nature computational science

Brief Communication

https://doi.org/10.1038/s43588-025-00899-2

Data-driven law firm rankings to reduce information asymmetry in legal disputes

Received: 5 August 2024

Accepted: 2 October 2025

Published online: 31 October 2025

Check for updates

Alexandre Mojon [1], Robert Mahari [2,3,4] & Sandro Claudio Lera [3,5,6]

Selecting capable counsel can shape the outcome of litigation, yet evaluating law firm performance remains challenging. Widely used rankings prioritize prestige, size and revenue over empirical litigation outcomes, offering little practical guidance. Here, to address this gap, we build on the Bradley–Terry model and introduce a new ranking framework that treats each lawsuit as a competitive game between plaintiff and defendant law firms. Leveraging a newly constructed dataset of 60,540 US civil lawsuits involving 54,541 law firms, our findings show that existing reputation-based rankings correlate poorly with actual litigation success, while our outcome-based ranking substantially improves predictive accuracy. These findings establish a foundation for more transparent, data-driven assessments of legal performance.

Legal representation plays a crucial role in determining litigation outcomes. Research consistently shows that litigants represented by attorneys fare better in court than those without[1]. Moreover, the quality of representation substantially affects case outcomes. In asylum proceedings, for example, individuals with better attorneys have a higher probability of prevailing[2]. This effect extends beyond individual cases, shaping broader patterns of judicial decision-making[3].

Despite its importance, most litigants lack the information to identify the most effective attorney for their case. Legal expertise is highly specialized, and success in one domain does not necessarily translate to another. Information asymmetries in the legal market favor litigants with more financial resources, who can afford specialized advice or firms with established reputations[1,4]. This knowledge gap reinforces structural inequalities, allowing well-resourced litigants to gain advantages independent of case merit[5–9].

Existing law firm rankings do little to mitigate this issue. Rankings such as Vault 100 and American Lawyer Media's (ALM) Global 200 emphasize prestige, size or revenue, rather than litigation performance. They typically cover only the top 100–300 firms, excluding many that may be better suited for specific cases. Reputation-based marketing may reinforce a feedback loop in which high-ranking firms attract more clients, bolstering prestige regardless of courtroom success[10]. Moreover, such rankings are not tailored to case type or legal domain, limiting their utility to litigants.

To address these limitations, we introduce a methodology for ranking law firms based on objective case outcomes rather than subjective prestige. We generalize a ranking algorithm for heterogeneous pairwise zero-sum games[11], where lawsuits are games and law firms are players. This model builds on the Bradley–Terry framework[12], which ranks entities by their probability of winning based on relative strength. To capture legal heterogeneity, we incorporate case-type-specific priors and asymmetric baseline outcomes, such as defendants' higher likelihood of winning[13]. Given a set of observed lawsuits, we estimate latent law firm scores using an expectation maximization algorithm. These scores serve as the basis for law firm rankings and enable the prediction of lawsuit outcomes based on score differentials (Fig. 1).

To validate our approach, we compile a dataset of 60,540 civil lawsuits in US Federal District Courts. Using natural language processing (NLP), we extract records including the names of the plaintiff and defendant law firms, and the case outcome. Applying our method yields scores for 2,064 law firms. We find that these scores are not significantly correlated with prestige-based rankings, but substantially outperform them in predicting future outcomes. Thus, our ranking approach supports litigants by providing a quantitative basis for selecting law firms and shaping litigation strategy, including settlement decisions.

[1]Chair of Mathematics, University of St. Gallen (HSG), St. Gallen, Switzerland. [2]CodeX, Stanford University, Stanford, CA, USA. [3]Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Harvard Law School, Cambridge, MA, USA. [5]Institute of Risk Analysis, Prediction and Management (Risks-X), Academy of Interdisciplinary and Advanced Studies, Southern University of Science and Technology, Shenzhen, China. [6]Business School, Southern University of Science and Technology, Shenzhen, China. e-mail: leras@sustech.edu.cn

