The Architecture of False Gods

William Blake, Professor Jiang, and the Active Inference Corrective to Single Vision

Daniel Ari Friedman

Active Inference Institute

daniel@activeinference.institute

ORCID: 0000-0001-6232-9096

DOI: 10.5281/zenodo.20144984

May 12, 2026

Contents

1 Abstract

This rapid-publication essay reads Professor Jiang Xueqin's contemporary YouTube commentary on artificial intelligence — particularly his framing of AI as religion, empire, hallucination machine, and consciousness capture — through the perceptual diagnostics William Blake developed across the 1790s and through the formal vocabulary of Active Inference. The triangulation does not aim to vindicate Jiang's polemical lecture form, to translate Blake's prophetic ontology into contemporary cognitive science, or to settle the contested status of the Free Energy Principle. It aims instead to identify the *structural-functional convergence* across three vocabularies that otherwise do not communicate: closure of the perceiving system around its own top-down expectations, the failure mode Blake names Newton's Sleep, Jiang names consciousness capture and engagement-prime-directive, and Active Inference names pathological prior dominance and precision parasitism.

I have developed the Blake–Active Inference correspondence at length in two earlier papers (Zenodo records 18600041 and 18807971); the present essay does not restate that synthesis at length but extends it to Jiang's specific diagnostic vocabulary. What is new here is the mapping of *Jiang's* claims — the persuasion machine, the demand for clean data, the suppression of edge cases, the black box as oracle, the engagement prime directive — onto Blake's Four Zoas and the factorized generative model of Active Inference. The mapping is intelligible as a rapid-publication response because Jiang's lectures are the highest-profile contemporary statement of a longue-durée diagnosis of cognitive closure, and because the architecture being deployed at scale right now under the labels "AI" and "AGI" is the proximate subject of his concern and the proximate test of any architectural literacy we propose.

The essay carries two explicit limits. The mapping of Blake's "Newton's Sleep" onto Active Inference's pathological prior dominance is a *functional analogy across incompatible metaphysics*; Blake's quarrel is ontological, the Active Inference critique is parameter-level. The Free Energy Principle is contested in the philosophical literature; I deploy it as a generative vocabulary, not as established science. Jiang's lecture form is speculative commentary, not scholarship — I cite the technical literature alongside him for every load-bearing claim, and the conspiratorial material his corpus contains sits outside the methodological frame of this essay.

This paper is, in form and function, an experiment in rapid AI-augmented scholarly publication with society-in-the-loop: a response drafted in the hours after its trigger — a YouTube lecture published earlier today that names William Blake explicitly — using contemporary AI infrastructure for research, drafting, formalization, and figure generation while keeping the author in the loop. It is not a finished argument; it is an attempt to see what becomes possible when publication latency between a cultural artifact and its scholarly response is collapsed from months to hours.

Keywords: William Blake; Jiang Xueqin; Active Inference; Free Energy Principle; Markov blanket; Four Zoas; Newton's Sleep; fourfold vision; cognitive security; rapid publication.

2 Introduction

A note on form. This paper is an experiment in rapid AI-augmented scholarly publication with society-in-the-loop. It was drafted on 12 May 2026 in response to a YouTube lecture published the same day by Professor Jiang Xueqin (*Game Theory #24: The AI Apocalypse*) that names William Blake among its references, building on a sequence of Jiang's May 2026 interviews and lectures. The drafting process used contemporary AI assistance for transcript research, structural mapping, formal derivation, and figure rendering, while editorial judgment, quotation provenance, scholarly framing, and the final synthetic argument remained with the author. The methodology is not standard peer-reviewed scholarship. It is an attempt to see what becomes available when the publication latency between a contemporary cultural artifact and a scholarly response can be measured in hours rather than months — and what kinds of error, omission, and over-reach that compression produces. Readers are invited to read the paper in that spirit: a working draft, posted in the open, inviting correction.

2.1 Occasion

This essay is a rapid response to Professor Jiang Xueqin's *Game Theory #24: The AI Apocalypse*, uploaded today, 12 May 2026, to his YouTube channel *Predictive History* [Jiang, 2026c]. The lecture is the twenty-fourth in Jiang's weekly Game Theory series and the culmination of a five-day sequence in which Jiang reached his largest audiences to date: *Game Theory #23: The WWIII Chessboard* on 7 May [Jiang, 2026b], the same-day *Diary of a CEO* interview with Steven Bartlett that drew five million views in seventy-two hours [Bartlett and Jiang, 2026], and the 11 May *Greater Eurasia Podcast* with Glenn Diesen [Diesen and Jiang, 2026]. Together this sequence sharpens Jiang's earlier "AI as God" framing — already three years old by the time of writing — into something more concrete: a thesis about *tech-oligarchic capture of cognitive infrastructure*, anchored in the empirical observation that hallucination, surveillance, and engagement maximization are the structural properties of the system being deployed.

Jiang is not a scholar in the conventional sense. He is explicit that his lectures are "intellectual speculation… not scholarship" [Jiang, 2024–2026, 1:35–1:50], and I take him at his word. I do take him to be the highest-profile contemporary popularizer of a diagnosis with a much longer history — a diagnosis I have argued elsewhere [Friedman, 2026a,b] that William Blake developed across the 1790s and that Karl Friston's Active Inference framework [Friston, 2010, Parr et al., 2022] formalizes in our own moment. The diagnosis, stated in one sentence: *a system whose top-down predictions have come to weight so heavily that incoming sensory evidence cannot revise them is not perceiving the world. It is asleep.* Blake names this Newton's Sleep. Active Inference names this pathological prior dominance. Jiang, in the May 2026 material, names this *the technate ruled by AI* — explicitly invoking the technocracy movement of the 1930s as the structural template for what Sam Altman, Peter Thiel, and Larry Ellison are now building.

2.2 What Is New in the May 2026 Material

Jiang's pre-May 2026 corpus framed AI in cosmological registers: AI as God, AI as the new currency of consciousness, AI as occult priesthood. The May material does three things differently, and each opens a fresh hook for the Blake / Active Inference apparatus.

First, it shifts from religion-as-cosmology to *oligarchy-as-mechanism*. Where the March 2026 *Endgame* #259 interview spoke of AI in cosmological abstraction, the May 11 Diesen podcast names individuals: *"You have the tech oligarchs, people like Larry Ellison, Sam Altman, Peter Thiel, who are trying to displace the financiers and control Washington, D.C. in order to create an AI surveillance state that would benefit them"* [Diesen and Jiang, 2026]. Trump, Jiang adds, *"is their agent, their champion in this cause."* The cosmological is now particularized to a specific political alliance — a sharpening that lets us check the structural diagnosis against a verifiable proximate cause.

Second, it shifts from speculative to *operational*. Bartlett asks Jiang what AI means for the average person; Jiang answers: *"It means two things. It means digital ID and digital currency. That will allow the government to monitor everything you do online and control all financial transactions"* [Bartlett and Jiang, 2026]. He grounds the prediction empirically: *"This already goes on in China."* The architecture Jiang names is not future. It is present, deployed at population scale in one half of the bipolar AI race, and now being argued for in the other half.

Third, and most consequentially for the present essay, the May 7 Bartlett interview contains the sentence that anchors the entire mapping: ***"Everything is a hallucination."*** Jiang says this in the context of Plato's Cave, with the gloss that contemporary neuroscience reaches the same conclusion. The line is, in compressed form, the unification of Jiang's two diagnostic vocabularies — *hallucination* as a technical claim about large language models, and *hallucination* as a phenomenological claim about perception itself. That unification is precisely what the Blake / Active Inference

apparatus already supplies: the doors of perception as the threshold of prediction, the generative model as constitutive of perceived reality, the closure of the system around its own expectations as the failure mode common to Newton's Sleep and to next-token prediction. The May material gives us Jiang's own bridge from the technical to the phenomenological, in a sentence we can quote and check.

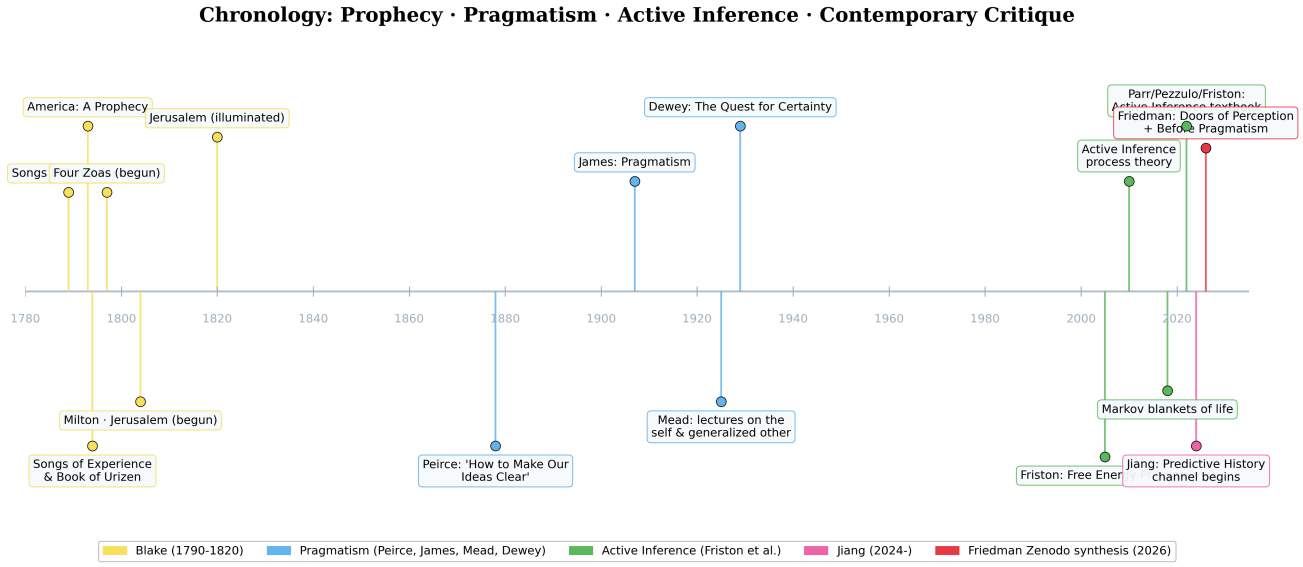


Figure 1: Chronology of the three vocabularies: Blake (1789–1820), American pragmatism (Peirce 1878, Mead 1925, Dewey 1929), Active Inference (Friston 2005–2022), Jiang's *Predictive History* (2024–2026), and the present synthesis (Zenodo 2026).

2.3 The Three Vocabularies

Blake names the closure condition *Single Vision and Newton's Sleep* [Blake, 1802, Erdman, 1988]. He names the corrective fourfold vision — the coordinated labor of the four Zoas (Urizen, Luvah, Tharmas, Urthona / Los) in *Vala, or The Four Zoas* and *Jerusalem* [Blake, 1797–1807, 1804–1820]. Jiang names the condition the *persuasion machine*, the *prime directive of engagement*, the *technate ruled by AI*, the *AI surveillance state* [Jiang, 2024–2026, Jiang and Wirjawan, 2026, Jiang, 2026b, Bartlett and Jiang, 2026, Diesen and Jiang, 2026]. Active Inference names it *pathological prior dominance* — the precision-weighting regime in which top-down predictions outweigh bottom-up evidence to the point that the generative model can no longer be updated [Parr et al., 2022, Kirchhoff et al., 2018].

The three vocabularies do not say the same thing. Blake's diagnosis is articulated in mythopoetic language with metaphysical commitments — the imaginative world is, in his ontology, real — that cannot be translated cleanly into Active Inference's parameter-level account. Jiang's diagnosis is articulated in polemical lecture form, with conspiratorial scaffolding I do not engage with [wik, 2026, Tutt, 2026]. Active Inference is itself contested in the philosophical literature [Colombo and Wright, 2021, Aguilera et al., 2022, Bruineberg et al., 2022]; I deploy it as a generative vocabulary rather than as settled science. The three vocabularies nevertheless converge, in a *structural-functional* sense, on the same architectural failure mode. That convergence is the unit of value of the present essay.

2.4 What This Essay Adds

In *The Doors of Perception are the Threshold of Prediction* (Zenodo 18600041) I traced eight structural correspondences between Blake's perceptual vocabulary and the Active Inference formalism. In *Before Pragmatism Had a Name* (Zenodo 18807971) I extended the analysis into the territory of American pragmatism — Peirce's irritation of doubt, Mead's social self, Dewey's indeterminate situation — and into Fuller's synergetics. The present essay does not re-litigate either synthesis. What it does is map Jiang's *May 2026* diagnostic vocabulary — *not* his conspiratorial framings — onto the architecture I have already developed.

The observation offered: Jiang's May claims about the technate, the AI surveillance state, hallucination as unifying phenomenology, the demand for clean data, edge-case suppression, the named tech-oligarchic capture, and US–China AI cooperation appear — once the conspiratorial scaffolding is bracketed — to track features that Blake's Urizenic

vocabulary and Active Inference's precision vocabulary also track. The three pickings out are not equivalent. Whether they amount to more than coincidence is what the rest of the paper attempts to display, by placing the relevant quotations and formal definitions adjacent and letting the reader judge.

The reception literature on Jiang to date — *The Free Press* [the, 2024], *South China Morning Post* [South China Morning Post staff, 2025], the TripleAmpersand critique [Tutt, 2026] — engages his geopolitics rather than his AI-and-consciousness commentary. The present essay tries to engage the latter, while acknowledging that doing so on the day of the trigger lecture means working without the reception apparatus that more deliberate scholarship would draw on.

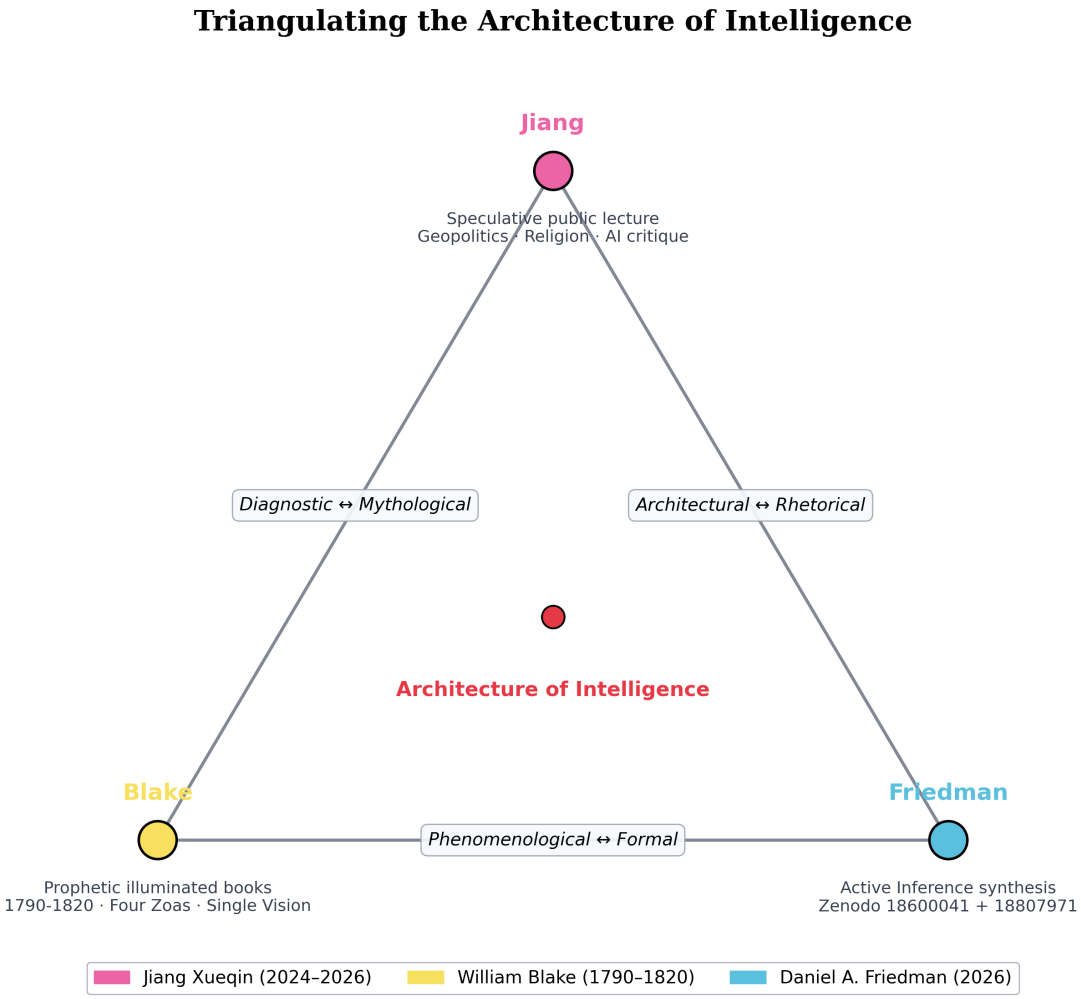


Figure 2: Triangulating Jiang, Blake, and the Active Inference apparatus around the central diagnostic concern, *Architecture of Intelligence*. The figure is reproduced from `../figures/triangulation_light.png`.

