

The Coordination Limit

The Internet of Everything Cannot Be Built Under Current Architectural Assumptions. Here Is the Math.



THOMAS ROCHA III

APR 01, 2026

The Atlantic published on March 26. “Welcome to a Multidimensional Economic Disaster.” The IEA projects that data center electricity will double to 945 TWh by 2030. Goldman Sachs calls AI’s 2025 GDP contribution “basically zero.” GTC 2026 formalizes tokens-per-watt as the governing CEO metric. The EU is actively drafting its Data Centre Energy Efficiency Package, with a consultation open through June.

The IETF archive is full of drafts describing session-persistence failures, Zero Trust validation gaps, and incremental state-synchronization problems. GitHub is full of proposed fixes. Standards working groups across telecom, cloud, and industrial automation are independently naming fragments of the same problem in their own domain vocabulary.

Nobody has named the architectural source at the interaction boundary, across domains, as a unified constraint.

The architecture is burning the watts before useful work begins.

In identity reconciliation, policy evaluation, and cross-boundary synchronization. Before the model runs. Before a token is produced.

When AI systems coordinate across independently governed identity, policy, modality, authority, and transport boundaries, the coordination cost scales as a Cartesian product of those dimensions:

$$\mathbf{C_frag} = \mathbf{k} \times \mathbf{P} \times \mathbf{M} \times \mathbf{F} \times \mathbf{A} \times \mathbf{T}$$

Where k is per-surface cost, P is participants, M is modalities, F is features, A is authorities, and T is transports.

In an interaction-scoped architecture, where all five dimensions bind to a single persistent session identity:

C_{ssoar} = k × (P + M + F + A + T)

The difference between a Cartesian product and a sum. At the IoE scale, that difference becomes the difference between buildable and not. **These constraints reduce the viable architecture class. That is the structural claim.**

The Numbers

Published December 2025. Before the Hormuz closure. Before GTC.

	Coord. Surfaces	Power (overhead)	SSOAR Surfaces	SSOAR Power
2025	~10 ¹²	~1 GW	~10 ¹⁰	<0.01 MW
2030	~10 ¹⁴	~100 GW	~10 ^{10.5}	<0.1 MW
2040	~10 ¹⁶	90-100 GW	~10 ¹¹	~0.9 MW

90 to 100 gigawatts for coordination overhead alone. Before payload runs. Roughly double today’s entire global data center capacity, consumed before a single token is produced.

These are scaling projections, not measured values. The conclusion rests on the structural divergence between multiplicative and additive coordination scaling. Conservative cadence assumptions were used deliberately. Even if every numeric estimate here is wrong by an order of magnitude, the structural conclusion holds: the divergence between a product and a sum dominates at scale regardless of the constant.

The Hormuz Acceleration

The architectural case stands independently of current events. February 28, 2026, did not create the problem. It removed the ability to mask it with more infrastructure spend.

Jensen Huang at GTC 2026:

Revenue = (Tokens per Watt) × (Available Gigawatts)

Hormuz capped the right side. QatarEnergy declared force majeure on long-term LNG contracts. Trains 4 and 6, 12.8 million tons per annum, are offline for a minimum of three to five years. The energy denominator against which every hyperscaler capital program was priced no longer exists.

The left side also has a floor. One imposed by architecture, not silicon.

Hardware efficiency improves the watts that reach useful work. Coordination overhead reaches nothing. No chip roadmap moves that floor.

Usable tokens are created by the loss of usable wattage.

What the Current Conversation Is Missing

The Atlantic piece correctly named the debt structure, the token deflation spiral, the supply chain fragility, and the physical exposure. The Spyker piece at a16z correctly translated Hormuz into infrastructure terms. The IEA correctly identified the energy ceiling. The EU correctly identified the regulatory gap.

What none of them have said: a meaningful fraction of every watt in this conversation is consumed before it produces anything. Not by the models. By the architecture holding fragmented systems together.

The IETF drafts describe the symptoms. The standards working groups are naming fragments. The proposed answer to the whole problem, an

interaction-scoped control plane that collapses the product to a sum, has been in the public record since December 2025.

