Our EA Testing Methodology

The full technical process behind every review on The Robotic Trader.

Version 2.5 · 28 October 2026 · Supersedes Version 2.4 (19 October 2026)

This document is published openly, with no email required, because we believe a testing methodology only means something if you can actually see it. It is the same standard we hold every Expert Advisor to before it appears anywhere on this site — including the ones that fail.

1. Scope: Which Tests This Governs

Every EA under review first goes through our standard Red-Flag Checklist. This regime is a separate, additional layer. There are two routes.

Route 1 — live record review:
We choose which EAs to review by what readers ask about. To be reviewed at all, an EA needs a real-money live record of at least 12 months and 300 closed trades attributable to it, and every criterion in this document must be assessable from it. Under 24 months it is an Early Review and cannot be Recommended. We run the full checklist against the record; no backtest is required. The record does not have to be third-party verified for us to review it. It does have to be verified for a Recommended verdict: monitored by Myfxbook, FXBlue, Show My Trades, or a linked MQL5 Signal monitored by MQL5's own servers in real time. A broker investor-password login on its own does not count as verification, because it is not a public record that readers can check. Every eligible EA we choose is reviewed and published, whatever the verdict.

Two review tiers:
a Full Review needs a real-money live record of at least 24 months and 300 closed trades, and any of the three verdicts is possible. An Early Review needs at least 12 months and 300 closed trades; the verdict can be Inconclusive or Not Recommended, never Recommended, because a year of trading can miss the market conditions that break a strategy. Every other rule applies to both, so a disqualifying problem such as positions added to a losing trade can end an Early Review in Not Recommended. An Early Review is labelled on the review itself, on Best Expert Advisors and in every summary we publish. Each quarter we check every Early Review's record; once it reaches 24 months, we assess it against the full record on the same page and remove the label. Below 12 months or 300 closed trades, we do not review an EA at all. Reviews published before these minimums were adopted stay published as dated records; they are not Early or Full Reviews, and when a verdict rule changes, their verdicts change at once.

Grid and martingale:
an EA whose trade history shows a martingale, grid, averaging or recovery strategy is Not Recommended, disclosed or not, whatever its profit. Where no trade history can be read, the developer's own description of such a mode as the default decides the verdict for the product as sold.

Backtest-only EAs are not reviewed:
We do not review an EA whose only evidence is a backtest, however good. Backtests are used only in rule-conformance tests (Route 2), never as the basis of a review.

Route 2 — rule-conformance tests:
Tests asking whether a specific configuration survives a specific external rule set — for example, whether a developer-supplied prop-firm settings file stays inside that firm's daily and overall loss limits. A conformance test is not a review and produces no verdict on the EA. Section 7 governs it.

Route 2 can apply even to an EA that is reviewed under Route 1: a live signal for the standard configuration tells you nothing about a different configuration.

Re-verification:
every published review is re-verified each quarter, and sooner after a material event, such as a vendor changing its settings or a monitored account being deleted or made private. Changes are dated on the review.

2. Environment Requirements

These apply to every test, at every tier. A run outside them is not admissible evidence in any form.

Realistic spread modelling:
variable spread data matching real market conditions, never a fixed, artificially tight spread.

Realistic slippage and commission:
typical execution costs for the broker and account type the EA is designed for, not a zero-cost idealised fill. Where a test models a prop firm account, it runs on that firm's own server data, so the spreads and commission are the firm's, not a stand-in broker's. Every report must show the commission charged; if a server ever charges none in the Strategy Tester, we add it afterwards from the report's deal list at the firm's rate. Every figure we publish includes it, and we state the total.

Leverage matched to real conditions:
and where the test models an external rule set, matched to that rule set's own cap. Testing at higher leverage removes the margin constraint the rules impose and invalidates the comparison.

Data from the broker the EA is intended to run on:
third-party tick data from another venue carries a different spread profile, and may be used only as a Tier 2 run, labelled as such.

Every setting documented:
broker, server build, symbol data source, spread model, commission, leverage, starting balance, account currency, test period and History Quality percentage — all recorded so the test is independently reproducible. The exact parameter file is archived with the result.

