Model Cards: What Each Card Contains, with a Sample

Criterica Intelligence Research

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

Published 8 November 2026 Updated 8 November 2026 Version 1.0 8-minute read

ABSTRACT

Every Criterica production model carries a card: what it estimates, for which population and horizon, how it was validated, the gates it cleared, its version and the policy for retiring it. This document sets out the fields a card contains and publishes one sample card built from fleet-level figures cleared for the research series, so a reviewer knows what to ask for and what they will receive under NDA.

IN ONE PARAGRAPH

A model card is the standing answer to what a model estimates, for which population and horizon, how it was tested and when it would be withdrawn. Every Criterica production model carries one in a fixed twelve-field format, validated on a later time period than it trained on and versioned for life. This document sets out the fields and publishes a sample card built from fleet-level figures.

SERIES	VERSION
Model Card MC-002	1.0
PUBLISHED	STATUS

8 November 2026 Published

AUTHORS

Criterica Intelligence Research

Corresponding author: Jack Cosentino, founder

WHO THIS IS FOR AND THE DECISION IT INFORMS

Reader. Model-risk reviewers, courts and regulators, platform partners, buyers and insurers who need to know what stands behind a figure.

Decision. Whether the models behind a report are documented well enough to rely on, and what to request for diligence.

Cost of getting it wrong. A figure from an undocumented model cannot be reviewed, challenged or defended, and every decision built on it inherits that gap.

01

Why every model carries a card

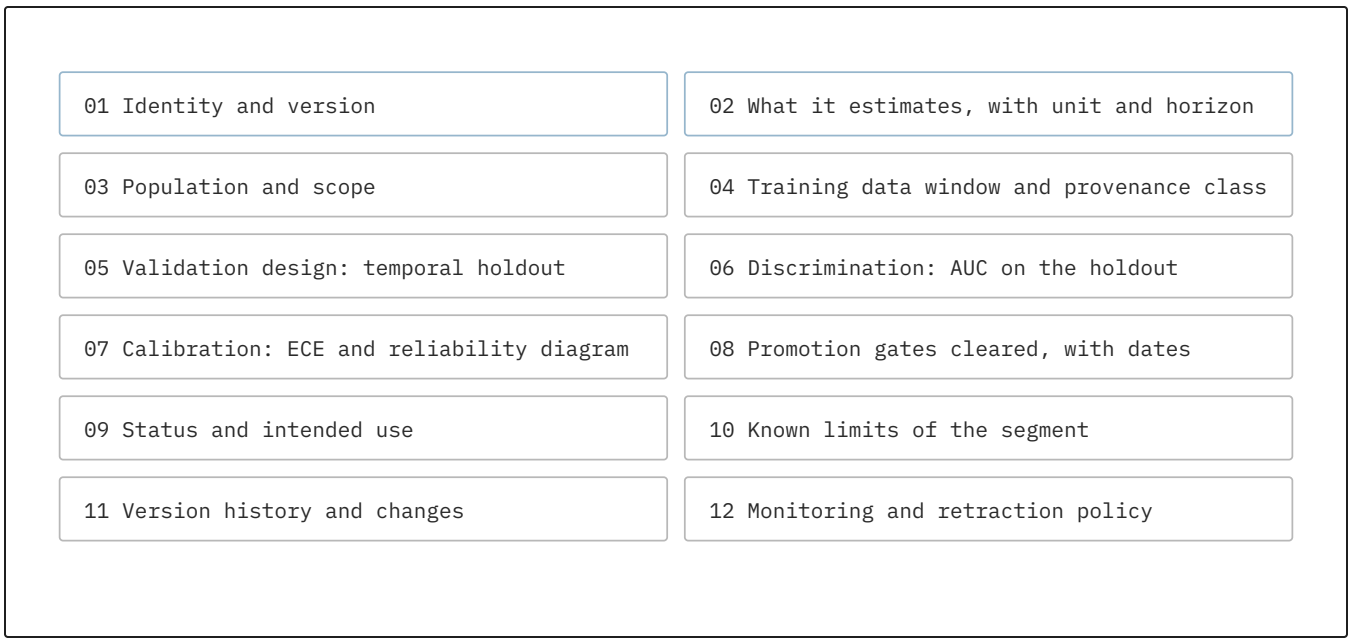
A reviewer looking at a Criterica report sees figures: a duration band, a resolution probability, a settlement range. Behind each figure is a model, and the reviewer's question is always the same: what is this model, what was it tested on, and how did it do? The model card is the standing answer. It is written when a model is promoted, updated whenever the model changes, and kept with every earlier version.

Cards follow one format across the production predictive fleet.[2] The card is designed for the person who has to sign off on reliance: a model-risk officer, an investment committee, a court-appointed expert or a regulator.

02

What a card contains

FIGURE 1 **The twelve fields of a model card**



Fields one and two identify the model and what it estimates; the rest record how it was built, tested, governed and changed. This figure is evidence of what a reviewer will find on every card.

