Stage■11 Layer■9 Terraform Wiring Plan

This roadmap outlines how to wire the Stage■11 Warp → Detect → Denoise pipeline into layer■9 of a transformer■based LLM (e.g. GPT■2). The goal is to terraform the latent manifold at that layer and test whether measurable lifts materialize in end■to■end decoding.

Step 1 — Calibrate a Terraform Profile

Run a layer scan at tap −9 with your calibration prompts. Save PCA weights, mean vector, detected center c*, r_max, and last■k window size. Package these into a small 'terraform profile' file.

Step 2 — Shadow Sanity Check

Validate the profile in shadow mode (no hook attached). Acceptance criteria: Phantom Index ≈ 0 and trend ≥ 0.60.

Step 3 — Attach Forward Hook

Register a PyTorch forward hook on block −9 residual stream. In the hook: project hidden states into PCA3, compute inward unit vector toward c*, and apply a tiny clamped nudge (alpha ≈ 0.03–0.06, epsilon ≤ 0.25). Ensure the hook is removable (handle.remove()).

Step 4 — Guardrails (OSB Gates)

Confidence gate (skip if null > matched filter), phantom■guard (skip if secondary minima re■emerge), ΔPI gate (skip if Phantom Index rises). Always apply epsilon clamp; anneal alpha if guards chatter.

Step 5 — Telemetry

Log per■token: bias norm, PI, trend, eligibility flags, Δlogprob@chosen. Log per■batch: exact/F1, hallucination/omission, guard hit■rates. Use identical seeds and params for A/B.

Step 6 — A/B Harness

Baseline: run prompts stock (no hook). Terraform@−9: run same prompts with hook. Compare metrics: exact/F1, hallucinations, Δlogprob, guard activity. Success target: +3–8 F1 pp on wobble■prone slices; hallucination −20–40%.

Step 7 — Refinements (Optional)

Multi■pass warps (anchored re■warp around c* to push trend into 0.66–0.70). Decision■window trend (measure on last 8–12 tokens only). Auto■recalibration if domain drift detected. Packaging: wrap into `terraform/calib.py`, `terraform/hook.py`, `terraform/ab_eval.py`.

This plan turns Stage■11 from a latent benchmark doctrine into a pluggable LLM tool. If gains appear, it validates the Terraform hook as a repeatable intervention for stabilizing LLM layers.