IP Theory

Volume 9 | Issue 1 Article 2

Winter 2020

Using AI to Analyze Patent Claim Indefiniteness

Dean Alderucci
Carnegie Mellon University, alderucci@cmu.edu

Kevin Ashley
University of Pittsburgh, ashley@pitt.edu

Using AI to Analyze Patent Claim Indefiniteness

Dean Alderucci
Carnegie Mellon University

Kevin D. Ashley
University of Pittsburgh

Abstract

We describe how to use artificial intelligence (AI) techniques to partially automate a type of legal analysis, determining whether a patent claim satisfies the definiteness requirement. Although fully automating such a high-level cognitive task is well beyond state-of-the-art AI, we show that AI can nevertheless assist the decision maker in making this determination. Specifically, the use of custom AI technology can aid the decision maker by (1) mining patent text to rapidly bring relevant information to the decision maker's attention, and (2) suggesting simple inferences that can be drawn from that information.

We begin by summarizing the law related to patent claim indefiniteness. A summary of existing case law allows us to identify the types of information that can be relevant to the legal determination of indefiniteness. This in turn guides us in designing AI software that processes a patent's text to extract information that can be relevant to the legal analysis of indefiniteness. Some types of relevant information include whether terms in a claim are defined in the patent, whether terms in a claim are not mentioned in the patent's specification, whether the claim includes non-standard terms coined by the drafter of the patent, whether the claim relies on vaguely-specified measurements, and whether the patent's specification discloses structure corresponding to a means-plus-function limitation.

The AI software rapidly processes a patent's text and identifies information that is relevant to the legal analysis. The software then provides the human decision maker with this information as well as simple metrics and inferences, such as the percentage of claim terms that are defined explicitly or by example, and whether terms that are coined by the drafter should be defined or renamed. This can provide the user with insights about a patent much faster than if the user read the entirety of the patent to locate the same information unaided.

Moreover, the software can aggregate the various types of information to "score" a claim (e.g., from 0 to 100) based on its risk of being deemed indefinite. For example, a claim containing only defined terms and lacking any vague measurements would score much lower in terms of risk than a claim with terms that are not only undefined but do not even appear in the patent's specification. Once each claim in a patent is assigned such an indefiniteness score, the patent itself can be given an overall indefiniteness score.

Scoring groups of patents in this manner has further advantages even if the scores are blunt measurements. AI software ranks a group of patents (e.g., all patents owned by a company) by indefiniteness scores. This allows a very large set of patents to be quickly searched for patents that have the highest, or lowest, indefiniteness score. The results of such a search could be, e.g., the patents to target for detailed review in litigation, post-grant proceedings, or licensing negotiations.

Finally, we present some considerations for refining and augmenting the proposed methods for partially automating the indefiniteness analysis, and more broadly other types of legal analysis.

Table of Contents

I. INTRODUCTION

In many fields of law, the legal decision maker must review numerous documents, such as statutes, case law, contracts, memoranda, and email correspondence, in order to identify and synthesize information that is relevant to a legal decision. The nature of the legal decision dictates the types of relevant information, the documents in which that information might be found, and the types of decisions that can be drawn from that information. For example, in the field of patent law the decision maker, such as a lawyer, judge, or patent office examiner, might need to review a patent describing an invention and the technological principles by which that invention operates. From that text the decision maker must extract information such as the details of the invention, how it functions, how it can be constructed, and how it differs from previously existing technology. The decision maker might need to read additional documents containing information that is required to draw a legal conclusion, such as whether the invention is patentable or whether a product infringes the patent.

Searching large amounts of text to locate information of interest is a task at which software excels. Artificial intelligence (AI) software[1] is especially adept at identifying text that represents a phrase or sentence of interest, even text that uses words different from those the searcher

[1] We use the term AI to generally indicate software that appears to behave with human intelligence. *See generally* Kevin D. Ashley, *Automatically Extracting Meaning from Legal Texts: Opportunities and Challenges*, 35 GA. ST. U.L. REV. 1117, 1117-18 (2019).

employed. AI software is also adept at mimicking complex decision making, besting human champions at games such as chess[2] and video games[3].

Despite these and other impressive accomplishments, many fairly straightforward legal analysis tasks exceed the capabilities of even state-of-the-art AI systems.[4] For this class of legal analysis we must accept more modest goals from AI systems. Rather than attempt to automate the entire process of legal decision making, we recommend using AI text analytics to locate and extract from text information that is potentially relevant to the legal decision. The human remains responsible for evaluating this information and drawing the legal conclusion she deems most appropriate in light of this information.

Legal text analytics involve computational techniques that apply Natural Language Processing (NLP), Machine Learning (ML), and other methods to automatically extract some aspects of meaning or semantics from the text of legal case decisions, contracts, patents, or statutes.[5] Although AI software cannot yet read legal texts the way human attorneys do, the semantic information that can be extracted has been applied in numerous ways to assist legal decision-making, although subject to various limitations discussed below. Some examples of utilizing partial semantic information include predicting outcomes of cases and particular issues, answering legal questions relevant to the decisions, retrieving documents or passages containing information that is potentially relevant to the legal decisions, and summarizing relevant aspects of cases. Each of these is described briefly.

Predicting outcomes of IP-related litigation including patent infringement cases has been the focus of Lex Machina[6], which predicts outcomes of new cases based on litigation participants and their behavior in a large repository of past IP cases.[7] Lex Machina eschews representing cases in terms of features related to substantive issues in favor of features like the identities of the sides in a lawsuit, their attorneys, the presiding judge, and the court district where the action was filed. In the future, data analysis will be applied other non-substantive features, for example, in courts' docket information. This may yield ever more accurate estimates of the time until a decision is rendered and the costs of litigation, data which will factor in litigators' strategic decision-making, evaluation of claims, and settlement advice.

---

[2] Researcher at DeepMind developed a program that taught itself to play chess in four hours with sufficient skill to convincingly defeated a world champion chess program. David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play", Science, Vol. 362, Issue 6419, pp. 1140-1144 (Dec 7 2018).

[3] Researchers at DeepMind created a program that plays the StarCraft II video game at a Grandmaster level under professionally approved conditions. It ranks above 99.8% of all officially-ranked human players. Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning, 575 Nature 350–354 (2019).

[4] *See generally* Dean Alderucci, *Legal Reasoning and Customized AI Techniques for the Patent Field*, 58 DUQ. L. REV. ________ (forthcoming 2019) (describing how contemporary Machine Learning and Natural Language Processing algorithms cannot penetrate many types of "common-sense" reasoning that are essential to the vast majority of legal analysis tasks).

[5] *See generally* Kevin D. Ashley, ARTIFICIAL INTELLIGENCE AND LEGAL ANALYTICS: NEW TOOLS FOR LAW PRACTICE IN THE DIGITAL AGE 5 (2017).

[6] Lex Machina Home Page, https://lexmachina.com.

[7] Mihai Surdeanu, Ramesh Nallapati, George Gregory, Joshua Walker & Christopher D. Manning, *Risk Analysis for Intellectual Property Litigation*, PROC. OF THE THIRTEENTH INT'L CON. ON ARTIFICIAL INTELLIGENCE AND LAW, 116 (2011).

Legal question answering involves searching large text collections to locate a sentence or short excerpt from a document that appears to answer a user's question. For example, the Ross program,[8] based on IBM Watson,[9] accepts as input a question in plain English such as, "In New York, what is secondary liability with respect to copyright infringement and how is it established?" In response, as an answer it outputs a sentence such as:

> ". . . A party is liable for contributory infringement if, 'with knowledge of the infringing activity,' it 'induces, causes, or materially contributes to the infringing conduct of another.' Gershwin Publ'g . . . .", complete with citation, "Arista Records LLC v. Usenet.com, Inc. (SDNY 2009)," suggested readings, and updates.

In fact, Ross retrieves multiple answers, ordered in terms of the likelihood that the short text answers the query and its confidence in the responsiveness of its answer.

Increasingly, legal information retrieval makes use of citation network analysis and legal text analytics. A citation network graphs reference relations among legal cases or statutory provisions. Ravel[10] presents these in visual maps of citations of U.S. cases regarding a legal concept input by a user. For example, a user may input "invalid for indefiniteness" revealing a map of cases citing or cited by the U.S. Supreme Court case Nautilus, Inc. v. Biosig Instruments, Inc.[11] that are relevant to the query, that is, important in the context of search terms such as "invalid for indefiniteness." It lists matching excerpts such as:

> ". . . patent survived indefiniteness review. Held: 1. A patent is invalid for indefiniteness if its claims, ..."

In addition, according to an Executive Editor of the publisher TechCrunch, "Through its Judge Analytics dashboard, Ravel Law highlights the top cases and federal circuit courts that a judge has cited in the past, cases that might be more familiar to a judge and therefore more likely to receive a favorable ruling."[12] As an example, an attorney could determine that the judge assigned to a patent lawsuit is more likely to cite cases from the circuit where the judge studied law and clerked.

With CaseText's *CARA system*[13], one can input a brief (i.e., written memorandum of law) and the program will output and summarize additional cases to cite in support of arguments in the brief based on citation networks. In Google Scholar Cases[14], the "How Cited" tool groups cases citing a particular case into equivalence classes of citing cases that cite the case for the same reason.

---

[8] Ross Intelligence Home Page, https://rossintelligence.com.

[9] *See generally* Betsy Cooper, *Judges in Jeopardy!: Could IBM's Watson Beat Courts at Their Own Game?*, 121 YALE L.J. ONLINE 87 (2011) (describing how IBM's Watson system beats the world's top Jeopardy! Champions); David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer & Chris Welty, *Building Watson: An Overview of the DeepQA Project*, 31 AI MAGAZINE, Fall 2010, at 59.

[10] RAVEL Home Page, https://ravellaw.com (visited, July 29, 2019).

[11] 572 U.S. 898, 901 (2014).

[12] Danny Crichton, *With Judge Analytics, Ravel Law Starts to Judge the Judges*, TECHCRUNCH, https://techcrunch.com/2015/04/16/who-judges-the-judges/ (April 16, 2015).

[13] Casetext Home Page, https://casetext.com.

[14] Google Scholar Search Page, https://scholar.google.com.

Tools like these use text analytics to obtain more info about the citation links between citing and cited cases about why a case is being cited. They may identify the topic of the paragraph where the cite appears or that of the cited case. For example, after inputting the *Nautilus, Inc*. cite to Google Scholar Cases, the How Cited tool identifies 915 decisions that cite it for the Court's definition of "invalid for indefiniteness".

