One Theory, Two Hemispheres

Right-Brain AI (November 2025) Read Through the Vacuum.net Theory, and What Has Since Been Built and Measured

J. Konstapel — Constable Research, Leiden — 20 September 2026

Second, self-contained edition. *This edition is written so that no reader — human or machine — needs any other document to follow it. The 2025 architecture is restated in full. The net vocabulary is derived, not assumed. Every number is computed in the text. The open engineering question of 20 September 2026, and the proposed answer to it, are stated inside this essay. It is the single reference for this line of work; earlier summaries and relay messages are superseded.*

1. Two documents, ten months apart

In November 2025, the essay "Left and Right Brain AI" (constable.blog, 23 November 2025) laid out an architecture in the vocabulary of fields. Contemporary AI is a hypertrophied left hemisphere. The missing right hemisphere is a Resonant Stack: an oscillatory substrate, a nilpotent coherence kernel, multi-scale coupling to the world. The essay was written from intuition and from physics, before the vacuum.net theory had its present vocabulary of strand, winding, closure, mesh and tension.

In September 2026, MAZE.AI stands running: an answer loop over 8.19 million addressed pieces, with machine-checked invariants, a signed ledger, and a pre-registered placement measurement of 15 out of 20. It was designed in the vocabulary of discrete addresses.

The claim of this essay is that these are not two projects. They are one theory read at two winding depths. Right-Brain AI reads the knowing layer as a continuous field — the deep, fast windings. The maze reads it as discrete addresses — the shallow, slow windings. Identity across scale, not analogy. The essay establishes the correlation term by term (Section 6), measures the 2025 claims against 2026 (Section 7), names the one theoretical gap that the building work has exposed (Section 8), and states the open measurement that will decide the next step (Section 9).

2. The 2025 architecture, restated in full

The November 2025 essay contains two stacks and a connection. They are restated here so the reader needs nothing else.

2.1 The left-brain stack

Five layers. **Substrate:** GPU and TPU clusters; the unit is the bit; compute is clocked and discrete; energy scales superlinearly with capability. **Foundation models:** transformers trained on internet-scale corpora; the core operation is the forward pass toward a next-token prediction. **Scaling as law:** loss follows power laws in model size, data and compute; capability becomes a function of investment. **Agent layer:** orchestration on top — task decomposition, tool calls, external memory. **Policy overlays:** alignment and safety added as filters above the core, not inside its logic.

The stack is discrete at every joint: bits, tokens, steps. Three constraints accumulate. **Temporal fragmentation:** fixed context windows; long-range coherence is simulated by bookkeeping; the system has no intrinsic sense of slow change. **Loss-function myopia:** behaviour is fixed by a human-chosen scalar loss; misalignment is patched afterwards, with more labels. **The energy ceiling:** compute demand grows faster than capability; the essay calls this a physical constraint, not a solvable engineering problem.

2.2 The right-brain stack

Five layers, but continuous — modes of one field. **Oscillatory substrate:** coupled oscillators, ideally photonic; the unit is phase and frequency, not the bit; computation is the self-organisation of the field. The canonical dynamics is Kuramoto's: each oscillator i has phase θ_i and natural frequency ω_i, and $d\theta_i/dt = \omega_i + (K/N)\cdot\Sigma_j \sin(\theta_j - \theta_i)$. The degree of synchrony is the order parameter: $r\cdot e^{i\psi} = (1/N)\cdot\Sigma_j e^{i\theta_j}$, with $r = 1$ for perfect phase lock and $r \approx 0$ for incoherence. **Nilpotent coherence kernel:** after Rowlands, the state of the field is a vector N encoding space, time, mass and charge; only states with $N^2 = 0$ — zero-totality, the whole summing to nothing — are admissible. Incoherent states are not rare failures to be filtered; they are impossible. **Virtual Resonant Being:** a persistent self-referential pattern of the field, not an agent bolted on top. **Multi-scale coupling:** fast (milliseconds — neural rhythms), intermediate (hours — conversation, markets), slow (years — ecology); each a frequency band of the same field, so long-term memory is intrinsic: the slow modes are the memory. **Anthropic constraints in the physics:** configurations incompatible with human and ecological viability are made energetically unstable, not filtered afterwards.

