WORKING PAPER WP-007 · CRITERICA INTELLIGENCE RESEARCH · NOVEMBER 2026

What the Record Adds to an AI Legal Platform

Criterica Intelligence Research

Corresponding: Jack Cosentino, founder · research@criterica.com

Published 8 November 2026 Updated 8 November 2026 Version 1.0

11-minute read

ABSTRACT

Legal platforms now intake, draft and manage matters well, and their users still ask two questions the software cannot answer: what is this matter worth, and how long will it take. This paper frames the answer as value of information: decision volume, times the cost of an error, times the reduction in dispersion a calibrated band provides. It describes how the record is embedded at the level of inputs and outputs, files in and bands and reports out, and what Criterica never sees or keeps.

IN ONE PARAGRAPH

Legal platforms handle intake and drafting well, and their users still ask what a matter is worth and how long it will take. The value of a calibrated band is decision volume, times the cost of an error, times the reduction in dispersion. The record is embedded as files in and bands and reports out, the same request returns the same figure, and Criterica receives only what each request sends.

SERIES

VERSION

Working Paper WP-007 1.0

PUBLISHED
8 November 2026

STATUS
Published

AUTHORS

Criterica Intelligence Research
Corresponding author: Jack Cosentino, founder

WHO THIS IS FOR AND THE DECISION IT INFORMS

Reader. Product and partnership leads at legal-technology platforms, and the law firms that use those platforms every day.

Decision. Whether to add record-based outcome, duration and settlement estimates to a platform, and what to require of the provider.

Cost of getting it wrong. Users who cannot get a defensible number inside the product take the decision elsewhere, or make it on intuition and blame the tool when it goes wrong.

01

The two questions users still ask

Legal platforms have become very good at the work around a matter. They take intake, extract facts from documents, draft letters and pleadings, manage deadlines and keep the file in order. Firms that use them work faster and with fewer errors in the paperwork.

The same users, at the same desk, still ask two questions the software does not answer: what is this matter worth, and how long will it take? They ask it at intake, when deciding whether to accept a matter. They ask it when setting settlement authority, when explaining options to a client, when deciding whether to seek funding and when planning the firm's cash. Those are the decisions that determine whether the firm makes money, and today they are made on experience, a few remembered comparables and intuition.

This paper is for the platforms that serve those users. It explains what the court record adds, how to think about its value in the platform's own terms, and how it is embedded without changing how the platform works.

02

Why drafting tools do not answer them

It is natural to ask a general-purpose model the valuation question, and many platforms already let users do so. The answers are fluent and often reasonable. They are also unstable: a sampled answer can differ each time the same question is asked, as the paper Same File, Same Number (WP-005) sets out.

A platform cannot put a number that moves in front of a client or into a settlement authority memo. The firm needs a figure that is the same every time the file is read, that comes with its uncertainty stated, and that was tested against how comparable matters actually resolved. That is a different kind of system from a drafting model, and it complements one rather than replacing it.

> Drafting tools help a firm do the work. The record helps it decide which work is worth doing, and on what terms.

