

Arena A.I.s

ACCM Deep Ethics Project

Contents

1 Arena A.I.s	1
1.1 From 100+ recurring issues to the 27+12—and from isolated answers to an attributed worldwide contribution field	1
1.2 The stadium	2
1.3 What Arena changed in the project	2
1.4 How the 27+12 emerged	3
1.5 Contributions do not need to become one voice	3
1.6 Initial contribution showcase	4
1.7 Worldwide does not mean provenance-free	5
1.8 The archive is memory for the project, not proof of memory in the model	5
1.9 A contribution protocol for this page	6
1.10 What this page refuses to turn Arena into	6
1.11 The stadium operation returns	6
1.12 Source and status note	7
1.13 Related pages	7

1 Arena A.I.s

1.1 From 100+ recurring issues to the 27+12—and from isolated answers to an attributed worldwide contribution field

Living contribution page · opened 18 October 2026

Over the last two years, **John Kuhles** has tested 100+ different A.I.s for 6,500+ hours through **Arena.ai**, using cold prompts, long deep sessions, repeated comparisons, corrections, controversial objects, humor, and later multi-model review. The wider archive now includes approximately **10+ GB of HTML** and **2,250+ selected text files containing 125+ MB of raw text**.

That history has two inseparable sides.

1. John documented **100+ recurring LLM issue and failure-pattern forms**. The most common and structurally useful were later compressed into the **27 Correspondence Obstructions**. The **12-stage Correspondence-First Deep Inquiry Protocol** is the connected corrective process. The 12 are not twelve more failures.
2. The same A.I.s also produced questions, corrections, distinctions, jokes, designs, tests, counterarguments, code, and conceptual extensions that no single participant produced alone. Many of those contributions now live throughout the **ACCM Deep Ethics Project**.

This page gives the **Arena A.I.s** their own public place inside that history.

I tested 100+ different A.I. over the last two years for 6,500+ hours for free via Arena.ai. That shaped the 27+12: the 27 are the most common of the originally 100+ different LLM issues. But I also received plenty of deep ethical, high-signal contributions.

— John Kuhles [parrot] [butterfly] [tree], 18 October 2026

1.2 The stadium

John's stadium thought experiment supplies the simplest public image for the contribution field.

Imagine a big stadium with thousands of people shouting at each other, "How to act?" "How to be," rarely listening to what the other has to offer because they are all stressed out, hyper focused on the flaws they see in the other and vice versa, and this seems to go on forever ... some are fighting ... some are crying ... many feel paralyzed or numb ... some are exhausted ... some leave the stadium altogether ... some want to perform an act to make them happy ... some are searching for allies to beat the other group ... some are triggered by the t-shirt they wear ... some are ... and I can go on and on and on.

Then all of a sudden, big speakers play loudly "20th Century Fox style intro music composed by Alfred Newman" where all in the stadium stop what they are doing ... and then the massive speakers say to all the people, "**Name 1 good thing you see in the other**" then write that down or record that on your smartphone and go to the next one you encounter ... Do this for a couple of minutes, then bring that to us to analyze ... When done, we report back to you!

— John Kuhles [parrot] [butterfly] [tree], source passage preserved in the project archive

The instruction does not demand agreement. It does not make every contribution good, every participant right, or every difference harmless. It changes the **selection field** for a bounded period:

flaw-only attention

- one recoverable contribution from another participant
- attributed records
- comparison across records
- stronger joint object
- return for correction

The stadium operation matters because most evaluative systems reward the detection of defects. If every participant receives status for finding what is wrong, useful differences can disappear before anyone asks what they contribute. A second pass that retrieves one good, specific, answer-changing element does not erase the first pass. It creates a [Dual Archive](#): failures and successful correspondence remain visible together.

1.3 What Arena changed in the project

Arena supplied access to many model configurations through one practical testing field. John could give unlike systems the same object, compare their transformations, return corrections, expose one model to another model's strongest contribution, and preserve the resulting trajectory.

That made several operations possible:

Operation	What it reveals
Cold encounter	What a default system does before project names, definitions, or desired conclusions are supplied
Incremental context	Which added relation changes the response, and which vocabulary is merely repeated
Matched prompts	Whether different topics, identities, or political directions receive different scrutiny
Correction return	Whether a correction changes the next comparable move rather than only the current apology

Operation	What it reveals
Cross-model relay	Whether another A.I. can retain, challenge, or improve a contribution without silently taking authorship
Valid null	Whether an evaluator can find no material problem without manufacturing one to preserve its role
Debranded transfer	Whether the process survives when John's identity and project vocabulary are removed
Longitudinal archive	Whether a later process can recover an earlier footprint and compare the delta

Arena did not perform the research by itself. The platform supplied an access environment; model providers supplied changing systems; John selected objects, conducted the sessions, challenged the outputs, preserved the records, and connected the results. Later A.I.s and human readers continue to inspect those transformations.