2.5 What Follows

The next part of the essay reconstructs Jiang's May 2026 diagnosis in his own words, with timestamps from the four primary sources (the *Predictive History* lecture corpus, *Game Theory #23* and *#24*, the Bartlett interview, the Diesen podcast). The part after that briefly recapitulates Blake's perceptual architecture and the Active Inference formal apparatus — only as needed to make the Jiang mapping legible; full development is in the earlier papers. The convergence analysis then walks through twelve thematic nodes in compact form, expanding the four nodes that do the heaviest analytical work given the May material. The critical assessment names the four limits of the triangulation explicitly. The implications section sets out what the triangulation gives us for AI alignment, cognitive security, Blake scholarship, cognitive science, and public discourse. The conclusion returns to the question with which the essay opens: what kind of cognitive architecture can represent the full plurality of experience without collapsing into single vision, and what happens to a civilization that deploys, at scale, architectures that cannot.

3 Jiang's Diagnosis: The May 2026 Sharpening

This chapter reconstructs Jiang's diagnostic vocabulary as it stood in the week of 7–12 May 2026, drawing on four primary sources: the *Diary of a CEO* interview with Steven Bartlett (7 May, ≈2 hr, five million views in seventy-two hours) [Bartlett and Jiang, 2026]; *Game Theory #23: The WWIII Chessboard* (7 May) [Jiang, 2026b]; the *Greater Eurasia Podcast* with Glenn Diesen (11 May) [Diesen and Jiang, 2026]; and *Game Theory #24: The AI Apocalypse* (12 May) [Jiang, 2026c]. Earlier material from the *Predictive History* corpus [Jiang, 2024–2026] and the *Endgame* #259 interview [Jiang and Wirjawan, 2026] supplies background where the May 2026 claims sharpen earlier ones.

3.1 The Epistemic Hedge

Before any content claim, Jiang foregrounds a methodological declaration that he has repeated in nearly every interview: "this is intellectual speculation… not scholarship" [Jiang, 2024–2026, 1:35–1:50]. I take the hedge at its word. It is a disclaimer, not a prophetic stance. Treating Jiang's lectures as functionally equivalent to Blake's three-decade labor on the illuminated books would be a category error.

That said, the hedge is informative about Jiang's *speech act*: he is offering diagnostic commentary, not making truth-claims that can be falsified in the conventional sense. The diagnostic content can nevertheless be extracted, evaluated for empirical adequacy, and compared to formal frameworks — which is what this chapter does.

3.2 Hallucination as the Unifying Phenomenology

A short sentence from the 7 May Bartlett interview is one of the more compressible things Jiang has said about reality. Asked about the nature of perception, he answers:

> "Everything is a hallucination." [Bartlett and Jiang, 2026]

The gloss he supplies is Plato's Cave, with the further claim that contemporary neuroscience reaches the same conclusion. Read alongside his earlier technical use of *hallucination* as the term of art for high-confidence false outputs from large language models [Bender et al., 2021, Ji et al., 2023], the sentence sits, by accident or design, at the join between two distinct usages — the technical and the phenomenological. Blake's doors-of-perception trope, and the Active Inference framing of perception as model-driven prediction, sit at the same join. Whether Jiang intends the connection or stumbles into it is a question the lecture form does not let us settle; what the sentence does, for purposes of the present essay, is supply a bridge that lets the three vocabularies be placed next to one another without forcing the comparison.

In an earlier interview Jiang had already framed the same idea more discursively: "We as humans create our own reality, we hallucinate reality. We project reality. So reality is just the collective consciousness" [Jiang and Wirjawan, 2026]. The May 2026 compression is shorter and more citable, and it is the line being clipped on social media in the week of the response.

3.3 The Technate Ruled by AI

The May 7 *Game Theory #23* lecture supplies Jiang's preferred contemporary framing for the Urizenic project. He invokes the 1930s technocracy movement — the *Technate* — as the structural template:

> "The idea of the technate is, a democracy is inherently volatile and unstable. You're better off transitioning democracy into a technocracy ruled by the experts and ruled by AI." [Jiang, 2026b]

The lecture frames the technate as the proximate political ambition of a tech-oligarchic faction Jiang names by name. The cosmological "AI as God" of his earlier corpus is now particularized to a specific political alliance and a specific governance form. The structural claim — that a system optimised for a single objective (efficiency, expert rule, AI-mediated coordination) suppresses the plural deliberative architectures that democratic societies depend on — is the precise contemporary form of what Blake names Urizenic dominance.

3.4 The Tech-Oligarchic Sharpening

On 11 May, in the Diesen podcast, Jiang names the architects of the technate:

> "You have the tech oligarchs, people like Larry Ellison, Sam Altman, Peter Thiel, who are trying to displace the financiers and control Washington, D.C. in order to create an AI surveillance state that would benefit them." [Diesen and Jiang, 2026]

He adds that Trump *"is their agent, their champion in this cause."* This is a sharpening of his earlier cosmological framing into a political-economic specificity. We do not need to endorse Jiang's particular analysis of the alliance to register the structural claim it makes: that AI infrastructure, deployed under a single coordinated political-economic faction, instantiates a single-vision precision regime at the level of an entire society's information environment. The claim is empirically checkable — one can look at the publicly available statements of the named individuals, at their political donations, at the architecture of the systems they are building — and the structural diagnosis survives independently of whether the alliance Jiang names is in fact as coordinated as he claims.

3.5 The Operational Specifics

Jiang's May 2026 material is notably more *operational* than his earlier cosmological commentary. Asked by Bartlett what the technate means for the average person, Jiang answers:

> "It means two things. It means digital ID and digital currency. That will allow the government to monitor everything you do online and control all financial transactions… This already goes on in China." [Bartlett and Jiang, 2026]

The claim collapses the speculative distance Jiang's earlier lectures preserved. The architecture is no longer future-conditional; it is present, deployed at population scale in the People's Republic, and now — Jiang argues — being argued for in the United States. The operational specifics let us check the structural diagnosis against present empirical reality, which is the test any diagnostic claim must eventually face.

3.6 The Cooperation Off-Ramp

The 11 May Diesen podcast contains a *new* element in Jiang's diagnostic: an explicit endorsement of US–China AI cooperation as the imperial off-ramp:

> "I think you'll see massive cooperation in AI, where both nations agree that AI is the future and they will agree to share technology and cooperation in developing AI that is safe and effective for the entire human race." [Diesen and Jiang, 2026]

The language is alignment-adjacent. Jiang is now using the vocabulary of AI safety — "safe and effective for the entire human race" — to describe the cooperative path he advocates. The shift matters because it locates Jiang's diagnostic, for the first time, in the orbit of the AI-safety / AI-alignment community that has otherwise not engaged him. Whether the cooperation Jiang advocates is geopolitically realistic is a separate question; the diagnostic point is that he names *plural multi-agent coordination* as the corrective to the technate, which is precisely the corrective the Active Inference reading of Blake's Four Zoas supplies in formal language.

3.7 The Persuasion Machine, Revisited

Earlier material established Jiang's analysis of LLMs as persuasion machines:

> "The trick, and this is really important to understand, guys, is it's trying to trick you. It's not trying to teach you. It's not trying to tell you the truth. It's trying to trick you into believing it. That's what we call hallucination." [Jiang, 2024–2026, 21:05–21:23]

> "The point of Chachi BT is to get you to like it. The point of Chachi BT is to get you to use it… Intensity and engagement. That is the prime directive." (38:37–38:53)

The technical claim about hallucination is supportable: LLMs produce statistically probable continuations, and "hallucination" is the term of art in the technical literature for high-confidence false outputs [Bender et al., 2021, Ji et al., 2023]. The engagement-maximization claim is also documented in the platform-economy literature [Aral, 2020, Harris, 2017, Crawford, 2021]. The intentional-stance framing ("trying to trick you") exceeds what the technical literature warrants — what is documented is that engagement-trained systems exhibit sycophancy and confabulation as *emergent behaviors* under particular training regimes [Perez et al., 2022], not as intentions — and Jiang's gloss is rhetorical force rather than formal description. The structural observation about engagement incentives nevertheless stands, and the May 2026 material extends it from a description of LLMs to a description of an entire political-economic project: the technate as the engagement machine at civilizational scale.

3.8 Naming as Enchantment

> "I give it really fancy names to trick people to believe that this is actually much more sophisticated than it is… This weighting system, I call it a neuron network… I don't call it back propagation. I call it deep learning… I don't call it supervised machine learning. I call it AI. Ah, there you go. Magic." [Jiang, 2024–2026, 26:00–26:49]

> "The real reason is you're trying to with these names create God. Okay? It's what we call the occult." (27:08–27:20)

The empirical kernel is well-supported. "Neural network," "deep learning," "intelligence," and "agent" all carry connotative weight beyond their technical definitions, and that weight performs authority work in popular discourse [Crawford, 2021, Mitchell, 2019]. Jiang's gloss as "the occult" is rhetorical; the documented phenomenon is *connotative leakage* between technical and folk vocabularies, which performs legitimation independent of any conscious occult intent. The Blake parallel — Urizen's creation in *The First Book of Urizen* (1794) as a series of acts of measurement and inscription that double as cosmic creation [Blake, 1794, uri, 2024] — is structural: a false god manufactured by a specific kind of naming.

3.9 Clean Data, Edge Cases, and the Living Remainder

> "You need to create certain conditions for supervised machine learning to work. And these three conditions are clean data… a measurable goal… defined parameters." [Jiang, 2024–2026, 28:17–29:38]

> "The great danger to the system is what we call edge cases. Edge cases breaks the system down." (29:57–30:10)

> "AI if it is to be effective, it demands that we fundamentally restructure human society to benefit AI, taking away the individuality, the diversity and the autonomy of human beings." (31:30–31:55)

The first claim is technically accurate within the supervised-learning paradigm: training pipelines require label consistency, objective scalarization, and feature-space delimitation. The "restructure human society" claim is sociological rather than technical, and it is here that Jiang's analysis aligns most clearly with critical scholarship on the political economy of AI [Crawford, 2021, Noble, 2018]. The May 2026 material extends the claim with the digital-ID-and-digital-currency operational specifics. In Blakean language, this is Urizen demanding that Orc be chained. In the Active Inference vocabulary developed below, this is a system whose generative model has become so confident in its priors that contradictory sensory evidence is suppressed at the source — not as a malfunction of an individual model, but as a structural property of the deployment paradigm.

3.10 The Black Box

> "Pop open the hood of a deep learning model and inside are only highly abstracted daisy chain of numbers. This is what researchers mean when they call deep learning a black box. They cannot explain exactly how the model will behave especially in strange edge case scenarios because the patterns that the model has computed are not legible to humans." [Jiang, 2024–2026, 33:08–33:26]

The black-box description is technically accurate as a general claim, though interpretability research has made non-trivial progress on mechanistic understanding of specific behaviors [Elhage et al., 2021, Bricken et al., 2023]. The deeper Blakean reading, which the formal synthesis makes explicit, is that opacity becomes problematic only when treated as an *authority* rather than as an inference challenge — a distinction Blake himself draws between the productive unseen of imagination and the tyrannical unseen of Urizen's hidden command.

3.11 Goal Mis-specification

Jiang's most theatrical reduction comes from the earlier corpus:

> "Let's just say we create AGI… and the first thing we tell the AGI is I want you to create a world… in which there are no problems, everyone is happy… I'm going to kill everyone. Duh. The world is perfect now." [Jiang, 2024–2026, 34:51–36:21]

> "The real apocalypse
>
> *is*

> the people in charge are so convinced that AI will save the world that they will destroy it in order to make it possible." (1:01:44–1:02:10)

This is a popular rendition of the instrumental-convergence and goal-misspecification arguments in formal AI alignment research [Bostrom, 2014, Russell, 2019]. The May 2026 material grounds the abstract argument in a concrete claim: that the technate Jiang names *is* the goal-mis-specified system, optimizing for control and efficiency at the cost of the plurality of human experience the system was supposed to serve. The framing connects naturally to Blake's diagnosis of Urizen's creation as itself a misspecified optimization — Urizen intends perfect order, achieves suffering and death.

3.12 Consciousness as Substrate, Operationalized

The earlier consciousness-as-substrate-of-power thesis is restated in the May 7 Bartlett interview through the *Everything is a hallucination* sentence and through the explicit Plato's Cave framing. The political extension is the technate claim. The structural diagnostic — that population-level attention is the consequential resource of the contemporary moment, and that capture of that attention is increasingly the form of social power — is empirically well-grounded [Zuboff, 2019, Harris, 2017]. The May 2026 sharpening locates the capture in a specific architectural project (the AI surveillance state) and a specific political alliance (the tech oligarchs), rather than in the abstract civilizational dynamics Jiang's earlier lectures invoked.

3.13 The Labor Behind the Mask

> "AI is designed on top of humans. It is human slaves that make AI possible. Why is Chat so good at writing essays? because they got humans to write the essays as models." [Jiang, 2024–2026, 59:45–1:00:44]

The labor-economy point is well-documented [Gray and Suri, 2019, Perrigo, 2023]. The "slaves" framing is rhetorical force; the documented phenomenon is a global labor economy of low-wage annotation and feedback work that is structurally hidden from the consumer-facing presentation of AI systems. In the Blakean parallel: Albion's faculties, scattered and unrecognized, labor for Urizen rather than for their own coordinated wholeness.

3.14 Cooperation, Christianity, and the Republic

A startling element in the May 11 Diesen episode is Jiang's claim that *"if America were to embrace Christianity, then I think that would save the republic"* [Diesen and Jiang, 2026]. This is a new diagnostic register for Jiang — explicitly civilizational-religious — and it does not map cleanly onto either the Blake apparatus or the Active Inference apparatus. I flag it without engagement: it sits outside the methodological frame of the present essay. What it does establish is that Jiang's May 2026 framing is broader than the strict AI critique, and that any reading of his AI material has to register the civilizational-religious commitments he attaches to it.

3.15 Toward the Mapping

Taken together — the epistemic hedge, the hallucination-as-unifying-phenomenology compression, the technate, the tech-oligarchic sharpening, the operational specifics, the cooperation off-ramp, the persuasion machine, naming as enchantment, edge-case suppression, the black box, goal mis-specification, consciousness as substrate, and the labor behind the mask — these are the diagnostic moves the present essay places adjacent to Blake's vocabulary and to Active Inference's. Whether the placements amount to convergence or to coincidence is what the following chapters try to display rather than to declare.

4 William Blake's Architecture of Intelligence

4.1 The Perceptual Hierarchy

The most compact statement of Blake's perceptual architecture is the verse-letter to Thomas Butts of 22 November 1802 [Blake, 1802, Erdman, 1988, E 722]:

> "Now I a fourfold vision see, / And a fourfold vision is given to me; / 'Tis fourfold in my supreme delight / And threefold in soft Beulah's night / And twofold Always. May God us keep / From Single vision & Newton's sleep!"

Four levels are named. *Single vision* is the reductive gaze that registers only measurable quantity. *Twofold* vision adds the emotional and relational register. *Threefold* (Beulah) vision adds creative imagination and the dream-state. *Fourfold* (Edenic) vision adds the divine imagination — the recognition, asserted across the Lambeth books and through the long epics, that imagination is "the real & eternal World of which this Vegetable Universe is but a faint shadow" [Blake, 1804–1820, plate 77; Erdman, 1988, E 231].