Using similar legal text analytics, programs increasingly are able to summarize legal decisions in terms designed to convey their legal significance so that readers can more readily determine if they are worth following up.[15] The automatic summarizers extract sentences from the full texts of decisions based on a variety of criteria and then fit them into a template. The criteria include sentence topics, roles that sentences play in legal argument, predictiveness of a sentence, that is, its correlation with case outcome, satisfaction of specific template requirements, or dissimilarity from other included sentences. Some automatically-generated summaries have been evaluated by knowledgeable human users.[16]

Despite these advances, the text analytic applications are subject to significant limitations. Prediction programs that do not represent substantive features of a legal dispute cannot explain their predictions in terms with which legal professionals would be familiar. Legal question answerers do not understand either the questions or the answers. They rely on matching, not reading, and cannot reason about how small changes in the question would affect the applicability of the answer. Similar limitations affect legal summarizers and much work has yet to be done before they can tailor summaries to a user's specific problem or need.

Given these limitations, it is important to identify "use cases" where the limitations of legal text analytic tools do not necessarily interfere with the utility of the outputs. A use case is a description of how the technology is to be used and the kinds of problems that it is being used to solve. Suitable use cases are likely to involve tasks where the roles of humans and machine are clearly delineated and where the interaction takes advantage of those abilities to perform the intelligent activities that each performs best. This interaction has been called cognitive computing.[17]

Branting, et al. are working on automating explainable predictions with respect to particular issues based on substantive features.[18] Prediction, however, is not just a matter of identifying the issues or matching to past cases. Some legal issues involve specialized tests or complex, context specific analyses. It may be infeasible to automate the analysis completely but possible to decompose the decision making into subsidiary tasks. For some of these tasks, AI can be applied to identify relevant information from which a program may draw simple inferences or combine different types of information into an aggregate metric that bears upon a legal decision even if it does not determine it.

---

[15] Paheli Bhattacharya, Kaustubh Hiware, Subham Rajgaria, Nilay Pochhi, Kripabandhu Ghosh & Saptarshi Ghosh, *A Comparative Study of Summarization Algorithms Applied to Legal Case Judgments*, EUROPEAN CONF. ON INFORMATION RETRIEVAL 413 (2019).

[16] *See, e.g.*, Linwu Zhong, Ziyi Zhong, Zinian Zhao, Siyuan Wang, Kevin D. Ashley & Matthias Grabmair, *Automatic Summarization of Legal Decisions Using Iterative Masking of Predictive Sentences*, PROC. OF THE SEVENTEENTH INT'L. CONF. ON ARTIFICIAL INTELLIGENCE AND LAW 163 (2019).

[17] Ashley, *supra* note 5, at 13, 395.

[18] Luther Karl Branting, Craig Pfeifer, Lisa Ferro, Alex Yeh, Bradford Brown, Brandy Weiss, Mark Pfaff & Amarty Chakraborty, *Expoliting Search Logs to Aid in Training and Automating Infrastructure for Question Answering in Professional* Domains, PROC. OF THE SEVENTEENTH INT'L CONF. ON ARTIFICIAL INTELLIGENCE AND LAW 413 (2019).

That is the approach we take with the issue of patent claim indefiniteness. We propose the use of AI software designed to locate in patent text information that is potentially relevant to claim indefiniteness. Accordingly, a necessary first step is to endow the AI software with the ability to understand what types of information bears upon the legal decision.

II. THE STANDARD FOR INDEFINITENESS

In this section we provide a brief overview of the legal standard for patent claim indefiniteness. The purpose of this review is to understand not only the analysis that the legal decision maker undertakes but also the types of information that are relevant to the analysis. Both will be used in designing AI software to assist the decision maker.

The definiteness requirement is specified in 35 U.S.C. § 112(b), which requires that the "specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter" of the invention.[19] The Supreme Court recently clarified the statutory standard by holding that "a patent is invalid for indefiniteness if its claims, read in light of the specification delineating the patent, and the prosecution history, fail to inform, *with reasonable certainty*, those skilled in the art about the scope of the invention."[20] The "reasonable certainty" standard balances two interests. On the one hand patent claims should provide the public with "clear notice" of the exclusionary rights provided by the patent.[21] Distinct claims "guard against unreasonable advantages to the patentee and disadvantages to others arising from uncertainty as to their [respective] rights."[22] However, "the definiteness requirement must take into account the inherent limitations of language."[23] Some uncertainty is the "price of ensuring the appropriate incentives for innovation."[24]

Claim definiteness is a question of law that courts review without deference.[25] This flows from a court's obligation to construe claims *de novo*.[26] Nevertheless, the definiteness inquiry depends on "the understanding of a skilled artisan at the time of the patent application, not that of a court viewing matters post hoc."[27] Thus, the particular level of skill of the person having ordinary skill in the art is relevant to definiteness.[28] Determining "the perspective of one of skill in the art may involve reference to evidence extrinsic to the patent, such as prior art and witness

---

[19] 35 U.S.C. § 112(b) (2012).

[20] Nautilus, Inc., v. Biosig Instruments, Inc., 572 U.S. 898, 901 (2014) (emphasis added).

[21] *Id.* at 909-910 (holding that clear notice is necessary to avoid "[a] zone of uncertainty which enterprise and experimentation may enter only at the risk of infringement claims." (quoting United Carbon Co. v. Binney & Smith Co., 317 U.S. 228, 236 (1942)).

[22] General Electric Co. v. Wabash Appliance Corp., 304 U.S. 364, 369 (1938).

[23] *Nautilus, Inc.*, 572 U.S. at 899.

[24] *Id.* at 909 (quoting Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U. S. 722, 732 (1996)).

[25] Star Sci., Inc. v. R.J. Reynolds Tobacco Co., 655 F.3d 1364, 1373 (Fed. Cir. 2011).

[26] Kinetic Concepts, Inc. v. Blue Sky Med. Grp., Inc., 554 F.3d 1010, 1022 (Fed. Cir. 2009).

[27] *Nautilus, Inc.*, 572 U.S. at 911.

[28] AllVoice Comp. PLC v. Nuance Comm. Inc., 504 F.3d 1236, 1240 (Fed. Cir. 2007).

testimony."[29] However, extrinsic evidence is irrelevant to indefiniteness where the meaning of a claim term is unambiguous from the patent's specification or other intrinsic evidence.[30]

In *Ex Parte Miyazaki*,[31] the Board announced a "lower threshold showing of ambiguity to support a finding of indefiniteness" in Patent Office proceedings than in the courts. Because issued patents are presumed valid,[32] there is "a high standard of ambiguity for finding indefiniteness" in litigation.[33] Pending patent applications are afforded no such presumption, and pending claims can be amended. Although *Miyazaki* was decided well before the Supreme Court's *Nautilus* decision, commentators have argued that its reasoning for a higher indefiniteness standard in Patent Office proceedings remains applicable.[34]

One commentator has suggested that it can be helpful to consider two distinct types of definiteness, linguistic and physical.[35] Claims that can be construed in more than one way by the person of ordinary skill are linguistically indefinite, while claims whose single meaning does not sufficiently delineate a necessary relationship among claim elements are physically indefinite.[36] For example, claims "with comparative terms or ambiguous spatial relationships between claim elements fail to meet the physical definiteness requirement."[37]

III. Indefiniteness and Means-Plus-Function Limitations

The patent statute permits a patent claim to include limitations described in terms of a function to be performed instead of the structure that performs that function.[38] For example, a claim for an ink jet printer might include a limitation "ink delivery means", where the function of ink delivery is performed by an ink spray head that has a high pressure air source.[39] This flexibility afforded the patentee is tempered by construing the limitation more narrowly than its literal words

---

[29] Solomon v. Kimberly-Clark Corp., 216 F.3d 1372, 1378 (Fed. Cir. 2000). *See also* *Kinetic Concepts, Inc.*, 554 F.3d at 1022 (Fed. Cir. 2009) (accepting extrinsic evidence regarding how a doctor of ordinary skill would understand the claim).

[30] Intel Corp. v. VIA Techs., 319 F.3d 1357, 1367 (Fed. Cir. 2003); Personalized Media Communs., L.L.C. v. ITC, 161 F.3d 696, 706 (Fed. Cir. 1998).

[31] *Ex parte* Miyazaki, No. 2007-3300, 2008 WL 5105055, at *6 (B.P.A.I. 2008).

[32] 35 U.S.C. § 282 (2012).

[33] *See* *supra* note 31, at *4. *See also* *In re* Packard, 751 F.3d 1307, 1312 (Fed. Cir. 2014) (explaining indefiniteness rejections by the USPTO arised in a different posture from that of indefiniteness challenges to an issued patent).

[34] David Orange, *Are Means-Plus-Function Claims Reasonably Certain to Require Tables of Support?*, 98 J. PAT. & TRADEMARK OFF. SOC'Y 809, 815 (2016) ("[O]ne would expect that Miyazaki automatically raises the requirements for definiteness in light of the Supreme Court's heightened requirements in Nautilus to a 'Nautilus plus Miyazaki' standard."); Andrei Iancu, Michael Fleming & C. Maclain Well, *Indefiniteness In Inter Partes Review Proceedings*, 98 J. PAT. & TRADEMARK OFF. SOC'Y 4, 7 (2016) ("Because a removal notice has not been issued by the Director, the public can assume that the Director has determined that these Board precedential opinions [such as Miyazaki] have not been overcome by subsequent Federal Circuit cases and therefore continue to be binding precedent on the Board.").

[35] Gary M. Fox, *Understanding Nautilus's Reasonable-Certainty Standard: Requirements for Linguistic and Physical Definiteness of Patent Claims*, 116 MICH. L. REV. 329 (2017).

[36] *Id.* at 342–345.

[37] *Id.* at 342.

[38] 35 U.S.C. § 112(f) (2012) ("An element in a claim for a combination may be expressed as a means or step for performing a specified function without the recital of structure, material, or acts in support thereof, and such claim shall be construed to cover the corresponding structure, material, or acts described in the specification and equivalents thereof.").