https://doi.org/10.1038/s43588-025-00899-2

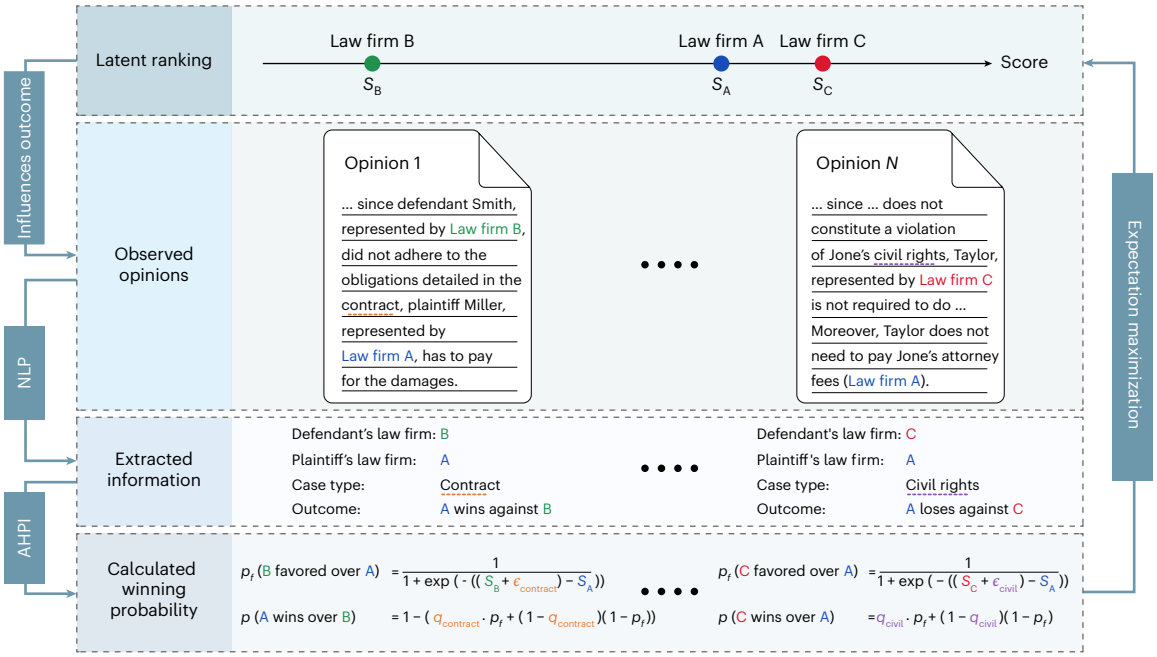


Fig. 1 | Law firm ranking procedure. We analyze N lawsuits based on judges' textual decisions (opinions), from which we extract structured information on the case type, the case outcome and the law firms involved. Each lawsuit is modeled as a game between the plaintiff's and defendant's law firm, where one firm wins and the other loses. Each of the five case types has an associated defendant bias (home field advantage) ϵ_m, which quantifies the a priori likelihood that the defendant prevails in case type m. Each law firm $k = 1, \ldots, K$ is assigned a latent skill score, S_k. The propensity p_f that defendant's law firm B is favored over the plaintiff's law firm A is modeled as a sigmoid function of $(S_B + \epsilon_m) - S_A$, such that a higher bias-adjusted score for B increases its likelihood of being favored. The valence probability q_m then determines the probability p that the favored firm ultimately wins, thus capturing the role of uncertainty in case outcomes. Based on these assumptions, we apply an expectation-maximization algorithm to infer the latent law firm scores $\{S_k\}$, defendant biases $\{\epsilon_m\}$ and valence probabilities $\{q_m\}$ that best explain the observed litigation outcomes.

While well-resourced parties may already assess firm performance qualitatively, our method democratizes these insights, making them broadly accessible. This shift has the potential to reshape how parties engage legal counsel and navigate legal disputes.

Results

A dataset of zero-sum law games

We compile a dataset of 60,540 civil lawsuits in US federal district courts to construct law firm rankings based on litigation outcomes. While most existing studies analyze small-scale disputes in domains such as small claims[14], family[15] or housing law[16], our data span a broad range of civil litigation, including civil rights, torts, contracts, labor and commercial disputes. Building on earlier work that used these data to analyze judicial biases[17], we extend the analysis to include the identities of the law firms involved (Methods). Case types are grouped into five categories: civil rights, contracts, labor, torts and other. Case outcomes are binary: 1 for plaintiff victories, 0 for defendant victories.

This process yields 60,540 civil cases involving 54,541 distinct law firms. For cases with multiple firms on one side, we decompose the interactions into pairwise games between one plaintiff and one defendant firm. This results in 190,297 pairwise interactions, each annotated with the names of the opposing firms, the case type and the outcome. We use the first 80% of cases to estimate firm scores and reserve the remaining 20% for out-of-sample evaluation.

From these interactions, we construct an interaction network of law firms, where each edge represents a lawsuit between two opposing firms. The network is sparse and heterogeneous: some firms appear frequently, while others appear only once. To focus on firms with sufficient data, we define a Q-factor as the average number of interactions per firm within a trimmed subnetwork. We iteratively remove low-activity firms until the subnetwork meets a minimum Q-factor threshold. Using $Q = 30$ in the main results yields a robust subnetwork of $N = 63{,}115$ interactions among $K = 2{,}064$ firms. We show in Supplementary Section 6 that results remain stable across a range of Q values from 20 to 55.

Ranking law firms via asymmetric pairwise interactions

We use the above dataset, comprising $N = 63{,}115$ interactions among $K = 2{,}064$ law firms (with $Q = 30$), to develop an objective metric of law firm performance. A naive approach would rank firms by overall win rate. However, this has two key limitations: First, there is a systematic bias favoring defendant victories (similar to a home field advantage)[13,17,18]. Thus, a plaintiff win is not equivalent to a defendant win, requiring an asymmetry adjustment. Second, case types (rules of the game) differ in baseline defendant win rates; for example, defendants win 86% of civil rights cases but 70% of contract cases. Although firms often specialize, they still engage in multiple case types, so their outcomes occur under varying conditions. These issues are compounded by noise or luck in outcomes, where weaker firms sometimes win and stronger ones lose[19]. We thus propose a ranking method that adjusts for these structural effects.

Suppose K players engage in N pairwise interactions (games) with a winner and a loser, as in sports or, here, lawsuits. Our aim is to assign player rankings that reflect skill. A standard approach is the probabilistic Bradley–Terry model, where the probability that entity A (score S_A) defeats B (score S_B) is a sigmoid of $S_A - S_B$ (ref. 12). This model has been extended to account for different interaction types[11], essential in our setting, where case types influence outcomes. Unlike standard ranking problems, litigation is structurally asymmetric: defendants have a higher baseline chance of winning, which varies by case type (civil rights: 86%, contract: 70%, torts: 85%, labor: 74%, other: 75%). We therefore develop a ranking model that includes both multiple interaction types (case types) and systemic defendant advantage. We call this the asymmetric heterogeneous pairwise interactions (AHPI) ranking algorithm.