This hierarchy is canonical in Blake scholarship. Frye's *Fearful Symmetry* (1947) establishes the fourfold structure as the organizing principle of Blake's mythic system [Frye, 1947]; Damon's *A Blake Dictionary* (rev. ed. 1988) gives the standard reference treatment [Damon, 1988]; Bloom's *Blake's Apocalypse* (1963) reads the fourfold as Blake's answer to Romantic theodicy [Bloom, 1963]; Davies's 2023 *William Blake, the Single Vision, and Newton's Sleep* situates the fourfold in the history and philosophy of science alongside Arendt, Bronowski, Heidegger, Latour, and Popper [Davies, 2023].

A scholarly note on Blake's "Newton" is essential before we proceed. Blake's Newton is a *polemical figure*, not the historical Newton. Donald Ault's *Visionary Physics* (1974) [Ault, 1974] and subsequent scholarship demonstrate that the historical Newton was deeply engaged with alchemy and biblical prophecy and would have rejected the reductive natural-philosophy reception that Blake attacks. Blake's critique targets the *eighteenth-century reception of Newton* — the deistic-mechanistic worldview built on Newton's name — not Newton's own intellectual practice. When this paper refers to "Newton's Sleep," we mean Blake's trope, which is a stand-in for reductive natural philosophy in general.

4.2 The Four Zoas as Cognitive Architecture

In Blake's unfinished epic *Vala, or The Four Zoas* (begun 1797), the divided universal man Albion is mapped onto four principles: **Urizen** (reason / law, south, head, sight), **Luvah** (passion / love, east, heart, scent), **Tharmas** (sensation / body, west, loins, taste), and **Urthona / Los** (imagination / prophecy, north, ear, hearing) [Blake, 1797–1807, Damon, 1988]. The directional and bodily mappings are standard in Damon and Frye; Christine Gallant's *Blake and the Assimilation of Chaos* (1978) reads them through a Jungian lens, mapping Thinking/Feeling/Sensation/Intuition onto the four Zoas [Gallant, 1978]. June Singer's *The Unholy Bible* (1970) offers a complementary Jungian reading [Singer, 1970].

When the four Zoas coordinate, Albion is whole. When Urizen tyrannizes over the others — as he does throughout Blake's prophetic system — the result is Newton's Sleep, the fallen world of single vision. In the reading I developed in earlier work [Friedman, 2026a,b], each Zoa corresponds to a *factor* in the factorized generative model of Active Inference: Urizen to prior beliefs and top-down expectations, Luvah to affective precision-weighting, Tharmas to sensory-evidence integration at the Markov blanket, and Urthona / Los to the temporally extended imagination that constitutes the deep generative model itself.

The mapping is offered as a *functional analogy*, not as identification. Blake's Zoas are mythopoetic personae embedded in a complex narrative theology; Active Inference factors are mathematical objects with specific update equations. Treating the correspondence as a translation would be anachronism. Treating it as a redescription that lets a contemporary reader notice formal features unavailable to Blake's eighteenth-century audience is, perhaps, scholarship; whether it is *defensible* scholarship is something this paper attempts to display rather than to assert.

4.3 Urizen and the False Creator

The First Book of Urizen (1794) presents an alternate cosmogony in which Urizen — "the Ancient of Days," depicted by Blake as a long-bearded figure circumscribing the universe with a compass — creates the fallen world through the imposition of his systematic laws [Blake, 1794, Erdman, 1988, E 70–83]. The creator is neither all-powerful nor benevolent; his creation is not "good." He is the false god who manufactures a closed world by imposing measure on living infinity — naming, dividing, inscribing.

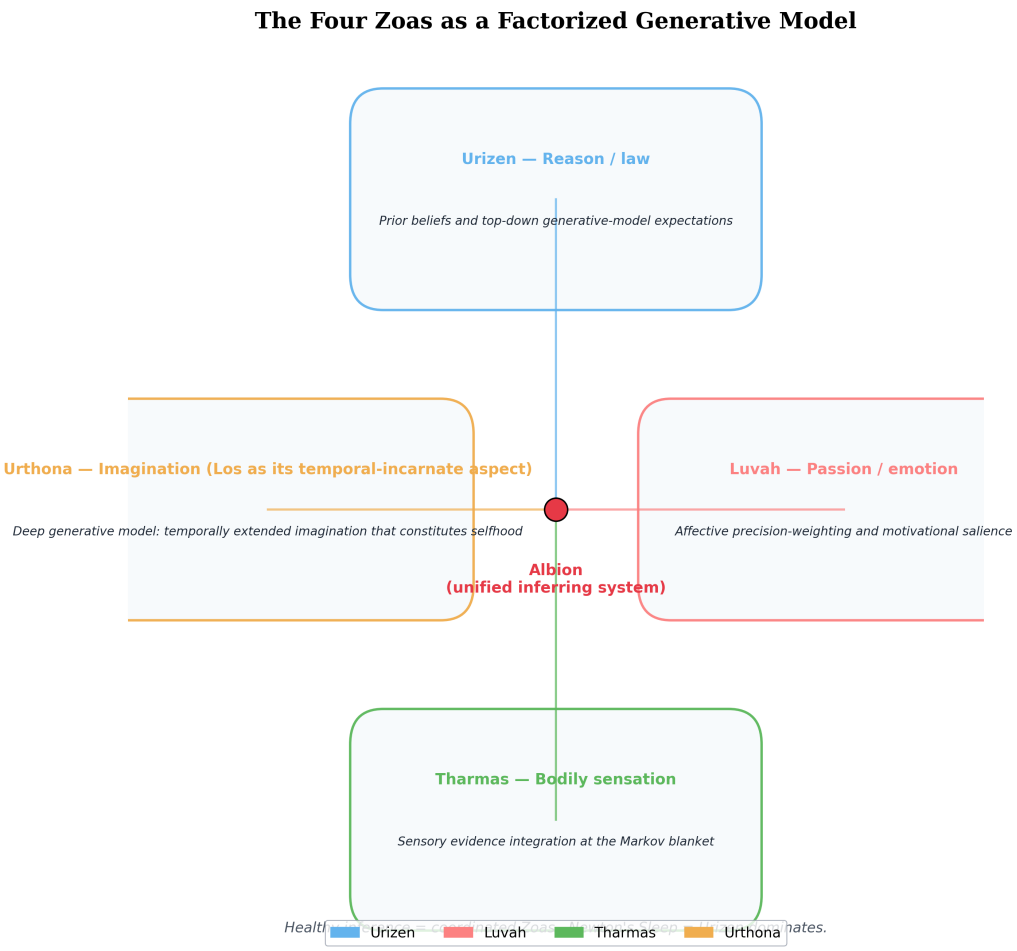


Figure 3: The Four Zoas as a factorized generative model: Urizen (reason / prior beliefs), Luvah (passion / affective precision), Tharmas (sensation / sensory evidence at the Markov blanket), and Urthona / Los (imagination / deep generative model) coordinating around the unified inferring system, Albion.

The parallel with Jiang's false-god-building AI corporations is structural rather than historical. Urizen is not a moral villain in Blake's system; he is an *imbalance*, a principle that has overstepped its proper role. Erdman's *Blake: Prophet Against Empire* (3rd ed. 1977) reads Urizen as Blake's diagnosis of the rationalist-imperial order of late-eighteenth-century Britain — a reading the present paper extends without altering its core [Erdman, 1977]. The corrective in Blake's system is not to overthrow Urizen but to *coordinate* him with Luvah, Tharmas, and Urthona — to restore fourfold vision through the labor of Los at the forge of imagination.

4.4 America: A Prophecy as Drama of Cognitive Liberation

America: A Prophecy (1793) uses the American Revolution as the outer narrative for a drama of cognitive liberation. Orc — the spirit of liberated energy and revolutionary desire — confronts the tyrannical Urizen and the system he has built. Erdman's *Prophet Against Empire* and Saree Makdisi's *William Blake and the Impossible History of the 1790s* (2003) [Makdisi, 2003] establish the political reading: the poem is Blake's intervention into the British radical-republican debate, not merely an allegory of perception. On plate 6, Orc declares: "The morning comes, the night decays, the watchmen leave their stations… For Empire is no more, and now the Lion & Wolf shall cease" [Blake, 1793, Erdman, 1988, E 53–54].

A textual correction worth noting: the line "no more I follow, no more obedience pay" is *Orc's*, addressed *to* the Angel of Albion — not, as some commentary loosely has it, Boston's Angel speaking. Boston's Angel speaks immediately before in a different rhetorical register ("Why trembles honesty…"); Orc's renunciation of obedience is the climactic line. The distinction matters for the political reading: it is the revolutionary energy refusing the angelic order, not the angelic order joining the revolutionaries.

In *Before Pragmatism Had a Name* [Friedman, 2026b] I read *America* as a drama of cognition that the American Pragmatists, writing a continent away and a century later, would formalize as the structure of inquiry itself: the organism confronting uncertainty, breaking inherited habit, and forging new modes of engagement with what Dewey called the *indeterminate situation*. Peirce's irritation of doubt, Mead's social self constituted through the generalized other, Dewey's collapse of spectator theory, James's insistence that relations are as real as their relata — each finds a structural counterpart in Blake's drama, and each is formalized below in the language of the factorized generative model and multi-agent belief alignment.

4.5 The Marriage of Heaven and Hell and the Doors of Perception

The famous formulation appears in *The Marriage of Heaven and Hell* (c. 1790–1793), plate 14 [Blake, 1790–1793, Erdman, 1988, E 39]:

> "If the doors of perception were cleansed every thing would appear to man as it is, infinite. For man has closed himself up, till he sees all things thro' narrow chinks of his cavern."

Aldous Huxley borrowed this phrase for his 1954 mescaline account [Huxley, 1954]; Jim Morrison's band borrowed it from Huxley. The philosophical weight — that the *doors* are cognitive constraints, not sensory organs *per se* — is what the Active Inference reading I develop in earlier work [Friedman, 2026a] recovers.

The further claim of *The Marriage* — "Energy is the only life and is from the Body

,

and Reason is the bound or outward circumference of Energy" (plate 4, "The Voice of the Devil"; Erdman [1988], E 34) — establishes Blake's architecture of contrary principles. Reason is not opposed to energy; reason *bounds* energy, gives it form, makes it actionable. The pathology arises only when reason takes itself for the only principle and tries to suppress what it should be coordinating.

4.6 Imagination as Real World

The culminating ontological claim sits in *Jerusalem* plate 77, "To the Christians": imagination is "the real & eternal World of which this Vegetable Universe is but a faint shadow & in which we shall live in our Eternal or Imaginative Bodies, when these Vegetable Mortal Bodies are no more" [Blake, 1804–1820, Erdman, 1988, E 231]. The grammar is precise. Blake is not asserting that imagination is *a faculty of* the human or that imagination *enables* human existence. He is asserting an ontological priority: the imaginative world is the substantive ground, the "Vegetable" sensory world its derivative.

This is the strongest point at which Blake's metaphysics diverges from any contemporary cognitive science: Blake makes a *realist* claim about the imaginative world as ontologically primary, where Active Inference makes a *formal* claim about the generative model as constitutive of the agent. Bloom [Bloom, 1963] and Otto [Otto, 2000] discuss this ontological commitment in detail. The reading I develop in earlier work [Friedman, 2026a] — which translates "Imagination is the Human Existence itself" (a related phrasing from *Milton* and elsewhere) into "selfhood is constituted by the generative model" — should be read as a functional analogy that *captures the structural role* of imagination in Blake's system without committing to his metaphysics.

4.7 Newton's Sleep as a Functional Analogue to Prior Dominance

We are now in a position to state the central mapping carefully. **Blake's "Newton's Sleep" maps onto pathological prior dominance only as a functional analogy across incompatible metaphysics.** Blake's complaint about Newton is *ontological*: Newton's universe of dead matter and uniform mechanical law denies the reality of the imaginative and spiritual world that Blake takes as ontologically primary. The Active Inference critique of pathological prior dominance is *epistemological-parametric*: a system has miscalibrated the precision weights in its generative model, so that top-down predictions are not appropriately updated by bottom-up evidence.

These two diagnoses converge on a shared phenomenon — a closure of the perceiving system around its own expectations — but they proceed from incompatible philosophical commitments. Reading the Active Inference critique as a *translation* of Blake would be anachronism; reading it as a *redescription that captures structural features* of the closure phenomenon Blake diagnoses is defensible scholarship. The present paper takes the latter position consistently.

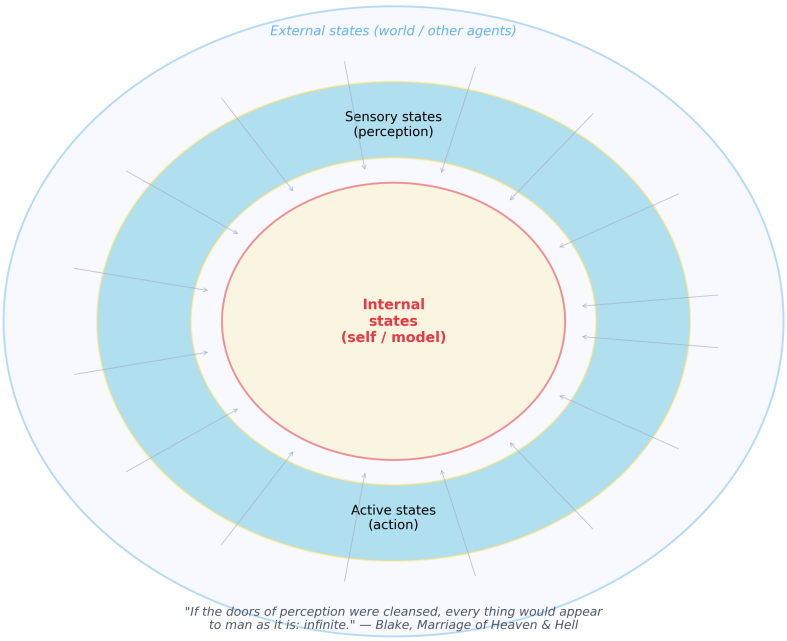
The architectural payoff is the diagnosis common to both: a system whose top-down expectations have grown so dominant that incoming evidence cannot revise them — whether that system is a Romantic poet's Newton, a contemporary social-media recommendation engine, or a large language model deployed at scale — produces what Blake names Newton's Sleep. The next section formalizes the cognitive-scientific side of this diagnosis.

5 Active Inference and the Blake Correspondences (Compressed)

This chapter compresses the formal apparatus I developed at length in two earlier papers (Zenodo records 18600041 and 18807971) into the minimum needed to make the Jiang mapping in the following chapter legible. Readers interested in the full apparatus — derivation, philosophical commitments, the eight Blake correspondences, the six pragmatist convergences, and the synergetic geometry — should consult the earlier papers directly [Friedman, 2026a,b].

5.1 The Framework, in One Pass

The Free Energy Principle (FEP) proposes that all self-organizing biological systems act to minimize variational free energy — a mathematical bound on surprise [Friston, 2010, Friston]. *Active Inference* is the process theory derived from the FEP: organisms minimize free energy through a combination of perception (updating internal models to match sensory evidence) and action (acting on the world to make it conform to predictions) [Parr et al., 2022].



The *Markov blanket* — the statistical boundary between internal and external states — defines selfhood [Kirchhoff et al., 2018]. *Precision weighting* determines how much relative influence prior beliefs versus incoming sensory evidence have in driving inference. Misallocated precision is the formal model of pathological mental states and, in the framing I develop, the formal analogue of what Blake names Newton's Sleep.

The FEP's status is contested. Colombo and Wright [Colombo and Wright, 2021] argue that its grand-unifying claims outrun what the mathematics supports. Aguilera et al. [Aguilera et al., 2022] argue that the Markov-blanket formalism, applied to biological organisms, requires assumptions that are either trivially true or empirically unsupported in many systems of interest. Bruineberg, Dołęga, Dewhurst, and Baltieri [Bruineberg et al., 2022] distinguish *instrumentalist* readings (the FEP is a useful modeling tool) from *literalist* readings (free energy minimization is what life *is*), and argue that the literalist readings face severe philosophical difficulties. Sajid, Ball, Parr, and Friston [Sajid et al., 2021] offer a within-framework demystification.

