

Melaninternet.Volume_1 — Technical Architecture



The evidence layer between foundation models and the work they do — live today, with staged layers for verification, assurance, agents, personal AI, health, world models, and Physical AI.

LIVE

1 Experience layer

Public site, mobile-first chat, guided demos, document upload, CTA forms. Planned: workspace and personal-assistant apps.

LIVE

2 Application layer

FastAPI service, routing, sessions, file handling, prompt compilation, validation, error handling.

LIVE

3 Policy layer

Versioned constitution, privacy rules, tool-use boundaries, grounding constraints, approval gates.

LIVE

4 Model layer

NVIDIA Nemotron 3 Super 120B A12B primary with automatic failover to Ultra; model-agnostic routing planned.

R&D

5 Verification layer

Citation checks today. R&D: claim-level verifier, shift-robust calibration, answer / qualify / abstain.

LIVE

6 Reliability + evidence

Health and model checks, regression, synthetic users, concurrency, canary grounding, PASS/FAIL reports; per-decision evidence records planned.

NEXT

7 Agents + assurance

Agents run in sandboxed runtimes (NVIDIA OpenShell-aligned); tool approvals, audit trails, behavioral baselines, anomaly flags, SIEM exports.

PLANNED

8 Personal + health apps

Evidence-cited daily brief, calendar, goals, approval-gated shopping; health research, protocol, and admin intelligence.

LONG-TERM

9 World model + Physical AI

Stateful enterprise/health models; plan verification and action gates for robots on open world models (e.g., Cosmos).

Current proof points	Release v1.9.7 (31 Aug 2026): 46/46 production checks, 25/25 synthetic users, 5/5 concurrent isolated sessions, P50 2.8s / P95 11.4s, 200-case regression corpus. Persistence: PostgreSQL telemetry and leads; Render Key Value session state; temporary document context.
Research program	Guarantees that survive model changes: model-swap-robust calibration and distractor-aware verification (NSF SBIR resubmission; NIH SBIR path).
Mission alignment	Pew: Black and Hispanic workers are more concentrated in occupations with lower AI exposure. Crunchbase: companies with a Black founder or co-founder received about 0.32% of U.S. venture funding in 2025.

Architecture principle: separate user content, policy, model calls, verification, telemetry, identity, lead capture, and evidence so the system can be demonstrated, measured, secured, and governed without overstating readiness.