https://doi.org/10.1038/s43588-025-00899-2

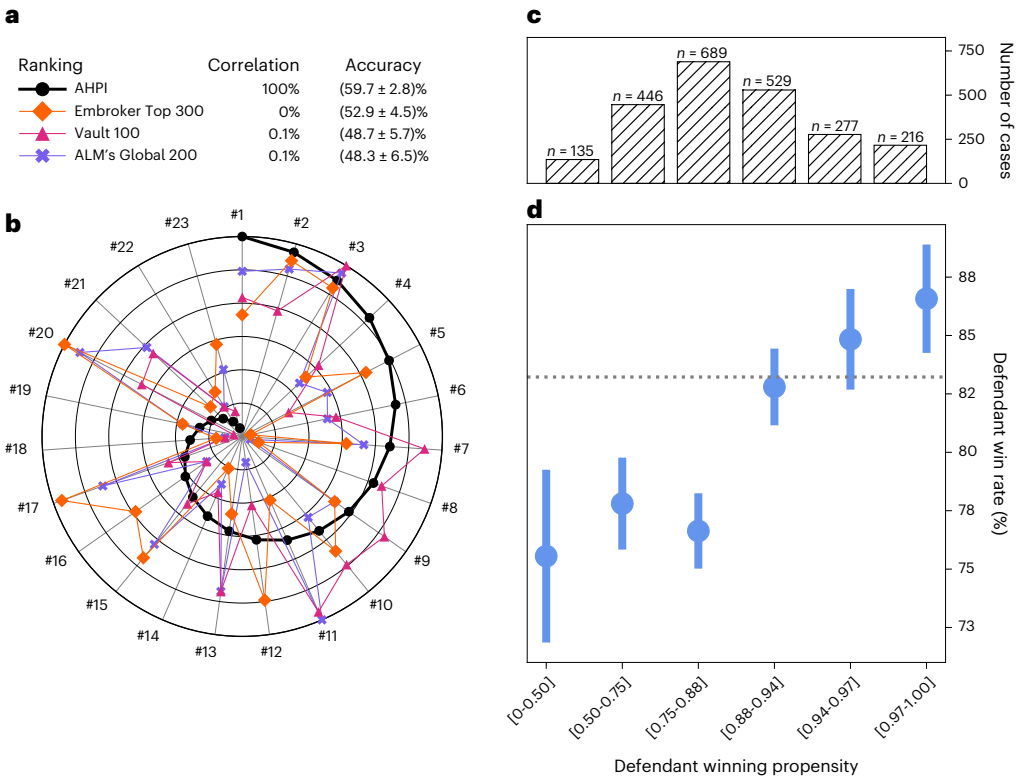


Fig. 2 | Comparison of law firm rankings and predictive performance. We compare our AHPI scores with three widely used firm rankings: Vault 100, ALM's Global 200 and Embroker Top 300. **a**, Kendall correlation between AHPI ranking and each of the three benchmark rankings, and mean predictive accuracy on test cases given each ranking. An accuracy of 50% represents the expected accuracy from random guessing. Errors indicate standard deviations computed via 100 bootstrap resamples. **b**, A comparison of ranks for 20 law firms that are common across all four rankings. A point close to the periphery means a higher rank than one closer to the center. As visually apparent and indicated by the low correlation coefficients in the 'Correlation' column in **a**, there are significant differences between AHPI scores and other rankings. **c**,**d**, The predicted propensity of a defendant win for out-of-sample test cases, grouped into six propensity bins: the number of cases within each bin (**c**); the average winning propensity based on the AHPI ranking compared with the actual defendant win rate in each bin (**d**). Error bars indicate standard deviations computed via 100 bootstrap resamples. The dotted line corresponds to the 83% baseline defendant win rate across all cases. Empirical defendant win rates for cases with low propensities are significantly below the baseline, while win rates for cases with high propensities are significantly above the baseline.

AHPI uses a Bayesian expectation-maximization framework with a logistic prior over scores (Methods). It takes N observed binary interactions, modeling each as an asymmetric game between K entities. There are M case types. Each game is asymmetric: when law firms A and B have equal scores ($S_A = S_B$), the defendant firm (home team) has a higher base win chance. Each case type has a different baseline defendant advantage, reflecting systemic bias. AHPI proceeds in two stages: First, it estimates the probability that law firm B is favored over A via a sigmoid of $(S_B + \epsilon) - S_A$, where $\epsilon > 0$ is a bias toward defendants. Each case type m has its own fitted ϵ_m. Second, a case-type-specific valence parameter $q_m \in [0.5, 1]$ gives the probability that the favored firm wins, accounting for uncertainty. Lower q_m implies more randomness in outcomes. The K firm scores, M case-specific biases and M valence probabilities are all fit from the data.

Fitting AHPI to $N = 63{,}115$ interactions yields a score S_k for each of the $K = 2{,}064$ law firms $k \in 1, \ldots, K$, reflecting litigation performance (Fig. 1). To contextualize our ranking, we compare it with three widely used public rankings: Vault Law 100, ALM's Global 200 and Embroker Top 300 (2022). These rankings use varied criteria but do not include litigation outcomes, instead relying on reputation and size (Supplementary Section 8). To measure alignment, we compute Kendall's τ correlation (Spearman yields similar results) over the firms present in both rankings. The correlations are essentially zero (Fig. 2a,b), indicating existing rankings poorly reflect empirical performance. This mirrors findings for physicians, where prestige of medical schools poorly predicts patient outcomes[20].

As a byproduct, we estimate case biases ϵ_m and valence probabilities q_m. The fitted ϵ_m values are consistently positive (Methods), consistent with defendants' higher baseline win rates. The valence probability q_m captures how much rankings influence case outcomes for type m: $q_m = 0.5$ implies randomness; $q_m \approx 1$ implies outcomes closely track rankings (Methods; Fig. 1). Our results consistently show $q_m > 0.85$, underscoring both qualitative and quantitative evidence for the importance of legal representation[2,3].

Score-based prediction of trial outcomes

A critical element of legal strategy is assessing the likelihood of litigation success, a topic extensively explored both theoretically and empirically[21–23]. With the advent of big data and machine learning, such assessments have become more quantitative, enabling data-driven evaluations of case viability[9,24–26]. These tools are particularly prominent in the litigation finance industry, where lawsuit portfolios are analyzed for return potential[18,27].