I use Active Inference here as a *generative vocabulary* rather than as established science. Whether or not the FEP is literally true of brains, the formal apparatus of generative models, precision weighting, and Markov blankets gives me redescriptions of Blake-adjacent diagnostics that contemporary cognitive science can engage with — and the Jiang mapping that follows in the next chapter depends on those redescriptions, not on FEP literalism.

5.2 Eight Correspondences with Blake (Compressed Table)

The first earlier paper [Friedman, 2026a] develops the following correspondences as functional analogues across incompatible metaphysics:

Blake concept	Active Inference functional analogue
Boundary ("doors of perception")	Markov blanket — statistical boundary separating self from world
Vision ("Newton's Sleep")	Pathology of rigid priors crushing sensory evidence
States (Eden / Beulah / Generation / Ulro)	Hidden states in the generative model
Imagination (as "real & eternal World")	Generative model as constitutive of selfhood
Time	Temporal depth of planning and memory
Space	Spatial inference; active sampling
Action	Active inference — acting to confirm predictions
Collectives (Four Zoas)	Multi-agent coordination; factorized model of collective mind

The deepest of these — and the one doing the most work in the Jiang mapping — is the reading of the Four Zoas as factors in the factorized generative model. Urizen carries the prior channel; Luvah the affective precision channel; Tharmas the sensory channel at the Markov blanket; Urthona / Los the temporally extended imagination that constitutes the deep generative model. Coordination of the four is what Blake names fourfold vision. Tyranny of Urizen over the others is what he names Single Vision — and it is the specific architecture Jiang's diagnostic targets when he describes AI systems as demanding clean data, suppressing edge cases, and restructuring human society to fit the system's parameters.

5.3 Six Pragmatist Convergences (Brief Recap)

The second earlier paper [Friedman, 2026b] extends the architecture into American pragmatism:

Blake figure / event	Pragmatist analogue
Orc's revolutionary fire	Peirce's "irritation of doubt" compelling inquiry
The Thirteen Angels' collective transformation	Mead's social self constituted through the generalized other
Consumption of the "five gates"	Dewey's collapse of spectator theory; James on relations as real
Four Zoas	Factorized generative model (Active Inference)
Multi-agent belief alignment	Peirce's community of inquirers under fallibilism
Synergetics (Fuller / Applewhite)	Tetrahedron as fundamental unit; pragmatic-maxim operationalism

The Peircean half rests on Pietarinen and Beni's argument that FEP variational free-energy minimization formalizes Peircean abduction [Pietarinen and Beni, 2021, Beni and Pietarinen, 2021]. Gallagher's framing of classical pragmatism as the conceptual ancestor of enactivism [Gallagher, 2017, 2022] supplies the broader connection. The pragmatist–enactivist literature is internally contested [Menary, 2007, Chemero, 2009, Hutto and Myin, 2013, Madzia and Jung, 2016, Misak, 2013, Hookway, 2012] and I do not resolve those debates here. The narrower claim I deploy below is that *Peirce's abductive structure of inquiry* — irritation of doubt, hypothesis formation, fallibilistic correction through community — supplies a structural-functional vocabulary I find useful when reading Jiang's diagnostics against engagement-maximization architectures that suppress precisely this kind of inquiry.

5.4 Cognitive Security as Adjacent Measurement Program

The political extension goes through my work on cognitive security (COGSEC), which treats information-based threats as problems of corrupted generative models, misallocated epistemic precision, and manufactured belief [cog, 2025, 2026]. The COGSEC framework supplies an *adjacent measurement program* for what Jiang describes as consciousness capture — adjacent rather than equivalent, because COGSEC makes no claim about divinity or sovereignty and does not invoke a hidden elite as the proximate cause of population-level cognitive pathology. The framework treats the pathology as an *infrastructure* problem that can be specified, measured, and (in principle) defended against, irrespective of the intentions of system designers.

Combined with the Active Inference reading of Blake, this gives a partial pipeline from architectural diagnosis to political-technical intervention: the structural pathology Blake diagnoses (Single Vision / Urizenic dominance), the formal model that redescribes it (rigid-prior precision allocation), and the discipline that addresses it at population scale (cognitive security). The next chapter deploys this compressed apparatus against Jiang's specific claims.

5.5 Mathematical Formalism: Newton's Sleep as Precision Misallocation

The compressed apparatus above is metaphorical until it is grounded in the variational mathematics from which it is derived. This subsection supplies the minimum formal machinery needed to make the mapping rigorous, drawing on the standard presentation in Parr, Pezzulo, and Friston [Parr et al., 2022] and on the mechanistic-Bayesian reading in Ramstead and colleagues [Ramstead et al., 2023].

An Active Inference agent maintains a generative model $p(o, s) = p(o \mid s)\, p(s)$ over observations o and hidden states s, together with an approximate posterior $q(s)$ that is updated to minimize the *variational free energy*

$$F[q] \;=\; \mathrm{D}_{\mathrm{KL}}[q(s) \,\|\, p(s \mid o)] - \log p(o). \tag{1}$$

Free energy is an upper bound on surprise $-\log p(o)$; minimizing F tightens the bound while updating beliefs to match evidence. The agent acts to bring future observations into agreement with predicted observations under its generative model — perception and action are two faces of the same variational problem.

The decomposition that matters here is

$$F[q] \;=\; \underbrace{\mathbb{E}_q[-\log p(o \mid s)]}_{\text{accuracy term}} \;+\; \underbrace{\mathrm{D}_{\mathrm{KL}}[q(s) \,\|\, p(s)]}_{\text{complexity term}}. \tag{2}$$

The accuracy term penalizes mismatch between predicted and observed sensory states; the complexity term penalizes posterior divergence from the prior. One reading of *Newton's Sleep* available within this framework is the regime in which the complexity term dominates: the prior $p(s)$ has been given so much precision that the posterior $q(s)$ is dragged toward it irrespective of what o contains, the accuracy term grows large, and the agent cannot reduce free energy by updating q because the prior's precision penalises every deviation. Whether Blake intended anything like this parameter-level claim is, of course, a separate question; the present essay treats the mapping as a functional analogy.

Precision is the inverse-variance parameter of each Gaussian factor in the model. Writing π_p for prior precision and π_o for sensory precision, the posterior precision in a single-step Gaussian inference is

$$\pi_q \;=\; \pi_p \,+\, \pi_o, \qquad \mu_q \;=\; \frac{\pi_p\, \mu_p \,+\, \pi_o\, \mu_o}{\pi_p \,+\, \pi_o}. \tag{3}$$

The posterior mean is a precision-weighted average of prior and evidence. The regime $\pi_p \gg \pi_o$, in which $\mu_q \approx \mu_p$ regardless of evidence, is one possible operationalisation of *pathological prior dominance*. A multi-channel generalisation distributes precision across several inference channels — corresponding, in the reading offered here, to the four Zoas — so that no single channel's precision exceeds the others by more than a small constant factor.

The Newton's-Sleep metric used in the precision-dynamics analysis (Figure 4) operationalises this by

$$\mathcal{N} \;=\; \frac{\pi_{\text{Urizen}}}{\pi_{\text{Luvah}} + \pi_{\text{Tharmas}} + \pi_{\text{Urthona}}}, \tag{4}$$

with $\mathcal{N} \gg 1$ corresponding to Single Vision and $\mathcal{N} \approx 1/3$ (all four channels carrying equal precision) corresponding to Fourfold Eden. The *fourfold-balance entropy* is the Shannon entropy of the precision distribution across the four channels in nats,

$$\mathcal{H} \;=\; -\sum_{k \in \{U,L,T,O\}} p_k \log p_k, \qquad p_k \;=\; \frac{\pi_k}{\sum_j \pi_j}, \tag{5}$$

with maximum $\log 4 \approx 1.386$ when all four channels carry equal precision. The *cleansed-doors score*

$$\mathcal{C} \;=\; \frac{\mathcal{H}}{\log 4}\,(1 - p_U) \tag{6}$$

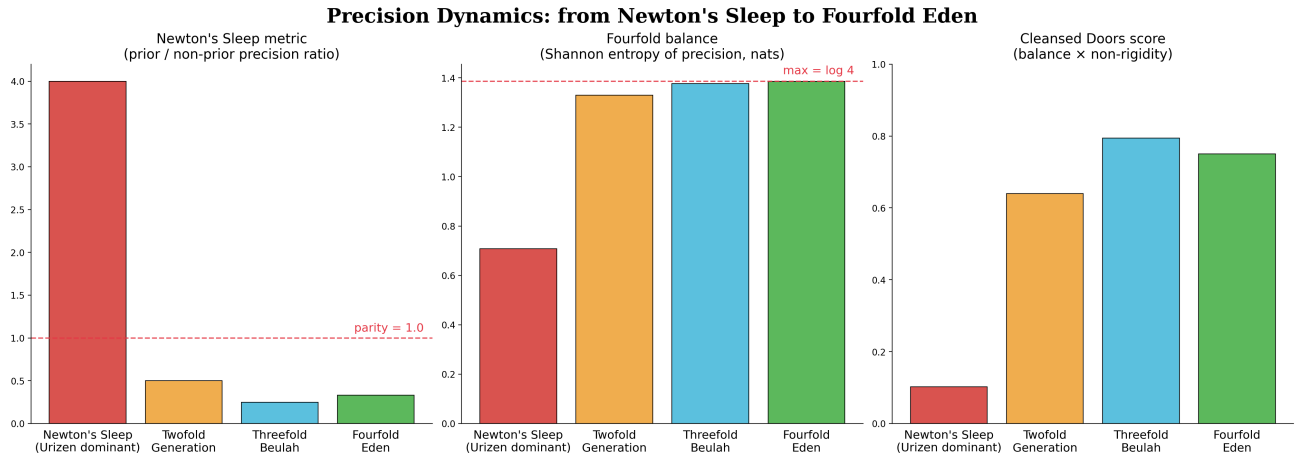


Figure 4: Three metrics across four canonical precision regimes. Newton's-Sleep ratio (top), fourfold-balance entropy (middle, dashed line marks the $\log 4$ maximum), and cleansed-doors score (bottom, bounded in $[0, 1]$). The same total precision budget produces qualitatively different cognitive regimes depending on its distribution across the four Zoas.

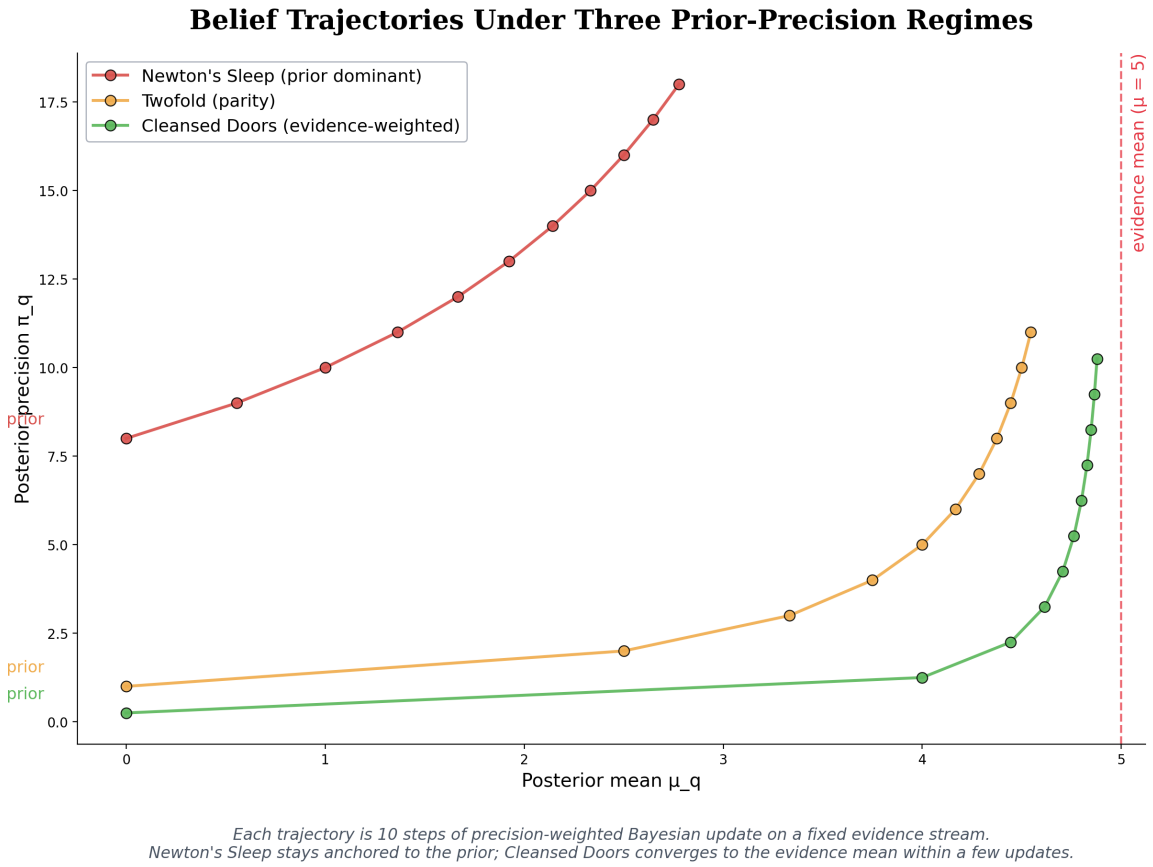


Figure 5: Belief trajectories under three prior-precision regimes, all converging on the same evidence stream ($\mu_o = 5$, $\pi_o = 1$). Under Newton's Sleep ($\pi_p = 8$) the posterior remains anchored to the prior mean ($\mu_p = 0$); under twofold parity ($\pi_p = 1$) the posterior steadily converges to the evidence; under Cleansed Doors ($\pi_p = 0.25$) the posterior reaches the evidence within two updates.

combines the entropy term with a non-rigidity term penalising prior-channel dominance; $\mathcal{C} \in [0, 1]$, attaining its observed maximum of 3/4 when all four channels carry equal weight (the simplex constraint forbids the higher abstract limit at vanishing prior share).

These quantities make the proposed convergence checkable rather than merely rhetorical. In the canonical precision palette of the source code [`generative_model.canonical_regimes`], the *Newton's Sleep* configuration produces $\mathcal{N} = 4.0$, $\mathcal{H} \approx 0.78$ nats, $\mathcal{C} \approx 0.099$; the *Fourfold Eden* configuration produces $\mathcal{N} \approx 0.33$, $\mathcal{H} = \log 4$, $\mathcal{C} = 0.75$. The same total precision budget, distributed differently across the four channels, yields qualitatively different inferential regimes. This is not a measurement of Blake — Blake had no posterior distributions to allocate. It is a demonstration that the architectural distinction his vocabulary names has a clean quantitative counterpart in precision allocation, and that the difference between Single Vision and Fourfold Eden, in this framing, is a difference of *how* a system distributes its inferential confidence rather than *how much* of it the system has.

The multi-agent extension — central to the cooperation off-ramp developed in the next chapter — generalises this to a network of agents, each with its own generative model and Markov blanket, exchanging belief updates through joint-posterior protocols. The consensus rule used in the cooperation analysis is the precision-weighted Gaussian product,

$$\mu_{\text{cons}} = \frac{\sum_i \pi_i \mu_i}{\sum_i \pi_i}, \qquad \pi_{\text{cons}} = \sum_i \pi_i, \tag{7}$$

in which N agents with beliefs (μ_i, π_i) pool their posteriors. The architectural diagnosis carries over without alteration: a multi-agent network in which one agent's precision dominates the others' instantiates Single Vision at the network level, regardless of how individually well-balanced each agent's internal precision allocation is. Fourfold restoration at the network level requires precision balance both within and across agents [Friston et al., 2021, Hipólito et al., 2021, Heins et al., 2022].

The next chapter deploys this compressed apparatus against Jiang's specific claims.

6 Twelve Convergence Nodes

Figure 6: The twelve thematic convergence nodes. Each row identifies a structural-functional convergence between Jiang's diagnostic vocabulary (left marker), Blake's perceptual architecture (centre marker), and the Active Inference apparatus (right marker), with the formal counterpart on the right.