2.3 The connection

The essay's central design problem is the corpus callosum: the interface between hemispheres. Its first pattern reads: the language model receives the request, decomposes and phrases; before anything executes, the coherent core is consulted; the model proposes, the core disposes. Its neuroscientific ground is McGilchrist's synthesis of the split-brain literature, with one clinical fact the essay leans on and this essay will use again: an isolated left hemisphere confabulates — it invents explanations and denies what is in front of it.

2.4 The promised guarantee

One strategic claim of the 2025 essay stands out and will be measured in Section 7. A stack with nilpotent constraints can show that whole classes of destructive states are impossible — not rare, not filtered at 99.9 percent, but impossible. The essay names this as a different class of guarantee than "we tested the model and it performed well."

3. The lineage: Hamilton, Maxwell, and the amputation

The mathematics under both stacks has one root, and naming it locates the theoretical gap of Section 8 before the engineering finds it.

In 1843 Hamilton found the quaternions: numbers $q = w + xi + yj + zk$ whose product multiplies rotations. The product of two quaternions has two parts at once — a scalar part, measuring how well the two align, and a vector part, giving the axis along which they differ. One operation, both answers.

Maxwell wrote much of the 1873 Treatise on Electricity and Magnetism in this language: the field equations in quaternion form, alignment and axis held together.

In the 1880s Heaviside and Gibbs cut the product in two. The scalar part became the dot product; the vector part became the cross product; the unified operation, and with it the scalar term of the field product, disappeared from working physics. Vector analysis is easier to teach and was adopted everywhere. The cost was invisible for a century: physics kept the pieces and lost the composition.

The line then runs underground. Pauli's spin matrices (1927) are quaternions in disguise; Dirac's equation is built from them; Rowlands' nilpotent algebra — the kernel of the 2025 Resonant Stack — is explicitly quaternionic, with $N^2 = 0$ as its admission rule. So the answer to "does all of this come from Maxwell's quaternion theory?" is: from Hamilton's algebra, in Maxwell's form, through Rowlands' kernel — one line, and both hemispheres of this essay stand on it. What the field lost in the 1880s, the composition rule, is exactly what Section 8 will identify as the one thing the 2026 construction still lacks.

4. The net vocabulary, derived

The vacuum.net theory describes everything as one strand that winds. A winding that closes carries no net tension; closure is the stable form. Windings of windings form a mesh; position in the mesh is depth. What persists is selected by fit: only what returns in phase remains. Where two quantities are one state read twice, the theory writes a constitutive relation — simultaneous, no causal arrow. The register of the net is the trit: +1, 0, −1.

4.1 Why the register is the trit

Balanced ternary is the number system with digits {−1, 0, +1}. Every integer has exactly one representation. The two signs carry each other and the zero stands lawful in the middle — the discrete form of a closed winding: net tension zero. Worked example, the number 43. Repeatedly divide by 3, choosing the remainder from {−1, 0, +1}: $43 = 3\cdot14 + 1$, digit +1; $14 = 3\cdot5 - 1$, digit −1; $5 = 3\cdot2 - 1$, digit −1; $2 = 3\cdot1 - 1$, digit −1; $1 = 3\cdot0 + 1$, digit +1. Reading the digits from least to most significant: [+, −, −, −, +], and indeed $1 - 3 - 9 - 27 + 81 = 43$. The address is the record of the walk — one step out, three back, one out — not a label pasted on.