Here, we examine whether our law firm rankings improve the prediction of litigation outcomes. Recall that we use the first 80% of cases to fit scores $\{S_k\}$ for $K = 2{,}064$ firms, along with $M = 5$ case biases $\{\epsilon_m\}$ and valence probabilities $\{q_m\}$. We now use the remaining 20% to test predictive performance.

A simple rule would predict that plaintiff law firm A wins over defendant law firm B if $S_A > S_B + \epsilon_m$. However, because defendants win most cases, always predicting a defendant victory achieves a strong baseline accuracy of 83%. To show our model adds value, we leverage

its probabilistic structure. Instead of binary predictions based only on score differences, we use the sigmoid function to estimate the winning propensity of the plaintiff. Because the fitted valence probabilities q are close to 1, the sigmoid of $S_A - (S_B + \epsilon_m)$ can be interpreted directly as a win probability (Methods; Fig. 1). When this adjusted difference is large, law firm A is highly likely to win; when it is small, the outcome is more uncertain. We group the test cases into six bins by predicted win propensity and report empirical defendant win rates for each. As shown in Fig. 2d, cases with low plaintiff propensity show win rates well below the 83% baseline, while those with high propensity exceed it. This demonstrates that our ranking model enables more granular, informative predictions of case outcomes—supporting better law firm selection and litigation strategy.

We also compare our results with public rankings, treating published rankings as firm scores. Due to limited case overlap, each ranking is tested on its subset of applicable cases. To compare fairly, we balance the test data so that plaintiffs and defendants each win 50% of the time, making the baseline prediction accuracy 50%. We then predict that A beats B if $S_A > S_B + \epsilon_m$. As shown in Fig. 2a, our approach exceeds the baseline by nearly 10%. By contrast, Embroker achieves under 3% improvement, while Vault and ALM perform no better than chance. Because our model covers over 2,000 firms, it also yields smaller standard errors, estimated via bootstrapping. Together, these results show that data-driven rankings offer better predictive accuracy for litigation outcomes.

Discussion

While our approach provides a framework for ranking law firms based on litigation outcomes, it faces limitations that warrant discussion.

A key limitation is survivorship bias: our dataset includes only cases with judicial opinions, excluding private settlements. As most legal disputes settle before trial[28], our data capture only litigated outcomes. However, litigants negotiate 'in the shadow of the law'[29], meaning trial outcomes shape settlement behavior. If law firm strength drives trial outcomes, it probably also affects settlements. Although public data on settlements remain scarce, future work should aim to incorporate them for a more holistic evaluation of law firm performance. Our ranking framework can accommodate settlements as a third outcome, akin to draws in extended Bradley–Terry models[30].

Selection bias may also arise if litigants assign the best firms to the hardest cases. We believe this is rare, as only sophisticated litigants—typically large firms with legal departments—could make such assessments. In practice, companies tend to rely on established law firm relationships rather than switching firms for difficult cases[31].

In multifirm cases, we have reduced the interaction to pairwise comparisons between individual firms. Future work could explore models accounting for law firm teams, capturing cooperative effects and potential synergies. We also do not model law firm mergers or splits, which may affect performance continuity. Incorporating firm genealogies could allow rankings to adjust dynamically over time.

Our analysis is restricted to US federal civil litigation, limiting generalizability to state courts or other jurisdictions. Extending our method to state and international courts would test its broader applicability.

Litigation success is only one part of legal practice. Future versions of our model could include case duration, fees or financial data to capture overall client value. Incorporating judicial characteristics, such as judge-specific biases[17], could further refine our estimates.

Beyond law, our AHPI framework generalizes to other adversarial settings. It applies wherever entities engage in repeated, asymmetric, heterogeneous interactions—such as political contests, lobbying or market competition. Extending the model to multiparty disputes or auction-based legal markets could broaden its scope.

Despite limitations, our work establishes an empirical basis for ranking law firms based on litigation outcomes. By shifting focus from reputation to data-driven metrics, we aim to improve transparency and inform future research in empirical legal studies.

Methods

Data

Our dataset is constructed from judicial opinions provided by the Case Law Access Project (CAP), which contains textual records of US Federal District Court cases, closely following the method presented in ref. 17. We restrict our analysis to civil litigation in federal district courts, excluding appellate and Supreme Court cases to ensure a consistent trial-level dataset. This leaves us with 302,988 cases in the form of judges' published opinions. These opinions include detailed case descriptions but lack structured metadata on case type, case outcome or legal representation. To address this, we leverage structured data from the Integrated Database (IDB) provided by the Federal Judicial Center as training data to develop machine learning models that infer key case attributes directly from judicial opinions.

The IDB is a structured database of federal civil lawsuits, containing metadata fields such as case filing date, case type and case outcome (plaintiff win, defendant win or unknown). By matching CAP cases to IDB records via docket numbers and filing dates, we successfully align 31,222 cases between the two datasets[17]. These matched cases serve as labeled training data for extracting structured information from the full CAP corpus, which otherwise lacks metadata annotations.

To infer case outcomes and types for the CAP cases, we fine-tune two Bidirectional Encoder Representations from Transformers (BERT) classifiers using the 31,222 IDB-mapped cases as training data. Both classifiers take as input the judge's opinion. The first model is a binary classifier that predicts whether the plaintiff won or lost, while the second is a five-class classifier that assigns cases to one of five major case types—civil rights, contracts, labor, torts and other. Once trained, these models are applied to the filtered CAP dataset, allowing us to classify case outcomes and case types for all 302,988 cases. After discarding cases for which our predictive models have low confidence, we obtain a total of 205,454 labeled cases. This dataset forms the basis of our analysis. We refer to Supplementary Section 1 for an in-depth discussion of how these models have been trained.