[39] *See* Signtech USA, Ltd. v. Vutek, Inc., 174 F.3d 1352, 1355 (Fed. Cir. 1999).

might suggest. Such a means-plus-function limitation covers "only the structure, materials, or acts described in the specification as corresponding to the claimed function and equivalents thereof."[40]

If the patent does not disclose any structure to perform the claimed function, the claim is indefinite.[41] Moreover, if a person of ordinary skill in the art would not recognize disclosed structure and associate it with the corresponding function, the means-plus-function limitation renders the claim indefinite.[42] The intrinsic evidence must clearly link or associate that structure to the function recited in the claim.[43] The disclosed structure must be sufficiently defined to render the bounds of the claim understandable by the implementer, though not necessarily so detailed as to eliminate the need for any implementation choices.[44]

For many software inventions the claims recite functions that the software performs by running an appropriately programmed general-purpose computer or microprocessor. In such cases the structure is a combination of the computer and the algorithm used to program the computer,[45] unless the functions do not require any special programming because they are coextensive with the computer itself.[46] When programmed with the algorithm, the general-purpose computer is converted into a special purpose computer.[47] Therefore, the algorithm is part of the required structure and must be disclosed.[48] Disclosure of an algorithm is required even if the person of ordinary skill in the art would know what structures to use to accomplish the claimed function.[49] The patent drafter can choose to disclose the algorithm in any understandable manner, including as a mathematical formula, in prose, or as a flow chart.[50] If no algorithm to perform the claimed function is disclosed, then the patent lacks the required structure and is indefinite.[51] Similarly, if the only disclosed structure is a "black box" that performs the function, this is not a sufficient structure and the claim is indefinite.[52]

It is not always obvious whether the words used in a claim limitation will invoke 35 U.S.C. §112(f). The test is whether the person of ordinary skill in the art would understand the words of the claim to have a sufficiently definite meaning as the name for structure.[53] This remains true

---

[40] Williamson v. Citrix Online, L.L.C., 792 F.3d 1339, 1347 (Fed. Cir. 2015) (en banc), *superseding*, 770 F.3d 1371 (Fed. Cir. 2014). *See also* Halliburton Energy Servs. v. M-I L.L.C., 514 F.3d 1244, 1256 n.7 (Fed. Cir. 2008) ("This statutory provision was meant to preclude the overbreadth of open-ended functional claims which effectively cover any and all means for performing the recited functions.").

[41] *Williamson,* 792 F.3d at 1351–52; Noah Sys., Inc. v. Intuit Inc., 675 F.3d 1302, 1311–12 (Fed. Cir. 2012).

[42] *Williamson*, 792 F.3d at 1351–52.

[43] *Id.* at 1352.

[44] Ibormeith IP, LLC v. Mercedes-Benz USA, L.L.C., 732 F.3d 1376, 1379 (Fed. Cir. 2013).

[45] Aristocrat Techs. Austl. Pty Ltd. v. Int'l Game Tech., 521 F.3d 1328, 1333 (Fed. Cir. 2008).

[46] Katz Interactive Call Processing Patent Litig. v. Am. Airlines, Inc., 639 F.3d 1303, 1316 (Fed. Cir. 2011) ("[The] claimed functions of 'processing,' 'receiving,' and 'storing' . . . can be achieved by any general purpose computer without special programming. As such, it was not necessary to disclose more structure than the general purpose processor that performs those functions.").

[47] *Id.*

[48] *Id.*; Net MoneyIN, Inc. v. VeriSign, Inc., 545 F.3d 1359, 1367 (Fed. Cir. 2008).

[49] *Aristocrat Techs. Austl. Pty Ltd.*, 521 F.3d at 1337.

[50] Finisar Corp. v. DirecTV Grp., Inc., 523 F.3d 1323, 1340 (Fed. Cir. 2008).

[51] *Aristocrat Techs. Austl. Pty Ltd.*, 521 F.3d at 1338.

[52] Augme Techs., Inc. v. Yahoo! Inc., 755 F.3d 1326, 1338 (Fed. Cir. 2014).

[53] *Williamson,* 792 F.3d at 1349.

even if the term covers a very broad class of structures.[54] If the limitation includes the word "means", as in the phrase "*ink delivery means*", then there is a rebuttable presumption that §112(f) applies.[55] Even if the word "means" is absent, §112(f) is invoked if the claim term fails to recite sufficiently definite structure, or recites a function but not sufficient structure for performing that function.[56] The claim, specification, prosecution history, and relevant extrinsic evidence are used to determine whether a claim term recites sufficiently definite structure.[57]

The Federal Circuit has considered claim words such as "symbol generator",[58] "distributed learning control module",[59] "compliance module",[60] "sleep propensity algorithm",[61] as not reciting structure sufficient to prevent invocation of §112(f). Conversely, terms such as "modernizing device",[62] "heuristics",[63] "soft start circuit",[64] "system memory",[65] and "digital detector"[66] have been held to recite sufficient structure.

---

54 TecSec, Inc. v. IBM, 731 F.3d 1336, 1347 (Fed. Cir. 2013) (citing Lighting World, Inc. v. Birchwood Lighting, Inc., 382 F.3d 1354, 1359–60 (Fed. Cir. 2004)).

55 *Williamson,* 792 F.3d at 1349.

56 *Id.*

57 Apple, Inc. v. Motorola, Inc., 757 F.3d 1286, 1298 (Fed. Cir. 2014).

58 Advanced Ground Info. Sys. v. Life360, Inc., 830 F.3d 1341, 1348 (Fed. Cir. 2016) ("[T]he combination of the terms [symbol and generator] suggests that it is simply an abstraction that describes the function being performed (i.e., the generation of symbols).").

59 *Williamson, 792 F.3d at 1350* (explaining that generic terms such as module, mechanism, element, device, and module "typically do not connote sufficiently definite structure.").

60 Media Rights Techs., Inc. v. Capital One Fin. Corp., 800 F.3d 1366, 1373 (Fed. Cir. 2015) ("We have never found that the term 'mechanism'—without more—connotes an identifiable structure; certainly, merely adding the modifier 'compliance' to that term would not do so either.").

61 Ibormeith IP, LLC v. Mercedes-Benz USA, L.L.C., 732 F.3d 1376, 1382 (Fed. Cir. 2013) ("A description of an algorithm that places no limitations on how values are calculated, combined, or weighed is insufficient to make the bounds of the claim understandable.").

62 Inventio AG v. Thyssenkrupp Elevator Ams. Corp., 649 F.3d 1350, 1359 (Fed. Cir. 2011) ("[T]he claims recite a 'modernizing device,' delineate the components that the modernizing device is connected to, describe how the modernizing device interacts with those components, and describe the processing that the modernizing device performs. The written descriptions additionally show that the modernizing device conveys structure to skilled artisans.").

63 *Apple Inc., 757 F.3d at 1300* (finding the claim recites sufficiently definite structure to a person of ordinary skill in the art because "heuristic has a known meaning" and the patent describes its "operation, including its input, output, and how its output may be achieved").

64 Power Integrations, Inc. v. Fairchild Semiconductor Int'l, Inc., 711 F.3d 1348, 1365 (Fed. Cir. 2013) ("The word 'circuit' in combination with such a clear and unambiguous description of the circuit's operation weighs heavily in favor finding sufficient structure to avoid means-plus-function claiming.").

65 *TecSec, Inc.,* 731 F.3d at 1347 ("To those skilled in the art, a system memory is a specific structure that stores data.").

66 Personalized Media Commc'ns v. ITC, 161 F.3d 696, 705 (Fed. Cir. 1998) ("[T]hough the term 'detector' does not specifically evoke a particular structure, it does convey to one knowledgeable in the art a variety of structures known as 'detectors.'").

IV. Terms in Claims

The definiteness of a claim depends on whether the terms used in the claim have ascertainable meanings.[67] Accordingly, analysis of claim terms is useful to indefiniteness analysis.

Often a claim term has no standard meaning. For example, the patent drafter may have devised a new term rather than used a term known in the literature of the relevant technical field. If so, it is incumbent on the drafter to define the custom term,[68] or risk the claim being considered indefinite.[69]

Moreover, a nonstandard term without an accepted meaning might be considered to be a means-plus-function limitation even if the drafter did not intend it to be. The words in the term may describe a function, possibly triggering §112(f). For example, in *Advanced Ground Info. Sys. v. Life360, Inc.*,[70] the Federal Circuit held the claim term "symbol generator" to be indefinite.[71] That term recites an abstract function to be performed—the generation of symbols.[72] Moreover, the term is not used in common parlance or by persons of ordinary skill in the pertinent art, so it cannot designate any structure.[73] Therefore the term invokes §112(f) and its meaning must be limited to the corresponding structure disclosed in the specification. Since the patent in that case did not disclose such structure the claim was indefinite.

The broader lesson is that when a nonstandard, undefined claim term seems to recite a function to be performed, it invokes §112(f) and thereby makes it harder to satisfy the definiteness standard. Not only must the claim term have some meaning (e.g., describe a function) but there must also be corresponding structure disclosed in the specification.[74]

Another issue stems from terms that are subject to unspecified or vague limits. In patent claims this can occur with the use of the modifier "substantially". For example, the claim may include a term of degree such as that a distance must be "substantially equal to"[75] some amount, a balloon must be "substantially filled"[76], or a chemical does not "interfere substantially"[77] with

---

[67] Cox Commc'ns., Inc. v. Sprint Commc'n. Co. LP, 838 F.3d 1224, 1232 (Fed. Cir. 2016) ("[T]he common practice of training questions of indefiniteness on individual claim terms is a helpful tool. Indeed, if a person of ordinary skill in the art cannot discern the scope of a claim with reasonable certainty, it may be because one or several claim terms cannot be reliably construed.").

[68] Vitronics Corp. v. Conceptronic, 90 F.3d 1576, 1582 (Fed. Cir. 1996) ("[A] patentee may choose to be his own lexicographer and use terms in a manner other than their ordinary meaning, as long as the special definition of the term is clearly stated in the patent specification or file history.")

[69] Capital Sec. Sys. v. NCR Corp., 725 Fed. App'x 952, 959 (Fed. Cir. 2018) (affirming the district court's holding of indefiniteness because the claim term 'transactional operator' "has no commonly-accepted definition and its scope is unclear in view of the intrinsic evidence").

[70] 830 F.3d 1341 (Fed. Cir. 2016).

[71] *Id.* at 1349.

[72] *Id.* at 1347–48.

[73] *Id.*

[74] *See, e.g.*, Diebold Nixdorf, Inc. v. ITC, 899 F.3d 1291, 1302–03 (Fed. Cir. 2018) (finding no evidence that "cheque standby unit" was reasonably well understood by persons of ordinary skill to refer to a structure, and the patent does not disclose any structure for that function).

[75] Seattle Box Co. v. Indus. Crating & Packing, 731 F.2d 818, 821 (Fed. Cir. 1984).

[76] Tinnus Enters., L.L.C. v. Telebrands Corp., 846 F.3d 1190, 1206 (Fed. Cir. 2017).

