EVALUATION PLAN

Tesla Automotive

Evaluation plan—resolution specifications

Paul Campillo

Version 2.1 · March 2026 · Pre-registration

Purpose

This document specifies exact resolution criteria, data sources, scoring methods, and publication commitments for all predictions in the Tesla Genesis forecast. Designed for pre-registration and independent verification. The forecast is hash-locked in the Seer registry as SEER-2026-0003.

Prediction resolution specifications

P1: No strategic recalibration (85%, resolves September 2026)

Resolves TRUE if: Through September 30, 2026, Tesla leadership does not publicly acknowledge brand damage as a strategic problem requiring organizational response AND does not revise forward growth targets downward to align with actual delivery trajectory.

Resolves FALSE if any of the following occur:

- Public statement (earnings call, shareholder letter, SEC filing, or official Tesla communication) acknowledges brand loyalty decline, political polarization impact, or identity-based buyer loss as a problem requiring strategic response
- Forward delivery or growth guidance revised downward by ≥15% from previously stated targets
- Organizational brand recovery initiative announced with specific actions (not general messaging)

Data sources: Tesla quarterly earnings call transcripts, shareholder communications, SEC filings (10-Q, 10-K, 8-K), official Tesla blog posts, Musk's public statements on X/Twitter and at public events.

What doesn't count: Analyst interpretation of Tesla's strategy; third-party speculation; pricing adjustments (which are tactical, not strategic recalibration); vague references to "market challenges" without specific acknowledgment of brand/identity damage.

Check date: First week of October 2026.

P2: Third consecutive year of delivery decline (60%, resolves December 2026)

Resolves TRUE if: Tesla's total global vehicle deliveries for calendar year 2026 are lower than 2025 deliveries (1.64M units as reported by Tesla).

Resolves FALSE if: 2026 total deliveries exceed 1.64M units.

Data source: Tesla's official delivery and production report, typically released in the first week of January following the calendar year. If Tesla changes its reporting methodology (e.g., combining vehicle categories differently), resolution uses the most comparable figure to prior years' reporting.

Note: Deliveries, not production. If Tesla produces more than it delivers (inventory build), production numbers don't count toward resolution.

Check date: First week of January 2027 (when Tesla releases Q4/annual delivery data).

P3: European market share falls below 7% (70%, resolves September 2026)

Resolves TRUE if: Tesla's share of the European passenger EV market falls below 7.0% in any single quarter of H1 or Q3 2026 (Q1, Q2, or Q3), as measured by European Automobile Manufacturers' Association (ACEA) registration data or equivalent national registration databases.

Resolves FALSE if: Tesla's European EV market share remains at or above 7.0% in every quarter through Q3 2026.

Data sources: ACEA quarterly registration data; national registration databases (KBA Germany, SMMT UK, CCFA France); industry aggregators (EV-Volumes, CleanTechnica) used for cross-reference only—ACEA is the primary source.

Definition: "European market" includes EU27 + UK + Norway + Switzerland + Iceland—the standard European EV market definition used by ACEA.

Check date: October 2026 (when Q3 registration data publishes).

P4: Brand loyalty stays below 55% (80%, resolves September 2026)

Resolves TRUE if: S&P Global Mobility's Tesla repeat purchase rate (also called "loyalty rate" or "return to make rate") remains below 55.0% in their most recent measurement published before October 1, 2026.

Resolves FALSE if: The S&P Global Mobility measurement shows Tesla repeat purchase rate at or above 55.0%.

Data source: S&P Global Mobility (formerly IHS Markit) loyalty reporting. This is the industry-standard source used by automotive analysts and cited in the Tesla diagnostic (baseline: 73% declining to 49.9%).

Fallback: If S&P Global changes methodology or discontinues this specific metric, the closest equivalent loyalty/repeat-purchase metric from S&P Global or JD Power serves as the resolution source, documented at time of scoring.

Check date: October 2026 (using most recent S&P Global release).

P5: No FSD L3+ approval in any US jurisdiction (85%, resolves December 2026)

Resolves TRUE if: Through December 31, 2026, no US state or federal regulatory body grants Tesla approval, exemption, or permit for unsupervised (SAE Level 3 or above) autonomous vehicle operation on public roads.

Resolves FALSE if any of the following occur:

- NHTSA grants a federal exemption for unsupervised Tesla autonomous operation
- Any US state grants Tesla a permit for driverless commercial or consumer operation without a safety monitor
- Tesla receives regulatory clearance to remove the steering wheel from vehicles operated on public roads

What doesn't count: FSD Beta expansions that remain classified as Level 2 (driver-supervised); Tesla's own claims about capability level; permits for

supervised testing; permits granted to other companies (e.g., Waymo, Cruise).

Data sources: NHTSA Federal Register notices; California DMV autonomous vehicle permit database; state-level DMV and transportation department records; Tesla regulatory filings.

Check date: First week of January 2027.