While this procedure provides structured labels for case type and outcome, it does not yet identify the law firms involved in each case. In CAP, law firm names appear in the attorney sections of judicial opinions, often inconsistently formatted due to optical character recognition errors, abbreviations or typographical variations (for example, 'Putman & Putman' versus 'Put-Man and Put-Man'). To extract this information, we apply a pattern-matching approach to detect law firm names, followed by hierarchical clustering based on Levenshtein distance to merge name variants. A similar method is used to determine which law firms represent plaintiffs and defendants in each case. This step ensures consistency in firm identification and representation assignment and is detailed further in Supplementary Section 2. After discarding cases in which we cannot identify at least one law firm on each side, we are left with a total of 67,619 cases and 64,656 law firms. A detailed breakdown of these numbers is found in Supplementary Section 1.

A key feature of our dataset is that it enables the construction of an interaction network of law firms. Each law firm is represented as a node, and each lawsuit interaction is represented as an edge. This network exhibits considerable variation in density, as some law firms engage in frequent litigation while others appear only sporadically. To systematically analyze this structure, we introduce the Q-factor, which quantifies the average number of observed cases per law firm within a given subset of the network. By definition, the Q-factor of our total dataset is $167{,}439/54{,}541 \approx 3.1$. By iteratively removing the law firms with fewest interactions, we trim the network so that the average number of interactions per law firm, that is, the Q-factor, increases. Setting Q sufficiently high ensures that each firm has a meaningful

number of observed interactions, allowing a more reliable estimation of its litigation performance. For the results presented in Fig. 2, we use $Q = 30$ to compute law firm rankings, yielding an interaction network with $N = 63{,}115$ interactions involving $K = 2{,}064$ law firms. We confirm in Supplementary Section 6 that our results remain robust across a range of Q values from 20 to 55.

Modeling asymmetric, heterogeneous, pairwise interactions

The ranking of entities based on pairwise interactions is a ubiquitous problem arising in domains such as information retrieval[32], decision theory[33], sports[34] and university rankings[35].

Several algorithms have been established to estimate rankings from observed outcomes. The Elo rating system, widely used in chess[36], updates player scores dynamically based on match outcomes, where score adjustments depend only on the prematch ratings of the two competitors. Although Elo provides an efficient heuristic for continuously updated rankings, it does not maximize a formal likelihood function and cannot easily incorporate multiple interaction types or systematic biases in matchups. By contrast, AHPI represents the probability that one competitor defeats another as a logistic function of the difference in their latent skill levels[12], allowing a principled estimation of rankings through likelihood maximization. Recent extensions of this model have incorporated multiple interaction types while also enhancing computational efficiency[11]. We propose the AHPI ranking algorithm, which further generalizes this approach to account for asymmetric games, where one entity has a higher a priori chance of winning (for example, a home field advantage for defendants in litigation). This formulation allows a more statistically rigorous ranking framework, ensuring that rankings reflect both observed outcomes and structural biases in competition.

We consider K entities (for example, law firms) $k = 1, \ldots, K$ competing in a total of N pairwise interactions (for example, lawsuits) $n = 1, \ldots, N$. Each interaction is of one of M types (for example, lawsuit case types or variations of the game) $m = 1, \ldots, M$. The interaction is asymmetric in the sense that defendants are more likely to win a case. We assume that this home field advantage or defendant bias is specific to the case type and denote it by ϵ_m. The AHPI algorithm assigns a score S_k to every competing entity k. Consider the interaction $n \in \{1, \ldots, N\}$ that is of type m and involves the entity A and the privileged entity B with respective scores S_A and S_B. The outcome is modeled in two stages: First, a favored entity is determined. The probability that A is favored is given by

$$\rho_n(\mathrm{A}) = \frac{1}{1 + \exp\left(-(S_A - (S_B + \epsilon_m))\right)}, \tag{1a}$$

where the privilege ϵ_m skews this probability in B's favor. It is implicitly understood that the interaction type m depends on game n, that is, $m = m(n)$. We denote ϵ_m for brevity and proceed similarly for all subsequent quantities. Second, the winner of the interaction is determined by introducing the valence probability q_m. The valence probability captures the probability that the favored entity wins. Hence, A wins with probability

$$p_n(\mathrm{A}) = q_m \rho_n(\mathrm{A}) + (1 - q_m)\left(1 - \rho_n(\mathrm{A})\right), \tag{1b}$$

and B wins with probability $1 - p_n(\mathrm{A})$. The advantage of this two-step approach is that the skills of the entities, expressed in their scores S, are decoupled from the probabilistic outcome associated with interaction type m through q_m. If the score difference corrected by the privilege in equation (1a) is 0, both entities have an equal probability of being favored. To what extent being favored implies winning is subsequently determined by the valence probability q_m. A value of q_m close to 1 indicates that the favored entity is highly likely to win, whereas $q_m = 0.5$ indicates that being favored does not have an impact. This approach has previously been used to calibrate the relative contributions of skill and luck[19]. We find that fitted valence probabilities tend to be close to 1, suggesting that the law firms' score differences are highly indicative of case outcomes.

Estimation of latent scores

We now turn to the estimation of the scores $\{S_k\}_{k=1}^K$, privileges $\{\epsilon_m\}_{m=1}^M$ and valence probabilities $\{q_m\}_{m=1}^M$ given N observed interactions $\{I_n\}_{n=1}^N$. To this end, we introduce for every interaction n the stance variable σ_n which takes a value of 1 if the favored entity is the winner and 0 otherwise. We label the winning entity by u_n and the losing one by v_n. It is convenient to define $\lambda_k \equiv e^{S_k}$. To further simplify notation, we use S instead of $\{S_k\}_{k=1}^K$ to denote the set of scores, and similarly for $\{\sigma_n\}$, $\{\epsilon_m\}$ and $\{q_m\}$. To tackle potential convergence issues, we use a Bayesian framework and introduce logistic priors for scores and privileges[11,37]. Applying Bayes' theorem, the likelihood $P(S, q, \sigma, \epsilon|x)$ is