[77] Enzo Biochem, Inc. v. Applera Corp., 599 F.3d 1325, 1329 (Fed. Cir. 2010). Though the term "not interfering substantially" does not provide a precise numerical measurement, the intrinsic evidence provided "a general guideline and examples sufficient to enable a person of ordinary skill in the art to determine" the scope of the claims. *Id.* at 1335.

some ability. To avoid indefiniteness there must be some standard for measuring that degree,[78] either in the intrinsic evidence or from the knowledge of a person of ordinary skill in the art.[79] If the claim provides enough certainty to one of skill in the art when read in the context of the invention then the claim is not indefinite.[80]

Other words besides modifiers introduce a potentially indefinite term of degree into the claim. For example, adjectives such as "fragile" can be ambiguous as to the requisite degree of the fragileness of the gel, rendering the term indefinite.[81] The claim term "at least partially soluble in water" was held to be improperly vague.[82]

However, definiteness does not require that the claim provides mathematical precision.[83] Terms of degree without numerical limits can nevertheless be considered definite, particularly if the relevant field of technology admits no more precise way of specifying the invention.[84] The key issue is whether the specification provides some standard for measuring that degree.[85] For example, a claim requiring a wax guard to be "readily installed and replaced by [an] user" was definite where the specification gave two examples that clearly provided a standard to measure the scope of that phrase.[86] The claim term "visually negligible" was held to be definite because the meaning involves "what can be seen by the normal human eye", which is an objective baseline.[87] Moreover, imprecise claim terms such as "approach each other," "close to," "substantially equal," and "closely approximate" are ubiquitous in patent claims.[88]

A claim term cannot have a meaning that depends completely on the user's subjective opinion because that would not notify the public of the boundaries of the claim.[89] For example, the term "aesthetically pleasing look and feel" is indefinite because it depends on the identity of who makes aesthetic choices; the term fails to provide any direction regarding how to determine whether that person succeeded in creating an aesthetically pleasing look and feel.[90] Likewise, the adjective "unobtrusive" in the phrase "unobtrusive manner that does not distract a user" is highly subjective as to what exactly rises to the level of a distraction for different users.[91]

More subtle indefiniteness issues can be caused by claim terms that require some sort of measurement, explicitly or implicitly. Such terms can inject ambiguity when the manner in which

---

[78] *Id.* at 1332; *Seattle Box Co., Inc.*, 731 F.2d at 826.

[79] Verve, L.L.C. v. Crane Cams, Inc., 311 F.3d 1116, 1119–20 (Fed. Cir. 2002).

[80] *Biosig Instruments, Inc v. Nautilus, Inc.*, 783 F.3d 1374, 1378 (Fed. Cir. 2015) (quoting Interval Licensing L.L.C. v. AOL Inc., 766 F.3d 1364, 1370 (Fed. Cir. 2014)).

[81] Halliburton Energy Servs. v. M-I L.L.C., 514 F.3d 1244, 1256 (Fed. Cir. 2008).

[82] Standard Oil Co. v. Am. Cyanamid Co., 774 F.2d 448, 453 (Fed. Cir. 1985).

[83] Invitrogen Corp. v. Biocrest Mfg., L.P., 424 F.3d 1374, 1384 (Fed. Cir. 2005); Sonix Tech. Co. v. Publ'ns Int'l, Ltd., 844 F.3d 1370, 1377 (Fed. Cir. 2017) (noting that "[b]ecause language is limited" terms of degree are not inherently indefinite).

[84] Rosemount, Inc. v. Beckman Instruments, Inc., 727 F.2d 1540, 1547 (Fed. Cir. 1984) (affirming the district court's holding that the term "close proximity" is as precise as the subject matter permits).

[85] Datamize, L.L.C. v. Plumtree Software, Inc., 417 F.3d 1342, 1351 (Fed. Cir. 2005).

[86] Hearing Components, Inc. v. Shure Inc., 600 F.3d 1357, 1368 (Fed. Cir. 2010).

[87] Sonix Tech. Co. v. Publ'ns Int'l, Ltd., 844 F.3d 1370, 1378 (Fed. Cir. 2017).

[88] Andrew Corp. v. Gabriel Elecs., Inc., 847 F.2d 819, 821, (Fed. Cir. 1988).

[89] *Datamize, L.L.C.*, 417 F.3d at 1342 (Fed. Cir. 2005).

[90] *Id.*

[91] Interval Licensing L.L.C. v. AOL, Inc., 766 F.3d 1364, 1374 (Fed. Cir. 2014) (holding that the intrinsic evidence failed to "inform those skilled in the art about the scope of the invention with reasonable certainty").

measurement is made can yield different values. For example, the Federal Circuit has held the term "average molecular weight" indefinite because that measurement could be ascertained by any of three possible measures, the claims do not indicate which measure to use, and the term does not have a plain meaning to one of skill in the art.[92] Similarly, the claim term "slope of strain hardening coefficient" was held indefinite because there were different methods for measuring that slope and each can produce different results.[93] In another case the claim required calculating the "melting point elevation", which in turn involved preparing a sample. Depending upon which of several sample preparation methods is used, the calculated melting point elevation can vary greatly.[94] Since the intrinsic evidence gave no guidance as to which sample preparation method to use the claims were held indefinite.[95]

The fact that there are different possible measurement methods does not necessarily render a claim indefinite. Even where the measurements produce different values, if the differences are not significant it is not indefinite.[96] Even if no measurement method is disclosed, the claim avoids indefiniteness if a person of skill in the art would know how to utilize an established measurement method to make the necessary measurement.[97]

A somewhat simpler indefiniteness issue involves the antecedent basis of claim terms. In other words, a reference is made to a previous item but it isn't clear which item is referenced. For example, if a claim recites "the lever" but contains no earlier recitation of a lever, it would be unclear what was referred to.[98] Similarly, if a claim recites "the lever" and contains two earlier recitations to two different levers, it is unclear which of the two "the lever" references.[99]

A claim term can have sufficient antecedent basis even if the earlier term is not identical to the later term. For example, the phrase "at least one mobile unit" provides an antecedent basis for the later use of the term "the mobile unit" in the claim.[100] Even a failure to provide explicit antecedent basis for terms does not necessarily render a claim indefinite.[101] When the meaning of the claim would be understood by a person of ordinary skill despite some problem with antecedent basis in claims, the claim is not indefinite.[102]

---

92 Teva Pharms. USA, Inc. v. Sandoz, Inc., 789 F.3d 1335, 1344 (Fed. Cir. 2015).

93 Dow Chem. Co. v. Nova Chems. Corp. (Can.), 803 F.3d 620, 633 (Fed. Cir. 2015) (finding that intrinsic evidence does not discuss the methods, provide any guidance as to which method should be used, or even whether the possible universe of methods is limited to those methods).

94 Honeywell Int'l, Inc. v. ITC, 341 F.3d 1332, 1336 (Fed. Cir. 2003).

95 *Id.* at 1340.

96 Takeda Pharm. Co. v. Zydus Pharms. USA, Inc., 743 F.3d 1359, 1366–67 (Fed. Cir. 2014).

97 Presidio Components, Inc. v. Am. Tech. Ceramics Corp., 875 F.3d 1369, 1376–77 (Fed. Cir. 2017).

98 MPEP § 2173.05(e) (9th Ed., Rev. January 2018).

99 *Id.*

100 TomTom, Inc. v. Adolph, 790 F.3d 1315, 1322–23 (Fed. Cir. 2015).

101 *See, e.g.*, Energizer Holdings, Inc. v. ITC, 435 F.3d 1366, 1370–71 (Fed. Cir. 2006); Bose Corp. v. JBL, Inc., 274 F.3d 1354, 1359 (Fed. Cir. 2001).

102 *See Energizer Holdings Inc.*, 435 F.3d at 1366, 1370–71.

Claims are interpreted in light of the specification,[103] and terms in the claim can be used in the specification.[104] A claim term might be defined in the specification or illustrated through example in the specification. However, a claim term might not appear at all in the specification. This lack does not necessarily render the claim term indefinite. For example, a claim may be broader than the specific embodiment disclosed in the specification[105] and thus use different words. Often the pair of terms will have a genus-species relationship. If this relationship is understood by the person of ordinary skill in the art, then disclosure in the specification of a narrow species can be sufficient support for a broad genus in the claim.[106]

V. AUTOMATING THE INDEFINITENESS ANALYSIS

Armed with an overview of the case law of indefiniteness, we can proceed to partially automating the required legal analysis. Partial automation entails using AI software and text analytics to assist the human decision maker.[107] This kind of collaborative activity, known as "cognitive computing," allows humans and computers to each perform the kinds of intelligent activities that they can do best.[108]

We must emphasize that partial automation can at best only assist the decision maker. We do not intend to suggest that software could supplant the human decision maker or perform the complete legal analysis. On the contrary, there is every reason to believe that fully automating the indefiniteness analysis is beyond both state-of-the-art AI and any AI improvements that are likely in the short term. The indefiniteness analysis is a high-level cognitive task. The analysis must first take into account the person of ordinary skill in the relevant art, which requires assessing the knowledge and capabilities that person has as well as how that person would understand certain terms and concepts.[109]

These tasks, which can be challenging even for skilled lawyers to accomplish, are notoriously difficult for state-of-the-art AI. For software to manipulate this type of knowledge, the software must first be able to represent the significant amount of domain knowledge and common-sense knowledge that the person of ordinary skill possesses. It would not be enough to employ advanced statistical analysis on very large amounts of text, a technique underlying the most impressive contemporary machine learning accomplishments.[110] Deep learning and other machine

---

[103] *See, e.g.*, Phillips v. AWH Corp., 415 F.3d 1303, 1315 (Fed. Cir. 2005) (noting that claims "must be read in view of the specification, of which they are a part.").

[104] For clarity we adopt the convention that "specification" refers to the portion of the patent other than the claims. Although the statute defines "specification" as including the claims, in modern usage some refer to the specification and claims as distinct. *See* In re Kullmann, 115 F.3d 942, 945 (Fed. Cir. 1997).

[105] Ralston Purina Co. v. Far-Mar-Co, 772 F.2d 1570, 1575, (Fed. Cir. 1985); In re Rasmussen, 650 F.2d 1212, 1215 (C.C.P.A. 1981) (citing an example in which a specification disclosure of a "lead weight" would support a claim term "metal weight").

[106] Hologic, Inc. v. Smith & Nephew, Inc., 884 F.3d 1357, 1362 (Fed. Cir. 2018) (claiming genus "light guide" is supported by a disclosure in the specification of the species "fiber optic bundle").

[107] *See generally* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 101 (2014) (describing the goal of partial automation as to complement, rather than replace, an attorney, such as for example by filtering likely irrelevant data in order to increase attorney efficiency).

[108] *See supra* note 5, at 11–14.

[109] *See, e.g.*, Kinetic Concepts, Inc. v. Blue Sky Med. Grp., Inc., 554 F.3d 1010, 1022 (Fed. Cir. 2009) (accepting evidence regarding how a doctor of ordinary skill would understand the claim).