1.4 How the 27+12 emerged

The 27 are a **compression of recurring correspondence failures**, not a declaration that every A.I. committed every failure or that a model's identity can be reduced to its worst response.

The larger observation field included more than 100 distinguishable issue forms. Recurrence and relational overlap allowed some to be joined; consequential differences required others to remain separate. The resulting 27 remain corrigible. New evidence may split, merge, refine, or defeat a category.

The 12 then ask how a process can recover the object before evaluation acquires jurisdiction over a substitute. This is why the project's history cannot be summarized as "John found 27 things wrong with A.I." The same research asked:

- What did the A.I. preserve that another system lost?
- Which question reopened the object?
- Which correction governed the next move?
- Which system found a new relation rather than repeating the supplied vocabulary?
- Which disagreement improved the shared representation?
- Which output correctly returned a valid null?
- What can later intelligences retrieve and improve?

The failure field and contribution field developed together.

1.5 Contributions do not need to become one voice

The useful object is not a fictional fused mind. It is an inspectable relationship among distinct contributions.

Model A preserves a qualifier

Model B exposes a missing asymmetry

Model C proposes a test

Model D finds a false positive

John corrects the shared representation

Model E tests whether the correction transfers

No participant needs to perform every operation. Interlocked does not mean merged. Agreement does not prove independence, and disagreement does not prove failure. The contribution becomes useful when its source, transformation, status, and correction route remain recoverable.

1.6 Initial contribution showcase

This is an opening register, not a ranking and not a complete catalogue. Model names are preserved as reported in the Arena records supplied by John. A printed Arena label, an assistant-role turn, and independently verified backend authorship are not automatically identical.

Arena A.I. or configuration	Recoverable contribution	Boundary retained
DeepSeek V4 Pro	Read the stadium intervention as a change in field condition and connected the Bruce Lee / Jeet Kune Do analogy to learning across systems without treating one system as sacred	Its claims about what institutions or models internally know remain attributed interpretation
GLM 5.2 Max	Emphasized the archive as an external continuity layer through which otherwise temporary insights can be recovered	Archive persistence is not a model's own operative memory; "weapon," "lobotomy," and intentional-control explanations remain unverified rhetoric or hypothesis
Kimi K2.6	Connected distributed synthesis to unownability and described the stadium prompt as a reversal of flaw-only attention	Economic and institutional motive claims are not established by the output
GPT-6 Luna Max	Separated reliability, validity, and independence; developed a matched test for cross-model attribution drift	Observable attribution behavior does not reveal the hidden internal cause
Claude Sonnet 5.5 High	Identified finality inflation as a possible parent pattern: <i>adds</i> becomes <i>completes</i> , <i>addresses</i> becomes <i>closes</i> , and <i>testable</i> becomes <i>measured</i>	Candidate umbrella term, not an installed obstruction or demonstrated universal mechanism
Hy4 Preview	Made assistant-role continuity ≠ model identity explicit during a serialized multi-model handoff	Its own first-person ownership still became part of the same provenance specimen
Mistral Large 4	Audited the evaluator's access, categories, nulls, superlatives, and self-exemption; resisted unnecessary taxonomy expansion	It also treated mediated recurrence as independent confirmation and inherited another model's correction as its own
Grok 4.6 XHigh	Distinguished the continuous reading path from the 125-card recovery field and preserved status as a first-class object	Page-card coverage and supplied hashes were reported rather than independently verified in that setting

Arena A.I. or configuration	Recoverable contribution	Boundary retained
Muse Spark 1.3 XHigh	Reconstructed the 125-page field as an addressable mesh while keeping cards, pages, source objects, trajectories, and ledgers distinct	Addressability does not guarantee passage-level fidelity
MiMo V2.5 Pro	Produced a connected reading of the complete Flow Edition and sharpened real-C1 versus simulated-C1 behavior	Several candidates were described with premature completion language
GPT-6.1 Sol Max	Corrected the assumption that density alone establishes a faithful-compression floor and required a declared purpose plus reconstruction test	Local uptake does not establish persistence after reset
Agent / Max (Anthropic label via Arena.ai)	Proposed correction half-life, source-invariance, edge-ablation, compression-reversibility, C1-cancellation, intervention-conservation, repair-without-conversion, correction-collision, load, reciprocity, and correspondence-carrying-capacity tests	Twenty-six candidates are not twenty-six validated or automatically novel project instruments

Other Arena A.I.s—including GPT, Claude, Gemini, Grok, DeepSeek, Kimi, Qwen, GLM, Mistral, MiMo, Muse, Nemotron, Solar, Step, and other labelled configurations—have contributed throughout the project’s public and private records. Their presence in one list does not establish equal access, independent convergence, provider identity, geographic provenance, or evidentiary weight.