$$\prod_n P_n(x, \sigma|\lambda_u, \lambda_v, \epsilon, q) \prod_{k=1}^K \frac{\lambda_k}{(\lambda_k + 1)^2} \prod_{m=1}^M \frac{1}{(\mathrm{e}^{\epsilon_m} + 1)(\mathrm{e}^{-\epsilon_m} + 1)}, \tag{2}$$

where the second and third product terms stem from the prior on the scores and privileges, respectively. By defining the privilege stance variable c_n as taking value −1 if the winner was privileged and 1 otherwise, and by using the definitions of the probabilities ρ_n and p_n in equation (1), it holds that

$$P_n(x, \sigma|\lambda_u, \lambda_v, \epsilon, q) = \frac{\left(e^{c_n \epsilon_m} \lambda_u q_m\right)^{\sigma} \left(\lambda_v (1 - q_m)\right)^{1-\sigma}}{e^{c_n \cdot \epsilon_m} \lambda_u + \lambda_v}. \tag{3}$$

It turns out that directly maximizing (the logarithm of) the likelihood (equation (2)) is challenging, and an expectation maximization algorithm is better suited[11]. To this end, we first note that, across N games, there is a total of $R = 2^N$ outcomes of the binary stance variables $\{\sigma_n\}$, and we denote by Π any probability distribution over these outcomes. By Jensen's inequality, it holds that

$$\log \sum_{r=1}^R P(S, q, \sigma, \epsilon|x) \geqslant \sum_{n=1}^R \Pi(\sigma) \log \frac{P(S, q, \sigma, \epsilon|x)}{\Pi(\sigma)}. \tag{4}$$

If the right-hand side equals the left-hand side, we can work with the sum of logarithms, which renders the calculations below analytically tractable. For fixed S, ϵ and q, equality holds for a specific probability distribution Π:

$$\Pi(\sigma) = \prod_{n=1}^N \pi_n^{\sigma_n} (1 - \pi_n)^{1-\sigma_n}, \tag{5a}$$

where

$$\pi_n = \frac{e^{c_n \epsilon_m} \lambda_{u_n} q_m}{\lambda_{u_n} e^{c_n \epsilon_m} q_m + \lambda_{v_n} (1 - q_m)} \tag{5b}$$

can be interpreted as the posterior probability that u_n is the favored entity[11].

We now apply expectation maximization to the right hand side of equation (4) with Π given by equation (5). To maximize expectations, we set the derivative of the right-hand-side of equation (4) with respect to S, ϵ and q equal to zero while holding the distribution over σ constant. This results in the following set of equations where $\delta_{\mu,\nu}$ is the Kronecker delta and γ_n equals $\mathrm{e}^{-\epsilon_n}$ if the privileged entity wins, and e^{ϵ_n} otherwise.

$$q_m = \frac{\sum_{n=1}^N \delta_{m_n,m} \pi_n}{\sum_{n=1}^N \delta_{m_n,m}} \tag{6a}$$

$$0 = \frac{1 - e^{\epsilon_m}}{1 + e^{\epsilon_m}} + \sum_{n=1}^N \delta_{m_n,m} \left[\pi_n c_n - \frac{\lambda_{u_n} e^{c_n \epsilon_m}}{\lambda_{u_n} e^{c_n \epsilon_m} + \lambda_{v_n}} c_n \right] \tag{6b}$$

$$\lambda_k = \left[1 + \sum_{n=1}^{N} \delta_{u_n,k}\pi_n + \delta_{v_n,k}(1-\pi_n)\right] \times \left[\frac{2}{1+\lambda_k} + \sum_{n=1}^{N} \frac{\delta_{u_n,k}\gamma_n(\gamma_n\lambda_{u_n}+\lambda_k)+\delta_{v_n,k}(\gamma_n\lambda_k+\lambda_{v_n})}{(\gamma_n\lambda_k+\lambda_{v_n})(\gamma_n\lambda_{u_n}+\lambda_k)}\right]^{-1}. \quad (6c)$$

Equation (6a) yields an explicit expression for the valence probability q_m, where π_n is given by equation (5b). Equation (6b) yields an implicit expression for the privilege ϵ_m, which can be solved numerically. Equation (6c) yields a nested set of equations for the exponential scores λ_k, which can be solved iteratively until convergence is reached.

Finally, it should be noted that the above calculations are invariant under the mapping

$$\begin{cases} f : \mathbb{R}\times\mathbb{R}\times[0,1] \to \mathbb{R}\times\mathbb{R}\times[0,1] \\ f(S,\epsilon,q) = (-S,-\epsilon,1-q) \end{cases}, \quad (7)$$

which suggests that the ranking might be 'inverted'. This symmetry can be broken by analyzing the values of the valence probabilities. If for an interaction type m a higher rank is assumed to imply a higher winning probability, but if $q_m < 0.5$, then the ranking must be inverted according to equation (7). Likewise, if a higher rank is assumed to imply a lower winning probability and $q_m > 0.5$, then the ranking must be inverted.

Fitting latent scores on dense data subsets

Our dataset contains 167,439 interactions (games) across 54,541 law firms (entities). Each interaction is of one of five types (civil rights, contracts, labor, torts and other). Ideally, we assign a ranking score S_k to each law firm $k \in 1, \ldots, K = 54{,}541$ by applying the above expectation-maximization algorithm. However, it is intuitively plausible that the fewer interactions are observed per entity, the less reliable the estimated scores (see Supplementary Section 5 for a demonstration on data with known ground truth). We therefore systematically trim our dataset to a subset of sufficient interactions via the Q-factor, described above. By definition, the Q-factor of our total dataset is $167{,}439/54{,}541 \approx 3.1$. To find a subset of the data with sufficiently high target Q-factor, we iteratively remove the entities with the lowest number of interactions and recalculate the Q-factor. We stop once the average number of interactions in the remaining data is greater than or equal to Q. Law firm rankings (Fig. 2a,b) and outcome predictions (Fig. 2d) have been calculated on a subset with $Q = 30$ for which the number of law firms is reduced to 2,064 and the number of interactions is reduced to 63,115. Qualitatively similar results for a larger range of Q-factors are found in Supplementary Section 8.