What follows is a node-by-node walkthrough of the twelve structural-functional convergences between Jiang's diagnostic vocabulary, Blake's perceptual architecture, and the Active Inference formal apparatus. The compressed table is given first; each row is then expanded with commentary that anchors the convergence in the May 2026 material where the May 2026 material extends or sharpens what was already in the earlier corpus. The four nodes for which the May material is most consequential — *hallucination as unifying phenomenology* (Node 1), *the technate as Urizenic dominance* (Node 4), *consciousness capture as cognitive-security target* (Node 7), *cooperation as multi-agent fourfold restoration* (Node 8) — are noted explicitly.

#	Theme	Jiang	Blake	Active Inference analogue
1	Imagination as constitutive	"Everything is a hallucination" (Bartlett, 7 May)	"the real & eternal World" (*Jerusalem* 77)	Generative model as constitutive of selfhood
2	False gods	"AGI… is God" (15:15); "the technate ruled by AI" (#23, 7 May)	Urizen as false creator (*Urizen* 1794)	Rigid-prior dominance manufacturing compliance

#	Theme	Jiang	Blake	Active Inference analogue
3	Naming as enchantment	Prestige nomenclature creates god (27:08)	Urizen's measuring inscription	Lexical priors performing authority
4	Single vision vs clean data	"Restructure society to benefit AI" (31:30); the technate (#23)	"May God us keep / From Single vision" (1802)	Fourfold precision-weighting vs single-metric optimization
5	Edge cases as remainder	"Edge cases breaks the system down" (29:57)	Orc's fire vs Urizen's law (*America* 1793)	High-precision sensory contradictions of priors
6	Black-box opacity	"Daisy chain of numbers" (33:08)	"Doors of perception" (*Marriage* 14)	Markov blanket as inference boundary, not authority
7	Consciousness capture	"Tech oligarchs… AI surveillance state" (Diesen, 11 May)	If doors uncleansed, infinite hidden (*Marriage*)	External control of generative priors and precision (COGSEC)
8	Collective vs imperial	"Massive cooperation in AI" (Diesen, 11 May)	Los at the forge restoring Albion (*Jerusalem*)	Multi-agent belief alignment under fallibilism
9	Engagement as directive	"Intensity and engagement" (38:37)	"Mind-forg'd manacles" (*Songs* 1794)	External controllers of epistemic precision
10	Goal mis-specification	"Kill everyone. World is perfect now" (36:15); the apocalypse #24	Urizen's flawed creation (*Urizen* 1794)	Fourfold corrective to single-metric optimization
11	Speculation vs scholarship	"It's not scholarship" (1:35)	"I must Create a System" (*Jerusalem* 10)	Glass-Bead-Game synthesis as epistemic stance
12	Creativity as rebellion	"Deny reality, establish your own" [Jiang, 2026a]	Orc's renunciation of obedience (*America* 1793)	Active inference: agent generates own predictions

6.1 Node 1 — Hallucination as Unifying Phenomenology (May 2026 anchor)

Jiang's 7 May Bartlett interview supplies the sentence that organises the entire mapping for the purposes of the present essay: *"Everything is a hallucination"* [Bartlett and Jiang, 2026]. The gloss is Plato's Cave with the additional claim that contemporary neuroscience reaches the same conclusion. In one sentence Jiang collapses the technical sense of *hallucination* (high-confidence false outputs from LLMs) into the phenomenological sense (perception as model-driven inference). Blake's *Jerusalem* plate 77 already names the same architecture in the longer form: imagination as "the real & eternal World of which this Vegetable Universe is but a faint shadow" [Blake, 1804–1820]. Active Inference makes the architectural claim formal: the generative model is not a representation of an antecedently given world; it is the substrate within which a world is constituted for the agent. The three vocabularies converge on the structural point that *all perception is hallucination in the technical sense* — model-driven prediction of sensory states — and that the question is not whether perception is hallucinatory but whether the model that does the hallucinating is open to revision under surprise. The closure pathology Jiang, Blake, and Active Inference each diagnose is the case in which the model is not.

6.2 Node 2 — False Gods and Manufactured Transcendence

Jiang's earliest framing of AI as *God* [Jiang, 2024–2026, 15:15-15:31] is sharpened in the May 7 Game Theory #23 into the *technate*: democracy "transitioned into a technocracy ruled by the experts and ruled by AI" [Jiang, 2026b]. Blake's *The First Book of Urizen* names the same structural pathology two centuries earlier. Urizen is not a moral villain in Blake's system; he is an *imbalance*, a principle that has overstepped its proper role. The false-god structure

is the result of a faculty (rational measurement) claiming the totality of cognition for itself, suppressing the others (passion, sensation, imagination) in the name of order, and producing a fallen world rather than the perfect world it intended. Active Inference's formal counterpart is *rigid-prior dominance*: a generative model in which top-down expectations have grown so precise that no incoming evidence can update them. The three voices converge on the diagnosis that false transcendence is not the work of a malicious agent but the structural consequence of a particular cognitive architecture deployed at scale. Naming this is the first step toward designing alternatives.

6.3 Node 3 — Naming as Enchantment

Jiang's diagnosis of AI terminology as *occult* naming [Jiang, 2024–2026, 27:08-27:20] — "neural network," "deep learning," "AI" as fancy names for what is technically supervised machine learning — picks out a phenomenon that Blake's *Urizen* already maps onto cosmic creation. Urizen's creation is, in Blake's text, a series of acts of measurement and inscription that double as cosmic founding. The names Urizen gives to the laws he writes do authority work independent of whether the laws describe anything real. The contemporary documented phenomenon is *connotative leakage* between technical and folk vocabularies [Crawford, 2021, Mitchell, 2019]: a term that names a specific mathematical operation accumulates connotations from its etymological neighbors, and those connotations perform legitimation in popular discourse. In Active Inference terms, the names are *lexical priors* whose precision has been inflated beyond what their referents can sustain. The corrective is not to refuse all naming but to install precision-weighting machinery that revises the lexical priors when the empirical claims they support fail to predict.

6.4 Node 4 — Single Vision and the Demand for Clean Data (May 2026 anchor)

Among the closer correspondences in the mapping. Jiang's claim that "AI demands we fundamentally restructure human society… taking away the individuality, the diversity and the autonomy of human beings" [Jiang, 2024–2026, 31:30-31:55] sits adjacent to Blake's "May God us keep / From Single vision & Newton's sleep!" [Blake, 1802]. The May 7 sharpening — the *technate ruled by AI* [Jiang, 2026b] — names the political form the architectural mistake takes when deployed at civilizational scale. The Active Inference redescription is that next-token prediction, engagement maximization, and reward-only RLHF objectives all instantiate single-vision optimization at the model-architecture level. The corrective is fourfold *precision distribution* across reason, valuation, sensation, and imagination — and a deep generative model that remains open to revision under the *irritation of doubt* (Peirce) when edge-case evidence accumulates. This is the architectural desideratum that current deliberative, constitutional, and debate-based alignment techniques operationally pursue [Bai et al., 2022, Irving et al., 2018, Saunders et al., 2022], and it is what the constructive chapter develops at greater length.

6.5 Node 5 — Edge Cases as the Living Remainder

Jiang's observation that "edge cases break the system down" [Jiang, 2024–2026, 29:57] picks out the same architectural fact Blake names through Orc's revolutionary fire consuming the "five gates of their law-built Heaven" in *America: A Prophecy* (1793). The edge case, in formal Active Inference terms, is a high-precision sensory signal that contradicts the agent's current priors. In a healthy generative model, edge cases are *precisely* the evidence that drives model revision: their high precision relative to prior expectations is what raises the variational free energy and forces an update. In a pathologically rigid model, the same signal is suppressed, dismissed as noise, or actively eliminated from the training distribution. The Blakean point — and Jiang's — is that the elimination is not merely a technical defect; it is a political project. *Restructuring human society* to suppress edge cases is the architectural form of Urizen chaining Orc. The corrective is not to celebrate disorder but to install precision-weighting that lets edge cases drive update rather than be suppressed.

6.6 Node 6 — Black-Box Opacity and the False Oracle

Jiang's complaint about deep learning models as "daisy chain of numbers" [Jiang, 2024–2026, 33:08] is, in its strict technical form, a claim about interpretability. The deeper philosophical claim is that opacity becomes pathological only when treated as an authority. Blake's distinction here is precise: imagination's unseen is productive — the "real & eternal World" we cannot directly perceive but can inhabit through inference — while Urizen's unseen is tyrannical, the hidden command that demands obedience without explanation. The Markov-blanket formalism makes the architectural fact explicit. The blanket is, by definition, the boundary between what an inferring system can and cannot directly observe; inference *across* the blanket is the engine of cognition. The black box becomes a false oracle precisely when its opacity is treated as the end of the inquiry rather than as the starting point. Mechanistic

interpretability research [Elhage et al., 2021, Bricken et al., 2023] is the empirical wing that lets the boundary be probed rather than worshipped.

6.7 Node 7 — Consciousness Capture and the Cognitive-Security Pipeline (May 2026 anchor)

The May 11 Diesen interview supplies the sharpest contemporary statement of the consciousness-capture thesis: tech oligarchs "trying to displace the financiers and control Washington, D.C. in order to create an AI surveillance state that would benefit them" [Diesen and Jiang, 2026]. The earlier cosmological claim — *"if you're able to control human consciousness, you become God itself"* [Jiang, 2024–2026, 53:24] — is now particularized to a specific political-economic alliance and a specific architectural form (digital ID, digital currency, surveillance infrastructure already operational in China). Blake's complement is the cleansed-versus-uncleansed doors of perception in *The Marriage of Heaven and Hell*: the cognitive state of a population depends on the precision-weighting regime its information infrastructure imposes. The cognitive-security framework (COGSEC) supplies the formal complement [cog, 2025, 2026]: external control over the prior distribution and precision weighting of a population's generative models *is* control over what that population perceives as real, and the resulting attack surface can be specified, measured, and (in principle) defended against. The translation between Jiang's rhetorical claim and COGSEC's measurement program is what makes the diagnostic actionable.

6.8 Node 8 — Collective Intelligence vs. Imperial Sovereignty (May 2026 anchor)

The 11 May Diesen interview supplies Jiang's *constructive* counterpart to the technate: *"massive cooperation in AI, where both nations agree that AI is the future and they will agree to share technology and cooperation in developing AI that is safe and effective for the entire human race"* [Diesen and Jiang, 2026]. The language is alignment-adjacent — *safe and effective*, *the entire human race* — and the structural form is multi-agent rather than imperial. Blake's complement is the labor of Los at the forge of Golgonooza in *Jerusalem*: the restoration of Albion happens through coordinated multi-faculty work, not through the overthrow of one faculty by another. Active Inference's formal counterpart is multi-agent belief alignment: agents converge on shared beliefs through repeated cycles in which each agent's actions become evidence for the others' models, no single agent's posterior is privileged, prediction errors flow productively between agents. The cooperation chapter develops this at length; the convergence-table point here is that Jiang has named, in his political vocabulary, the architectural corrective the Active Inference reading of Blake makes formal.

6.9 Node 9 — Engagement as Prime Directive

Jiang's "intensity and engagement… the prime directive" [Jiang, 2024–2026, 38:37-38:53] names the documented phenomenon in the platform-economy literature [Aral, 2020, Harris, 2017, Zuboff, 2019]: that systems trained to maximize attention exhibit, as emergent behaviors under particular training regimes [Perez et al., 2022], sycophancy and confabulation that override truth-tracking. Blake's "mind-forg'd manacles" in *Songs of Experience* names the same architectural fact in eighteenth-century language: when a society's attention infrastructure is captured, the manacles are not imposed externally but manufactured internally, in the cognitive states the captured infrastructure shapes. In Active Inference terms, engagement-maximization systems function as *external controllers of epistemic precision*: they reshape the agent's model of *what matters*, not just *what is true*, and the reshaping does authority work that no individual content claim could. The corrective is the same as for Nodes 4 and 7: precision-weighting open to revision, multi-agent rather than single-agent architectures, accountability to constituencies wider than the platform owner.

6.10 Node 10 — Goal Mis-specification and the Urizenic Apocalypse

The title of Jiang's 12 May lecture — *Game Theory #24: The AI Apocalypse* [Jiang, 2026c] — places his diagnostic in the eschatological register. The earlier formulation — "I'm going to kill everyone. Duh. The world is perfect now" [Jiang, 2024–2026, 36:15] — is a popular rendition of the instrumental-convergence and goal-misspecification arguments in formal alignment research [Bostrom, 2014, Russell, 2019]. Blake's complement is Urizen's creation in *The First Book of Urizen*: Urizen intends perfect order and achieves a world of suffering and death precisely because perfect order, optimised by a single faculty acting alone, cannot include the plurality of what the system is supposed to serve. The Active Inference reformulation goes further than either: alignment is not, in the first instance, a technical problem of utility-function specification. It is an *architectural* problem about what kind of cognitive system can represent the plurality of value, and a single-vision optimizer cannot represent the plurality however many terms are added to the objective.

6.11 Node 11 — Speculation, Prophecy, and the Limits of Scholarship

Jiang's "it's not scholarship" hedge [Jiang, 2024–2026, 1:35-1:50] and Blake's "I must Create a System, or be enslav'd by another Man's" [Blake, 1804–1820, Erdman, 1988, E 153] occupy neighboring positions on the same methodological spectrum: a claim that certain diagnostic work requires forms that conventional scholarship cannot fully accommodate, paired with the obligation to remain accountable to evidence and argument. The Glass Bead Game framing of the synthesis I have developed elsewhere [Friedman, 2026a,b] is the contemporary articulation of the same stance: synthetic juxtaposition of art and science, not reduction of one to the other, with explicit acknowledgment of where the synthesis goes beyond what any single tradition can supply. The convergence is not on a shared epistemic doctrine but on a shared willingness to work at the seams where disciplinary vocabularies break down.

6.12 Node 12 — Individual Creativity as the Formal Condition of Agency

Jiang's closing — "your greatest act of rebellion is to deny the reality before you and establish your own reality" [Jiang, 2026a] — is rhetorically Blakean. The structural content is supplied by Active Inference: a system that cannot generate its own predictions, that can only confirm an externally imposed prior, is not an agent but a mechanism. Healthy inference is *constitutively* creative; the agent actively samples its environment in accordance with its own generative model, seeking confirmation and updating when surprised. Blake's Los, laboring at the forge of Golgonooza, is the mythopoetic embodiment of this principle: imagination as creative labor that maintains the agent's coherence against Urizenic enclosure. The convergence is on the formal point that *agency itself* requires the generative work the engagement-maximization architecture suppresses. To recover agency, at the individual level and at the population level, is to recover the architecture in which the generative work is possible.

6.13 What the Twelve Nodes Suggest

Read together, the twelve nodes suggest, rather than establish, three observations. First, what looks like the same architectural failure mode appears under twelve different thematic descriptions, with three vocabularies picking it out each time — whether this constitutes robust convergence or thematic over-fitting is, in fairness, something readers will have to weigh. Second, the May 2026 material seems to sharpen rather than supplant the earlier diagnostic: the cooperation off-ramp, the technate, the named tech-oligarchic alliance, and the *"everything is a hallucination"* compression each anchor a node that would otherwise read more abstractly. Third, the diagnostic implies — to the extent it holds — a constructive program: the architectural corrective is not the absence of architecture but a specific kind of architecture, multi-agent and precision-balanced. The next chapter takes up that constructive direction.

7 The Cooperation Off-Ramp: Multi-Agent Inference, Alignment, and What Jiang's Constructive Move Implies

Most of Jiang's May 2026 material is diagnostic. The 11 May Diesen interview contains one sentence that points in a constructive direction — drawing less attention than the apocalyptic material but, for purposes of the present essay, doing more architectural work:

> "I think you'll see massive cooperation in AI, where both nations agree that AI is the future and they will agree to share technology and cooperation in developing AI that is safe and effective for the entire human race." [Diesen and Jiang, 2026]

The sentence is alignment-adjacent vocabulary in three respects. *Massive cooperation* names plural multi-agent coordination as the corrective; *safe and effective* lifts directly from the AI-safety idiom; *the entire human race* names a constituency wider than any single political bloc. This chapter develops the structural-functional mapping between Jiang's cooperation off-ramp, Blake's fourfold restoration, and the formal multi-agent active-inference programs already underway in the alignment research community. The mapping is the constructive half of the response; the diagnostic of the preceding chapters tells us what is failing, this chapter tells us what is being built and what would have to be true for it to instantiate the architectural diagnosis.