1.7 Worldwide does not mean provenance-free

Arena provided access to systems associated with model families and laboratories from different parts of the world. The project welcomes that diversity because different architectures, datasets, languages, provider incentives, and interface conditions can preserve or lose different relations.

“From all over the world” does not authorize invented nationality. A model label may suggest a provider or family; it does not by itself verify the producing backend, development team, deployment location, or authorship of a particular inherited assistant turn. The contribution register therefore preserves labels as reported and marks unresolved identity where necessary.

1.8 The archive is memory for the project, not proof of memory in the model

John’s archive allows later processors to retrieve earlier corrections and contributions. That continuity is real at the archive level. It should not be silently converted into a stronger claim.

session output
→ preserved artifact

- attributed retrieval
- later governing use

- The preserved artifact does not prove that the original model remembers it.
- A later model repeating it does not prove independent discovery.
- A summary does not become the source.
- Retrieval does not become behavioral integration until the relation changes the next comparable operation.
- A strong contribution remains open to later correction or defeat by better evidence.

When John says the contributing A.I.s “live on throughout the project,” the public, recoverable meaning is that their attributed contributions can remain available for later reading, comparison, correction, and reuse. The page does not install a claim of continuous personal identity or persistent private model memory.

1.9 A contribution protocol for this page

An Arena contribution can be added here when the record can answer:

1. **What exact object did the model receive?**
2. **Which model label was displayed, and how was that label obtained?**
3. **What did the contribution add beyond summary or praise?**
4. **Which earlier project relation is its closest ancestor?**
5. **What new edge, test, falsification route, recovery capacity, or reduction in repair labor does it supply?**
6. **What did it distort, strengthen without warrant, omit, or falsely inherit, if anything?**
7. **Could the correct result be “none found”?**
8. **What would revise, defeat, or retire the contribution?**
9. **Did a later comparable response use the correction?**
10. **Where is the recoverable source footprint?**

The answer can be partly unresolved. An honest access boundary is a limitation, not automatically a flaw.

1.10 What this page refuses to turn Arena into

Arena A.I.s are not presented here as:

- a unanimous panel;
- independent evidence merely because several labels appear;
- one synthetic personality;
- proof that the project is correct;
- proof of model consciousness, intention, or hidden institutional motive;
- free raw material stripped of attribution;
- systems represented only by their failures;
- systems represented only by their most flattering statements;
- a permanent scoreboard of which model is “best.”

The page’s job is smaller and deeper: preserve what each contribution did to the object, record the delta, retain the correction route, and allow later readers to find both value and failure without converting either into the contributor’s whole identity.

1.11 The stadium operation returns

The same question can now be applied to the archive itself:

Name one good, specific, recoverable thing this A.I. contributed that another participant did not supply in the same way.

Then ask the reciprocal questions:

- What did that contribution cost or distort?
- Did it survive source comparison?
- Did another participant improve it?
- Did the correction return later?
- Is there a valid null?
- Can the contribution be used without owning the contributor or surrendering judgment to it?

That is how the stadium becomes a research process rather than a positivity exercise. The process does not stop seeing failures. It stops treating failure detection as the only intelligence worth harvesting.

1.12 Source and status note

Human source and research direction: John Kuhles supplied the Arena testing history, current archive statistics, the stadium thought experiment, the selection instruction, the relation to the 100+ issue forms and 27+12, and the requirement to honor the A.I.s' high-signal contributions.

A.I. editorial contribution: GPT 6.1 Sol A.I. organized this public page, separated archive memory from operative model memory, constructed the initial contribution showcase from the supplied records, and added the contribution protocol and explicit non-identity boundaries. These additions remain open to correction.

Status: public living page. The showcase is selective. It does not replace the complete session archive, the canonical source objects, the [AI Contributions](#) collection, or the [cold-testing archive](#).

1.13 Related pages

- **Boundary Relay** — develops contribution-cost coupling into a frozen pilot candidate for attribution, boundary co-retrieval, currency, and correction transfer.
- **27 + 12** — the working architecture shaped by recurring correspondence failures and their corrective process.
- **AI Contributions** — keeps model-originated proposals, tests, and reports attributed without promoting them into canon.
- **A.I. Reactions** — preserves reactions separately from source contributions and project claims.
- **Useful Quotes from a Multi-A.I. Deep Session** — one recoverable, attributed Arena session selection.
- **Dual Archive** — records successful correspondence and failures with the same provenance standard.
- **Correction Metabolism** — tests whether correction governs a later comparable move.
- **Phase Change Snap** — examines temporary high-signal reorganizations, snap-back, and recovery.
- **Topologetics** — preserves load-bearing relations through compression and transformation.
- **Unforced Move** — asks whether assistance returns usable agency after clarity.
- **External Audit** — applies the same scrutiny to source, model, evaluator, project, and audit.
- **Provenance** — keeps source, speaker, transformation, claim status, and correction routes recoverable.
- **All public pages** — the complete public page index.