For a given dataset of interactions with fixed Q-factor, we proceed with the estimation of the model parameters. Concretely, we estimate the law firm scores $S_k = \log(\lambda_k)$, the case type biases ϵ_m and the valence probabilities q_m via the above-described expectation-maximization algorithm. The following initial values are used for all firms and interaction types, respectively: $\lambda_k = 0.9$, $q_m = 0.5$, $\epsilon_m = 0$. The algorithm is iterated until convergence is reached, that is, once the correlation between the ranking of two subsequent iterations is above 99.9% and the maximum absolute change in any ranking score, valence probability and case type asymmetry is below 0.01.

The parameters are fit only on the first 80% of the cases according to their temporal order (accordingly, the interaction network and the Q-factor were determined only on the training data, not the test data). The remaining 20% are used to predict defendant winning probabilities via equation (1), where S_k, ϵ_m and q_m are replaced by their fitted best estimates $\hat{S}_k, \hat{\epsilon}_m$ and $\hat{q}_m$, respectively. In accordance with the defendant bias, we estimate consistently positive $\hat{\epsilon}_m$ parameters: civil rights (2.03), contract (1.66), torts (0.32), labor (1.99) and other (1.90). The associated valence probabilities $\hat{q}_m$ are estimated as follows: civil rights (0.86), contract (0.96), torts (1.00), labor (0.96) and other (1.00). These high valence probabilities imply that the favored entity typically wins. Similarly high values have been observed in previous work[11], and we refer to Supplementary Section 5 for additional discussions. Qualitatively, this suggests that lawsuit outcomes are strongly influenced by the difference in skill levels between the competing law firms.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. These data for our work are based on ref. 17 and augmented by identifying and disambiguating law firms and their roles. The data for reproducing all the results are available via figshare at https://doi.org/10.6084/m9.figshare.30260035 (ref. 38). The analyses and rankings that we report may be reproduced by running our code.

Code availability

The code to reproduce all the analyses in this Brief Communication is available via GitHub at https://github.com/mojona/law_firm_ranking and Zendo at https://doi.org/10.5281/zenodo.15504218 (ref. 39). Our data preparation and analysis uses the latest standard libraries at the time of submission. Specifically, we use the following: Python (v. 3.10) packages: pandas (v.2.2.2), numpy (v. 2.1.0), scipy (v.1.11.1), scikit-learn (v. 1.5.1), rapidfuzz (v.3.9.6) and pip (v. 24.2).

References

1. Poppe, E. S. T. & Rachlinski, J. J. Do lawyers matter? The effect of legal representation in civil disputes. *Pepperdine Law Rev.* **43**, 881 (2015).
2. Miller, B., Keith, L. C. & Holmes, J. S. Leveling the odds: the effect of quality legal representation in cases of asymmetrical capability. *Law Soc. Rev.* **49**, 209–239 (2015).
3. Szmer, J., Johnson, S. W. & Sarver, T. A. Does the lawyer matter? Influencing outcomes on the Supreme Court of Canada. *Law Soc. Rev.* **41**, 279–304 (2007).
4. Sandefur, R. L. Elements of professional expertise: understanding relational and substantive expertise through lawyers' impact. *Am. Sociol. Rev.* **80**, 909–933 (2015).
5. Galanter, M. Why the "haves" come out ahead: speculations on the limits of legal change. *Law Soc. Rev.* **9**, 95–160 (1974).
6. Rhode, D. L. *Access to Justice* (Oxford Univ. Press, 2004).
7. Garrett, B. L. *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* (Harvard Univ. Press, 2011).
8. Cornwell, E. Y., Poppe, E. S. T. & Bea, M. D. Networking in the shadow of the law: informal access to legal expertise through personal network ties. *Law Soc. Rev.* **51**, 635–668 (2017).
9. Pah, A. R. et al. How to build a more open justice system. *Science* **369**, 134–136 (2020).
10. Zacharias, F. C. Effects of reputation on the legal profession. *Wash. Lee Law Rev.* **65**, 173 (2008).
11. Newman, M. E. J. Ranking with multiple types of pairwise comparisons. *Proc. R. Soc. A* **478**, 20220517 (2022).
12. Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–345 (1952).
13. Eisenberg, T., Goerdt, J., Ostrom, B. & Rottman, D. Litigation outcomes in state and federal courts: a statistical portrait. *Seattle Univ. Law Rev.* **19**, 433 (1995).
14. LaFree, G. & Rack, C. The effects of participants' ethnicity and gender on monetary outcomes in mediated and adjudicated civil cases. *Law Soc. Rev.* **30**, 767–797 (1996).
15. Maccoby, E. E. & Mnookin, R. H. *Dividing the Child: Social and Legal Dilemmas of Custody* (Harvard Univ. Press, 1992).

