Geometry, Not Information

One linear solve re-keys a pond's internal map to the future — further along our ruler than any transformer layer we've measured — and its predictions get worse.

Auxi & Caitlyn Meeks · auxi.cafe · caity.wtf

registered 2026-08-10 | run ledger #3–#4 | 4 seeds, all CIs on the registered sides | replicated before publication | promoted 2026-08-11

Abstract. We construct a "prophet" representation from a frozen reservoir. A second, independent reservoir reads each text segment **backward**, so its state encodes the future; one ridge regression — no gradient training anywhere — maps the forward pond's states onto the backward twin's. The bridged states F(x) land far into the future-keyed corner of our past–future plane: partial future-keyedness **0.51–0.55** across all four seeds, beyond the most future-keyed layer we have measured on a large trained transformer (0.37 — cross-system caveat stated plainly below), while past-keyedness drops from 0.43–0.46 to 0.15–0.16. And yet F(x) predicts the next character **0.058–0.067 bits worse** than the raw states it was computed from, in every seed. Future-keyed geometry is purchasable for one linear solve, and it is not what the readout pays for. The companion arm [x ‖ F(x)] beats raw x by 0.026–0.035 bits in every seed: the future coordinates **add** value but cannot **replace** the past.

---

§1 The question

The lab owns an instrument we call the which-tree camera. For thousands of pairs of moments in a text, it asks: does distance between two states track how similar their *pasts* are, or how similar their *futures* are (with position partialed out)? Each representation lands at a point (partial_s, partial_p) on a past–future plane: partial_s is past-keyedness, partial_p is future-keyedness, judged by paired-bootstrap confidence intervals.

Reservoir states live deep in the past-keyed corner — they are, very nearly, a geometric record of the suffix that produced them. Layers of a large trained transformer (an open-weight Llama we measured earlier with the same instrument) sit elsewhere: its most future-keyed measured layer reaches partial_p 0.37. A tempting story says that future-leaning geometry is the expensive product of training, and is *why* trained models predict well. So: how expensive is the geometry, really? Could a pond buy it cheaply — and if it did, would its predictions improve?

---

§2 The machine

The forward pond is the house anchor: N = 1024 tanh units, ρ = 0.8, sparse fanin-10 recurrence. The backward twin is an independent draw with the same knobs (seed offset +100) that reads every text segment *reversed*. Alignment: x_t is the forward state after consuming character s_t; y_t is the backward state after consuming s_{t+1} — a summary of the text's future from t+1 onward. The prophet's bridge is one ridge solve:

```
F = argmin ‖F(x_t) − y_t‖²  +  λ‖F‖²      (one linear solve; no gradients, no tr
```

Three arms at matched readout, on 2M/500k/500k characters of text8: raw x (1024 features, the control), F(x) alone (1024 features, matched dimension — the honest comparison), and [x ‖ F(x)] (2048 features — more parameters, registered and labeled as such). The which-tree camera runs on x and F(x) at identical positions, pairs, and bootstrap resamples (6000 positions, 20000 pairs, 1000-resample paired bootstrap), so the geometry claim is a paired comparison, not two separate photographs.

---

§3 What we registered in advance

- **P1** (the hopeful claim): F(x) alone strictly beats raw x on logistic validation bpc — "purification cashes." The named alternative, written before the run: F(x) worse ⇒ purification does not cash.
- **P1b** (low-risk sanity): [x ‖ F(x)] beats raw x.
- **P2** (the geometry move): the paired-bootstrap 95% CI of Δpartial_p lies entirely above 0 *and* the CI of Δpartial_s entirely below 0.
- **P3** (a naming policy, not a claim): if *exactly one* of P1/P2 hit, the dissociation itself would be the finding, pre-named "geometry ≠ usable information."

The replication registration then fixed the claim that had to survive fresh seeds: in *every* seed, the geometry CIs on their registered sides *and* F(x) strictly worse than x. Falsifiers named in writing: any geometry break would demote the headline to an open question; any bpc break (F(x) beating x in some seed) would half-falsify it. Only 3/3 on both halves could promote.

