Stage 10 vs Stage 11: Synthetic, Latent-ARC, and LLM Integration

This report compares Stage-10 and Stage-11 performance across three contexts: synthetic traces, Latent-ARC benchmark, and a first LLM integration attempt. The contrast highlights why Stage-11's Warp → Detect → Denoise doctrine is required.

Stage-10 (Synthetic)

• Benchmark: ARC-like synthetic traces (flip_h, flip_v, rotate). • Parser: perpendicular energy, smoothing, matched filter. • Outcome: Clean lobes, no phantoms. • Result: 100% exact accuracy (10/10).

Stage-11 (Latent ARC, n=100)

Method	Accuracy	Precision	Recall	F1	Hallucination	Omission
Stock	0.49	0.89	0.78	0.80	0.11	0.22
Geodesic (Stage-10)	0.64	0.85	1.00	0.90	0.16	0.00
Stage-11 Denoiser	1.00	0.998	0.999	0.998	0.005	0.002

The Stage-11 denoiser path converged deterministically across all five latent wells. Hallucinations and omissions are near the noise floor. This demonstrates the necessity of Warp → Detect → Denoise for robustness.

Stage-11 (LLM Integration Attempt, n=12)

• Model: GPT-2, layer 8 hidden states. • Setup: Stage-10 parser + Stage-11 residual/denoise extensions. • Outcome: 0/12 exact accuracy. Predictions collapsed to a single primitive. • Interpretation: Raw LLM latents are too noisy and irregular; doctrine pieces are necessary but tuning and warp-fit are not yet sufficient.

Conclusion

Stage-10 works only in controlled synthetic settings. Stage-11 achieves deterministic reasoning on Latent-ARC by suppressing phantoms and stabilizing trajectories. Direct LLM integration remains challenging: the doctrine is necessary, but additional tuning and manifold warping are required for success.