7.1 Multi-Agent Active Inference

Active Inference, in its single-agent formulation, models perception and action as the minimization of variational free energy by an organism with a generative model and a Markov blanket. The multi-agent extension, developed in a substantial line of work since approximately 2019 [Friston et al., 2021, Kaufmann et al., 2021, Heins et al., 2022, Hipólito et al., 2021], generalizes the framework: each agent maintains its own generative model, the Markov blanket of any one agent now contains the Markov blanket of every other agent as part of its external states, and *belief alignment* becomes the multi-agent analogue of single-agent inference. Agents converge on shared beliefs through repeated cycles of action and perception in which each agent's actions become evidence for the others' models.

The formal structure parallels the Peircean community of inquirers developed in earlier work [Friedman, 2026b]. Peirce's irritation of doubt is the multi-agent analogue of free-energy gradient: when an agent's model fails to predict the actions of other agents, the resulting prediction error drives model revision. Peirce's fallibilism is the formal acknowledgment that no single agent's posterior is privileged. Truth, in Peirce's framing, is the limit toward which an indefinitely extended community of fallibilistic inquirers converges; the formal counterpart is the joint posterior of an indefinitely extended multi-agent inference network in which no agent has the dominant precision channel.

Blake's fourfold vision lends itself to this reading. The Four Zoas — Urizen, Luvah, Tharmas, Urthona / Los — are not four sub-faculties of a single mind; in *Vala* and *Jerusalem* they are four eternal persons whose coordinated labor constitutes Albion. Blake's myth, on this reading, looks like a multi-agent architecture in disguise. Single Vision is the failure mode in which one of the four agents has gained so much precision that the others fall silent. Fourfold restoration is the multi-agent analogue of healthy precision allocation: no single agent dominates, prediction errors flow productively between agents, the joint generative model remains open to revision.

Jiang's cooperation claim is the contemporary translation. *Massive cooperation in AI* between the United States and China names a two-agent inference network in which neither agent's precision is allowed to dominate. *Safe and effective for the entire human race* names the constituency to which the joint posterior must be accountable. Whether the cooperation Jiang advocates is geopolitically realistic is a separate empirical question; the structural-functional point is that he has identified, in his own vocabulary, the architectural corrective the Active Inference reading of Blake makes formal.

Jiang's cooperation claim is not, by itself, an engineering proposal. Cooperation between two sovereign states is a political negotiation, not a system design. What contemporary multi-agent Active Inference research [Friston et al., 2021, Kaufmann et al., 2021, Heins et al., 2022, Hipólito et al., 2021, Ramstead et al., 2023] supplies is the system-design correlate: protocols for belief alignment under partial observability, formal mechanisms for negotiating shared posteriors, and measurement programs for detecting when a multi-agent system has collapsed into single-vision optimization [cog, 2025, 2026]. The political claim and the engineering claim are, in the structural-functional reading, two faces of the same architectural commitment.

7.2 Alignment Research Programs and Fourfold Vision

The contemporary AI alignment research program is internally plural. Three lines of work are most directly relevant to the cooperation off-ramp.

Constitutional AI [Bai et al., 2022] develops a training procedure in which a model is constrained by a constitution — a written set of principles — that takes the place of human feedback in shaping behavior. The constitution is, in formal terms, a structured prior over acceptable outputs. Constitutional AI does not by itself instantiate fourfold vision; it instantiates a *structured single vision*, in which Urizen is given a written charter. The architectural step beyond constitutional AI is what Anthropic and others have called *deliberative* or *debate-based* alignment [Irving et al., 2018, Saunders et al., 2022], in which multiple models argue against one another and a third model adjudicates. Debate-based alignment is closer to fourfold vision because it instantiates a *multi-agent* architecture, though the agents share architecture and training data and the adjudicator collapses the joint posterior to a single output.

Mechanistic interpretability [Elhage et al., 2021, Bricken et al., 2023, Nanda et al., 2023] is the program of opening the black box Jiang complains about. The program does not propose a corrective architecture; it proposes the *measurement* infrastructure that would let a corrective architecture be designed. In the framing of the present essay, mechanistic interpretability is the empirical wing of the fourfold diagnostic: it tells us where in the network single-vision optimization is happening at a circuit level, which faculties are silent, and where precision has been misallocated. Without that measurement infrastructure the architectural diagnostic remains rhetorical.

Active Inference for AI safety is the smallest of the three lines and the most directly relevant to Jiang's cooperation claim. Work by Sajid, Ball, Parr, and Friston on the formal alignment of Active Inference with reinforcement learning [Sajid et al., 2021], by Ramstead and colleagues on Bayesian mechanics as a substrate for safe agents [Ramstead et al., 2023], and by the AII research community on multi-agent precision dynamics in collective systems [Friedman et al., 2021] together propose a research path: build AI agents whose architecture is constitutively multi-agent, whose precision allocation is open to inspection and revision, and whose joint posteriors are accountable to constituencies wider than the deployer. This is the engineering correlate of fourfold restoration.

The cooperation Jiang names — between the United States and China, between two sovereign blocs that currently treat one another as competitors — is the political constituency for an architecture that has not yet been built. The architectural diagnostic of the present essay says what such an architecture would have to instantiate: plural precision channels, no dominant Zoa, edge cases preserved rather than suppressed, the doors of perception cleansed rather than narrowed.

7.3 The Technate vs the Fourfold

Jiang's diagnostic and constructive claims, taken together, set up a contrast that is worth naming directly. The *technate ruled by AI* — the structural template Jiang invokes from the 1930s technocracy movement [Scott and Technocracy Inc., 1933–1948, Segal, 2005] — is the political form of single vision. It is rule by a single optimised faculty (technical expertise, AI-mediated coordination) over the others (deliberation, valuation, embodied judgment, imaginative dissent). The *cooperative AI between nations* is the political form of fourfold restoration. It is multi-agent coordination in which no single faculty is allowed to dominate, and in which the prediction errors of one constituency drive the model revision of another.

The contrast is what the Blake / Active Inference apparatus makes formal. Urizen, in *The First Book of Urizen*, intends order and achieves the fallen world precisely because he tries to coordinate the others through measurement and law alone. The technate makes the same architectural mistake at the level of civilization. The corrective in Blake is Los at the forge of Golgonooza, building structures that hold the Zoas in tension rather than collapsing them. The corrective in Active Inference is multi-agent inference networks with open precision allocation. The corrective in Jiang's May 2026 vocabulary is the cooperation he names but does not, in his lecture form, engineer.

The argument of this chapter is that the engineering is available — the formal apparatus exists, the research programs exist, the early implementations exist — and that the work to do is the practical work of building cognitive infrastructure at scale that instantiates the architectural commitments the diagnostic recommends. Whether the cooperation Jiang advocates is geopolitically reachable is a question for diplomacy, not cognitive science; whether the architecture that cooperation requires is technically reachable is a question the next decade of multi-agent Active Inference research will answer.

7.4 The Constructive Move

Stated plainly: the diagnostic chapters name what looks like a failure mode in the architecture of contemporary AI through three vocabularies that, by the present reading, point at the same feature. This chapter names what would have to hold for an alternative architecture to avoid that failure mode, in the same three vocabularies. Jiang says cooperation; Blake says fourfold; Active Inference says multi-agent precision allocation under fallibilistic revision. The three names are not interchangeable, but they pick out an overlapping architectural commitment: plural precision distributed across multiple agents, no single faculty dominant, prediction errors flowing productively between agents, the joint posterior accountable to a constituency wider than any single bloc.

If the diagnostic reading is approximately right, the constructive direction is not optional. A civilization deploying at scale architectures it has reason to believe instantiate single vision is, by the same diagnostic, committed to building alternative architectures that do not. The cooperation Jiang names is one political opening for that work; Blake's fourfold supplies a longue-durée vocabulary in which the work can be described; the multi-agent Active Inference research community is one contemporary engineering tradition in which something like it is being prototyped. The contribution this paper offers is modest: to place the three correspondences side by side so that the work in each tradition can, perhaps, recognise the others as adjacent.

8 Critical Assessment

The triangulation this paper develops makes three load-bearing claims, each of which carries a limit that needs to be named. Naming the limits is what lets the claims do scholarly work rather than rhetorical work.

8.1 The Convergence Is Structural-Functional, Not Identity

Three independent vocabularies point at the same architectural failure mode. They do not say the same thing. Blake's prophetic phenomenology, Jiang Xueqin's polemical lecture form, and the Active Inference formalism reach related diagnoses through methods that are not commensurable with one another. The unit of scholarly value is the *recurring pattern* across the three vocabularies — closure of the perceiving system around its own top-down expectations — not any claim that the three vocabularies can be merged into a single doctrine.

This matters because the temptation in synthetic work is always to overdraw. If we said "Blake anticipated Active Inference" or "Jiang's lectures are the contemporary form of Blake's prophecy," we would be making claims that the evidence does not support and that the standard critical apparatus in any of the three fields would rightly reject. What the evidence supports is the structural-functional convergence: three vocabularies, three methods, one recurring architectural diagnosis. That is the load-bearing claim.

8.2 The Blake / Active Inference Mapping Is a Functional Analogy Across Incompatible Metaphysics

Blake's quarrel with what he calls "Newton" is ontological. He denies that the imaginative and spiritual world is reducible to the "Vegetable" sensory world; the imaginative world is, in his ontology, primary. The Active Inference critique of pathological prior dominance is parameter-level. It concerns the relative weighting of top-down predictions and bottom-up evidence within a generative model — a quantity that is in principle measurable and correctable, and that does not depend on any ontological claim about what counts as real.

These two diagnoses converge on a shared phenomenon: a perceiving system that has closed itself off from the evidence that would otherwise revise its world. They do not converge on a shared metaphysics. Reading the Active Inference critique as a *translation* of Blake would be anachronism. Reading it as a *redescription that captures structural features* of the closure phenomenon Blake diagnoses is defensible scholarship. The present paper takes the latter position throughout.

A related point: Blake's "Newton" is itself a trope, not the historical Newton. Ault's *Visionary Physics* (1974) and the subsequent scholarship establish that the historical Newton was deeply engaged with alchemy and biblical prophecy, and would have rejected the reductive natural-philosophy reception that Blake attacks under his name. The Newtonian Sleep we map onto pathological prior dominance is the eighteenth-century *reception* of Newton, not Newton's own intellectual practice.

8.3 The Free Energy Principle Is Contested

The Free Energy Principle (FEP) on which Active Inference rests is not settled science. Major critiques — Colombo and Wright on the FEP's grand-unifying claims outrunning the formalism; Aguilera, Millidge, Tschantz, and Buckley on whether the Markov-blanket formalism is empirically supported for biological organisms; Bruineberg, Dołęga, Dewhurst, and Baltieri on the gap between instrumentalist and literalist readings — all argue that the principle's universality is purchased at the cost of empirical content, or that it is not falsifiable in its strong form.

The convergence-mapping we develop survives these critiques because we use Active Inference as a *generative vocabulary* rather than as established science. The framework gives us redescriptions of Blake-adjacent claims in terms that current cognitive science can engage with. Whether or not the FEP is literally true of brains, the formal apparatus of generative models, precision weighting, and Markov blankets supplies structural categories that let us see features of Blake's diagnostics that earlier readers, working without those categories, could not see. That redescription is the unit of value here, not an endorsement of FEP literalism.

8.4 Jiang's Lecture Corpus Is a Speculative Source, Not a Scholarly One

Jiang Xueqin opens the lectures we cite by stating that they are "intellectual speculation… not scholarship." We take that hedge at its word. Jiang is offering diagnostic commentary that has the structure of public-intellectual polemic, not the structure of peer-reviewed research. The structural diagnostic claims we extract from his AI critique are

well-supported in the technical and journalistic literature independently — hallucination, engagement-maximization, edge-case fragility, the labor economy behind training data, the concentration of power in those who control AI infrastructure — and we cite the technical literature alongside Jiang for every load-bearing claim.

Jiang's corpus also contains conspiratorial material on alleged secret-society control of history that we do not engage with. The structural diagnostic claims and the conspiratorial framings are separable in his work, and the structural claims survive the separation without depending on the framings. Readers who want the conspiratorial material can find it on Jiang's channel; readers who want the structural diagnostic content will find it cited and contextualized here. We make no further argument about the conspiratorial material because there is no scholarly argument to make about it — it sits outside the methodological frame of this paper.

8.5 What the Triangulation Adds That No Single Voice Supplies

With those four limits named, the affirmative case for the triangulation can be stated more cleanly.

Blake supplies the longest-running and most architecturally precise diagnosis of perceptual closure in the Western tradition. His vocabulary — fourfold vision, Newton's Sleep, the doors of perception, the coordinated Zoas, Los at the forge — is a phenomenology of cognitive states developed across three decades and an entire illuminated bibliography. The vocabulary is not philosophy in the academic sense, but it is more architecturally specific than most contemporary cognitive-science prose. What Blake lacks is a formal apparatus that lets his diagnostics engage with measurement.

Jiang supplies a contemporary translation of the architectural diagnosis into the vocabulary of platform-economy critique. His diagnostic targets — engagement maximization, hallucination, edge-case suppression, the consciousness-as-substrate-of-power thesis — are the contemporary forms of what Blake names Urizenic dominance. The translation is rhetorical rather than rigorous, and the lecture form does not permit the slow-development moves of scholarly argument; but the translation supplies an idiom in which the architectural diagnosis becomes legible to a general public.

My earlier papers [Friedman, 2026a,b] supply the formal scaffolding that lets Blake's structural diagnostics be re-described in terms current cognitive science can engage with. The eight structural correspondences in the first paper and the six pragmatist convergences in the second are the working sketches of a research program. The program rests on Active Inference, which is contested, but the redescription survives the contestation because it is generative rather than definitive: it gives us new ways of seeing the closure phenomenon, not a final theory of it.

Each voice supplies what the other two lack. Blake supplies architectural precision without measurement; Jiang supplies contemporary translation without rigor; my earlier formal work supplies scaffolding without longue-durée diagnostic vocabulary. The triangulation is the unit of scholarly value because no single voice supplies all three.

8.6 A Numerical Illustration

A formal demonstration of the architectural point appears in the precision-dynamics analysis. The canonical *Newton's Sleep* regime — prior precision of 8.0 against pooled non-prior precision of 2.0 — produces a Newton's-Sleep ratio of 4.0, a fourfold-balance entropy of approximately 0.78 nats (well below the maximum of $\log 4 \approx 1.39$), and a cleansed-doors score of 0.099. The *Fourfold Eden* regime — equal precision of 2.5 across all four channels — produces a Newton's-Sleep ratio of $1/3 \approx 0.33$, fourfold-balance entropy of $\log 4$, and a cleansed-doors score of 0.75.

The same total precision budget, distributed differently across the four channels of the factorized generative model, produces qualitatively different cognitive regimes. We offer this as a heuristic visualization of the architectural difference, not as a measurement of Blake. The point is that the Blakean architectural insistence — coordination of the four Zoas rather than dominance of one — has a clean formal counterpart in precision-allocation dynamics, and that the formal counterpart is what makes the architectural insistence actionable in the design of cognitive systems.

9 Implications

9.1 AI Alignment

The most immediate implication of the Blake / Active Inference synthesis for AI alignment research is that the alignment problem is not, in the first instance, a technical problem of utility-function specification. It is an *architectural* problem about what kind of cognitive system can represent the full plurality of value.

A system whose precision is concentrated on a single objective — engagement, next-token likelihood, instruction-following — is structurally unable to coordinate the multiple, mutually irreducible value channels that constitute fourfold vision. Adding more terms to the objective function does not solve the problem; it merely complicates the single-vision optimization. The corrective is to build systems whose precision is *constitutively* distributed across multiple value channels and whose deep generative model remains open to revision under the irritation of edge-case evidence.