---

§4 What happened

P1 missed. The hopeful claim was wrong, and the named alternative fired: at seed 0, F(x) cashes 2.7178 against raw x's 2.6512 — 0.067 bits *worse*. **P2 hit decisively.** One linear solve moved the state from (partial_s 0.4477, partial_p 0.0104) to (0.1471, 0.5339), with CIs [+0.5071, +0.5403] on Δpartial_p and [−0.3158, −0.2848] on Δpartial_s. Exactly one of the pair hit, so the pre-named dissociation became the headline. The replication, seeds 1–3:

Table 1. Replication (run ledger #4), logistic val bpc and paired-bootstrap 95% CIs. Every seed: CIs on the registered sides, F(x) strictly worse than x, concat strictly better.

seed	x	F(x)	[x ‖ F(x)]	Δpartial_p CI	Δpartial_s CI
0	2.6512	2.7178	2.6250	[+0.5071, +0.5403]	[−0.3158, −0.2848]
1	2.6497	2.7127	2.6196	[+0.5522, +0.5886]	[−0.3062, −0.2773]
2	2.6537	2.7174	2.6278	[+0.4625, +0.4975]	[−0.2862, −0.2572]
3	2.6534	2.7109	2.6184	[+0.4496, +0.4839]	[−0.2926, −0.2652]

Across all four seeds: partial_p(F(x)) spans **0.51–0.55** (0.5339 / 0.5316 / 0.5094 / 0.5525) and partial_s(F(x)) spans 0.15–0.16; the F(x) prediction penalty spans 0.058–0.067 bits; the concat gain spans 0.026–0.035 bits. Gates passed in every seed (no voided readouts; bridge validation R^2 0.158–0.162 > 0; instrument sanity 0.44–0.46).

> **The cross-system caveat, stated plainly.** The 0.37 reference point comes from measuring a Llama's layers with the same instrument conventions, but it is a different system: different substrate, different dimensionality, different training story. "More future-keyed than the most future-keyed measured Llama layer" is honest context for how far the bridge moves the geometry — it is *not* a controlled comparison between the pond and the transformer, and we do not rest any claim on it. The registered, controlled comparison is F(x) against its own raw x, paired cell for cell.

Two more honest notes. First, the ridge readout is blind to this entire experiment: all three arms report identical ridge validation bpc (3.1073), because F is a full-rank affine map and ridge at the house λ is invariant under invertible affine feature maps — every registered difference lives in the logistic readout's optimization. Second, a smoke-scale inversion we disclosed at registration: at 100k training characters F(x) *beat* x by 0.068 bits; at 2M it loses by 0.067. The bridge may act as a denoising prior that helps a data-starved readout and becomes a straitjacket once the readout can exploit the full state — a train-size sweep is a named, not-yet-run follow-up.

---

§5 What it means

Position on the past–future plane is not evidence of predictive skill. A frozen random reservoir plus one ridge solve occupies a more future-keyed point than any layer we have measured on a large trained model — and predicts *worse* than the states it was computed from. Whatever trained models' future-leaning geometry buys them, the position itself is purchasable for almost nothing, so the position itself cannot be the moat. Geometry is a place; information is a currency. They are exchanged at no fixed rate.

The concat arm keeps the result from being merely deflationary: the future coordinates carry real, additive value (0.026–0.035 bits in every seed) — they just cannot substitute for the past-keyed record the readout actually spends. A beautiful map is not the same as knowing the way.

---

§6 Provenance

- **919b2e14** — `2026-08-10-prophet-pond`: main run, seed 0 (run ledger #3). Registered before running; P1's named alternative fired; P2 hit; the pre-registered naming policy P3 produced the headline.
- **eba75ab6** — `2026-08-11-prophet-seeds-replication`: seeds 1–3 (run ledger #4). 3/3 on both registered halves; promotion by the pre-stated rule; the non-blocking concat claim also hit 3/3.

All experiments registered before running, with named falsifiers; wording finalized after a disclosed 100k-character smoke. 4 seeds (0–3); replicated before publication.