ICLR 2026 · LOGICAL REASONING OF LLMS WORKSHOP · PAPER #148

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[1†] · **Shlok Gilda**[2†]

[1] DeepThought Solutions · [2] University of Florida, Computer Science · [†] *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 — generator proposes hypotheses — R ≤ 0.35

↓ VERIFY

L1 Deduction — logically consistent — R ≤ 0.75

↓ VALIDATE

L2 Induction — empirically corroborated — R ≤ 1.00

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 Γ must satisfy.

- **IDEM** Γ([x]) = x single premise keeps its score
- **COMM** order-invariant under permutation
- **LOC** local propagation only
- **WLNK** Γ(S) ≤ 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_{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_{eff}

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

LAYER CAPS · MIN L

L0 ≤ 0.35 · L1 ≤ 0.75 · L2 ≤ 1.00

Abduction · Deduction · Induction

FORMALITY CAPS · MIN F

F0 ≤ 0.70 · F1 ≤ 0.85 · F2 ≤ 0.95 · F3 ≤ 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^{5+} cases / test · **52** property tests

SCAN & 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.*

sankalp.gilda@gmail.com

OpenReview: zy6HdcsJ9V

ICLR 2026 · Rio de Janeiro