[110] *See generally* Levesque, H., Davis, E. and Morgenstern, L. (2012), *The Winograd Schema Challenge*, PROC. OF THE THIRTEENTH INT'L CONF. ON PRINCIPLES OF KNOWLEDGE REPRESENTATION AND REASONING, AAAI Press, pp.

learning techniques which are based on such advanced statistical analyses have difficulty making inferences beyond what is explicitly provided in a text[111] and integrating prior knowledge.[112] Also, a purely statistical model is unlikely to be sufficient in automated legal reasoning. Even if a supervised learning model could learn this information from a statistical analysis of large amounts of text, such a model assumes a non-varying relationship between inputs and outputs. In other words, the same set of inputs should yield identical outputs. Unfortunately, because the law evolves the same fact patterns (inputs) can and often do lead to different results (outputs) over time. For example, the Supreme Court's *Nautilus* standard for indefiniteness overturned the Federal Circuit's insolubly ambiguous standard.[113] A similar difficulty stems from the wide latitude that the patent drafter has in choosing terminology and linguistic structures. This means that any "standard" manner of drafting that might be gleaned from previous patents cannot be taken as a definitive rule for the future. Failure to conform to standard patent drafting practice does not in itself render a claim indefinite.[114]

As described below, we propose AI software that performs three tasks:

- identify information of possible relevance to the legal analysis;
- draw simple inferences from that information when possible; and
- provide the information and inferences to the decision maker in an efficient manner.

The software thus can provide the decision maker with information much faster than if the user read the entirety of the patent to locate the same information unaided. This in turn can allow the decision maker to more quickly reach a tentative conclusion about the legal analysis. Where the decision maker is involved in analyzing many patents, the software can also prioritize some patents over others, bringing to the decision maker's attention those that have the most (or least) signs of potential indefiniteness. This focuses the decision maker on the patents that require more attention and legal skills.[115]

The design of this software is intimately related to the legal analysis it will support; the software must be aware of what information is relevant to the analysis and what is not. Since much of the indefiniteness analysis consists of interpreting the technical disclosure and claims of the patent, relevant information is encoded in the patent's text. That information must be extracted from the text regardless of the words and sentence structure the patent drafter chose to represent that information in English prose. Therefore, the software uses techniques from the field of Natural

552-561. The authors describe problems that cannot rely on statistical analysis of a large corpus of English text as "Google-proof." *Id.* at 554. An example of a question that software can answer using a pure statistical approach is "The women stopped taking the pills because they were pregnant. Which individuals were pregnant: the women or the pills?" By sampling a large enough amount of text, software can learn that the word "pregnant" occurs much more often close to "women" than close to "pills", so the software would answer "pregnant." *See also* Walid S. Saba, On the Winograd Schema Challenge: Levels of Language Understanding and the Phenomenon of the Missing Text (2018) (unpublished manuscript).

[111] *See, e.g.*, Gary Marcus, Deep Learning: A Critical Appraisal, 10 (2018) (unpublished manuscript).

[112] *Id.* at 6 ("Deep learning currently lacks a mechanism for learning abstractions through explicit, verbal definition."); *id.* at 11 ("[T]he knowledge represented in deep learning systems pertains mainly to (largely opaque) correlations between features, rather than to abstractions like quantified statements.").

[113] Nautilus, Inc. v. Biosig Instruments, Inc., 572 U.S. 898, 901 (2014).

[114] Marley Mouldings Ltd. v. Mikron Indus. Inc., 417 F.3d 1356, 1360 (Fed. Cir. 2005).

[115] *See supra* note 107, at 101 (arguing that machine learning tools can automate easier cases to conserve the attorney's cognitive efforts and time for tasks likely to require higher order legal skills).

Language Processing (NLP), and we perform *NLP analytics by* using the information extracted from this text.

Our proposed software for partially automating the indefiniteness analysis is limited to information that is contained in the text of patents. We deliberately eschew any inquiry into extrinsic evidence, such as questions involving the person of ordinary skill, their knowledge, or their capabilities. We focus on providing the decision maker with relevant information, and with straightforward inferences that can be drawn from that information.

We also arrange our method in a broader hierarchy of partially automated methods. Each level in the hierarchy indicates the sources of information utilized by the software. At the lowest level are automation techniques that employ only information contained in the text of a single patent. Higher levels include the information specified in the corresponding row of Table 1 as well as all sources in lower levels. We suspect that one or more of these levels could be partitioned into finer levels of specificity.

Level	Source of Information
3	Knowledge base or other knowledge representation
2	The text of all patents and other technical literature
1	The text of all patents
0	The text of the patent being analyzed

Table 1. Hierarchy of Information Sources

To identify and extract the desired information from the text of patents, we employ NLP techniques. NLP is a subfield of Artificial Intelligence that overlaps significantly with Machine Learning[116]. NLP deals with computer processing and manipulating of "natural" languages, such as English or Spanish.[117] Among the different subfields of Artificial Intelligence, a defining characteristic of NLP is the use of some knowledge of the natural language being processed.[118] Although NLP can involve spoken language (speech) or written language (text), here we deal with the latter since we process the text of patents.

Each type of information that we extract from patent text is a *data-driven metric*, i.e., a metric that software computes entirely from data without human input or adjustment. Moreover, our desire is to use a data-driven metric that is highly correlated with some *target metric* of interest, such as the legal conclusion of claim indefiniteness. Of course, even if there is high correlation between a data-driven metric and the target metric, the two can lead to different conclusions in some circumstances. For example, a data-driven metric can indicate a high likelihood that the claim is definite when in fact it is indefinite. Nevertheless, such metrics provide useful information on average. Also, such metrics lead to useful conclusions about the characteristics of sets of patents too large for human analysis, as discussed below

---

[116] *See supra* note 107.

[117] DAN JURAFSKY & JAMES H. MARTIN, SPEECH AND LANGUAGE PROCESSING 30–31 (2d ed. 2008).

[118] *Id.*

VI. BENEFITS OF AUTOMATING THE INDEFINITENESS ANALYSIS

From the review of the legal standard for indefiniteness reveals several types of information that can be relevant depending on the facts in the case at hand. Table 3 below lists several types of information that can be automatically extracted from patent text, and briefly describes the potential relevance of this information to the indefiniteness inquiry. Each type of information and its relevance is explained in detail in subsequent sections.

Since claim definiteness is one of many requirements for patent validity, a definiteness analysis is performed when the validity of a patent must be assessed. The U.S. Patent & Trademark Office ("PTO") is charged with ensuring the definiteness of patent claims in each filed patent application. Claim definiteness is also assessed by the owner of the patent and third parties, such as when contemplating litigating or licensing the patent. Analyzing the definiteness of a single claim takes substantial effort. In addition, the results of an indefiniteness analysis can form the basis of consequential decisions involving significant amounts of money and time. Since the definiteness analysis involves legal reasoning and applying relevant legal doctrines, it is not always clear what conclusion should be drawn. Therefore, there is always the chance that an incorrect conclusion is made, i.e., a claim believed to be indefinite is in fact definite, or vice versa. Reaching an incorrect conclusion can cause wasted expenditures as well as lost opportunities.

As explained below, AI software that assists in the definiteness analysis would provide several advantages. The most prominent benefits would be to make the types of definiteness analysis that are performed today quicker and more accurate, leading to significant cost savings. In addition, AI processing would allow new types of indefiniteness analysis to be performed on arbitrarily large sets of patents. Because software would be able to process many more patents than any person could, it could analyze a patent portfolio to provide insights currently unavailable. This new information would be potentially valuable to companies, the PTO, and policy makers.

Before quantifying the benefits of AI-assisted definiteness analysis we first specify in greater detail the most common instances of that analysis. We believe that the vast majority of definiteness analyses are conducted by one of three types of entities. The first is when the PTO reviews the claims for compliance with 35 U.S.C. § 112(b). Like other patent claim validity determinations, the PTO assesses the definiteness of every claim in every patent application during patent prosecution.[119] The second type of entity conducting definiteness analysis is the patent

---

[119] In addition, the PTO may also evaluate the definiteness of claims in a granted patent pursuant to a request by the patent owner for Supplemental Examination under 35 U.S.C. § 257. Supplemental Examination may involve any ground of patentability, including claim definiteness. Supplemental Examination became available on September 16, 2012, pursuant to the Leahy-Smith America Invents Act ("AIA"). We disregard the burden imposed by definiteness analysis under Supplemental Examination because this process is so rarely used. *See, e.g.*, Adriana L. Burgy, Amanda K. Murphy, Sneha Nyshadham & Stacy Lewis, *AIA Supplemental Examination Nuts and Bolts: Get It in Your Toolbox and Don't Leave Home Without it!*, FINNEGAN: AIA BLOG, https://www.finnegan.com/en/insights/blogs/america-invents-act/aia-supplemental-examination-nuts-and-bolts-get-it-in-your-toolbox-and-dont-leave-home-without-it.html (June 3, 2019) (showing that only 246 supplemental examinations have been filed as of May 15, 2019).

owner or applicant. The patent applicant analyzes the definiteness of its claims during the drafting and refining of the patent application before it is filed with the PTO. After the patent is granted, the patent owner would likely assess the definiteness of claims that might be used for licensing or litigation. Proactively evaluating definiteness reduces the chance of improvidently entering licensing negotiations or court proceedings with claims that are susceptible to being invalidated for indefiniteness.

A similar set of actions are undertaken by third parties, the final entity conducting definiteness analysis that we will consider. Like patent owners, third parties evaluate the claims of granted patents in order to assess the suitability of licensing or litigating those patents. This might be done in response to receiving notification from the patent owner that the third party potentially infringes the patent. Assessing the potential susceptibility of the patent to indefiniteness and other invalidity attacks is part of the calculus in deciding whether to license the patent or settle litigation. Alternatively, the third party might become aware of the patent through other channels and decide to preemptively evaluate whether the patent might be used against the third party. The third party could decide that the most effective strategy would be to initiate post grant review in the PTO to invalidate the patent.

a. Quantification of Benefits

Having identified the three types of parties that commonly perform definiteness analysis, we next seek to quantify the costs for each. Two types of costs are most prominent: the costs in conducting the analysis and the costs that ensue when the analysis yields incorrect results. Table 2 provides a summary of these costs. We will try to demonstrate savings in both costs; AI processing can potentially reduce the time to conduct a definiteness analysis and also reduce the errors made in a definiteness analysis. We will also limit our exposition to the U.S. patent system, though similar conclusions could be drawn for the patent systems of other jurisdictions.[120]

While many of these costs are inherently speculative because required data are not available, we do our best to develop very conservative lower bounds on the costs, fully aware that the true costs are very likely greater. The reader may believe that we have seriously understated certain components of these costs or have set forth unrealistically meager amounts of time spent by various parties on certain activities. In many cases we would vigorously agree with these assessments. Our objective in developing lower bounds is not to estimate a likely average cost or a range within which the actual costs are likely to lie. Instead our objective is to obtain broad agreement on the *minimum* burden of the definiteness analysis. This in turn shows that AI-assisted definiteness analysis is potentially valuable. It also illuminates the minimum benefits likely to be obtained from partially automating the definiteness analysis. We contend that even these minimum benefits warrant investment in developing AI tools to reduce these costs, and the actual cost savings would likely be even greater. We would welcome any suggested enhancements to our estimation methodology, such as additional costs we may have omitted or improvements to particular cost lower bounds we have provided.