Sora Is the Live Exhibit

OpenAI shut down Sora on March 24, 2026. Inference costs: approximately \$15 million per day. Total lifetime revenue: \$2.1 million. Disney had committed \$1 billion to the partnership and found out the product was being killed less than an hour before the public announcement. The deal died with it.

The immediate reading is computer economics. Video generation is expensive. The unit economics never worked.

Apply the formula.

Dimension	Sora at launch
P: Participants	Users, OpenAI platform, Disney as licensor, moderation systems, safety systems, API consumers: 6
M: Modalities	Text input, video output, audio, face scanning, image-to-video: 5
F: Features	Generation, audio, cameos, Disney character licensing, content moderation, IP compliance, safety filtering, gallery, API, editing: 10
A: Authorities	User consent, Disney IP, platform policy, content safety, copyright compliance, DMCA: 6
T: Transports	Mobile app, web interface, API, sora.chatgpt.com, community gallery: 5
C_frag	$k \times 6 \times 5 \times 10 \times 6 \times 5 = k \times 9,000$ coordination surfaces
C_ssoar	$k \times (6 + 5 + 10 + 6 + 5) = k \times 32$ coordination surfaces
Ratio	281 to 1 — per session, every user, every frame

For EU users, the number is worse. The European Accessibility Act, which took effect on June 28, 2025, three months before Sora’s public launch, required synchronized captions, audio descriptions, and signing avatars as a precondition for market access. That adds three modalities and four governing authorities to the EU-user session surface.

Dimension	EU user adjustment
M: Modalities (EU)	Add caption stream, audio description, signing avatar: 8 total
A: Authorities (EU)	Add EAA, WCAG, member state implementations, conformance validation: 10 total
C_frag (EU)	$k \times 6 \times 8 \times 10 \times 10 \times 5 = k \times 24,000$ coordination surfaces
Ratio (EU)	750 to 1

The architecture lacked a mechanism to scope accessibility governance to users who required it without applying it universally, because there was no session-level identity capable of carrying jurisdictional context. Every session paid the coordination cost, whether the user was in Dublin or Dallas.

EU users accounted for roughly 20-25% of the consumer AI user base. The accessibility obligation applied to that fraction. The coordination overhead is applied to everyone.

Every Disney character invocation triggered IP authority verification against platform policy, against content moderation, against user consent, and against API governance. Independently. Multiplicatively. The cameos feature alone hit five of those six authority dimensions simultaneously at every session initiation.

The economics could not close because the coordination surfaces kept expanding faster than revenue. Not because video is expensive to generate. Because the architecture made every new partner, every new modality, every new governance requirement a multiplier rather than an addition.

Disney found out less than an hour before the public because there was no unified session governance capable of propagating a state change across all participants with a bounded blast radius. That is not a communication failure. That is the coordination limit, live, in a product that had Bob Iger's signature on it six months earlier.

The Validation Landscape

These are not corroborating opinions. They are independent communities arriving at the same wall from different directions, in the same 30-day window, without a shared vocabulary for what they are hitting.

- IEA Energy and AI report: demand doubling, efficiency gains overwhelmed by growth
- Goldman Sachs: AI's economic contribution near zero due to structural leakage
- GTC 2026: tokens-per-watt formalized as the primary infrastructure metric
- Atlantic, March 26: polycrisis named, architectural source not reached
- a16z, March 27: Hormuz translated for the tech audience
- EU Data Centre Energy Efficiency Package: active, vocabulary for coordination overhead as a distinct category does not yet exist in the draft

The problem is visible. The source is not named. The standards vocabulary does not yet exist for it.

The Position

The filing dates do not move.

A proposed architectural answer is in the public record: the SSOAR primitive, Session-Scoped Orthogonal Authority and Responsibility. IETF Internet-Drafts timestamped December 2025. SSRN preprint. Ten patent filings, the foundational application in Track One, accelerated examination.

The EU consultation on post-2030 energy efficiency frameworks closes June 12, 2026. The codes of conduct required by the EU AI Act for energy-efficient AI do not yet exist. The technical vocabulary for writing them is not yet in the standards community.

That is the gap. It has a location and a timestamp.

hermes-echo.com