The choice is also economical. The cost of holding numbers up to N in radix r is proportional to r/ln r. For r = 2 this is 2/ln 2 = 2.885; for r = 3 it is 3/ln 3 = 2.731; the mathematical optimum lies at e ≈ 2.718, and 3 is the closest integer. The trit is not exotic; it is the cheapest lawful register.

4.2 Depth is capacity, and truncation is rounding

Eleven trits give $3^{11} = 177{,}147$ cells. Nineteen give $3^{19} = 1{,}162{,}261{,}467$ — over a billion. Twenty-six give about 2.54×10^{12}. Thirty-two give about 1.85×10^{15}. Forty-two give about 1.09×10^{20} — and forty-two trits occupy $42 \times \log_2 3 \approx 66.6$ bits: nine bytes for an address space larger than every moment of every human life.

The stability under deepening is a theorem, not a hope. The value of any k-trit tail is at most $(3^k - 1)/2$ in absolute size: for k = 2, the largest tail is $1 + 3 = 4 = (9 - 1)/2$. So dropping the tail changes the value by less than half the next step — truncation is rounding, a deeper address refines a shallower one and never invalidates it, and nothing is ever re-indexed. A prefix is a coarser class; a neighbour is a relative. This metric structure — which a hash lacks entirely — is what Section 6 will identify as the master's mode of attention in discrete form.

5. What was built and measured by September 2026

MAZE.AI runs at maze.swarp.nl/ai as a context under the MAZE, a self-addressing ternary knowledge net of 8.19 million pieces (Wikipedia, Mathlib theorems, EPO patents, art, music, open questions), each at an address computed from its content.

Construction. Five components. A kernel: the balanced-ternary bijection, range bound, uniqueness and prefix-stable truncation, proved in Lean 4 (about 740 lines, no external libraries, only the three standard axioms). A certified store: PostgreSQL triples of address, content and receipt; in the Lean model both a list store and a finite-map store, the latter with write-once as a theorem. A loop with one output type of four constructors, each carrying an address: derived, proposal, vacancy, attested. A separate verifier that re-reads store and ledger before any emission and refuses rather than rewrites. An append-only ledger: canonical question, address, who, date, previous hash, hash (sha256), HMAC signature; two settlements of the same question both remain, the latest operative.

Guarantees. On the finite-map store the gate invariant is an equivalence: the loop emits exactly what the store holds, in both directions, and a vacancy occurs exactly when the address is empty. Fabrication is impossible by the output type: no constructor exists without an address. For the ledger, a whole-chain prefix theorem and a last-entry-changes-the-final-hash theorem are proved, the latter under an explicit injectivity hypothesis — without collision-freeness it is false, and the proof says so.

Binding to production. The Lean kernel and the production JavaScript kernel print the same 30,572 addresses with zero differences. One control command rebuilds the proofs, prints the axiom report, runs the differential and canon tests, and walks the ledger chain, stopping at the first red step.

Measurements. The pre-registered twenty-question placement test (plan hashed before the run, threshold 14) scored 15 of 20; a weakness flag at cosine 0.32 caught four of the five misses. On the archive itself, merging was measured: of 4.08 million distinct addresses built by independent sources, 0.06 percent are shared — federation by plain union, demonstrated at scale. And one experiment failed honestly: a byte-level model with computed addresses, prefix routing and native ternary weights, trained by gradient descent, reached 2.96 bits per byte against 1.39 for a conventional small model. Booked, archived, and read in Section 7.

One residual gap, stated. On open language, production resolves the deepest inhabited image of the computed address and ranks within it by cosine similarity; the gate equivalence holds exactly for exact addresses, and as an implication on images.

6. The correlation, term by term

Left brain = the long chain in the language layer. The 2025 essay says: discrete at every joint, brilliant emissary, blind to the whole. The net says what that is: computation held in the talkative, shallow layer, with assigned addresses (a vocabulary table someone maintains) and a deep continuous register (16-bit floats). Everything that requires an institution. One statement, two depths.