3. Test Windows: The Two-Tier Structure

Data quality and coverage trade off directly against each other on any retail broker. Quality asks whether the numbers are true; coverage asks whether they are representative. Both cannot be maximised, and this standard does not pretend otherwise.

Tier 1 — the evidence:
the maximum available real-tick history from the intended broker, whatever length that turns out to be, at 99% or above real ticks. No exceptions. Every figure we publish — return, drawdown, trade count, win rate, profit factor and any verdict — comes from Tier 1.

Tier 2 — the coverage run:
the longest run the available data allows, at whatever quality that is, stated as a percentage every time it is referenced. It answers one question only: does the strategy break under market conditions Tier 1 does not contain? It is never a source of published figures.

Why a lower-quality run still tells us something — in one direction only

Generated ticks understate the drawdown of grid, martingale and basket strategies, because they do not reproduce spread widening, gaps or stress-period slippage.

So a Tier 2 run that shows trouble is meaningful: if a strategy breaks on flattering data, it breaks. A Tier 2 run that shows no trouble proves nothing, and we never cite it as evidence of robustness.

Coverage, stated in events rather than years:
the combined windows must include at least one documented high-volatility episode relevant to the instruments traded, named with its dates; more than one market condition, not a single regime throughout; and a period the seller's own marketing never showed. Ideally more than one interest-rate regime, since rate cycles materially change currency pair behaviour. Where the available data cannot span one, we say so.

The minimum that cannot be waived:
Tier 1 must produce at least 100 closed trades on its own. A window too short for that is not a test, and no amount of Tier 2 coverage repairs it.

4. Pass / Fail Criteria

Objective thresholds a backtest must clear, assessed on Tier 1 figures, regardless of headline returns. A backtest is never the basis of a review: we do not review an EA whose only evidence is a backtest (Section 1), and clearing every threshold here cannot earn a “Recommended” verdict.

Minimum 100 closed trades:
before any verdict can be drawn — fewer is insufficient evidence either way.

Drawdown judged in context:
no universal number, but a drawdown materially larger than the seller’s own marketing implies is an automatic fail.

No catastrophic stress-period losses:
poor performance during stress is not automatically disqualifying — a wipeout or near-wipeout is.

Consistency out of sample:
a sharp, total breakdown outside the seller’s own window is a strong indicator of curve-fitting.

Low parameter sensitivity:
small changes to inputs should not produce wildly different outcomes.

No open drawdown held beyond 90 days:
a floating loss that is still open and unrecovered 90 days after it began, at any point in the record, rules out a Recommended verdict — even if it later recovers, because anyone running the EA would have had to sit through it. The same rule applies to live records, and where a record cannot show it, a Recommended verdict cannot rest on it.

5. Red Flags Specific to Backtesting

Suspiciously smooth equity curves:
with no visible drawdown — a strong sign of curve-fitting, or a hidden martingale or grid mechanism only visible over a longer window.

Martingale or grid masked by a short window:
these strategies can look flawless for a long stretch before a single catastrophic loss.

Performance concentrated in a handful of trades:
if removing a few exceptional trades collapses profitability, the strategy is not robust.

Unrealistic win rate with poor risk-reward:
very high win rates often signal tail risk the win rate alone conceals. We check the average win against the average loss — a high win rate with an average loss larger than the average win is a grid signature, not an edge.

A headline figure carried by the worst data:
if the return across a long run is strong but flat or negative inside the high-quality window, the strong figure is an artefact of generated ticks and we do not publish it.

A result that improves as data quality falls:
if drawdown is lower on the low-quality run than the high-quality one, the low-quality run is being flattered.

6. What We Publish With Every Test

So a reader can independently verify our conclusion rather than take it on trust:

- The full equity curve from the Tier 1 run.
- The tier, the History Quality percentage and the exact window — stated on the page, not in a footnote.
- Maximum drawdown, stated plainly, with the period it covers.
- The maximum daily drawdown as a series, wherever the test is measured against a daily limit. A single headline drawdown figure cannot answer a daily-limit question.
- Total trades, win rate, profit factor, average win and average loss.
- The longest time any trade or basket stayed open, and the longest open, unrecovered floating drawdown.
- All environment settings from Section 2.
- Where Tier 1 does not meet the coverage requirement, an explicit statement of what it does not cover and why.
- The conformance result, tied directly to the criterion passed or failed.

Where any figure draws on a longer, lower-quality run, we say so explicitly and state the quality percentage of both windows.

7. Rule-Conformance Tests

A conformance test asks whether a named configuration survives a named external rule set — at present, FTMO's published challenge rules. It produces a conformance result (Pass, Fail or Inconclusive), never a verdict on the Expert Advisor.

We establish the rules first, from the firm's own published terms:
never from an aggregator or affiliate summary, and always dated. Every limit the configuration could breach is recorded with its measurement basis — balance or equity, whether floating losses count, and when it resets — along with any profit target. A configuration that never breaches but never reaches the target over the full test window has failed.

One firm, tested on its own data:
every Expert Advisor is tested against the same firm's rules (FTMO), in the Strategy Tester logged into an FTMO Free Trial account, so the price history, spreads and commission are FTMO's own. Every tick based on real ticks, 99% history quality or above, a £10,000 GBP account at FTMO's leverage cap (1:100), and one common window for every test (currently 1 January 2025 to 1 October 2026). Any reader with a free FTMO trial can repeat the test.

Only where the developer makes a prop-firm claim:
we run a conformance test only on an Expert Advisor whose developer says it can pass a prop firm challenge, or supplies prop-firm settings. No claim, no test.

Which settings we test, in order:
(1) the developer's own prop-firm settings file, supplied by the developer or published by them (including in their MQL5 product comments); (2) if none exists, the developer's default settings as installed, and we say no prop-firm settings were available. We never test settings files from third parties or settings we built ourselves: neither can be traced to the developer, so a result on them proves nothing about the product.

How we obtain the Expert Advisor:
as the free demo from the MQL5 Market wherever the product is listed there. Free demos run only in the Strategy Tester, and nothing is bought for a test.

We test the configuration unmodified:
exactly as the supplier provides it. Altering it before establishing a baseline destroys the only claim worth making. A second test with settings changed may follow, but only after the unmodified baseline is published, and always labelled as a configuration the supplier did not supply.

We report every limit separately, with the margin:
a near miss is reported as a near miss, with the figure. A result that clears a limit by a fraction of a percentage point is never reported simply as "passed".

Where we cannot measure what the rule measures, we say so:
FTMO's daily loss limit is measured on equity, including floating losses, and the Strategy Tester report gives realised profit and loss only. We publish the worst realised day and an upper estimate with every open position at its stop loss, and record the daily-limit outcome as not established unless it can be measured. We never present the obtainable figure as the required one.

Curve-fitting: separating what a developer could have tuned from what they could not

A developer who re-optimises an Expert Advisor "to current market conditions" can tune its default or prop-firm settings to the very history we then test it on. So every test records the version tested and its release date, the settings file's issue date where one is used, and the date the developer last re-optimised the strategy or its settings, taken from the developer's own release notes. Bug-fix releases do not count.

We split the test window at that date. Before it is **in-sample**: data the settings could have been fitted to. After it is **out-of-sample**: data they could not. If the last re-optimisation predates the window, the whole window is out-of-sample and the rules above apply as normal.

How the result is decided

- **Fail anywhere in the window is a Fail.** Tuning can only flatter a result, so failing on data the configuration could have been fitted to is conclusive.
- **A pass over the full window is judged on the out-of-sample segment alone.**

- Fewer than 100 out-of-sample closed trades (for example, optimised up to or near the test date): **Inconclusive** — insufficient out-of-sample data.
- 100 or more out-of-sample trades that meet the pass criteria on their own: **Pass**.
- 100 or more out-of-sample trades that lose money or breach a limit: **Fail**. The in-sample pass is reported, but the result is the out-of-sample one.
- 100 or more out-of-sample trades that are profitable with no breach, but have not yet had time to reach the profit target: **Inconclusive** — a slow but sound configuration is not marked down for time alone.

Every conformance test publishes the optimisation date, the length and trade count of the out-of-sample segment, and the figures for both segments, so a reader can check the split.

8. Conflicts of Interest

Affiliate-blind vetting:
we check every EA against these criteria before knowing whether it has an affiliate programme. Affiliate availability is never a search criterion or part of the verdict — it is disclosed after the fact.

The same rule applies to prop firms and brokers:
any assessment is written and published before a commercial relationship with that company is entered or factored in.

Our own live accounts:
we run an EA ourselves on a real-money account only after it has passed on its own merits. The account is published in full on the EA's review, verified by a third party, with any commercial relationship with its broker disclosed, and it is never offered for copying.

Material supplied by a vendor:
where a settings file or other material is supplied by a developer after a commercial relationship exists, we disclose that on the page, even where the verdict predates it.

9. Why This Version Changed

Version 1.0 of this regime required 7 to 10 years of real-tick data. In practice no retail broker supplies that — real tick history typically runs two to five years — so every test would have invoked the exception clause. A standard waived on every use is not a standard, so we changed it rather than keep excusing it.

The two-tier structure replaces it. It is more demanding, not less: every published figure must now come from a window at 99% real ticks or above, and a longer run at lower quality can never supply a headline number. Version 1.0 would have permitted figures this version forbids.

Version 2.1, 1 October 2026, adds one criterion: how long a floating loss stays open. An average holding time can hide a few positions held open in loss for months, and a reader copying an EA lives through every one of those days. A floating drawdown still open and unrecovered after 90 days now rules out a Recommended verdict, and the longest open drawdown is published with every test.

Version 2.2, 2 October 2026, sets out how we test Expert Advisors against prop firm rules in full: which settings we test, how we obtain the Expert Advisor, and how a result is decided. It adds the curve-fitting split above. The split itself was adopted before the first test it applied to had been analysed; the wording that separates Fail from Inconclusive was settled the same day, after that first result, and we say so here rather than present it as having been fixed in advance.

Version 2.2 also corrects how our first conformance tests were run. They used another broker's price data and charged no commission. We have re-run all of them on FTMO's own server data, with FTMO's spreads and commission. Every result is still a fail, but most of the figures changed — the same unmodified settings file can make money on one broker's prices and lose it on another's, which is why a conformance test must run on the firm's own data. Where a published figure changed, the page that carries it says so.

Version 2.3, 14 October 2026, tightens what counts as evidence. A backtest on its own, even one we re-run ourselves, can no longer lead to a Recommended verdict: the most it can reach is Inconclusive until a live record exists. A broker investor-password login on its own no longer counts as live evidence. Show My Trades joins the third-party monitors we accept, because it verifies an account through the broker's investor password and confirms that the publisher controls it. This version also publishes our minimum record of 24 months and 300 closed trades, our re-verification schedule, and how we publish our own live accounts.

Version 2.4, 19 October 2026, changes how Expert Advisors are chosen for review and what qualifies. We choose which EAs to review by what readers ask about. To be reviewed at all, an EA needs a real-money live record of at least 24 months and 300 closed trades, with every criterion assessable from it; the record does not have to be third-party

verified to qualify, but it does for a Recommended verdict. We no longer review an EA whose only evidence is a backtest, so the backtest-only review route (formerly Track 2) and its standard disclosure are retired: backtests are used only in rule-conformance tests. Every eligible EA we choose is reviewed and published, whatever the verdict.

Version 2.5, 28 October 2026, adds the Early Review tier: real-money records of 12 to 24 months can be reviewed, with Inconclusive or Not Recommended the only possible verdicts, and an Early Review becomes a Full Review at 24 months. It also makes any martingale, grid, averaging or recovery strategy an automatic Not Recommended, disclosed or not, where before it only ruled out a Recommended verdict.

See EAs that have already been through this process on our **Best Expert Advisors** page.

therobotictrader.com/best-expert-advisors

This document is provided for educational and transparency purposes and does not constitute financial advice. Trading forex and CFDs carries a high level of risk and may not be suitable for all investors. Past performance, including backtested performance, is not a guarantee of future results.