FIELD	WHAT THE REVIEWER LEARNS
Identity and version	A stable identifier and version, so any figure can be traced to the exact model that produced it.
Estimate	The measure, its unit and its horizon, for example probability of resolution before trial within 24 months.
Population and scope	The case type, venue and stage the model is valid for, and the populations it is not used for.
Training window and provenance	The filing period the model learned from and the class of source.
Validation design	The time cut and the later period the model was scored on.
Discrimination	AUC on that later period.
Calibration	ECE and the reliability diagram on that later period.
Gates	Each promotion gate, the result and the date it was cleared.
Status and intended use	Production or otherwise, and the decisions the model is meant to inform.
Known limits	Thin sub-segments, recent regime changes and other conditions a reader should weigh.
Version history	Every earlier version, with what changed and why.
Monitoring and retraction	How the model is watched and the conditions under which it is withdrawn.

The card states results and design. It does not publish the model's internal construction; that is reviewed under NDA where a reviewer needs it. TN-003 explains how to read the calibration fields.[1]

03

Sample card

The sample below follows the card format exactly. In place of a single model's results it carries fleet-level figures cleared for publication, so the format can be shown in full.

SAMPLE MODEL CARD, FLEET-LEVEL FIGURES

FIELD	SAMPLE ENTRY
Identity and version	SAMPLE-FLEET, card format version 1.0
Estimate	Binary outcome probabilities for production predictive models across the fleet
Population and scope	Production predictive fleet; each model valid only for its own case type, venue and stage
Training window and provenance	Real filed records; each model trained on matters filed before its time cut
Validation design	Temporal holdout: scored on matters filed after the time cut
Discrimination	Median AUC 0.739; upper-quartile AUC 0.853 (July 2026 registry extract)
Calibration	Every production classifier clears ECE ≤ 0.05 on its holdout
Gates	AUC ≥ 0.70 and ECE ≤ 0.05 on the temporal holdout; both required
Status and intended use	Production; informs pricing, reserving and strategy alongside duration and settlement bands
Known limits	Segments without enough resolved matters are held at stub status and not issued as production figures
Version history	Each model version kept; changes recorded in the public changelog
Monitoring and retraction	Monitored against new resolutions; withdrawn to experimental status on the record if it stops clearing its gates

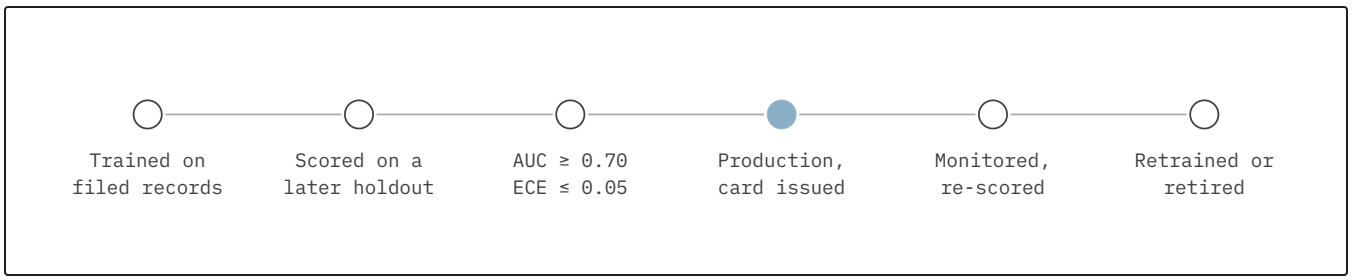
Reading the sample

A median AUC of 0.739 means that, for the typical production model, a randomly chosen matter with the outcome is ranked above a randomly chosen matter without it about 74% of the time, on matters from a period the model never saw. A quarter of the fleet does better than 0.853. Every model in the fleet has also cleared the 0.70 floor and the 0.05 calibration gate; the card for each one states its own figures.

04

Versioning and retraction

FIGURE 2 **The life of a model**



A card is issued at promotion and updated at every later step. Retirement is recorded, not deleted. This figure is evidence that a card follows the model through every later step.

- A retrained model is a new version with its own card. The earlier card is kept, so a figure issued under it can still be traced.
- A model that stops clearing its gates on new resolutions is withdrawn from production and the withdrawal is recorded with its reason.
- Every promotion, demotion and retirement appears in the public changelog maintained with the model registry (MC-001).

Corrections are treated the same way as improvements: recorded with a date and a reason. A registry that only ever shows additions would not be telling the reader what changed. See the model registry and changelog.

05

How to request cards

Institutional reviewers receive the cards for every model that contributes to a report they are relying on, under NDA, as part of the method pack. A request names the report or the matters concerned; the research team returns the cards with their version history.

06

The question, as the reader has it

A model-risk officer, an investment committee member or a court-appointed expert is asked to rely on a figure from a model they did not build. Their problem is the standard one in model governance: what is this model for, what population does it describe, how was it tested, and under what conditions would it be withdrawn? A single headline accuracy number cannot answer that. It is silent on population, on the period it was tested over and on calibration, so it supports reliance for no particular decision.