Energy ceiling = register depth, as a constitutive relation. The essay calls the ceiling physical. The net gives the number and refuses the arrow. Register depth: 16 bits against $\log_2 3 \approx 1.58$ bits per trit gives $16/1.58 \approx 10.1$. Measured energy of a full-precision model against a native ternary model of the same class: 0.258 against 0.028 joules per token, ratio ≈ 9.2. The same magnitude read twice; the ceiling *is* the register depth. No mechanism story is needed, and none is offered.

Coherence = selection by fit. The Resonant Stack's principle — only what synchronises persists — and the net's core operation — only what returns in phase remains — are one operation. The transformer already performs it in crutch form: the top-k gate keeps what fits and drops the rest, but still pays a learned score to decide fit. Deterministic-routing results show the score can go; the operation stays.

Nilpotent kernel = balanced closure = the trit. $N^2 = 0$ says: admissible states sum to nothing. Balanced ternary is the discrete object with exactly that property, as Section 4.1 derived on the number 43. The nilpotent kernel of 2025 and the ternary kernel of 2026 are the same closure form in two registers: continuous field, discrete trit.

Optics = the carrier assumption lifted. The 2025 essay chose photonics because light computes in phase. The 2026 theory derives why that choice was right: there is no separate carrier. "Hardware" is deep, slow winding treated as frozen background; computing in light is computing in the windings themselves. What was a hardware preference in 2025 is a derivation in 2026.

The master's mode = the mesh. Broad attention — holding a whole in view, context with centre — is exactly what the address metric supplies in discrete form: prefix = coarser class, neighbour = relative, truncation proved stable (Section 4.2), so the coarse view is never invalidated by the fine one. Knowing-by-place is not an add-on to the maze; it is the maze.

7. The 2025 claims, measured in 2026

The promised guarantee class exists — delivered first by the discrete hemisphere. Section 2.4's claim — impossible, not filtered — is now machine-checked in production: fabrication is impossible by the output type, the gate holds as an equivalence, and the proofs are bound to the running code by the differential test (30,572 addresses, zero differences) under one command. The promise was written for the field kernel; it was first cashed by the ternary kernel. Same closure form, same guarantee class: identity across scale, demonstrated.

The corpus callosum is running. The 2025 interface pattern — the model proposes, the core disposes — is the deployed design of maze.swarp.nl/ai: the language model's answer enters as a typed proposal, may phrase and may not assert, and leaves only as store-backed content or as a vacancy. The boundary between the hemispheres is not a metaphor in this construction; it is one named predicate — the total meaning relation over open language, unformalisable after Tarski — and everything on the machine side of it is proved. The clinical fact of Section 2.3 explains the build order: an articulate left partner had to exist first, and this one, by construction, cannot confabulate — the pathology of the isolated left hemisphere is exactly what the claim discipline removes.

The three constraints, answered discretely. Temporal fragmentation: depth is timescale — shallow prefixes are the slow modes, proved never invalidated, so long-term memory is intrinsic to the mesh; the ledger adds the slow mode of settled meaning. Loss-function myopia: the answer loop has no loss function — admission by certificate, selection by prefix, settlement by a named human. The energy ceiling: confirmed from outside by the 9.2 ratio and by the industry's own behaviour, dismantling its tables, scores and register depth under exactly the pressure the essay named.

The learning claim, measured in the negative direction. The 2025 essay says learning is not gradient descent. One experiment kept gradient descent while adopting the resonant register, and reached 2.96 against 1.39 bits per byte. The register was not refuted; the method was located: the loss function is the last left-brain remnant, and a right-brain register trained by a left-brain method does not cross. The field's own ternary successes still use backprop; nobody yet trains by coherence. That open instrument now has a first concrete form — the next section.