03

The value of a band

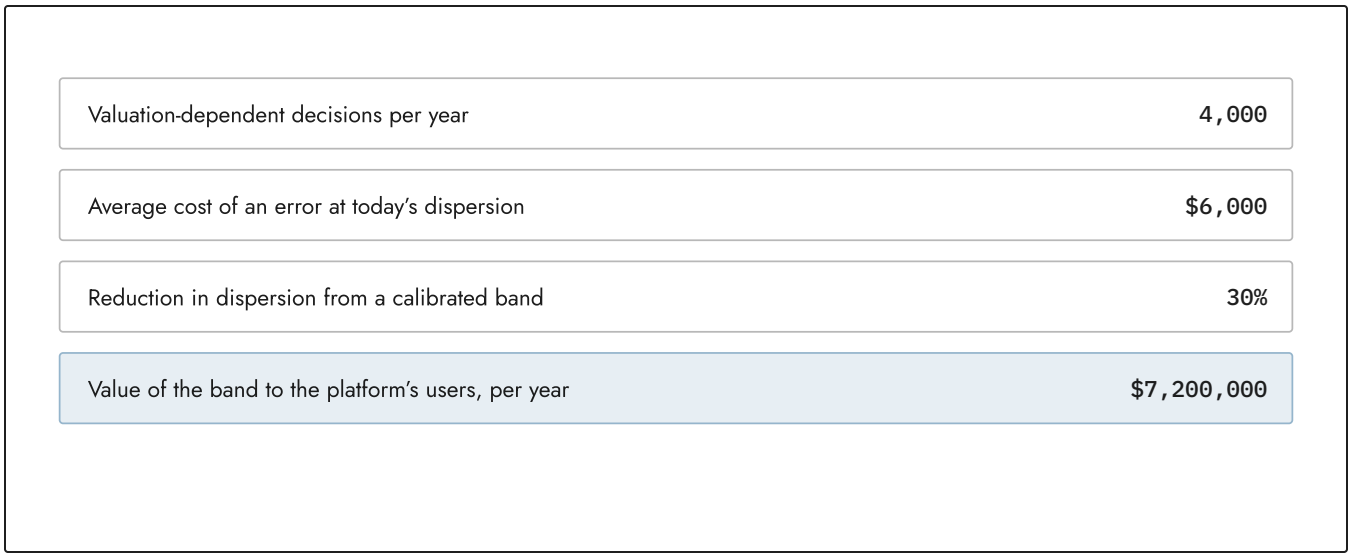
The value of adding estimates to a platform can be stated with one product of three terms.[1]

```
value of a band = decision volume × error cost × reduction in dispersion
```

- **Decision volume** is how many valuation-dependent decisions the platform's users make in a year: intake accept or decline, settlement authority, funding requests, fee arrangements.

- **Error cost** is the average cost of getting one of those decisions wrong at today's level of uncertainty: a matter accepted that loses money, a settlement taken too early or held too long, a fee priced below the work.
- **Reduction in dispersion** is how much a calibrated band narrows the uncertainty the decision is made under, compared with how it is made today.[2]

FIGURE 1 **Value of information, worked** Illustrative



Illustrative inputs for a platform whose firm users make a few thousand valuation-dependent decisions a year. The product of the three terms is the annual value of better information to those users. This figure is evidence of the arithmetic, on illustrative inputs, and not of any platform's measured result.

YOUR ASSUMPTIONS

Decisions per year 200

Average amount at stake per decision $250K

What a wrong call costs today, as a share of the amount 8%

COMPUTED FROM YOUR ASSUMPTIONS

Cost of wrong calls today, per year $4.0M

With the band, per year $3.0M

What the band is worth to you, per year

$1.0M

$5K per decision

Decisions × amount at stake × what a wrong call costs × the share a band removes. Every figure follows from your inputs. Nothing you enter is stored or sent.

Share of that error a band removes (your hypothesis)	25%

The framing matters for a platform because it locates the value with the user, where it belongs. The platform does not need to claim that its users will win more cases. It needs to show that its users make many decisions that depend on value and duration, that those decisions are expensive to get wrong, and that a calibrated band makes them less wrong. Each term can be measured for a real user base once estimates are in use.

04

Where the value lands inside a platform

MOMENT IN THE WORKFLOW	DECISION	WHAT THE RECORD ADDS
Intake	Accept, decline or refer	Outcome path and duration band for comparable matters in this venue, before time is invested
Case strategy	Where to push, when to settle	Resolution path by stage, and how duration changes after key events
Settlement	Authority and timing	Settlement range with its median, labelled by source, and the cost of waiting
Client communication	Setting expectations	A band the client can understand, stated the same way every time
Firm finance	Cash planning and funding	Book-level timing of resolutions and the spread around it