A card answers it in a fixed order, the same for every model, so reviewers can compare cards and a missing field is visible as a gap.

07

What a card measures

QUANTITY	UNIT	POPULATION
Discrimination	Area under the receiver operating characteristic curve, between 0 and 1	Matters filed after the time cut, scored by a model trained only on matters filed before it
Calibration	Expected calibration error and a reliability diagram	The same later-period matters
Interval coverage	Share of outcomes inside the stated band, for models that state one	The same later-period matters
Estimate	Probability over a stated horizon, a duration in months or a range in US dollars	Closed matters of one case type, venue and stage, from filed court records with a recorded termination

Inclusion is stated on the card at the level of a recorded termination: only matters whose record shows how they ended contribute to an observed outcome. The card names the case type, venue and stage a model is valid for and, as importantly, the ones it is not.

08

How a number travels, and what the card records at each step

1. **Record to structured event.** The card states the class of source the model learned from, and the window, so the reviewer can match the model to the population it is used for. Entities are resolved on a knowledge graph before any model sees them.
2. **Event to model class.** The card names the class: survival analysis for durations (Kaplan-Meier and proportional-hazards framing), calibrated probabilistic classification for event probabilities, and hierarchical partial pooling where a venue is sparse. It states why that class fits the estimate.
3. **Model class to calibrated output.** The card records that scores were recalibrated on held-out matters and that bands are quantile or conformal intervals with a tested coverage.
4. **Output to band shown.** The card ties the band a reader sees to the version that produced it, so any figure can be traced to its card.

The card publishes design and results. It does not publish a model's internal construction, which a reviewer sees under NDA when it is needed.

09

Validation design

Every card describes the same design. What is frozen: the model and its recalibration at a time cut. What is compared: each stated probability or band with how the matter closed in the later period. How leakage is prevented: the point-in-time boundary applies to every input, so a field recorded

after the cut is not visible to an earlier estimate. The metric families are discrimination, calibration and interval coverage, each reported with the date of the holdout beside it.

The public promotion gates are stated in the sample card and in TN-003. Figure 1 is evidence of what a reviewer will find on a card: twelve fields in a fixed order. Figure 2 is evidence of the card's place in the life of a model: issued at promotion, updated at every later step, and retained at retirement.

10

Scope and precision

A card describes a model that estimates what the court record supports: probabilities, durations and ranges for comparable closed matters. It does not describe a prediction of the realized dollars of one matter, and none of the cards states who prevails from a procedural disposition code.

- Data lag: the card states the date of its record and is updated as new resolutions arrive.
- Censoring: open matters are not outcomes; the duration models account for them.
- Venue sparsity: a segment without enough resolved matters is held at stub status and the card says so.

The sample card reports fleet-level figures, which describe the fleet as a whole and not any one model.

11

What a reader can reproduce

The card format is public, so a reader can check any report against it: a card is complete when every field is present. The reading of the fleet-level sample, what a median AUC means and how to interpret a gate, can be reproduced from the definitions in TN-003. The results for a specific model can be reproduced by an institutional reviewer from the held-out evaluation described on its card,

which is supplied under NDA with the version history, because it depends on matters held on the platform.

What this number means

The sample card reports fleet-level figures from the July 2026 registry extract cleared for the research series: across the production predictive fleet, median AUC 0.739 and upper-quartile AUC 0.853 on temporal holdouts. They describe the fleet as a whole, not any one model. A card for a specific model states that model's own results and is shared with institutional reviewers under NDA.

REFERENCES

1. 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
2. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. doi.org/10.1145/3458723
3. Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency (2011). Supervisory Guidance on Model Risk Management (SR 11-7 / OCC 2011-12). www.federalreserve.gov/supervisionreg/srletters/sr1107.htm
4. 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
5. 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
6. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning, PMLR 70, 1321–1330. arxiv.org/abs/1706.04599
7. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning, 625–632. doi.org/10.1145/1102351.1102430
8. 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

9. Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society, Series B, 34(2), 187–220.
10. Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.
11. 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
12. Romano, Y., Patterson, E., & Candès, E. J. (2019). Conformalized quantile regression. Advances in Neural Information Processing Systems, 32. arxiv.org/abs/1905.03222

NOTES

1. Provenance class records the kind of source a model was trained on, such as federal court records or agency decisions, so a reviewer can match the model to the population it is used for.
2. The registry counts 10,470 production predictive models. Governance artifacts and reference tables are recorded separately and do not carry predictive cards.

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). Model Cards: What Each Card Contains, with a Sample (Model Card MC-002, version 1.0). Criterica. https://critericaintelligence.com/research/model-cards

```
@misc{criterica_mc002_2026,
  author = {{Criterica Intelligence Research}},
  title = {Model Cards: What Each Card Contains, with a Sample},
  publisher = {Criterica},
  howpublished = {Model Card},
  number = {MC-002},
  year = {2026},
  month = nov,
  note = {Version 1.0},
  url = {https://critericaintelligence.com/research/model-cards}
}
```

DOI: pending