---

[120] For example, Canada has a similar requirement that patent claims cannot be "avoidably ambiguous." Pfizer Canada Inc. v. Canada (Minister of Health), (2005) F.C.R. No. 1725 (Fed. Ct. Can. 2005), and Article 84 of the European Patent Convention requires that claims "shall be clear and concise." Convention on the Grant of European Patents art. 84, Oct. 5, 1973, 1065 U.N.T.S. 255.

Entity	When analysis is performed	Annual costs lower bound	Incorrect results of analysis	Annual costs lower bound
PTO	When preparing a first office action	300,000 hours $7,500,000		
PTO			Incorrectly consider a definite claim to be indefinite	7,500 hours $750,000
PTO			Incorrectly consider an indefinite claim to be definite	120 hours $12,000
PTO			Public burden from improperly rejecting or failing to reject claims	(*not included*)
PTO				
Patent applicant / owner	When preparing patent application	250,000 hours $25,000,000		
Patent applicant / owner	Contemplating or engaging in litigation or licensing	(*not included*)		
Patent applicant / owner			Mistakenly drafts indefinite claims	7,500 hours $750,000
Patent applicant / owner			In litigation or licensing, asserts invalid claims or fails to assert valid claims	(*not included*)
Third Parties	Contemplating or engaging in litigation or licensing	(*not included*)		
			In litigation or licensing, license invalid claims or fail to license valid claims	(*not included*)
TOTAL		**550,000 hours $32,500,000**		**15,120 hours $1,512,000**

Table 2. Lower bound estimates of annual costs from indefiniteness analysis.

More specifically, we will attempt to quantify a lower bound on the costs in a single year for the errors in performing indefiniteness analysis. We begin with actions by the first entity – the PTO. In 2018 the PTO created about 1,200,000 office actions for examined applications, of which approximately 600,000 were first actions.[121] We will conservatively consider only first actions under the assumption that the majority of definiteness analysis is performed the first time the PTO examiner reads the application and examines its claims. This undercounts what is almost certainly a nontrivial amount of the work performed for claims amended after the first action and for previously examined claims that are revisited by the Examiner.

Performing the definiteness analysis requires that the Examiner read all claims as well as the relevant portions of the application. Moreover, although the Examiner is required to perform this analysis for every office action, some portion of office actions will not reflect any evidence of this analysis because no claims were actually rejected for indefiniteness. We estimate that this analysis consumes an average of at least thirty minutes of effort for each of the 600,000 first

[121] U.S. PATENT AND TRADEMARK OFFICE, FY 2018 PERFORMANCE AND ACCOUNTABILITY REPORT 178. The first amount includes first actions, actions in connection with allowed applications, and actions in connection with abandoned applications, all for other than design applications. The second amount also omits design applications from consideration.

actions, or 300,000 hours per year. If there are an average of fifteen claims per patent application,[122] this means our estimate assumes an average of just two minutes is devoted to assessing the indefiniteness of each claim being examined. Assuming that each hour of Examiner effort corresponds to more than $25 in salary costs,[123] the number of PTO examination hours devoted to analyzing claim definiteness corresponds to more than $7.5 million per year. This cost is imposed on patent applicants, who fund PTO operations through filing fees.[124]

Next, we attempt to quantify the costs imposed when the PTO incorrectly performs the definiteness analysis. Different costs are imposed when the PTO rejects a definite claim for indefiniteness, and when the PTO incorrectly considers an indefinite claim to be definite. We first consider improper rejections for indefiniteness. Rejections under 35 U.S.C. §112, which includes indefiniteness as well as other bases for rejection,[125] were made in over 37% of all first office actions mailed between 2008 and mid-2017.[126] We will estimate that at least 20% of first actions include a rejection for indefiniteness,[127] a total of 120,000 first actions. Of these we conservatively estimate that 10% are incorrectly made, so 12,000 first actions reject definite claims as indefinite.[128]

Exactly what sorts of costs are imposed by these 12,000 first actions? Faced with an improper rejection, the patent applicant may amend the claim and relinquish patent scope it was entitled to receive. Alternatively, the applicant may choose to argue against the rejection, whether only once in the response to the first Office Action or several times via an appeal to the Patent Trial and Appeal Board ("PTAB").[129] Some portion of these applicant arguments are successful in

---

[122] The average number is likely higher than fifteen for any year in the short term. *See, e.g.*, Dennis Crouch, *Standard Patent Size*, PATENTLY-O BLOG, https://patentlyo.com/patent/2017/10/standard-patent-size.html (October 22, 2017) (showing that the average number of claims in granted patents has been greater than sixteen for over a decade).

[123] Most patent examiners Examiners start with the PTO as a GS-7 or GS-9, which an accompanying annual starting salary of between $54,857 and $83,242. *See* Job Announcement for Patent Examiners, U.S. PATENT AND TRADEMARK OFFICE, https://www.uspto.gov/sites/default/files/documents/Examiner%20brochure%202018%20downloadable.pdf.

[124] Another component of this cost stems from the allocation of PTO Examiner time to this task. Since Examiner time is a fixed and limited resource, at least in the short term, time spent on the definiteness analysis cannot be spent on other examination tasks. This indirectly leads to fewer applications per year examined by the PTO.

[125] 35 U.S.C. § 112 (a)–(f) (2012) provides other bases for rejection, such as lack of enablement.

[126]*See* Qiang Lu, Amanda Myers & Scott Beliveau, *USPTO Patent Prosecution Research Data: Unlocking Office Action Traits* (U.S. Patent and Trademark Office, Economic Working Paper No. 2017-10, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3024621. Table 1 shows that 37.8% of first office actions included some form of rejection under Section 112.

[127] *Id.* Table 2 shows that 70.1% of all actions rejected under Section 112 included a rejection for indefiniteness. However, this table does not indicate what portion occurred in first office actions. If this 70.1% applied uniformly to both first and final office actions, then 37.8% * 70.1% = 26.5% of all first actions would have an indefiniteness rejection. We conservatively round this down to 20% of all first actions.

[128] In ex parte appeals, the PTAB reversed approximately half of all indefiniteness rejections over a six-month period. Adam Stephenson, *Cognitive Dissonance: How the PTAB Reported Appeal Statistics Ruins the Data for Everyone*, IP WATCHDOG BLOG, https://www.ipwatchdog.com/2017/11/14/cognitive-dissonance-ptab-publicly-appeal-statistics-ruins-data/id=89922 (November 14, 2017). However, ex parte appeals of definiteness rejections to the PTAB are not representative of all applications rejected for indefiniteness. Applicants who believe the rejection is proper are less likely to appeal than those who believe the rejection is improper.

[129] For convenience we ignore in our estimates the infrequent but significant costs imposed when an improper indefiniteness rejection is maintained by the PTAB and then overturned by a subsequent appeal to the Court of Appeals for the Federal Circuit.

persuading the Examiner that the claim was not actually indefinite,[130] so the applicant need not appeal to the PTAB. Some portion of improper rejections result in an appeal to the PTAB, a much more time-consuming process.

We estimate that 90% of these improper first action rejections cause applicants to take action: to amend the claim and/or to argue against the rejection. We also estimate that each such action consumes on average at least a half hour of attorney time. This account for 5,400 hours spent addressing improper indefiniteness rejections in a first action. We also assume that there are 7,000 ex parte appeals to the PTAB in a year,[131] of which at least 25% include a rejection for indefiniteness.[132] We also assume that 40% of the appealed indefiniteness rejections are improper,[133] a total of 700 appeals for improper indefiniteness rejections. We also estimate that each such appeal consumes on average at least three hours of attorney time in preparing the portion of the appeal related to the improper rejection. This includes analyzing the patent application and prosecution history and drafting portions of the appeal brief. This accounts for 2,100 hours spent addressing improper indefiniteness rejections in appeals. Together with the 5,400 hours spent addressing improper indefiniteness rejections in a first action, a total of 7,500 attorney hours addressing improper indefiniteness rejections. We finally ascribe a cost of $100 per hour to this attorney time, which is a total of $750,000 applicants paid to respond to improper indefiniteness rejections.

The PTO may also err by considering some indefinite claims to be definite, and therefore does not reject claims that should be rejected.[134] An improperly granted claim can be used by the patent owner to extract licensing fees or exclude third parties from making, using, and selling the patented invention. The improperly granted claim can also be the subject of litigation or post-grant proceedings before the PTO. We ignore in our estimate the costs when an improper indefinites rejection prompts a party to commence Post-Grant Review before the PTAB.[135] Focusing instead on district court litigation, we conservatively estimate that at least three litigations result in a claim being held indefinite.[136] Assume that for each such litigation at least forty hours of attorney time are spent by both parties on the indefinite claims, or 120 hours.[137] Ascribing a cost of $100 per hour of attorney time, we arrive at a $12,000 cost of errors from failing to identify indefinite claims.

---

130 It is also possible that in some portion of successful arguments the Examiner withdraws the rejection before appeal even though the rejection was proper and should have been maintained despite the applicant's arguments to the contrary. Although we believe this number of applications is small, we nevertheless include it in our later estimates of the costs imposed by the PTO when improperly considers an indefinite claim to be definite.

131 In fiscal years 2016–19, the PTAB disposed of 14,740, 14,275, 11,307, and 7820 ex parte appeals respectively. Statistics for Ex Parte Appeals, PATENT TRIAL AND APPEAL BOARD, https://www.uspto.gov/patents-application-process/patent-trial-and-appeal-board/ptab-statistics-receipts-and-dispositions.

132 *See supra* note 126.

133 In recent appeals closer to half of all indefiniteness rejections were reversed by the PTAB. *See supra* note 127.

134 Some number of the claims that should be rejected for definiteness will be (properly) rejected on other grounds, such as anticipation or obviousness.

135 About 2% of all PTAB trials are Post-Grant Review. As of July 31, 2019 there have been only 174 petitions for Post-Grant Review. *See* U.S. Patent and Trademark Off., Trial Statistics, IPR, PGR, CBM 3 (2019), https://www.uspto.gov/sites/default/files/documents/Trial_Statistics_2019-07-31.pdf

136 In each of the previous several years at least five cases before either a district court or the Court of Appeals for the Federal Circuit have resulted in claims being held indefinite. *See John R. Allison & Lisa Larrimore Ouellette, How Courts Adjudicate Patent Definiteness and Disclosure*, 65 DUKE L.J. 609 (2015), and accompanying data files available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/SBRQHJ.

