (1651) (as Hobbe's wrote, life is "nasty, brutish and short").
609. James Morwood, Oxford Latin Desk Dictionary, 174-75 (2005) (defining Latin to English translations of *simplex* and *sigillum*).
610. Carmine Gallo, The Storyteller's Secret 112 (2016).

APPENDIX A. SUMMARY OF NOTATION

Summary of Notation

Notation	Meaning
π^*	Optimal policy.
$Q(s,a)$	Q-function.
(s,a)	State-action pair.
ϕ	Q-function parameters.
γ	Discount factor.
$\mathbb{E}[x]$	Expectation of random variable.
$\arg\max_a f(a)$	A value of a, at which $f(a)$ takes its maximal value.
r	Reward.
θ_k	Policy parameters for k experiment.
$L(s,a,\theta_k,\theta)$	Objective function.
$A^{\pi_{\theta_k}}$	Advantage estimate for policy given parameters.
$\pi_\theta(a\|s)$	The policy given parameters.
ϵ	Hyperparameter defining how far away the new policy is allowed to go from the old.
$a^*(s)$	Optimal action-value function.
$\mathcal{D}$	Replay Buffer.