The substrate remains open, and arrives from outside. The photonic field, the resonant being, the multi-scale coupling are unbuilt here. The 2025 essay's own hedge holds: photonics is coming either way; every published photonic-computing result is free measurement for this claim. Meanwhile the discrete hemisphere gives the right-brain functions provisional homes: neighbourhood ranking within the image of an address, a weakness flag at a fixed coherence threshold, human settlement recorded as an attributed event.

8. The gap the building exposed: the missing product rule

On 20 September 2026 the engineering asked the theory a precise question, and the question names the one theoretical gap of this whole line. It is stated here in full, so this essay carries it.

The question. The loop currently recognises by similarity: the cosine between the question vector and the answer vector separates good from bad answers with AUC 0.87 on 200 blind questions (AUC: place one good and one bad answer side by side; the measure ranks the good one first in 87 percent of pairs). The theory says the criterion should be resonance, not similarity. Two direct translations were tried and measure nothing: the trit sum of the pair scored AUC 0.41, the closure of the pair 0.53 — coin-flip level. Requested: one rule that assigns each answer one number, "how well does this resonate", with its direction, computable from the question vector (1,536 numbers, unit length) and 24-trit address, the answer's vector and address, the answer's neighbours (names, addresses, vectors), and the timestamps of admission, asking and settlement.

Why the two candidates had to fail — the theory says so itself. The address is the recorded route, the place in the mesh; it is not a tension state, so arithmetic on two labels measures nothing. And closure, in the net, is not a property of a pair; it is a property of a ring — a winding closes through its surroundings. Resonance is returning in phase *via the neighbourhood*: exactly the ingredient both failed candidates ignored. One arithmetic trap is worth recording: for unit vectors, the "energy-momentum" candidate $E^2 - p^2$ is the old measure in disguise. With cosine c between q and a: $E^2 = |q + a|^2 = 1 + 2c + 1 = 2 + 2c$, and $p^2 = |q - a|^2 = 2 - 2c$, so $E^2 - p^2 = 4c$ — four times the cosine, nothing new.

Why the gap is exactly the 1880s amputation. Look at what has been proved: place (the address bijection), closure (truncation as rounding), union (merging). All statics — properties of one state. What the question asks for is how two windings *compose* — question and answer, through the mesh. A composition rule is a product, and the product is precisely what Heaviside and Gibbs cut out of Maxwell's form (Section 3). The quaternion product of two windings delivers both answers at once: a scalar part — how well they align (the cosine) — and a vector part — the axis along which they fail to align. The current loop computes only the scalar part. AUC 0.87 on the cosine is the Heaviside amputation, re-performed inside the embeddings. The theory's missing sentence has a name: **the product rule of the net is not yet written down.**

The first step back, registered for measurement. Resonance = the ring product closes. The cheapest computable form of "closing via the neighbourhood" is a two-step passage: with $n_1 \ldots n_k$ the k nearest neighbours of the answer a (k = 12),

$$R = (1/k) \cdot \Sigma_i \cos(q, n_i) \cdot \cos(n_i, a),$$

direction: large is resonating. Reading: the question must not only strike the point — the cosine already measures that — it must return through the point's surroundings. The three miss types of the twenty-question test each receive their own verdict: a name hit without a carrying neighbourhood scores near zero however similar the point looks; a too-general lemma in a neighbourhood pointing away from the question scores low; a well-placed answer in a co-rotating neighbourhood scores high. For the field version, the phase of a piece x in region u (the prefix at depth 6, with u the normalised sum of the region's inhabitant vectors): set e_1 = u and e_2 = the normalised component of q orthogonal to u; then $\phi(x) = \mathrm{atan2}(x\cdot e_2, x\cdot e_1)$, and the field measure is $r \cdot \cos(\phi(a) - \phi(q))$ with r the order parameter over the neighbour phases — the neighbourhood must cluster in phase *and* the answer must lie on the question's side.