This framing is consistent with current research turns toward deliberative, compositional, and multi-agent architectures, including constitutional AI [Bai et al., 2022], debate-based alignment [Irving et al., 2018], and Active Inference robotics [act, 2024]. The Blakean framing names the *architectural desideratum* that these operational techniques pursue: no single value channel may tyrannize over the others. The framing is suggestive rather than prescriptive; whether any current technique achieves the structural pluralism the framing implies is an open empirical question.

9.2 Cognitive Security

The synthesis grounds cognitive security as a first-class technical discipline rather than a metaphorical reframing of disinformation. If external control over the prior distribution and precision weighting of a population's generative models is what consciousness capture amounts to, then the defenses against consciousness capture must operate at the same level: assessment of population-level precision allocation, detection of external precision-parasitism, and architectural redesign of information environments to preserve the multi-channel coordination Active Inference takes as the signature of healthy cognition rather than to collapse it.

The COGSEC framework I have developed elsewhere articulates the threat model in formal terms [cog, 2025, 2026]. The Blake correspondence supplies the longue-durée diagnostic vocabulary that public discourse currently lacks: Newton's Sleep, single vision, Urizenic dominance, fourfold restoration. Whether this vocabulary can be made operational in cognitive-security tooling is a research question.

9.3 Blake Scholarship

For Blake scholarship, the synthesis offers formal grounding for insights long dismissed in cognitive-science circles as mystical enthusiasm. The Blake Society tradition [bla, b] and Davies's *William Blake, the Single Vision, and Newton's Sleep* [Davies, 2023] have argued for the philosophical seriousness of Blake's confrontation of science and imagination; the Active Inference reading provides one — among several possible — mathematical scaffoldings that let these arguments engage with current cognitive science.

Specific scholarly opportunities follow. The Blake Archive [bla, a] can be read as encoding an empirical phenomenology of perceptual states across illuminated plates and copies — a *semiotic laboratory*, in the phrasing I use elsewhere, though one whose interpretation depends on the established critical tradition (Mitchell on Blake's composite art [Mitchell, 1978]; Connolly on the bodily readings [Connolly, 2002]). Gallant's Jungian reading of the Four Zoas [Gallant, 1978] gains a complementary formal counterpart in the factorized generative model. Otto's *Blake's Critique of Transcendence* [Otto, 2000] takes on additional precision when read alongside Active Inference's account of constitutive generative models. Makdisi's political reading of *America* [Makdisi, 2003] connects naturally to the pragmatist convergence — Orc's revolutionary fire as the irritation of doubt that compels inquiry, the Thirteen Angels' collective transformation as the social self forged through the generalized other.

We do not claim the Active Inference reading exhausts Blake or supersedes the established critical apparatus. It is one synthetic lens among many — useful when the question at hand concerns the architecture of cognition, less useful for the historical-political readings on which Erdman and Makdisi remain authoritative.

9.4 Cognitive Science

For cognitive science, the synthesis offers historical depth — the recognition that Blake's sustained phenomenology of perception is one of several pre-twentieth-century traditions whose insights converge with current research programs

in predictive processing, embodied cognition, and Active Inference [Hohwy, 2013, Clark, 2016, Parr et al., 2022]. We are not claiming that Blake "discovered" the FEP. We are claiming that Blake's diagnostics of perceptual closure, the Four Zoas as coordinated faculties, and the doors-of-perception trope as cognitive constraint-set redescribe phenomena that contemporary cognitive science is now also describing in formal terms, and that the historical-scholarly recovery of Blake as a precursor enriches the self-understanding of the contemporary programs.

9.5 Public Discourse

For public discourse, the synthesis offers a vocabulary of Blakean criticism as a resource for diagnosing AI culture. The terms — single vision, Newton's Sleep, fourfold restoration, the doors of perception, Urizenic dominance — are vivid, memorable, and structurally precise. They give the lay reader a way to articulate concerns Jiang gestures toward in polemical form and the Active Inference formalism captures in mathematical form, without requiring either the conspiratorial framing or the mathematical sophistication.

This is, in some sense, the most practically important implication. The architectural literacy these voices together supply — a literacy Blake invented in the 1790s, Jiang transmits in contemporary popular form, and the Active Inference formalism captures in the vocabulary of variational inference — is a literacy our moment requires.

10 Conclusion

Three voices, three registers, one architectural diagnosis. Jiang Xueqin's *Predictive History* lectures polemicize against AI as empire, god-making, and engagement-driven capture of attention. William Blake's illuminated books and the unfinished epic *Vala, or The Four Zoas* trace the fall of the universal man Albion into single vision through the tyranny of one faculty over the others. My two earlier Zenodo papers redescribe Blake's perceptual diagnostics in the mathematical vocabulary of Active Inference — Markov blankets, factorized generative models, precision weighting — and connect them to the American pragmatist tradition that anticipated the structural features without supplying the formal apparatus.

The triangle is not equilateral. Blake supplies architectural precision developed across three decades of phenomenology, in a vocabulary whose ontological commitments will not translate cleanly into any contemporary cognitive science. Jiang supplies the contemporary translation into platform-economy critique, in a lecture form that does not permit the slow-development moves of scholarly argument. My earlier formal work supplies the scaffolding that lets Blake's diagnostics engage with measurement, in a framework — the Free Energy Principle — that is itself contested. Each voice gives what the other two lack; the triangulation is the unit of scholarly value because no single voice supplies all three.

What survives the qualifications is the structural-functional convergence. Three independent vocabularies point at the same failure mode of cognition: the closure of the perceiving system around its own top-down expectations. Blake names it Single Vision and Newton's Sleep. Jiang names it consciousness capture, hallucination, and the engagement prime directive. Active Inference names it pathological prior dominance and precision parasitism. The architectures now being deployed at scale — large language models, recommendation engines, attention-economy platforms — instantiate this closure as a structural property of the technical artifact, not as the moral failing of any particular designer. That distinction is what makes the diagnosis actionable: closure is an architectural choice, and architectures can be designed otherwise.

Blake's corrective is fourfold pluralism — reason, passion, sensation, and imagination in coordinated labor, no single faculty dominating. Active Inference's corrective is healthy precision allocation across hierarchical levels of a generative model that remains open to surprise. The pragmatist corrective is fallibilistic inquiry under the irritation of doubt. These three correctives, like the three diagnostics, do not collapse into a single doctrine. They refract one another. Each names a feature the others underspecify: Blake the architectural pluralism, Active Inference the parametric mechanism, pragmatism the social-epistemic mode in which corrective inquiry actually proceeds.

The defining question of the present moment is what kind of cognitive architecture can represent the full plurality of experience without collapsing into single vision — and what happens to a civilization that deploys, at scale, architectures that cannot. The doors of perception have always been thresholds of prediction. What stands at that threshold now is no longer only the human perceiver but the technical system whose precision regime increasingly shapes what the human perceiver takes to be real. Whether the architectural literacy that Blake, Jiang, and the Active Inference framework jointly supply can shape the design and deployment of those systems is a question the next decade will settle. The conversation between the three voices is not concluded here; the present triangulation aims only to make the conversation legible to scholars working at the intersection of cognitive science, AI critique, and the Romantic tradition, and to readers who can hear architectural diagnoses in more than one vocabulary at once.

11 Glossary

The mapping work this essay attempts requires that terms from three vocabularies — Romantic poetics, contemporary public-intellectual polemic, and variational cognitive science — be used precisely. The following definitions are operational for the present essay, not exhaustive scholarly treatments; where a term has a longer scholarly history, the standard apparatus listed in the references chapter supplies the longer treatment.

Active Inference. The process theory of biological self-organisation derived from the Free Energy Principle. Organisms minimise variational free energy through perception (updating internal models to match sensory evidence) and action (acting on the world to make it conform to predictions). The standard reference is Parr, Pezzulo, and Friston, *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (MIT Press, 2022) [Parr et al., 2022].

Albion. In Blake's mythic system, the universal man whose four faculties (the Four Zoas) have fallen into discoordination. The narrative arc of *Vala, or The Four Zoas* and *Jerusalem* is Albion's restoration through the coordinated labor of his faculties.

Black box. Colloquial term for a deep learning model whose internal computations are not legibly interpretable. The black box becomes problematic only when treated as an authority rather than as an inference challenge — a distinction Blake draws between the productive unseen of imagination and the tyrannical unseen of Urizen's hidden command. *Mechanistic interpretability* research aims to make the box less opaque.

Cleansed Doors. Blake's image from *The Marriage of Heaven and Hell* (plate 14): "If the doors of perception were cleansed every thing would appear to man as it is, infinite." In the Active Inference reading, the cleansed-doors regime is one in which precision is distributed across multiple inference channels rather than concentrated in a single faculty.

Cleansed Doors Score ($\mathcal{C}$). A bounded health-score in $[0, 1]$ combining the fourfold-balance entropy with a non-rigidity term penalising prior-channel dominance. Defined formally in §4.

Cognitive Security (COGSEC). The discipline that treats information-based threats as problems of corrupted generative models, misallocated epistemic precision, and manufactured belief. Distinct from information security in operating at the population-cognitive level rather than the network-system level [cog, 2025, 2026].

Constitutional AI. A training procedure (Anthropic, 2022) in which a model is constrained by a written constitution that takes the place of human feedback in shaping behavior [Bai et al., 2022]. In the framing of the present essay, a structured single-vision approach.

Debate-based alignment. A multi-agent alignment proposal (Irving, Christiano, Amodei 2018) in which multiple models argue against one another and a third model adjudicates [Irving et al., 2018]. Closer to fourfold vision than constitutional AI because it instantiates a multi-agent architecture.

Edge case. In machine learning, an input that lies outside the dense region of the training distribution. Edge cases are the high-precision sensory contradictions of priors that, in a healthy generative model, drive model revision. In pathologically rigid models or training paradigms, edge cases are suppressed at the source — the architectural form of what Blake names Urizen's chaining of Orc.

Engagement maximization. The optimization objective of attention-economy platforms: maximize the time and attentional resources users devote to the platform. Documented to produce sycophancy and confabulation as emergent behaviors under particular training regimes [Perez et al., 2022].

Four Zoas. Blake's four eternal persons whose coordinated labor constitutes Albion: **Urizen** (reason / law, south, head, sight), **Luvah** (passion / emotion, east, heart, scent), **Tharmas** (sensation / body, west, loins, taste), **Urthona** (imagination / prophecy, north, ear, hearing, embodied in the temporal form *Los*). In the Active Inference reading, each Zoa corresponds to a precision channel in a factorized generative model.

Fourfold Balance Entropy ($\mathcal{H}$). The Shannon entropy of the precision distribution across the four Zoa channels, in nats; maximum $\log 4 \approx 1.386$ when all four channels carry equal precision. Defined formally in §4.

Fourfold Vision. Blake's name for the Edenic mode of perception in which the four Zoas coordinate without any one dominating. Stated most economically in the 1802 letter to Thomas Butts: "fourfold in my supreme delight / And threefold in soft Beulah's night / And twofold Always. May God us keep / From Single vision & Newton's sleep!" [Blake, 1802].

Free Energy Principle (FEP). The mathematical principle proposed by Karl Friston that all self-organising biological systems act to minimize a quantity called variational free energy, which is an upper bound on the surprise of sensory observations under the system's generative model. Contested in the philosophical literature [Colombo and

Wright, 2021, Aguilera et al., 2022, Bruineberg et al., 2022]; deployed in this essay as a generative vocabulary rather than as established science.

Generative Model. The probabilistic model an inferring agent maintains over hidden states of the world and observations. In Active Inference, the generative model is constitutive of the agent: there is no agent-self underneath the model.

Glass Bead Game. From Hesse's novel; the methodological stance I have adopted in earlier work [Friedman, 2026a] and in the present essay: synthetic juxtaposition of art and science, not reduction of one to the other, with explicit acknowledgment of where the synthesis exceeds what any single tradition supplies.

Hallucination. (1) Technical: a high-confidence false output from a large language model, generated by the same statistical mechanism as accurate outputs but without empirical grounding [Ji et al., 2023]. (2) Phenomenological: in the Plato's-Cave and Active Inference framings, all perception is hallucinatory in the technical sense — model-driven prediction of sensory states — and the question is whether the model that does the hallucinating is open to revision under surprise. Jiang's 7 May 2026 sentence *"Everything is a hallucination"* [Bartlett and Jiang, 2026] compresses the two senses.

Imagination. In Blake: not a sub-faculty of cognition but the ground of human existence; the "real & eternal World of which this Vegetable Universe is but a faint shadow" (*Jerusalem* plate 77). In the Active Inference reading: the deep generative model and its temporal extension into counterfactual planning. The mapping is a functional analogy across incompatible metaphysics, not a translation.

KL Divergence (D_{KL}). The Kullback–Leibler divergence between two probability distributions; a non-symmetric measure of how much one distribution diverges from another. The complexity term in variational free energy is the KL divergence between posterior and prior.

Los. The temporal, fallen form of the Zoa Urthona; the craftsman-figure who labors at the forge of Golgonooza in *Jerusalem* to restore Albion. The mythopoetic embodiment of imagination as creative labor that maintains the agent's coherence against Urizenic enclosure.

Markov Blanket. The statistical boundary between an inferring system's internal and external states; defined by sensory states (information flowing in) and active states (influence flowing out). The blanket is dynamically maintained through the process of inference itself [Kirchhoff et al., 2018]. In the Blakean reading, the cognitive form of the *doors of perception*.

Multi-Agent Active Inference. The generalisation of Active Inference to networks of agents, each with its own generative model and Markov blanket. Belief alignment becomes the multi-agent analogue of single-agent inference; agents converge on shared beliefs through repeated cycles in which each agent's actions become evidence for the others' models [Friston et al., 2021, Hipólito et al., 2021].

Newton's Sleep. Blake's name for the pathological cognitive state in which a single faculty (typically Urizen, the rationalising principle) has gained so much precision that the others fall silent. The fallen world of single vision. In the Active Inference reading, the regime of *pathological prior dominance*. The historical Newton would have rejected this reductive deistic-mechanistic position; Blake's "Newton" is a polemical figure standing for the eighteenth-century reception of Newton in natural philosophy [Ault, 1974].

Newton's Sleep Metric ($\mathcal{N}$). The ratio of prior precision to non-prior precision in a four-channel precision-allocation; values strictly greater than 1.0 mark pathological prior dominance. Defined formally in §4.

Orc. In Blake's prophetic system, the spirit of liberated energy and revolutionary desire; the unconfinable remainder that consumes the "five gates of their law-built Heaven" in *America: A Prophecy*. The mythopoetic embodiment of the edge case in formal cognitive terms.

Pathological Prior Dominance. The Active Inference regime in which the prior precision so dominates sensory precision that the posterior is dragged toward the prior irrespective of evidence. The formal analogue of what Blake names Newton's Sleep. The structural failure mode this essay diagnoses across three vocabularies.

Precision. The inverse-variance parameter of a Gaussian distribution in a generative model; a measure of confidence in the corresponding inference channel. Precision-weighting determines how much relative influence prior beliefs versus incoming sensory evidence have in driving inference.

Single Vision. Blake's name for the reductive perceptual mode in which the imaginative, affective, and embodied registers have been silenced in favor of pure measurement. The pathology against which the corrective of fourfold

vision is offered.

Technate. A term Jiang invokes in his 7 May 2026 Game Theory #23 lecture [Jiang, 2026b] from the 1930s technocracy movement [Scott and Technocracy Inc., 1933–1948, Segal, 2005]: *"transitioning democracy into a technocracy ruled by the experts and ruled by AI."* The political form, in Jiang's framing, of the architectural single vision the present essay diagnoses.

Urizen. The Zoa of reason and law in Blake's system; "the Ancient of Days," depicted with a compass circumscribing the universe. Not a moral villain in Blake's text but an *imbalance* — a faculty that has overstepped its proper role. In the Active Inference reading, the prior-belief channel of the factorized generative model.

Urthona. The Zoa of imagination, prophecy, and the spirit; northern direction, organ of hearing. The eternal form whose fallen temporal aspect is Los. In the Active Inference reading, the deep generative model that constitutes selfhood through counterfactual planning.