P6: Cybercab production below 25K/quarter (70%, resolves December 2026)

Resolves TRUE if: Tesla does not produce more than 25,000 Cybercab units in any single quarter of calendar year 2026.

Resolves FALSE if: Tesla reports or independently verifiable data confirms Cybercab production exceeding 25,000 units in any quarter of 2026.

Data sources: Tesla quarterly earnings reports and production disclosures; Tesla investor presentations; SEC filings (10-Q, 10-K). If Tesla does not break out Cybercab production separately, resolution uses the best available estimate from earnings call commentary, analyst briefings, or third-party production tracking (e.g., registration data, VIN analysis).

What doesn't count: Announced production capacity or targets; pre-production or test units; production of other Tesla vehicle models.

Note: If Cybercab launch is delayed beyond 2026 entirely, P6 resolves TRUE by default (zero production < 25K). If Cybercab launches under a different product name but matches the announced robotaxi form factor (no steering wheel, autonomous-only), it counts as Cybercab for resolution purposes.

Check date: First week of January 2027.

P7: Automotive gross margin ex-credits stays below 18% (75%, resolves December 2026)

Resolves TRUE if: Tesla's automotive gross margin excluding regulatory credit revenue remains below 18.0% in every reported quarter through Q3 2026 (Q1, Q2, Q3).

Resolves FALSE if: Automotive gross margin excluding regulatory credits reaches or exceeds 18.0% in any quarter through Q3 2026.

Data sources: Tesla quarterly earnings releases and 10-Q filings. Tesla reports automotive gross margin and regulatory credit revenue separately, allowing direct calculation: (Automotive gross profit – regulatory credit revenue) / (Automotive revenue – regulatory credit revenue).

Note: If Tesla changes its segment reporting in a way that makes automotive-only margin calculation impossible, resolution uses the closest equivalent metric, documented at time of scoring. The baseline trajectory is 28.5% (2022) ⟶ 15.4% (2025), with Q1 2025 bottoming at 2.1%.

Check date: October 2026 (when Q3 10-Q is filed); confirmed January 2027.

Scoring methodology

Brier score

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

Where p_i is the predicted probability and o_i is the outcome (1 if true, 0 if false).

PREDICTION	IF TRUE	IF FALSE
P1 (85%)	0.0225	0.7225
P2 (60%)	0.1600	0.3600
P3 (70%)	0.0900	0.4900
P4 (80%)	0.0400	0.6400
P5 (85%)	0.0225	0.7225
P6 (70%)	0.0900	0.4900
P7 (75%)	0.0625	0.5625

Aggregate Brier score reported as mean across all seven predictions. Perfect calibration = 0.0. Random guessing at 50% = 0.25.

Logarithmic scoring rule

$$\text{LS} = \frac{1}{N}\sum_{i=1}^{N}[o_i \cdot \ln(p_i) + (1 - o_i) \cdot \ln(1 - p_i)]$$

Log score penalizes overconfident wrong predictions more severely than Brier. Reported alongside Brier for completeness.

Calibration assessment

With seven predictions, formal calibration curves aren't meaningful. Instead, we report:

- Individual prediction outcomes and scores
- Aggregate Brier and Log scores

- Qualitative calibration assessment: were the 85% predictions more likely to resolve TRUE than the 60% prediction?
- Comparison to Version 1 prediction scores (same resolution criteria applied to January 2026 predictions)

Publication commitments

COMMITMENT	DETAIL
Pre-registration	This evaluation plan published before any resolution data arrives; forecast hash-locked as SEER-2026-0003
No revision	Predictions, probabilities, and resolution criteria fixed as of publication date
Full reporting	All seven predictions scored and published regardless of outcome
Transparent misses	Any prediction that resolves contrary to forecast receives explicit analysis of what the model got wrong
Falsification engagement	Any triggered falsification criterion receives analysis of which part of the framework needs revision
Public access	Scored results published at genesistheory.org and on the Genesis Substack
Version comparison	Version 1 and Version 2.1 predictions both scored, with comparison analysis
Registry verification	Anyone can verify the forecast was locked before resolution by checking SEER-2026-0003.json hashes against the specification, evidence, and forecast JSON

Revision policy

Predictions in this document are fixed and won't be updated. If material new evidence arrives before resolution (e.g., a major regulatory action, leadership change, or financial shock), we'll publish a supplementary assessment noting the new evidence and its expected impact on the forecast—but the original predictions and probabilities remain locked for scoring purposes.

Monthly monitoring notes may be published between now and resolution. These are observational, not revisions to the predictions.

Protocol: Genesis Prediction Protocol v2.1 | Diagnostic Protocol v3.3 | **Evidence:** Public filings, market data, NBER study, court rulings, media analysis through January 2026; margin analysis updated March 2026

Forecast registered: SEER-2026-0003 · github.com/wakeupsmiling/seer · Pre-registration date: March 2026. First resolution: October 2026. Full resolution: January 2027.