

AI Can't Accurately Predict Case Length And Cost — Yet

By **Andrew Russell**

In a recent [Law360 guest article](#), Joseph Avery, the CEO of Claudius Legal Intelligence, argued that predicting case length and fee totals at the very start of a case — at the time attorneys are retained — is no longer, as the New Jersey Supreme Court suggested, a "difficult, if not impossible, task."

He suggested that artificial intelligence and advanced statistics provide a solution that allows any attorney to precisely estimate how long each case will take and the potential fees when the attorney is retained. I disagree.



Andrew Russell

Modern AI techniques are incredible at many tasks, but precisely predicting fee amounts and case duration at the time an attorney is retained is not yet one of them. To be most effective at a prediction task, AI typically needs large datasets, a narrow range of possible conclusions, and ample factual data about the situation at hand.

Those things rarely exist at the start of a typical case. The idea that attorneys should be using AI right now to predict potential fees and case duration at the time of retention — or that they should have an ethical obligation to do so — is unsupportable.

It is trendy these days to suggest that AI, machine learning and deep neural networks can solve every prediction problem. They cannot. These techniques can be exceptionally useful, but work best on narrow prediction tasks in areas with big datasets.

For example, a large and well-curated dataset of pictures of faces can be used to train an AI model that can later determine whether a new picture matches a known face.[1] The accuracy may be no better than that of a person,[2] but the benefit derives from the model's ability to search a database of photos far faster than any human could.

AI is well-suited for some purposes in the legal world, such as technology-assisted document review. An algorithm can review documents far faster than any human, and can often be trained to offer useful results that can accelerate the review process.[3]

It can also potentially help, when deployed carefully, with narrower tasks such as reviewing a specific type of contract,[4] or assisting in legal research.[5] Further, many attorneys already use AI more than they know, in the form of online search and recommendation engines, email spam filters, optical character recognition software, and the multitude of other fields where AI has already been implemented and can provide accurate and precise results.

There have also been efforts at predicting judicial outcomes under narrow circumstances, such as predicting the behavior of the U.S. Supreme Court in a specific case,[6] or predicting decisions of the European Court of Human Rights.[7] But by that stage the facts have been developed, positions staked out and issues briefed.

Possible outcomes are often limited (e.g., to affirmance, reversal and so on). Further, large datasets of judicial opinions are often freely available, enabling researchers to train useful algorithms.

Using AI to accurately and precisely predict case duration and fee amounts at the start of the case, when an attorney is retained, is a much more difficult task. There are too many unknowns.

The judge may not be assigned. Discovery has not taken place, experts have not offered opinions, the parties' positions have not been fully developed, and countless critical facts remain undetermined.

Moreover, large and reliable datasets about case resolution can be difficult to collect. Duration and fee totals depend on case outcomes, and training an AI system to predict every possible outcome of a case requires numerous sample cases where each outcome occurred.

But court docketing systems typically charge fees for access, which can rapidly become prohibitive when collecting data. Free alternatives, such as the Free Law Project's RECAP system,[8] are helpful but incomplete.

Changes in the law also impede the collection of useful datasets, and the law changes regularly. For example, in the last 10 years, the law of patent eligibility has changed dramatically, leading parties to pursue different claims and strategies.

If a dataset includes too many older cases, they may pollute the effectiveness of the algorithm. A case that would have lasted two years a decade ago may be resolved in six months under today's law, or vice versa. Keeping a dataset up to date for all cases, in all areas of law, is impractical.

Thus, any dataset on case outcomes will likely be too sparse to enable accurate, precise predictions of fees and duration at the time an attorney is retained.

Recent suggestions to the contrary are unconvincing. In his recent guest article, Avery challenged the New Jersey Supreme Court's conclusion that, for certain types of discrimination cases, it is a "difficult, if not impossible, task" for attorneys to provide prospective clients "examples of how much hourly fees [and costs] have totaled in similar cases." [9] The court was specifically concerned with whether and how attorneys could supply prospective clients with information about fees in cases that were litigated by competitors. [10]

The author disagreed with the court's conclusion and argued that his analytics firm, and others in the industry, can "predict various case features, including how long a case will take" using AI and advanced statistics. He urged the court to "hold attorneys accountable" by imposing an ethical obligation to provide such estimates, using AI or advanced statistics, to their clients at the time of retention.

But he offered no specifics on how an attorney is supposed to do that. [11] He did not address the practical problems involved, including that cases turn on countless unknown factors, [12] or the difficulties in obtaining the kind of datasets necessary for training a truly accurate AI system. He cited no studies showing that any available AI-based prediction model can overcome the difficulties discussed by the New Jersey Supreme Court. [13]

The bottom line is that, at present, there is no generally available way for attorneys to apply AI to correctly predict case outcomes, durations and fee totals with specificity at such an early stage, particularly as to competitor fees. There are just too many variables and too

few available datasets. These issues will likely constrain the development of truly precise and accurate AI predictions in this area for years to come.

Andrew Russell is a partner at Shaw Keller LLP.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Yaniv Taigman et al., DeepFace: Closing the Gap to Human-Level Performance in Face Verification, 2014 IEEE Conference on Computer Vision and Pattern Recognition 1701 (2014).

[2] Id. at 1707.

[3] See Agnieszka McPeak, Disruptive Technology and the Ethical Lawyer, 50 U. Tol. L. Rev. 457, 463 (2019).

[4] Id. at 466-467.

[5] Drew Simshaw, Ethical Issues in Robo-Lawyering: The Need for Guidance on Developing and Using Artificial Intelligence in the Practice of Law, 70 Hastings L.J. 173, 193-194 (2018).

[6] Daniel Martin Kats, Michal J. Bommarito II, and Josh Blackman, A General Approach for Predicting the Behavior of the Supreme Court of the United States, PLoS ONE 12(4): e0174698(2017), <https://doi.org/10.1371/journal.pone.0174698>, code available on GitHub at <https://github.com/mjbommar/scotus-predict-v2/>.

[7] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiu-Pietro, and Vasileios Lamos, Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective, PeerJ Computer Science 2:e93, <https://doi.org/10.7717/peerj-cs.93> (2016).

[8] <https://www.courtlistener.com/recap/>.

[9] Balducci v. Cige, 2020 N.J. LEXIS 25, at *51 (Sup. Ct. N.J. Jan. 29, 2020) (quoting Balducci v. Cige, 456 N.J. Super. 219, 242-243 (Super. Ct. App. Div. 2018)) (alteration in original).

[10] Id. The Court's opinion addressed whether an attorney must "refer a potential client to a competitor" with lower fees or inform a potential client of the fees typically charged by "other competent counsel," and questioned how an attorney would get that information. Id. at *51-53.

[11] His company's home page offers few details about what datasets or techniques it may use.

[12] In its opinion, the court itself noted that, to provide fee information for a "similar" case litigated by another, "[t]he attorney would have to know whether the 'similar case' settled

or was tried, the nature and length of the discovery process, the number of depositions conducted and expert witnesses retained, the overall complexity of the litigation, and many other factors." *Id.* at *51.

[13] See, e.g., *id.*, at 51 (noting the question of "how are [attorneys] to acquire meaningful information about comparable hourly fees and costs?").