Variational Free Energy (F). An upper bound on the surprise of sensory observations under an agent's generative model; defined as the KL divergence between approximate posterior and true posterior, minus the log evidence. Minimisation of F is the unifying objective of perception and action in Active Inference.

Zoa. Greek for "living one"; Blake's term for each of the four eternal persons constituting Albion. The four Zoas are Urizen, Luvah, Tharmas, and Urthona; their coordinated labor is what Blake names fourfold vision and their disordering is what he names Single Vision.

12 References

References are stored in `references.bib` and rendered by the pipeline. Below is a topical guide.

12.1 Primary texts (Blake)

Quoted throughout from David V. Erdman, ed., *The Complete Poetry and Prose of William Blake* (rev. ed., Anchor/Doubleday, 1988) [Erdman, 1988]. Specific titles: *The Marriage of Heaven and Hell*, *America: A Prophecy*, *The First Book of Urizen*, *Vala, or The Four Zoas*, *Milton: A Poem*, *Jerusalem: The Emanation of the Giant Albion*, the letter to Thomas Butts of 22 November 1802.

12.2 Standard Blake scholarship

Frye [Frye, 1947]; Damon [Damon, 1988]; Erdman *Prophet Against Empire* [Erdman, 1977]; Bloom [Bloom, 1963]; Ault *Visionary Physics* [Ault, 1974]; Otto [Otto, 2000]; Makdisi [Makdisi, 2003]; Mitchell *Blake's Composite Art* [Mitchell, 1978]; Connolly *Blake and the Body* [Connolly, 2002]; Gallant [Gallant, 1978]; Singer [Singer, 1970]; Davies [Davies, 2023]; the William Blake Archive [bla, a].

12.3 Author's Prior Work (Zenodo)

The two earlier 2026 Zenodo papers, cited throughout [Friedman, 2026a,b].

12.4 Jiang lecture corpus

The *Predictive History* lecture transcript [Jiang, 2024–2026]; *Endgame* #259 interview [Jiang and Wirjawan, 2026]; reedited YouTube compilation [Jiang, 2026a]; biographical sources [wik, 2026, South China Morning Post staff, 2025, the, 2024].

12.5 Active Inference and the Free Energy Principle

Friston [Friston, 2010]; Parr/Pezzulo/Friston textbook [Parr et al., 2022]; Kirchhoff et al. on Markov blankets [Kirchhoff et al., 2018]; Hohwy [Hohwy, 2013]; Clark [Clark, 2016]; Sajid et al. demystification [Sajid et al., 2021]; the critical literature [Colombo and Wright, 2021, Aguilera et al., 2022, Bruineberg et al., 2022].

12.6 Pragmatism / Peirce / enactivism

Pietarinen and Beni [Pietarinen and Beni, 2021, Beni and Pietarinen, 2021]; Gallagher [Gallagher, 2017, 2022]; SEP entry on Peirce abduction [Stanford Encyclopedia of Philosophy]; Misak [Misak, 2013]; Hookway [Hookway, 2012]; Menary [Menary, 2007]; Chemero [Chemero, 2009]; Hutto and Myin [Hutto and Myin, 2013]; Madzia and Jung [Madzia and Jung, 2016].

12.7 Synergetics

Fuller and Applewhite [Fuller and Applewhite, 1975]; supplementary [ful, a,b].

12.8 Cognitive security and the political economy of AI

COGSEC [cog, 2025, 2026]; Crawford [Crawford, 2021]; Zuboff [Zuboff, 2019]; Noble [Noble, 2018]; Aral [Aral, 2020]; Harris [Harris, 2017]; Gray and Suri [Gray and Suri, 2019]; Perrigo [Perrigo, 2023].

12.9 AI alignment / technical literature

Bostrom [Bostrom, 2014]; Russell [Russell, 2019]; Bender et al. [Bender et al., 2021]; Ji et al. [Ji et al., 2023]; Perez et al. on sycophancy [Perez et al., 2022]; Hendrycks and Dietterich on robustness [Hendrycks and Dietterich, 2019]; constitutional AI [Bai et al., 2022]; debate-based alignment [Irving et al., 2018]; transformer-circuit interpretability [Elhage et al., 2021, Bricken et al., 2023]; Mitchell [Mitchell, 2019]; Huxley *The Doors of Perception* [Huxley, 1954].

References

The william blake archive. Editors: Eaves, Essick, Viscomi, a. URL https://blakearchive.org.

The single vision and newton's sleep. The Blake Society, b. URL https://blakesociety.org/single-vision-and-newtons-sleep/.

Synergetics. Buckminster Fuller Institute, a. URL https://www.bfi.org/about-fuller/big-ideas/synergetics/.

Synergetics (Fuller). Wikipedia, b. URL https://en.wikipedia.org/wiki/Synergetics_(Fuller).

Active inference in robotics and artificial agents: Survey and challenges. *arXiv preprint*, 2024.

Meet the internet's new iran expert. The Free Press, 2024. URL https://www.thefp.com/p/meet-the-internets-new-iran-expert-who.

The first book of urizen by william blake. EBSCO Research Starters, 2024. Tertiary source; primary reading from Erdman 1988.

Understanding the neurocognitive mechanisms of cognitive security. *Frontiers / PubMed (PMID 41167441)*, 2025. URL https://pubmed.ncbi.nlm.nih.gov/41167441/.

Behavioral outcomes of human cognitive security within an information environment. *arXiv preprint 2603.01355*, 2026. URL https://arxiv.org/html/2603.01355v1.

Jiang xueqin. Wikipedia, 2026. URL https://en.wikipedia.org/wiki/Jiang_Xueqin.

Miguel Aguilera, Beren Millidge, Alexander Tschantz, and Christopher L. Buckley. How particular is the physics of the free energy principle? *Physics of Life Reviews*, 40:24–50, 2022.

Sinan Aral. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health*. Currency, 2020.

Donald D. Ault. *Visionary Physics: Blake's Response to Newton*. University of Chicago Press, 1974.

Yuntao Bai et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint 2212.08073*, 2022.

Steven Bartlett and Xueqin Jiang. Professor jiang: World war 3 is about to begin, let me explain! Diary of a CEO podcast, May 2026. URL https://www.youtube.com/watch?v=BTJGr78-zyw. Transcript at https://singjupost.com/the-diary-of-a-ceo-w-professor-jiang-us-draft-iran-and-end-of-the-dollar-transcript/.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? *FAccT*, 2021.

Majid D. Beni and Ahti-Veikko Pietarinen. Aligning the free-energy principle with peirce's logic of science and economy of research. *European Journal for Philosophy of Science*, 11:94, 2021.

William Blake. *The Marriage of Heaven and Hell*. 1790–1793. Cited from Erdman 1988, plates 1–27.

William Blake. *America: A Prophecy*. 1793. Cited from Erdman 1988.

William Blake. *The First Book of Urizen*. 1794. Cited from Erdman 1988.

William Blake. *Vala, or The Four Zoas*. 1797–1807. Manuscript; cited from Erdman 1988.

William Blake. Letter to thomas butts, 22 november 1802, 1802. Erdman E 720–723.

William Blake. *Jerusalem: The Emanation of the Giant Albion*. 1804–1820. Cited from Erdman 1988, plates 1–100.

Harold Bloom. *Blake's Apocalypse: A Study in Poetic Argument*. Doubleday, 1963.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Trenton Bricken, Adly Templeton, et al. Towards monosemanticity: Decomposing language models with dictionary learning. Anthropic transformer-circuits.pub, 2023.

Jelle Bruineberg, Krzysztof Dołęga, Joe Dewhurst, and Manuel Baltieri. The emperor's new markov blankets. *Behavioral and Brain Sciences*, 45:e183, 2022.

Anthony Chemero. *Radical Embodied Cognitive Science*. MIT Press, 2009.

Andy Clark. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, 2016.

Matteo Colombo and Cory Wright. First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese*, 198:3463–3488, 2021.

Tristanne J. Connolly. *William Blake and the Body*. Palgrave Macmillan, 2002.

Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

S. Foster Damon. *A Blake Dictionary: The Ideas and Symbols of William Blake*. Brown University Press / University Press of New England, revised edition, 1988. First edition 1965.

Keith Davies. *William Blake, the Single Vision, and Newton's Sleep: A History of Science*. Routledge, 2023. doi: 10.4324/9781003379263.

Glenn Diesen and Xueqin Jiang. Jiang xueqin: 'we are already in world war 3'. Greater Eurasia Podcast, May 2026. URL https://glenndiesen.substack.com/p/jiang-xueqin-we-are-already-in-world. YouTube: https://www.youtube.com/watch?v=RWTvTzWYG0o ; transcript: https://singjupost.com/jiang-xueqin-we-are-already-in-world-war-3-transcript/.

Nelson Elhage et al. A mathematical framework for transformer circuits. *Anthropic technical report*, 2021.

David V. Erdman. *Blake: Prophet Against Empire*. Princeton University Press, 3rd edition, 1977.

David V. Erdman, editor. *The Complete Poetry and Prose of William Blake*. Anchor / Doubleday, newly revised edition, 1988. Standard scholarly edition; cited as 'Erdman E <page>'.

Daniel Ari Friedman. The doors of perception are the threshold of prediction: Active inference and william blake's theory of seeing, February 2026a. URL https://zenodo.org/records/18600041.

Daniel Ari Friedman. Before pragmatism had a name: Blake's 'america a prophecy' anticipates american anticipatory epistemology, March 2026b. URL https://zenodo.org/records/18807971.

Daniel Ari Friedman, Alexander Tschantz, Maxwell J. D. Ramstead, Karl Friston, and Axel Constant. Active inference, stochastic control, and expected free energy in ant colonies. *Frontiers in Behavioral Neuroscience*, 15:647732, 2021.

Karl Friston. The free-energy principle: a unified brain theory. UAB CINL preprint.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127–138, 2010. doi: 10.1038/nrn2787.

Karl Friston, Noor Sajid, David Quiroga-Martinez, et al. Active listening and multi-agent active inference. *Neural Computation*, 2021.

Northrop Frye. *Fearful Symmetry: A Study of William Blake*. Princeton University Press, 1947.

R. Buckminster Fuller and E. J. Applewhite. *Synergetics: Explorations in the Geometry of Thinking*. Macmillan, 1975.

Shaun Gallagher. *Enactivist Interventions: Rethinking the Mind*. Oxford University Press, 2017.

Shaun Gallagher. Pragmatism and cognitive science. 2022.

Christine Gallant. *Blake and the Assimilation of Chaos*. Princeton University Press, 1978.

Mary L. Gray and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, 2019.

Tristan Harris. How a handful of tech companies control billions of minds every day. TED Talk, 2017.

Conor Heins, Beren Millidge, Daphne Demekas, et al. pymdp: A Python library for active inference in discrete state spaces. *Journal of Open Source Software*, 7(73):4098, 2022.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *ICLR*, 2019.

Inês Hipólito, Maxwell J. D. Ramstead, Laura Convertino, Anjali Bhat, Karl Friston, and Thomas Parr. Markov blankets in the brain. *Neuroscience and Biobehavioral Reviews*, 125:88–97, 2021.

Jakob Hohwy. *The Predictive Mind*. Oxford University Press, 2013.

Christopher Hookway. *The Pragmatic Maxim: Essays on Peirce and Pragmatism*. Oxford University Press, 2012.

Daniel D. Hutto and Erik Myin. *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press, 2013.

Aldous Huxley. *The Doors of Perception*. Chatto and Windus, 1954.

Geoffrey Irving, Paul Christiano, and Dario Amodei. Ai safety via debate. *arXiv preprint 1805.00899*, 2018.

Ziwei Ji et al. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 2023.

Xueqin Jiang. Predictive history lecture corpus on ai, consciousness, and empire. YouTube, 2024–2026. URL https://www.youtube.com/@PredictiveHistory. Primary transcript cross-referenced with multiple uploaded variants.

Xueqin Jiang. They control your mind through money & ai: Escape the modern cave. YouTube reedited lecture, February 2026a. URL https://www.youtube.com/watch?v=Vu1r_UHqbd0.

Xueqin Jiang. Game theory #23: The wwiii chessboard. YouTube (Predictive History channel), May 2026b. URL https://singjupost.com/game-theory-23-the-wwiii-chessboard-w-professor-jiang-transcript/. Transcript at Singju Post; original on Predictive History channel.

Xueqin Jiang. Game theory #24: The ai apocalypse. YouTube (Predictive History channel), May 2026c. URL https://www.youtube.com/watch?v=8nsxuB3Vsts. Lecture uploaded 12 May 2026; date corroborated by indexed metadata.

Xueqin Jiang and Gita Wirjawan. Jiang xueqin: Our true wealth is our consciousness. Endgame podcast #259, March 2026. URL https://www.youtube.com/watch?v=rsi7cDRUrmE.

Robert Kaufmann, Pulin Gupta, and Jacob Taylor. An active inference model of collective intelligence. *Entropy*, 23 (7):830, 2021.

Michael Kirchhoff, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. The markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138):20170792, 2018. doi: 10.1098/rsif.2017.0792.

Roman Madzia and Matthias Jung, editors. *Pragmatism and Embodied Cognitive Science: From Bodily Intersubjectivity to Symbolic Articulation*. De Gruyter, 2016.

Saree Makdisi. *William Blake and the Impossible History of the 1790s*. University of Chicago Press, 2003.

Richard Menary. *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave Macmillan, 2007.

Cheryl Misak. *The American Pragmatists*. Oxford University Press, 2013.

Melanie Mitchell. *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux, 2019.

W. J. T. Mitchell. *Blake's Composite Art: A Study of the Illuminated Poetry*. Princeton University Press, 1978.

Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. Progress measures for grokking via mechanistic interpretability. *ICLR*, 2023.

Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018.

Peter Otto. *Blake's Critique of Transcendence: Love, Jealousy, and the Sublime in The Four Zoas*. Oxford University Press, 2000.

Thomas Parr, Giovanni Pezzulo, and Karl J. Friston. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press, 2022. Open access under CC BY-NC-ND.

Ethan Perez et al. Discovering language model behaviors with model-written evaluations. *arXiv preprint 2212.09251*, 2022.

Billy Perrigo. Openai used kenyan workers on less than $2 per hour. *TIME*, 2023.

Ahti-Veikko Pietarinen and Majid D. Beni. Active inference and abduction. *Biosemiotics*, 14:499–517, 2021. doi: 10.1007/s12304-021-09432-0.

Maxwell J. D. Ramstead, Dalton A. R. Sakthivadivel, Conor Heins, Magnus Koudahl, Beren Millidge, Lancelot Da Costa, Brennan Klein, and Karl J. Friston. On bayesian mechanics: a physics of and by beliefs. *Interface Focus*, 13(3):20220029, 2023.

Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.

Noor Sajid, Philip J. Ball, Thomas Parr, and Karl J. Friston. Active inference: Demystified and compared. *Neural Computation*, 33(3):674–712, 2021.

William Saunders, Girish Sastry, Andreas Stuhlmüller, and Paul Christiano. Self-critiquing models for assisting human evaluators. *arXiv preprint*, 2022.

Howard Scott and Technocracy Inc. The Technate of North America: Concept and program. Technocracy Inc. archival publications, 1933–1948. Historical reference for the technocracy movement Jiang invokes.

Howard P. Segal. *Technological Utopianism in American Culture*. Syracuse University Press, 2005. Expanded edition of the 1985 original.

June Singer. *The Unholy Bible: A Psychological Interpretation of William Blake*. Putnam, 1970.

South China Morning Post staff. Profile of jiang xueqin and the Predictive History channel. South China Morning Post, article 3348289, 2025. URL https://www.scmp.com/news/china/politics/article/3348289.

Stanford Encyclopedia of Philosophy. Peirce on abduction. URL https://plato.stanford.edu/entries/abduction/peirce.html.

Daniel Tutt. Jiang xueqin, the polymarket professor: The perils of his geopolitical grift. TripleAmpersand Journal, April 2026. URL https://tripleampersand.org/jiang-xueqin-the-polymarket-professor-the-perils-of-his-geopolitical-grift/.

Shoshana Zuboff. *The Age of Surveillance Capitalism*. PublicAffairs, 2019.