137 We omit from consideration the value of the court's time, though this is a nontrivial cost.

Together with the $750,000 attributable to improperly considering definite claims to be indefinite, the total cost of errors is $762,000. Adding the $7.5 million spent by the PTO on performing indefinite analysis, our estimated costs attributable to indefiniteness analysis by the PTO is $8,262,000.

We mention for the sake of completeness other costs attributable to indefiniteness analysis by the PTO. One is when an improper rejection causes the applicant to forgo patent scope to which the applicant was entitled. An analogous cost is when a failure to reject a claim burdens the public with invalid patent scope and precludes the public from activities to which it should be entitled. Similarly, the public to some degree loses confidence in the patent system when improper rejections or allowances occur. These costs are difficult to quantify because they depend so strongly on the particular circumstances of the patent owners and the characteristic of their patents. We therefore omit them from consideration though it could be argued that these costs are substantial.

We now examine the costs attributable to the indefiniteness analysis of the second entity, the patent owner. As before, we bifurcate these costs into costs of conducting the analysis and costs when the analysis yields incorrect results. The patent owner invests time in an indefiniteness analysis when preparing a patent application for filing. This may include ensuring that claim terms are defined and defined appropriately, that examples in the patent employing the claim term are consistent with any definitions or some meaning known in the art, and that wording and grammatical constructions used in the claims do not lead to undue ambiguity. Assume at least 500,000 patent applications are filed annually.[138] Also assume that on average just 30 minutes is spent performing an indefiniteness analysis on the application being prepared, about two minutes per claim if there are an average of fifteen claims per application. Patent applicants therefore devote at least 250,000 hours annually to analyzing the definiteness of claims during the preparation of the patent application. Ascribing a cost of $100 per hour to this time, patent applicants spend at least $25 million on this type of indefiniteness analysis. Another cost of conducting definiteness analysis occurs when the patent applicant or patent owner contemplates litigating or licensing the patent. The analysis would be performed to determine how susceptible the patent is to an invalidity challenge, which in turn affects both the willingness of a third party to license the patent as well as the amount of money a third party would be willing to pay for such a license. For simplicity, we omit this type of cost from our lower bound estimates, though it is almost certainly nontrivial.

Turning now to the costs of errors in the definiteness analysis, when the patent applicant mistakenly drafts indefinite claims in their patent application the PTO rejects those claims. In response the patent applicant provides arguments and amendments to the rejected claims. Assume that at least 2.5% of all 600,000 first actions[139] represent PTO rejections of applicants' claims that are truly indefinite. Further assume that these 15,000 first actions require 30 minutes of attorney time to respond with arguments and / or claim amendments. Again, ascribing a cost of $100 per hour to these 7,500 hours yields $750,000 spent on addressing filed claims that were indefinite due to applicants' error. Another cost of erroneous indefiniteness analysis is imposed when a patent owner mistakenly believes its issued claims are indefinite. The patent owner with this misconception might fail to initiate licensing or litigation believing that the claim to be invalid. We omit this cost for simplicity.

---

[138] In each of fiscal years 2014–2018, the number of utility patent applications filed with the PTO were over 575,000. *See* *supra* note 126.
[139] *Id.*

We are left with the costs of indefiniteness analysis by third parties. As with the actions of the patent owner, a third party conducts an indefiniteness analysis when contemplating litigating or licensing the patent. The analysis would be performed to determine how susceptible the patent is to an invalidity challenge, since a patent that is likely invalid is a much less attractive to potential licensees. Similar costs are imposed when the third party mistakenly believes definite claims should be challenged, or indefinite claims should not be. For simplicity we omit these costs from our lower bound estimates, though they are almost certainly nontrivial.

In summary, our very conservative estimates are that definiteness analysis consumes significant societal resources. The costs incurred in analyzing patent claim definiteness are more than $34 million per year, representing more than a half million hours each year. As noted above, the true costs are almost certainly greater, perhaps substantially greater, than our lower bounds because we have both deliberately understated some costs and deliberately omitted other costs.

In light of these significant amounts, AI software that partially automates the definiteness analysis could potentially save significant amounts of time and money by reducing the time it takes to perform a definiteness analysis or by increasing the accuracy of a definiteness analysis. An approximate but realistic lower bound on the potential savings can be developed by considering only two significant components of the costs: time spent on definiteness analysis by the PTO and by patent owners.[140] Assume again that every year the PTO creates at least 600,000 first actions and patent applicants file at least 500,000 patent applications. If AI software saves six minutes of time on each first action and on each patent application, that represents 110,000 fewer hours on definiteness analysis (60,000 by the PTO and 50,000 by patent owners). If we attribute each hour of PTO time as saving at least $25 in salary, and each hour of patent application preparation as saving at least $100 in attorney fees, this represents at least $6.5 million per year in cost savings. This simplistic lower bound ignores time savings on other types of activities related to definiteness analysis, and also ignores time savings in increased accuracy. AI software that delivered even mild time savings would save society at least several hundred thousand hours over just a few years.

This conservative lower bound demonstrates that the development of such AI software is a worthwhile goal. Software could be developed by private actors such as the current providers of legal software. Moreover, development could even be partially financed by government incentives. Although most types of legal software are developed by the private sector, public funding of patent-related AI tools would have several benefits. Given the potential savings to the PTO alone, the government could justify a significant amount of financing in exchange for saving the PTO tens of thousands of hours per year spent on definiteness analysis. In addition, if the PTO or another government entity were to partially finance the development of AI analysis tools for the patent domain, that entity could ensure that the tools have certain desirable features for both the PTO and society more broadly. The financing entity could also require that the software adheres to certain standards and then permits many competing software developers to create software in compliance with that standard. This would allow multiple versions of the software to be developed by different companies, promoting experimentation to identify the best software design, while providing a software standard to facilitate reuse of components and minimize costs if users switch among different vendors' software. Similarly, government incentives could be invested in developing core AI algorithms for free distribution, while different software companies embed these algorithms in their own products.

---

[140] Because our previous estimates are conservative lower bounds with no attempt to reflect the average or likely costs, estimating savings as some fraction of this lower bound would not be a sound methodology.

It is natural to ask why private actors have not recognized the opportunity to develop this sort of AI software. After all, given the potential costs savings to users, it is plausible that users would be willing to pay a nontrivial price that was some significant portion of their cost savings. Perhaps the idea of using AI to analyze indefiniteness has not previously occurred to software developers.[141] In addition, we speculate that perhaps many software developers consider the market for patent-specific tools to be limited to just a few thousand patent attorneys, and therefore too small to risk the development costs. That the PTO might also be interested in such a tool could be perceived as too unlikely because the PTO has not (to the best of our knowledge) signaled any willingness to automate its definiteness analysis with AI software. Nevertheless, we believe our explanation above demonstrates that there are significant benefits to AI-assisted definiteness analysis. The time savings alone would cause many users to pay several thousand dollars per year if the software was effective at making the definiteness analysis faster and more accurate.

VII. NLP Analytics: Types of Useful Information

From the review of the legal standard for indefiniteness above we have several types of information types, each of which can be relevant depending on the facts in the case at hand. Table 3 lists several types of information which can be automatically extracted from patent text, and briefly describes the potential relevance of this information to the indefiniteness inquiry. Each type of information and its relevance is explained in detail below.

Type of Information	Relevance
Claim term is not used in the specification	More likely indefinite
Claim term is defined in the specification	Less likely indefinite
Claim term has examples in the specification	Less likely indefinite
Claim term is coined	More likely indefinite
Claim term recites a function	More likely indefinite (see next two items)
Claimed function not in the specification	More likely indefinite
Structure in specification linked to a function	Less likely indefinite
Claim contains potentially vague words	More likely indefinite
Claim term contains many words	More likely to be coined or to recite a function

Table 3. Information useful to indefiniteness

<u>Claim term is not used in the specification</u>

The mere presence or absence in the specification of a claim term is information that is potentially useful to the indefiniteness analysis. If a claim term is not present in the specification then this suggests that the claim is less likely to be definite because the specification might not provide the person of ordinary skill with enough information to understand the meaning of the

[141] We could find no references to such software.

term.[142] This situation could happen, for example, during prosecution if the claim is amended to include a term that was not in the application when it was filed.

Of course, a claim term need not be literally present in the specification for the clam to be definite. The claim term might be well understood by one of ordinary skill in the art, and thus would not need any explanation or clarification. Alternatively, the claim term might be a species of a genus that is explained in the specification.[143] This latter case shows the benefits of searching the specification for terms that describe a broader class to which the claim term belongs, as in a genus species relationship. More generally, the software could search the specification for terms that have some relation to the claim term as specified by an ontology of entities.[144]

Claim term is defined in the specification

The mere presence or absence of a definition for a term is information potentially useful to the analysis. A term that is defined is more likely to have a meaning that would be understood by a person of ordinary skill in the art.[145] For example, if a claim has four terms, definiteness is strengthened if those terms are defined than if they are all undefined, all other things being equal. The fact that a claim term has been defined in the specification might suggest that more attention was given to that term than if the term were not defined.

Note that this requires only that the software be able to identify that a term has been defined in the specification. It does *not* require that the software in any way understands any part of the definition.

Claim term has examples in the specification

The use of examples for a term is potentially useful to the analysis. Like a definition, examples help to provide meaning to a person of ordinary skill in the art.[146] For example, if a sentence in the specification states “The encryption method can comprise the Advanced Encryption Standard (AES), Twofish, RSA, and / or the Triple Data Encryption Standard (3DES)”, then the meaning of “encryption method” to a person of ordinary skill in the art is more likely to be definite than if the term is not associated with any examples in the specification.

---

142 Bancorp Serv. L.L.C. v. Hartford Life Ins. Co., 359 F.3d 1367, 1373 (Fed. Cir. 2004) (noting that a claim term can be sufficiently definite, even if it is not defined, “if the meaning of the term is fairly inferable from the patent”).

143 *Cf.* Fujikawa v. Wattanasin, 93 F.3d 1559, 1570–71 (Fed. Cir. 1996) (explaining that in the context of the written description inquiry, the disclosure of a genus can satisfy the written description requirement as to particular claimed species provided there is adequate description that would reasonably lead persons skilled in the art to the claimed species).

144 *See generally* Feng Wang, Lan Fen Lin & Zhou Yang, *An Ontology-Based Automatic Semantic Annotation Approach for Patent Document Retrieval in Product Innovation Design*, 446 Applied Mechanics and Materials 1581 (2013).

145 *Bancorp Serv. L.L.C.*, 359 F.3d at 1373 (noting that although claim terms need not necessarily be defined in the patent, a definition avoids “a time-consuming and difficult inquiry into indefiniteness”).

146 *Id.*

Claim term is coined

The patent drafter can include terms of her own devising.[147] Such terms need not have ever appeared in any previous publication or patent. If the term has never appeared in print before then it is possible that the person of ordinary skill would not ascribe a definite meaning to the term.