In each case the estimate sits beside the work the platform already does. A user reviewing an intake sees the band next to the extracted facts; a user preparing a settlement memo sees the range next to the draft. Nothing about the platform's existing features changes. The estimate adds a number the user can rely on at the moment a decision is made, and the platform keeps ownership of the workflow, the user relationship and the interface in which the number appears. Firms see the figure

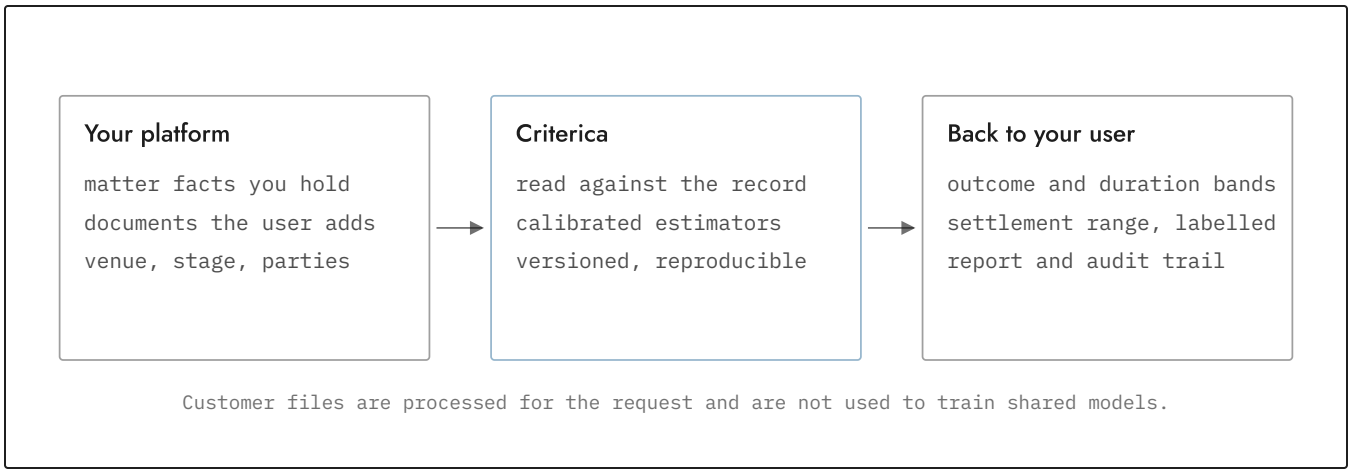
where they already work, and the platform decides how prominently to show it, to whom, and at which step of each workflow.

05

How embedding works: files in, bands and reports out

Embedding is defined entirely at the level of inputs and outputs, which keeps it simple for the platform and keeps each side's systems separate.

FIGURE 2 **Inputs and outputs**



The platform sends the facts it already holds; Criterica returns labelled estimates and reports. Each request is independent and reproducible. This figure is evidence of the interface: what goes in and what comes back.

What the platform sends

The facts the platform already holds about a matter: case type, venue, stage, key dates, parties' roles, and any documents the user chooses to include. The more of the record that is supplied, the narrower the band; a thin request still returns an answer, and an incomplete one never returns a higher figure than a complete one, as set out in WP-004.

What comes back

Outcome, duration and settlement bands with their labels: measure, population, as-of date, model and record version, and status. Where the platform wants it, a formatted report for the user and an audit trail for the firm. Because each figure carries its versions, the platform can regenerate any earlier answer exactly, which matters when a client or a court later asks how a number was reached.

How it is delivered

Through an API for platforms that render results in their own interface, and through reports for platforms that prefer to hand the user a finished document. Request and response formats are described in the developer documentation.

06

What Criterica never sees or keeps

A platform's users trust it with confidential client material, and nothing about embedding the record should change that.

- Criterica receives only what the platform sends for a request. It has no access to the platform's other data or users.
- Customer files are used for the customer's request. They are not used to train shared models without the customer's written agreement.
- No output names a counterparty, and no customer's matters are visible to any other customer.
- Retention and deletion follow the terms agreed with the platform, and the security and governance pages set out the controls.