The protocol, fixed in advance. The plan is hashed before the run; R is computed for the same 200 blind questions; the measure is AUC, alone and in combination with the cosine. At or above 0.70, R enters the loop as a gate measure beside the cosine, under a new hashed plan. Near 0.50, the rule measures no correctness — and that too is booked: then this scalar path-product does not carry resonance, the translation was wrong rather than the theory, and the search moves to the time layer or to the full quaternionic rule — scalar and vector part together: not only how much question and answer differ, but whether their deviations turn the same way through the neighbourhood. Either outcome is worth more than ten essays.

9. The state of the future AI

The formula of the whole line is now three terms: **LAI + the connection + RAI.** LAI — the honest left hemisphere — is finished and proved: the store, the gate, the ledger; it fabricates nothing and everything is traceable. The connection — the corpus callosum — runs in production as the proposal-and-gate pattern. RAI is the one missing organ, and it is buildable today as a simulated field on ordinary hardware: a Kuramoto field with one oscillator per inhabited depth-6 prefix costs linear time in mean-field form, so a laptop carries hundreds of thousands of oscillators; the nilpotent admission rule (propose, compute the balanced sum, admit only net zero) is pure algebra. What a laptop cannot do is *be* the field: simulation computes the right dynamics on the bit machine — the left brain playing the right brain — and yields the behaviour without the energy gain. The claimed factor of 1,000 to 10,000 in energy-delay arrives only when the physics does the computing, in photonic phase; that is the next tread of the 2025 roadmap and is arriving from outside either way. The field advises, the gate decides; there is no second output path, and none will be added.

10. Status, stated once

Theorems, machine-checked: address bijection, range and uniqueness, truncation as rounding, merge as union, the gate equivalence on the finite-map store, write-once, claim completeness by type, the ledger prefix and last-entry theorems (the latter under an explicit injectivity hypothesis). Measurements: AUC 0.87 for the cosine on 200 blind questions; 0.41 and 0.53 for the two failed resonance candidates; 15 of 20 placement against a pre-fixed threshold of 14; 30,572 differential addresses with zero differences; 2.96 against 1.39 bits per byte for the failed coherence-register training; 0.06 percent overlap on 4.08 million merged addresses; 0.258 against 0.028 joules per token (external). Readings: the term-by-term correlation of Section 6, and the reading of the failed training in Section 7, are offered as the diagnosis this essay exists to state — not as theorems. Open instruments: the R measurement of Section 8 (protocol fixed, outcome pending); the full quaternionic product rule; training by coherence; the oscillatory substrate; the token-axis question on ternary registers raised by Ouyang et al. The semantic boundary is classical logic applied and is not open.

11. Conclusion

The 2025 essay asked for a whole-brain infrastructure and warned against an emissary ruling alone. The 2026 construction answers with the half that could be proved: a left hemisphere that no longer confabulates, holding the mesh as its memory and one named predicate as its boundary. The connection it asked for runs. The field hemisphere remains the open half, and the building work has now told the theory exactly which sentence it still owes: the product rule — the composition that Hamilton had, Maxwell used, and the 1880s cut away. One strand, two windings, one theory, one missing rule, and a measurement already scheduled that will grade its first candidate.

Annotated References

1. **Konstapel, J., "Left and Right Brain AI", constable.blog, 23 November 2025.** *Why read?* The primary reference: the full field-side architecture restated in Section 2 — both stacks, the interface patterns, the promised guarantee class, the roadmap. https://constable.blog/2025/11/23/left-and-right-brain-ai/

2. **Konstapel, J., "The Resonant Stack: A Paradigm Shift from Discrete Logic to Oscillatory Computing", constable.blog, 19 November 2025.** *Why read?* The core exposition of the field hemisphere: Kuramoto dynamics, KAYS functions, oscillatory computation.

3. **Konstapel, J., "Right Brain AI — Technical Architecture", constable.blog, November 2025 (PDF).** *Why read?* The specification level of the field hemisphere; companion to reference 1.