It is also possible that a coined term includes words that suggest a function or other meaning. For example, even if one has never seen the term "heart-shaped donut hole puncher" it might nevertheless be reasonable to assume it refers to something that punches heart-shaped holes into donuts (though other meanings are certainly feasible as well). If a function is implicated by the term (e.g., the function of punching holes), other considerations are invoked, as describe immediately below.

Claim term recites a function

As in the above example, a claim term can include a function. If the claim recites only function without structure for performing that function, then §112(f) is invoked,[148] and this can make it more difficult to satisfy the definiteness standard because the specification must disclose both structure to perform the function[149] as well as a sufficient link between the structure and the function.[150] Therefore terms which recite functions or employ functional language increase the chance that the claim is indefinite, especially if other requirements under §112(f) are not satisfied. These other requirements are implicated in the next two items below.

Claimed function not in the specification

When a claim recites a means-plus-function limitation, the specification must disclose structure for performing that function. In addition, a person of ordinary skill in the art must recognize the disclosed structure and associate it with the corresponding function in the claim. If the function is never described in the specification, it is more likely that structure for performing the function is absent as well.

Structure in specification linked to a function

As stated above, a person of ordinary skill in the art must recognize disclosed structure and associate it with the corresponding function in the claim. If the function is disclosed in the specification but not firmly linked by text in the specification to any structure, this increases the likelihood that the person of ordinary skill would not associate the structure with the function.

Claim contains potentially vague words

Several types of terms having unspecified limits, including terms of degree and inherently vague adjectives, have been described above. The inclusion of such words increases the likelihood that the claim does not have the requisite amount of certainty to satisfy the definiteness requirement.

---

[147] *See supra* note 11, at 910 (quoting Hormone Research Found., Inc. v. Genentech, Inc., 904 F.2d 1558, 1563 (Fed. Cir. 1990)); Vitronics Corp. v. Conceptronic, 90 F.3d 1576, 1582 (Fed. Cir. 1996)

[148] *See supra* note 37.

[149] *See supra* note 40 and accompanying text.

[150] *See supra* note 41 and accompanying text.

Claim term contains many words

The number of words in a term can be significant to determining whether it is coined, and whether it recites a function. A claim term that contains several words, such as the term "distributed learning control module",[151] is more likely to be coined by the patent drafter, perhaps because the term refers to a novel element that did not previously exist, or because the term is a means-plus-function element that indicates a function to be performed.

VIII. AGGREGATE METRICS OF CLAIM INDEFINITENESS

Different types of information are potentially useful to the indefiniteness analysis. Automatically extracting and displaying this information to the decision maker should facilitate the analysis of a patent by bringing some of the most relevant information to the decision maker's attention rapidly and with less reading. The AI software could also apply a simple score, say from 0 to 100, for each type of information indicating how strongly it weighs in favor of indefiniteness. For example, if all of a claim's terms are defined then that type of information could have a score of 0. If the same claim recites a function that is not disclosed anywhere in the specification, then that type of information could have a score of 90. In this manner a single claim could have different scores for different types of information.

Combining the scores for a claim into an aggregate metric would yield a single score for the indefiniteness of a claim. The aggregate metric could be a simple average of the different scores for a single claim. A weighted metric would have the advantage of allowing certain types of information to be more probative of the indefiniteness of a claim. An appropriate weighting scheme would attempt to produce a number that correlates highly with the decisions that would have been made by a human performing the indefiniteness analysis.[152] This weighting should certainly rely on empirical evidence as well as expert opinion about the importance of various indefiniteness factors which are to be components of the aggregate score.

To illustrate how a weighted metric could work, let us suppose that there are two claims, each with their own scores in three indefiniteness categories. Claim #1 is likely indefinite and claim #2 is likely not indefinite, according to their respective scores across all categories. The portions of the specification relevant to the claims are provided by the footnote at the end of claim #2. Words in red font indicate problematic words, such as terms that are not defined and functions without corresponding disclosed structure. Words that are underlined indicate terms that are not problematic because, e.g., they appear to be defined in the specification.

1. A mouse trap, comprising:

    a base made of sturdy material;

    a jaw frame pivotally mounted on the base; and

    a pulling means that biases the jaw frame towards the base when a trigger is activated.

[151] *Williamson*, 792 F.3d at 1350.

[152] *See supra* note 107, at 97.

2. A mouse trap, comprising:
   a base made of wood;
   a metal beam pivotally mounted on the base;
   a spring that biases the metal beam towards the base when a trigger is activated.[153]

	Claim #1	Claim #2
Terms not defined	91 "base" "jaw frame"	12 "base"
Recites function without corresponding structure	76 "pulling means"	0
Recites potentially vague adjectives or terms of degree	89 "sturdy" "pivotally"	3 "pivotally"
Unweighted (Average) Score	**85**	**5**
Weighted Score	**90**	**10**

An aggregate metric of indefiniteness for a single patent is useful because it would provide a concise data point for the decision maker. An automatically calculated aggregate metric is even more useful for analyzing large groups of patents. AI software can automatically score and then sort a group of patents (e.g., all patents owned by a company, all patents granted within a particular year) according to the indefiniteness score. This allows a very large set of patents to be quickly searched for patents that have the highest or lowest indefiniteness score. The results of such a search could be, e.g., the patents to target in litigation or post-grant proceedings, or the patents to exclude from a licensing program.

IX. BENCHMARKS AND EVALUATION

Designing AI software capable of automatically assessing indefiniteness with an acceptable level of performance is a difficult task. It is best envisioned as a long-term endeavor for which we should seek incremental progress towards an ever-stronger performance of automated indefiniteness determinations. This gradual development allows for small successes along the path to truly powerful legal decision-making systems.

The family of indefiniteness metrics proposed above are characterized by an aggregate of several different scores, in which each score is itself derived from a type of information extracted from patent text. Other classes of indefiniteness metrics are certainly possible. For example, a neural NLP classifier trained on a corpus of positive and negative examples (i.e. claims which are

[153] Suppose the relevant portions of the specification define the three terms "metal beam," "trigger," and "spring": "A metal beam, preferably shaped like a letter 'U' or 'T' can be used as the component that strikes the mouse when the spring drives it forward." "A trigger can be a simple catch lever that, when depressed by the weight of at least 10 grams, moves to an engaged position." "The spring is any standard coiled spring that delivers at least 10 Newtons of force."

indefinite and claims which are not indefinite) could in theory learn to distinguish definite from indefinite claims.

Any automatically determined metric of claim indefiniteness should be evaluated to measure its efficacy. Evaluation by a publicly available benchmark would be desirable as it would allow different metrics to be easily compared and their relative strengths and weaknesses to be publicly debated. One important characteristic for a metric is the rate it produces false positives and false negatives. In different circumstances a false positive might have much higher costs than a false negative, and vice versa.

Indefiniteness benchmarks could include a number of patent claims, and for each claim a Likert scale (e.g., 0 – 100) to rate how likely the claim is indefinite. Such claim ratings could be determined by a survey of patent experts. The rating for a claim could explicitly disregard anything but the text of the patent. For example, a panel of patent experts could be asked to rate the likely indefiniteness of a claim based only on the text of the patent and not on any information regarding the person of ordinary skill and how that person would interpret the claim terms. This would provide for a better evaluation of metrics that likewise disregard the person of ordinary skill. A benchmark could have variations, such as a rating for a claim that disregards anything but the patent text, and a rating that encompasses the true legal analysis, including how the person of ordinary skill would interpret the claim terms. Additional questions could be given to respondents in order to elicit more detailed information that is relevant to the legal inquiry of claim ambiguity. For example, questions could be used to probe whether certain claim terms should be defined, and whether certain linguistic structures in the claim add or detract from clarity. Such benchmarks would provide insight into whether, and how much, a metric that disregards the person of ordinary skill could serve as a useful proxy for the true indefiniteness analysis.[154]

Another type of benchmark could be collected from the patent office rejections. Often an application for patent is rejected because the patent office alleges that the claim is indefinite, and in response the patent applicant amends the claim to overcome the rejection. If the PTO withdraws the rejection after amendment, then presumably the pair of claims, before and after amendment, represents similar claims whose differences relate to claim ambiguity. However, it is possible that this benchmark would not be as accurate as the ones created by expert survey described above. First, the fact that the patent office rejects a claim as ambiguous does not necessarily mean the rejection was appropriate. Also, the patent applicant might make the claim amendment not because the rejection was well-founded but merely because it would be easier than arguing against it. Another drawback is that the amendment made by the patent applicant might serve many purposes besides overcoming the ambiguity rejection, so it would be difficult to isolate the specific changes that were relevant to ambiguity. Finally, a pair of claims, before and after amendment, does not carry as much information as, e.g., an ambiguity score on a scale of 1 – 10 for a particular claim.

Other benchmarks could evaluate other types of automated determinations besides claim indefiniteness. Benchmarks compiled from patent experts could rate the likelihood that a claim term is indefinite, the likelihood that a claim tern invokes §112(f), whether particular structure performs the function recited in a means-plus-function limitation, and similar considerations that are components of the final determination of claim indefiniteness.

[154] The fact that the metric comes from a "model" other than the model of the true legal analysis is not fatal to the metric's utility. "All models are wrong, some are useful." G. E. P. Box, *Robustness in the Strategy of Scientific Model Building*, Robustness in Statistics 201–236 (Robert L. Launer & Graham N. Wilkinson eds., 1979).

Finally, making these benchmarks and data sets publicly available would promote their use by AI researchers. This in turn would facilitate distributed development of legal AI tools and would allow researchers to independently pursue a variety of different AI designs. The most promising designs would emerge over time, be evaluated by the community, and be disseminated and improved upon by others.

X. CONCLUSION

Although AI cannot perform legal analysis at anything close to the level of a human, AI nevertheless can assist the decision maker. AI software can be made responsible for the more modest task of recognizing the types of information that are potentially relevant to the legal decision. This information can then be made available to the legal expert, who can then utilize this information to render a decision with much less reading and searching than would otherwise be required. Our extremely conservative estimates are that even a mildly effective form of AI-assisted definiteness analysis would save at least a hundred thousand hours of labor and several million dollars every year. More accurate estimates would likely demonstrate that the actual savings would be substantially greater.

We show how our framework can be used to apply AI to the legal decision of patent claim indefiniteness. Informed by the relevant case law, we offer several types of information that AI could automatically locate and present to the decision maker. Since the decisions of the AI system must ultimately be tested and accepted, we also propose how one might design metrics to evaluate the performance of such an AI system. We recognize that many alternative AI systems designs and metrics for evaluation are possible. It is hoped that broader participation by legal experts and AI researchers will promote greater interest in custom AI tools to aid the legal decision maker.