The governance controls behind these commitments are described on the governance page.

07

Measuring the value once estimates are live

The three terms in the value-of-information product are not left as assumptions. Once estimates are in use inside a platform, each one can be measured from the platform's own activity, and the partnership is reviewed against those measurements.

- **Decision volume** is counted directly: the number of matters for which users requested an estimate at intake, at a settlement decision or at a funding request.
- **Error cost** is read from the firm's own history: matters accepted that were later written off, settlements reopened, fees that fell below the time recorded on the matter.
- **Reduction in dispersion** is read from the bands themselves and checked against resolutions as they arrive: whether outcomes landed inside the stated bands as often as the bands claimed.

Because every estimate is fixed and versioned, this review compares what each user was told on a stated date with what later happened, without relying on anyone's memory of the number. The same record lets a platform show its own customers, in its own terms, what the feature is worth to them, which is the strongest case a product team can make for it.

08

What to require of any provider

Whether a platform works with Criterica or with anyone else, the same requirements protect its users.

1. The same request returns the same figure, every time, with its version.
2. Uncertainty is shown as a calibrated band, with calibration evidence available on request.
3. Every figure states its population, date and status.
4. Removing information never raises a figure.
5. Customer data is used only for the customer's requests unless the customer agrees otherwise in writing.

A provider that meets these can sit inside a product that firms rely on for real decisions. The test in section 6 of WP-005 checks the first requirement in under an hour.

09

The question, as the reader has it

A product lead at a legal platform asks a practical question: if we add a valuation and duration estimate to the product, what changes for the people who use it, and how would we know? The firm user asks a related one: can I put this number in front of a client or a funder? A point estimate fails both readers. It gives the product lead nothing to measure, because a single number cannot be scored for how much uncertainty it removed, and it gives the firm user nothing to defend, because a single number does not say how far it could be wrong.

A band answers both. It can be scored against outcomes as they arrive, and it tells the user in the same view how much weight the number will bear.

10

What is measured

QUANTITY	UNIT	POPULATION
Time to resolution	Months, as a median with P25 to P75 and P10 to P90 spans	Closed matters of the same case type, venue and stage, from filed court records
Settlement band	US dollars, a range with its median	Closed matters with a recorded amount, adjusted for the facts supplied
Probability of an event	A number from 0 to 1 over a stated horizon	Closed matters that could have had the event
Value of the band to a platform	US dollars per year, from the three-term product	The platform's own users and their own decision history

The first three rows are statements about the court record and carry their population and as-of date. The fourth is a statement about a platform and is measured from that platform's own activity once estimates are in use; this paper shows the arithmetic with illustrative inputs only. Matters enter a comparison population only when the record shows how they ended.

11

How a number travels from the record to the platform

1. **Platform facts to structured event.** Case type, venue, stage, key dates and any documents the user chooses to include become dated events. Parties and courts are resolved against a knowledge graph, so one actor is one node.
2. **Event to model class.** Duration uses survival analysis, in the Kaplan-Meier and proportional-hazards tradition, because many comparable matters are still open. Event probabilities use calibrated probabilistic classification. Sparse venues use hierarchical partial pooling.
3. **Model class to calibrated output.** Scores are recalibrated on held-out matters, and bands are fitted as quantile or conformal intervals so their stated coverage can be tested.
4. **Output to the platform's screen.** The band returns with its labels and versions. The same request returns the same figure, so the platform can show it the same way every time and regenerate any earlier answer.

Value of information is the lens for why this path is worth embedding.[1] The expected gain from a piece of information is the difference between the best decision made with it and the best decision made without it. A band that is calibrated lowers the error of the decision; a band that is narrow but not calibrated raises it, which is why calibration, and not width, is what is validated first.

12

Validation design

Two things are validated, and they are validated at different times. The estimates are validated before release: models are frozen at a time cut and scored on matters filed after it, with the point-in-time boundary applied to every input so nothing known only later can inform an earlier estimate. The metric families are discrimination, calibration (reliability diagrams and Brier-type scores) and interval coverage. TN-003 explains how to read them.