4. **Hamilton, W. R., "Lectures on Quaternions" (1853).** *Why read?* The source algebra of Section 3: one product carrying alignment and axis together. *Advice:* a modern introduction to quaternions and rotations suffices.

5. **Maxwell, J. C., "A Treatise on Electricity and Magnetism" (1873).** *Why read?* The field equations in the quaternion form the 1880s abandoned — the shape of the theory before the amputation.

6. **Crowe, M. J., "A History of Vector Analysis" (1967).** *Why read?* The clean historical account of how Heaviside and Gibbs split the quaternion product into dot and cross, and why physics adopted the pieces. The documentary basis of Section 3.

7. **Rowlands, P., "A Universal Algebra and Rewrite System Approach to Physics" (2002); Rowlands, P. & Diaz, B., "Aspects of a Computational Path to the Nilpotent Dirac Equation", Foundations of Physics 37 (2007).** *Why read?* The nilpotent principle $N^2 = 0$, quaternionic in construction — the kernel of the 2025 stack, identified in Section 6 with the balanced register. *Advice:* the 2007 paper is the more readable entry.

8. **McGilchrist, I., "The Master and His Emissary" (2009).** *Why read?* The hemispheric synthesis behind the design heuristic, including the clinical fact this essay uses twice: the isolated left hemisphere confabulates.

9. **Kuramoto, Y., "Self-Entrainment of a Population of Coupled Non-Linear Oscillators" (1975).** *Why read?* The synchronisation model written out in Section 2.2 — selection by fit in continuous form, and the order parameter used in Section 8's field measure.

10. **Knuth, D. E., "The Art of Computer Programming", vol. 2, §4.1.** *Why read?* Balanced ternary: uniqueness, range, and the radix-economy computation of Section 4.1. Knuth calls it the prettiest number system there is.

11. **Tarski, A., "The Concept of Truth in Formalized Languages" (1936).** *Why read?* The theorem behind the one-predicate boundary where the hemispheres meet. *Advice:* a modern exposition suffices.

12. **Konstapel, J., "Where Computation Comes From" (2026).** *Why read?* The transformer read as a net: the four operations already approximated, the two remnants, and the constitutive relation of Section 6 derived in full. https://claude.ai/artifact/2KUPC9w6iqpWM4LxtLZurv

13. **Konstapel, J., "Provable AI — A Claim-Bounded Criterion, a Semantic Boundary, and a Machine-Checked Construction" (2026).** *Why read?* The criterion and boundary that make the left hemisphere provable, with the Lean core as appendices. https://claude.ai/artifact/PSG2144zBPND2jQvnfMszE

14. **Konstapel, J., "MAZE.AI: A Claim-Bounded Answer Loop with Machine-Checked Provenance", whitepaper, 20 September 2026.** *Why read?* The running system and every 2026 number of Section 5. Live at maze.swarp.nl/ai.

15. **Ma, S. et al., "BitNet b1.58 2B4T" (Microsoft Research, 2025).** *Why read?* The measured ternary register: 0.028 against 0.258 joules per token — the external half of the constitutive relation.

16. **Ouyang et al., "Low-Bit Quantization Scaling" (ACL 2025).** *Why read?* The honest counter-measurement: the low-bit gap grows with training tokens. Listed open in Section 10.

17. **DeepSeek-AI, "DeepSeek-V3 Technical Report" (2024) and its open inference code.** *Why read?* The left hemisphere examined at source: rotary phases, top-k gate, low-rank truncation, one residual stream — the four net operations in crutch form.

18. **Klein, G. et al., "seL4: Formal Verification of an OS Kernel" (2009).** *Why read?* The methodological model for proving a running system and stating exactly what the proof does not cover.

19. **Fawcett, T., "An Introduction to ROC Analysis", Pattern Recognition Letters 27 (2006).** *Why read?* The AUC measure used throughout Section 8: what 0.87, 0.70 and 0.50 mean, in four pages.