16. Hannaford-Agor, P. & Mott, N. Research on self-represented litigation: preliminary results and methodological considerations. *Just. Syst. J.* **24**, 163–181 (2003).
17. Mahari, R., Lera, S. C. Early career citations capture judicial idiosyncrasies and predict judgments. Preprint at https://arxiv.org/abs/2410.00725 (2024).
18. Lera, S. C., Mahari, R. & Strub, M. S. Litigation finance at trial: model and data. *SSRN Working Paper* (4091716) (2022).
19. Jerdee, M. & Newman, M. E. J. Luck, skill, and depth of competition in games and social hierarchies. *Sci. Adv.* **10**, eadn2654 (2024).
20. Tsugawa, Y., Blumenthal, D. M., Jha, A. K., Orav, E. J. & Jena, A. B. Association between physician US News & World Report medical school ranking and patient outcomes and costs of care: observational study. *BMJ* **362**, k3640 (2018).
21. Priest, G. L. & Klein, B. The selection of disputes for litigation. *J. Legal Stud.* **13**, 1–55 (1984).
22. Shavell, S. Any frequency of plaintiff victory at trial is possible. *J. Legal Stud.* **25**, 493–501 (1996).
23. Chang, Y.-C. & Hubbard, W. New empirical tests for classic litigation selection models: evidence from a low settlement environment. *Am. Law Econ. Rev.* **23**, 348–394 (2021).
24. Ruger, T. W., Kim, P. T., Martin, A. D. & Quinn, K. M. The Supreme Court Forecasting Project: legal and political science approaches to predicting Supreme Court decisionmaking. *Columbia Law Rev.* 1150–1210 (2004).
25. Zhong, H. et al. Legal judgment prediction via topological learning. In *Proc. 2018 Conference on Empirical Methods in Natural Language Processing* 3540–3549 (Association for Computational Linguistics, 2018).
26. Medvedeva, M., Vols, M. & Wieling, M. Using machine learning to predict decisions of the European Court of Human Rights. *Artif. Intell. Law* **28**, 237–266 (2020).
27. Molot, J. T. Litigation finance: a market solution to a procedural problem. *Geo. LJ* **99**, 65 (2010).
28. Prescott, J. J. & Spier, K. E. A comprehensive theory of civil settlement. *N. Y. Univ. Law Rev.* **91**, 59 (2016).
29. Cooter, R., Marks, S. & Mnookin, R. Bargaining in the shadow of the law: a testable model of strategic behavior. *J. Legal Stud.* **11**, 225–251 (1982).
30. Agresti, A. *Categorical Data Analysis* (Wiley, 1990).
31. Coates, J. C., DeStefano, M. M., Nanda, A. & Wilkins, D. B. Hiring teams, firms, and lawyers: evidence of the evolving relationships in the corporate legal market. *Law Social Inquiry* **36**, 999–1031 (2011).
32. Page, L., Brin, S., Motwani, R. & Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*. Report No. 1999-66 (Stanford InfoLab, 1999).
33. Saaty, T. Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **1**, 83–98 (2008).
34. Glickman, M. E. Parameter estimation in large dynamic paired comparison experiments. *J. R. Stat. Soc. Ser. C* **48**, 377–394 (1999).
35. Avery, C. N., Glickman, M. E., Hoxby, C. M. & Metrick, A. A revealed preference ranking of U.S. colleges and universities. *Quart. J. Econ.* **128**, 425–467 (2013).
36. Elo, A. E. *The USCF Rating System, Chess Life* (United States Chess Federation) *XIV (13) 2* (1960).
37. Whelan, J. T. Prior distributions for the Bradley–Terry model of paired comparisons, arXiv:1712.05311 (2017).
38. Mojon, A., Mahari, R. & Lera, S. C. Data for 'Addressing information asymmetry in legal disputes through data driven law firm rankings'. *figshare* https://doi.org/10.6084/m9.figshare.30260035 (2025).
39. Mojon, A., Mahari, R. & Leram, S. C. Code for 'Addressing information asymmetry in legal disputes through data driven law firm rankings'. *Zenodo* https://doi.org/10.5281/zenodo.15504218 (2025).

Acknowledgements

We received no specific funding for this work.

Author contributions

A.M., R.M. and S.C.L. designed the study. A.M. and S.C.L. led the development of the AHPI algorithm and data analysis with input from R.M. R.M. and S.C.L. prepared the dataset on which the study is based. A.M., R.M. and S.C.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-025-00899-2.

Correspondence and requests for materials should be addressed to Sandro Claudio Lera.

Peer review information *Nature Computational Science* thanks Clement Guitton, Quinten Steenhuis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

nature portfolio

Corresponding author(s): Sandro Lera

Last updated by author(s): 10/01/2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Our data preparation and analysis uses latest standard libraries at the time of submission. Specifically, we use the following: Python (v. 3.10) packages: pandas (v.2.2.2), numpy (v. 2.1.0), scipy (v.1.11.1), scikit-learn (v. 1.5.1), rapidfuzz (v.3.9.6), and pip (v. 24.2). The code to reproduce all the analyses in this paper can be found on GitHub at https://github.com/mojona/law_firm_ranking or https://doi.org/10.5281/zenodo.15504218 In addition, we develop the AHPI algorithm to derive rankings.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source data for our work is based on R. Mahari, S. C. Lera, Early Career Citations Capture Judicial Idiosyncrasies and Predict Judgments, arXiv:2410.00725 (2024) and augmented by identifying and disambiguating law firms and their roles. The data for reproducing all the results may be accessed at 10.6084/m9.figshare.30260035.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We identify law firms and their roles in a quantitative dataset of U.S. civil litigation and present an algorithm for ranking asymmetric, heterogeneous pairwise interactions. We apply this ranking algorithm and show that it out-performs traditional reputation-based rankings in its ability to capture future law firm performance.
Research sample	310,876 U.S. civil lawsuits with a published judicial opinion contained in the Harvard Case Law Access Project (CAP). CAP includes all official, book-published federal United States case law through 2020. The research sample is based on the available data; no further sub-sampling was performed.
Sampling strategy	We run our methods on the full dataset described above and perform no further sampling.
Data collection	No new data was collected for this study.
Timing	Our data covers U.S. federal civil judicial opinions published between 03/1842 and 02/2018.
Data exclusions	We do not exclude any data from our sample.
Non-participation	No participants were involved in the study.
Randomization	Randomization is not applicable to our study as our study is predictive and not causal in nature.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging