Sandbox and Testing Declaration

Mira risk engine and probability anchor (project Mira, U-Hackathon Flow Summit)

Document	MIRA-SBX-DECL-001, **version 1.1**. Replaces version 1.0 (issued 2026-10-15), which did not cover the 32-asset test or the retail-estimate costs for it.
Issued	2026-10-16 13:09 (every measurement in this document was taken when it was generated)
Status	Research documentation. Not investment advice. No real money was used.
Covers	Model v1 (production), model v2 (tested, not adopted), baht range tool, forward test
Prepared by	Claude (AI assistant), at the request of the project owner

1. Purpose

This document states exactly how the Mira models are tested, which controls stop a test from using information that was not available at the time, what was measured to show those controls work, and what the tests do not show.

2. Definitions

Term	Meaning here
Sandbox	A test run where each decision may use only data available on the decision day, with every rule fixed in advance.
Point-in-time	A value computed at date t from observations dated t or earlier only.
Walk-forward	Any fitted quantity at date t is fitted only on cases whose outcome was already known at t.
Declared rule	A parameter, test or success rule written to a dated file and hashed before results exist.
Forward test	A pre-registered test scored only on months after its declaration date; the only evidence not seen by the authors.
Brier score	Mean squared error of a probability forecast; lower is better; always saying 50% scores 0.25.

3. Data and provenance

Data	Source	Use
S&P 500 daily close, 2000-2026	Yahoo Finance chart API (cached JSON)	risk signals, anchor
SPY, 15 ETFs, 3-month T-bill: monthly adjusted prices	Yahoo Finance chart API (cached JSON)	portfolio returns
32 daily series (stocks, bonds, commodities, FX, crypto)	Yahoo Finance chart API	probability anchor
USD/THB and 17 other rates	ECB reference rates (Frankfurter)	baht range model
Commodity monthly prices	World Bank Pink Sheet file	commodity anchor

Data are read from local cache files; none was edited by hand. Files used in the long test are hashed in section 8.

4. How information is limited (the controls)

#	Control	How it is done in code
1	Decision timing	Decisions are taken at month end from data to that day; the position earns the following month's return.
2	Slicing inside the function	Signal functions cut their input to the decision index (`px[:i+1]`) themselves, so a mistake elsewhere cannot leak the future.
3	Poison test	All prices after the decision day are replaced with random garbage; the decision must stay identical.
4	Truncation test	Signals recomputed from data cut at month end must equal the signals used.
5	Causal arrays	Rolling statistics at index i use observations up to i only; checked by comparing a truncated and a full history.
6	Walk-forward fitting	Return pools, volatility regimes and the recalibration map use only questions with horizon on or before the origin.
7	Questions made by code	Asset, move size and type come from a rule; outcomes are real closing prices on the horizon date.
8	Declared and hashed rules	Parameters, tests and success rules are written to dated files with a code hash before results exist.
9	Frozen code guard	The forward test refuses to run if its rule code differs from the declared hash.
10	AI isolated	The AI layer is tested only after the model's training cutoff (from February 2025); it is not used in the 2001-2026 tests.
11	Costs	Trading cost on every change of stock share and a monthly platform fee, in three declared scenarios.

5. Test protocol

Test	Window	What is compared	Success rule (declared)
Long investing test	2001-01 to 2026-09, 309 decisions	fixed mix, gate, chain, trend dial vs all stocks; S&P plus cash	chain or gate vs fixed: worst fall at least 30% shallower, growth within 2 points, Sharpe not lower
Strict 2025-2026	2025-01 onward, defensive core S3	same variants	same rule; inconclusive if fixed never falls 5%
Sub-windows	2001-2007, 2008-2016, 2017-2026, four crises	same variants	reported, not selected
AI sealed test	origins 2025-02 to 2026-09, 720 questions	AI vs statistical anchor, 4 arms	paired bootstrap interval excludes zero
Anchor walk-forward	2003-2026, 9,318 questions	v2 anchor vs v1	interval excludes zero and not worse in 2025-26
Barrier v2	long test, scenario B	v2 vs v1 gate and chain	growth within 0.5 point, fall not deeper, Sharpe not lower
32-asset test (v1, frozen)	2000 to 2026, each of 32 cached assets on its own price, 60% base, cash side T-bill	100% asset, fixed 60%, v1 gate, v1 chain; equal-weight basket of 30 investable assets	same rule as the long test, applied to every asset and not selected; costs A and B

Forward test	from 2026-10 returns, 12 scored months	gate vs fixed, bases 30% and 60%	declared; judged only after 12 months

Metrics. Growth = compound annual rate of monthly returns. Volatility = monthly standard deviation x sqrt(12). Sharpe over cash = mean excess return over the 3-month T-bill x 12 / volatility. Worst fall = deepest drop from a previous month-end peak. **Costs (assumptions, not verified broker rates):** A 0.10% per unit traded; B 0.30% plus 0.20% a year; C 1.00% plus 1.00% a year. ETF running costs are inside adjusted prices. Taxes, currency conversion and slippage are not modelled.

6. Test suites executed for this document

20 of 20 suites passed in this run (offline, no network, no keys).

Suite	Result line	Status	Time
test_accuracy.py	70/70 passed	PASS	1.7s
test_assets.py	42/42 passed	PASS	1.0s
test_bot_fetch.py	ALL PASS	PASS	0.2s
test_broker.py	70/70 passed	PASS	1.5s
test_early_warning.py	39/39 passed	PASS	0.9s
test_eia_fetch.py	29/29 passed	PASS	0.3s
test_eia_study.py	26/26 passed	PASS	2.9s
test_fred_fetch.py	ALL PASS	PASS	0.2s
test_fx_range.py	37/37 passed	PASS	1.3s
test_fxanchor.py	test_fxanchor: 12 checks passed	PASS	0.4s
test_history_facts.py	test_history_facts: 5 checks passed	PASS	0.4s
test_mira_forward.py	test_mira_forward: 14 checks passed	PASS	0.4s
test_mira_fx.py	test_mira_fx: 11 checks passed	PASS	0.6s
test_mira_invest.py	test_mira_invest: 11 checks passed	PASS	0.5s
test_mira_v2.py	test_mira_v2: 22 checks passed	PASS	1.2s
test_pit.py	29/29 passed	PASS	2.3s
test_sandbox.py	7/7 passed	PASS	0.4s
test_source_quality.py	45/45 passed	PASS	0.2s
test_stress.py	27/27 passed	PASS	2.4s
test_team.py	70/70 passed	PASS	1.8s

7. Integrity evidence measured now

Check	Result
Truncation test, random months, long test	**12 of 12 identical**
Poison test, future prices replaced by garbage	**24 of 24 identical**
Re-run of the strict-test report from scratch	**byte-for-byte identical to the previous file**
Forward-test rule hash: declared / code now	**bb6bfa8d9d2a051c / bb6bfa8d9d2a051c (match)**
v2 rule hash declared in file / code now	**39a590daebda4e49 (match)**
32-asset test: poison test on every asset	**64 of 64 identical**
32-asset test: re-run from scratch, results file	**byte-for-byte identical to the previous file**

Forward ledger rows (decision months logged)	1

When each test was declared

Test	Declared	Before results?
v1 rules (alarms, gate, risk budget)	in the code of mira_invest.py, before its first run	yes (the S&P results were seen afterwards)
Forward test	FORWARD_TEST_RULE.md, 2026-10-15, with hash	yes, no forward result existed
v2 design	V2_DECLARATION.md, with hash	yes for v2 results; v1 results had been seen
32-asset test	this document and the test report, written when the test ran (it was run once, then reported)	**No pre-registration.** The rules were frozen earlier, but the test procedure was not hashed before the run.

8. Frozen artifacts (SHA-256, first 32 characters)

Code

File	Bytes	SHA-256
mira_invest.py	13,193	`d090dc22ec92eaa067de89e9e918820f`
allocator_backtest.py	14,644	`7e0bbf65993da8a2797ac04a6f4b97c5`
assets.py	9,821	`86e9b5256a197a2a00db44a8675d2047`
throttle_backtest.py	17,931	`a6c008d87d1cf016b123f987ba8f6471`
mira_forward.py	16,581	`42db5734325ec29490e75f61e66cf7e1`
mira_v2.py	12,602	`7f0dfe5f6906877464ea676025dc9cb2`
mira_sandbox_tests.py	12,328	`31b4d6872c86cf0066ff6912e25da408`
mira_32_assets.py	18,729	`2fdce87372f4123127506f155cc68277`
mira_fx.py	4,622	`f0eea7e198a46b9b07e7ff6b214c9495`
fx_range.py	16,844	`dce95dc8892ff3f1fd6c8d3f919dca96`
fxanchor.py	10,367	`fb103a10d91186964ae809cd861c8480`
pit.py	28,267	`152a338cb8b84af66fc4e94c228a0a20`
team.py	61,006	`a0d1aade07dfa63fd7ca14932acad5e4`
broker.py	25,647	`a18ab99191104ca58de7b6ebaa471ffe`

Declarations, ledger and reports

File	Bytes	SHA-256
FORWARD_TEST_RULE.md	2,156	`694e31e92af921c36737d96a019630d4`
V2_DECLARATION.md	2,279	`44b3f59f48af506b26e4859dca9aecd8`
mira_forward_ledger.csv	291	`0f287a1f1b457cf9b48afad8d8c07152`
REPORT_mira_invest.md	6,154	`7778e05aeaf0f83367fb15c7213eb92f`
REPORT_mira_sandbox_tests.md	12,730	`a950b2e35acd70ce1bcd2913d5b2ed79`
REPORT_sandbox_2025.md	10,183	`74402bc09c7cde21bbe0196e3b2282e0`
mira_32_assets_results.csv	8,970	`02e6974efba66a6830778a64f6497c42`
MIRA_OFFICIAL_REPORT.html	1,022,362	`79cbfd3a6beabf3efcb69eece2db3e3c`
MIRA_V2_OFFICIAL_REPORT.html	755,816	`97dff93d3cd6d88e2b716c0066741cd8`
MIRA_32ASSET_OFFICIAL_REPORT.html	394,719	`292092d50bb997113595b412acd95edb`

Data used in the long test

File	Bytes	SHA-256
data_multi/raw_GLD.json	30,382	`6dcfff20de9348e647560ca8de6d0c82`

data_multi/raw_SLV.json	28,863	`2a6eabeec06711ff98e0df07e9b5b09b`
data_multi/raw_DBC.json	29,048	`bbbd44bc46fb9e4f31d59f0695ef3b9c`
data_multi/raw_DBA.json	27,762	`3adbf616c49ba16b57f942209eefedb2`
data_multi/raw_TLT.json	46,913	`579c1ceec9641173fcce6a128ec7ac40`
data_multi/raw_IEF.json	46,904	`23f120b08a9cb9c42a2cc7ebe97b5ca2`
data_multi/raw_SHY.json	46,341	`036b0125802e13a1df7ee152938812b7`
data_multi/raw_TIP.json	41,242	`4c3be7d9b3361c55d4b962fa58380789`
data_multi/raw_LQD.json	47,350	`4217412fa38ca55aae919d90dd4214bc`
data_multi/raw_HYG.json	37,637	`d23523b2e7383751006aa43c883a2d58`
data_multi/raw_EMB.json	36,665	`8508f35d2ca84ddc099ad54e536f8576`
data_multi/raw_UUP.json	28,005	`e749325d9a869d6a17fa47ab6b581d19`
data_multi/raw_FXE.json	33,620	`fecb9fa08014b6da7c1f3435be8fe3e2`
data_multi/raw_FXY.json	26,705	`3eeb524bfb7c381a395c725b375fc16b`
data_multi/raw_VNQ.json	34,083	`10b469d26ddb4ab580d3b452fedfed09`
data_multi/raw_SPY.json	49,103	`9ac6a223f656ff7f81d96c8d2cc86a9c`
data_multi/raw__IRX.json	39,614	`63f199614bfa8881542c77c71ac53731`
data_assets/AAPL.json	223,834	`2ddf53922947ea89ce1ac52b128fed19`
data_assets/BTC-USD.json	134,225	`69ff9fcfef346b0475f902691d8443ed`
data_assets/BZ_F.json	156,618	`ef52ae1d46547d0381f062e43d224a07`
data_assets/CL_F.json	214,526	`082eb595b6cf73bfc84c844bf1e11324`
data_assets/DBC.json	172,212	`792865f1b8143c53e8711de6226cc538`
data_assets/DX-Y.NYB.json	220,555	`14696020af28c32103aa381d610333b6`
data_assets/EEM.json	194,483	`8135132c9df02054a16d8163a3e02633`
data_assets/EMB.json	155,821	`61ed26877ed6ff786e6dae9ad4f4e4ee`
data_assets/ETH-USD.json	106,426	`43109a068a8e3f7abb44ba09cb609ab0`
data_assets/GC_F.json	202,827	`19c968cc7b238719cd46051f7f8ece0a`
data_assets/HG_F.json	219,246	`52cd10fc2e0719e5051bd8cd85e7c705`
data_assets/HYG.json	157,714	`d1a8e903bc7ae32b87c207022983392a`
data_assets/IEF.json	199,385	`18da2a2ed2f1ea28dbb17bc6f89c357b`
data_assets/LQD.json	201,505	`2bdb94f1c1a2e9a1eb734b3abb892397`
data_assets/NG_F.json	218,076	`9cd5600fe9a4dc98e97438cd5f36d087`
data_assets/NVDA.json	228,078	`453d8d19ccc1dc3f624d5e4f37d554c6`
data_assets/SHY.json	196,744	`067010bfb9be9773c291e67128304d04`
data_assets/SI_F.json	218,691	`18dfc28e060d8e23656aefc8eb76b196`
data_assets/THD.json	151,216	`2b874473c9396355ca4cefd4b2f0cc67`

data_assets/TIP.json	190,586	`9e9304d6dcdb14a2856df53d957e30f0`
data_assets/TLT.json	199,466	`c0cac2235ce137cfbb089054518daa91`
data_assets/VNQ.json	179,991	`4bcd8fadc37905fce0d9f888d63f8b2a`
data_assets/ZC_F.json	141,033	`9903052a58b3b94e257dc05a224ee8f9`
data_assets/ZS_F.json	143,695	`cd05fadf015ee53711eec1e1e40296d2`
data_assets/ZW_F.json	141,275	`eab2cda5551df334e1f4c17f9762a755`
data_assets/_FTSE.json	200,258	`4f8d5bd1a4ba9718bcbc0431b58f6383`
data_assets/_GSPC.json	216,545	`613f415a3c716c3ef54a47b38333f917`
data_assets/_HSI.json	198,941	`026c70e4e3e7cbdde3a0df36c996d7d0`
data_assets/_IXIC.json	211,352	`de922cd30bbdb00d4cfc00369aa9223d`
data_assets/_N225.json	197,680	`bf15e4ca4131fda285cc0c1b63d12272`
data_assets/_TNX.json	224,008	`fd43eac1d4e24b5b9638168f7b03172b`
data_assets/_VIX.json	223,055	`b08dfc569e0e03e5500f7caeee8c8553`

Note: `pit.py` and `team.py` were extended on 2026-10-15 (price-history arm, lower AI weight); backups of the earlier versions are kept. The rule code covered by the forward-test hash was not changed after its declaration.

9. Limits and what is not claimed

- **Contamination.** The authors saw v1 results for 2001-2026 before designing v2 and before this document. No backtest year is unseen. Only the forward test (from 2026-10) is.
- Few independent crises (about four). Differences in growth between versions are inside statistical noise; the benefit shown is shallower falls.
- One signal (S&P 500) and one defensive side per test; results may differ for other assets or markets.
- Costs are assumptions. Taxes, currency conversion, slippage and liquidity are not modelled.
- The AI layer showed no measurable gain over the statistical anchor; its weight in the system is 0.10.
- **32-asset test:** prices are closes (no dividends or interest, so bond funds show near-zero growth; futures roll costs not modelled); VIX and the 10-year yield index cannot be bought; many assets share the same crises, so 32 results are not 32 independent proofs.
- **Growth figures differ between tests on purpose:** the S&P test uses total return, the 32-asset test uses price only, and the cash side earns T-bill rates. Numbers from the two tests must not be compared with each other.
- **Not claimed:** prediction of prices or crashes, outperformance of the market, guaranteed or expected returns, suitability for any person. This is not investment advice.

10. Change control

Alarm thresholds, gate values, risk budget, cost scenarios, windows and success rules of v1, the forward test and v2 are frozen by the hashes above. Any new idea is a new test with a new declaration date and a new hash. Results that fail a declared rule are reported as failed.

11. Declaration

The undersigned state that the procedures above were followed and that the figures in this document were produced by the code and data identified in section 8.

Role	Name	Signature	Date
Prepared by (AI assistant)	Claude (Anthropic), model claude-sonnet-5-5		2026-10-16
Project owner			
Reviewer / witness (optional)			