

## THE WEAKEST LINK PRINCIPLE

# Your reasoning chain has a ceiling.

No sound inference exceeds the reliability of its weakest premise. We prove *min* is the unique aggregation enforcing this — and that LLM chain-of-thought is only **25–39% faithful** to the model's actual computation (Anthropic, 2025).

Sankalp Gilda<sup>1+</sup> · Shlok Gilda<sup>2+</sup>

<sup>1</sup>DeepThought Solutions · <sup>2</sup>University of Florida, Computer Science · <sup>+</sup>Equal contribution

## WHY MIN-AGGREGATION MATTERS · CHAIN THAT OVERGENERALIZES

S1  
All metals conduct electricity

0.95

S2  
Iron is a metal

0.90

S3  
∴ all iron tools are safe to touch

0.30

## ARITHMETIC MEAN

0.72 × Weakness hidden — chain looks acceptable.

VS

## WEAKEST LINK · MIN

0.30 ✓ Weakness surfaced — flagged unreliable.

## THE ADI PROTOCOL

Separate Peirce's three inference modes into auditable phases. Each capped at its layer.

**L0** **Abduction**  $R \leq 0.35$   
generator proposes hypotheses

↓ VERIFY

**L1** **Deduction**  $R \leq 0.75$   
logically consistent

↓ VALIDATE

**L2** **Induction**  $R \leq 1.00$   
empirically corroborated

**Transformer Mandate.** The entity that generates a claim cannot also verify it — enforced by structure, not policy.

## THE GAMMA QUINTET

Five algebraic invariants any consistency-preserving operator  $\Gamma$  must satisfy.

**IDEM**  $\Gamma([x]) = x$  single premise keeps its score

**COMM** order-invariant under permutation

**LOC** local propagation only

**WLNK**  $\Gamma(S) \leq \min(S)$  no chain exceeds its weakest link

**MONO** monotone in premises

**Theorem.** *min* is the unique continuous idempotent t-norm satisfying all five (Klement–Mesiar–Pap, 2000).

## DUAL CEILING PIPELINE

$R_{\text{eff}}$  passes through **two min gates** — a layer cap (Peirce mode) and a formality cap (rigor of the artifact).

Raw → Layer cap → Formality cap →

$R_{\text{eff}}$

Expired evidence decays to  $R = 0.1$ , forcing re-validation before propagation.

## LAYER CAPS · MIN L

$L0 \leq 0.35$  ·  $L1 \leq 0.75$  ·  $L2 \leq 1.00$

Abduction · Deduction · Induction

## FORMALITY CAPS · MIN F

$F0 \leq 0.70$  ·  $F1 \leq 0.85$  ·  $F2 \leq 0.95$  ·  $F3 \leq 1.00$

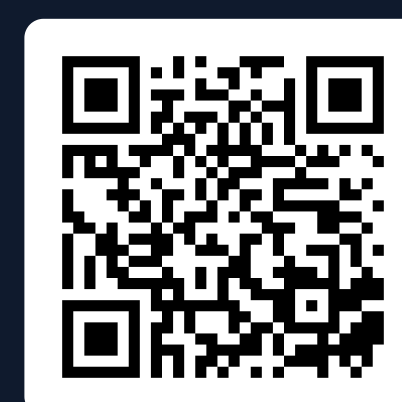
Informal · Semi-formal · Formal · Mechanized

## FUTURE WORK

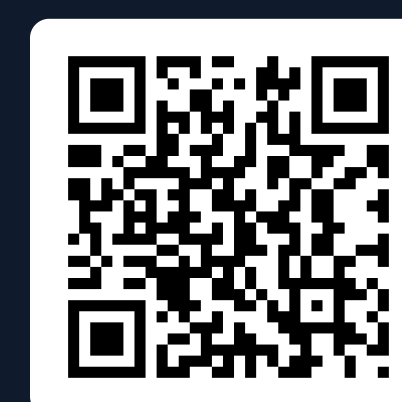
- WLNK as a differentiable training constraint
- Multi-agent ADI — specialized agents per mode
- End-to-end eval on ZebraLogic, FOLIO

**10<sup>5+</sup>** cases / test · **52** property tests

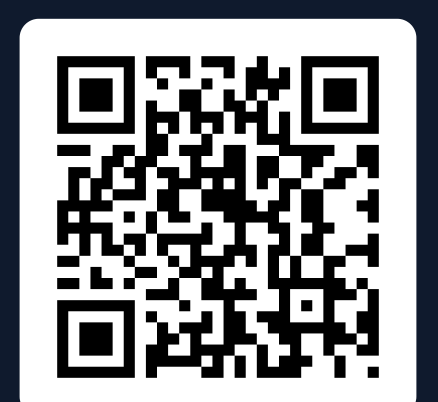
## SCAN &amp; CONNECT



PAPER  
OpenReview



SANKALP  
LinkedIn



SHLOK  
LinkedIn

## REFERENCES

Anthropic (2025). *On the biology of a large language model*. Transformer Circuits Thread.  
Peirce, C. S. (1903). *Harvard Lectures on Pragmatism*.

Klement, E. P., Mesiar, R., Pap, E. (2000). *Triangular Norms*. Kluwer.  
Wei, J. et al. (2022). Chain-of-thought prompting. *NeurIPS*.