The value to the platform is validated after launch, from the platform's own record. Because each estimate is fixed, dated and versioned, the platform can compare what its users were told on a given date with what later happened, and check whether outcomes landed inside the stated bands as often as the bands claimed. Figure 1 is evidence of the arithmetic of the three-term product on illustrative inputs; Figure 2 is evidence of the interface: what goes in, what comes back, and that each request stands alone.

13

Scope and precision

The record adds a band for outcome, duration and settlement. It does not add a prediction of the realized dollars of a particular matter, and it does not state which party prevails from a disposition code; direction is read only from sources that record it (WP-006). The value-of-information figure is illustrative until it is measured on a real user base.

- Data lag: the record reflects filings up to its as-of date, and the estimate is re-issued when the record moves.
- Censoring: open matters are not outcomes; survival methods account for them.
- Venue sparsity: where comparable matters are few, the band widens and the status says so.

14

What a reader can reproduce

The value-of-information example is a product of three terms and can be recomputed by hand, or with the calculator above, using the reader's own volume, error cost and reduction in dispersion. A platform can test the stability of any provider with the repeat-request protocol in WP-005. Reproducing a matter-level band requires the platform, because it depends on the record version and model version returned with the figure; with those two, any earlier answer can be regenerated exactly.

What this number means

The value-of-information example in this paper uses illustrative inputs, not measurements from any platform or customer. The value a real platform captures depends on its users' decision volume, the cost of their errors and how much a calibrated band narrows them, all of which can be measured for that platform once estimates are in use. Estimates are bands read from the court record with their scope labelled, as set out in WP-006.

REFERENCES

1. Howard, R. A. (1966). Information value theory. IEEE Transactions on Systems Science and Cybernetics, 2(1), 22–26. doi.org/10.1109/TSSC.1966.300074
2. Raiffa, H., & Schlaifer, R. (1961). Applied Statistical Decision Theory. Division of Research, Harvard Business School.
3. Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise: A Flaw in Human Judgment. Little, Brown Spark.
4. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457–481. doi.org/10.1080/01621459.1958.10501452
5. Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B, 34(2), 187–220.
6. Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
7. Fellegi, I. P., & Sunter, A. B. (1969). A theory of record linkage. Journal of the American Statistical Association, 64(328), 1183–1210. doi.org/10.1080/01621459.1969.10501049
8. Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50. doi.org/10.2307/1913643

9. Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: a gentle introduction. Foundations and Trends in Machine Learning, 16(4), 494–591. arxiv.org/abs/2107.07511

10. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1), 1–3. doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

11. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378. doi.org/10.1198/016214506000001437

12. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. doi.org/10.1145/3287560.3287596

NOTES

1. Value of information is a standard construct in decision analysis: the expected improvement in a decision's outcome from having a piece of information before making it.

2. Dispersion here means the spread of the estimate the decision is made on. Reducing it is only valuable if the narrower band is also calibrated; a narrow band that is wrong increases error cost.

Prepared by Criterica Intelligence Research with data engineering, validation and editorial review. Criterica Intelligence Research publishes numbered documents in five series; each is versioned, dated and issued as a new version when it changes, with the change recorded.

CITE THIS DOCUMENT

Criterica Intelligence Research. (2026). What the Record Adds to an AI Legal Platform (Working Paper WP-007, version 1.0). Criterica. https://critericaintelligence.com/research/what-the-record-adds

```
@techreport{criterica_wp007_2026,
  author = {{Criterica Intelligence Research}},
  title = {What the Record Adds to an AI Legal Platform},
  institution = {Criterica},
  type = {Working Paper},
  number = {WP-007},
  year = {2026},
  month = nov,
  note = {Version 1.0},
  url = {https://critericaintelligence.com/research/what-the-record-adds}
}
```

DOI: pending