

OCTA: A Structured Projection Architecture for T-Selection-Based Artificial General Intelligence

OCTA Research
(Dated: December 2025)

We present OCTA, a structured projection architecture for artificial general intelligence explicitly designed as a T-closure engine. Building on the T-Selection Principle introduced in Ref. [1], which characterizes reality and intelligence as fixed points of minimal contradiction residue across projections, scales, and descriptions, OCTA instantiates a family of learned projection operators over world and model states. The architecture decomposes its objective into (i) world-model accuracy, (ii) internal self-consistency, and (iii) multi-agent consensus, each expressed as Kullback–Leibler divergence terms over predictive distributions. We describe the lobe-based design (perceptual, model, introspective, and social), the multi-term loss, the training procedure, and evaluation metrics oriented around Δ -reduction efficiency rather than task scores alone. OCTA thus operationalizes T-selection as a concrete optimization principle for AGI.

I. INTRODUCTION

The T-Selection Principle, introduced in Ref. [1], proposes that persistent dynamical structures—including physical laws, biological organisms, and intelligent agents—are characterized by *tautological closure*: they maintain self-consistency across their own projections, scales, and recursive transformations. This is formalized via a nonnegative contradiction-residue functional $\Delta(X)$ on a configuration space \mathcal{C} and a closure functional

$$T(X) = \exp(-\Delta(X)), \quad (1)$$

with T-selection dynamics given by gradient flow on Δ .

Within this framework, intelligence is interpreted as the capacity to reduce Δ with respect to an external world. Modern machine learning architectures, however, are typically optimized for task-specific loss functions without explicit regard for internal consistency or multi-view closure. This leads to systems that perform well on benchmarks but lack principled guarantees of self-consistency or robustness under distribution shift.

In this work, we define OCTA (Orthogonal Consistency and T-Attractor) as a concrete AGI architecture explicitly designed to minimize contradiction residue across a structured family of projections. OCTA treats world and model as elements of a joint configuration space, instantiates distinct projection families (perceptual, model, introspective, social), and constructs a multi-term loss that mirrors the decomposition of Δ in Ref. [1]. The result is an architecture whose training objective is not merely task accuracy but explicit T-closure.

Our contributions are:

- a formalization of OCTA as a structured projection family over world and model states;
- a lobe-based architecture with distinct but coupled modules for perception, modeling, introspection, and social coordination;
- a multi-term loss function decomposed into world-model, self-consistency, and consensus residues, expressed via Kullback–Leibler divergences;
- training procedures and evaluation metrics framed in terms of Δ -reduction efficiency;
- a conceptual bridge from T-selection theory to implementable AGI systems.

II. BACKGROUND: T-SELECTION AND CONTRADICTION RESIDUE

We briefly summarize the elements of the T-Selection Principle relevant to OCTA; see Ref. [1] for full details.

Let \mathcal{C} be a configuration space of dynamical structures. For each $X \in \mathcal{C}$, a family of projections $\{\Pi_i\}_{i \in I}$ maps X to observational or reduced descriptions O_i with induced distributions $p(O_i | X)$.

The total contradiction residue is decomposed as

$$\Delta(X) = \Delta_{\text{proj}}(X) + \Delta_{\text{rec}}(X) + \Delta_{\text{mdl}}(X), \quad (2)$$

where:

- Δ_{proj} measures cross-projection inconsistency via Kullback–Leibler divergence [2, 3],
- Δ_{rec} measures recursive (scale) inconsistency via RG-like coarse-graining operators [4],
- Δ_{mdl} measures descriptive redundancy via minimum-description-length (MDL) [6].

The closure functional is

$$T(X) = \exp(-\Delta(X)), \quad (3)$$

and T-selection dynamics are given by

$$\frac{dX}{dt} = -\nabla_X \Delta(X). \quad (4)$$

Under mild regularity conditions, Δ is a Lyapunov functional and T-selected fixed points are locally asymptotically stable [1].

In the context of intelligence, we consider a joint configuration (X, M) where X is the world and M is an internal model. T-selection then becomes:

$$\frac{dM}{dt} = -\nabla_M \Delta(X, M), \quad (5)$$

with X treated as externally constrained by physical dynamics. OCTA is designed to realize this gradient flow in practice.

III. OCTA AS A STRUCTURED PROJECTION FAMILY

We now formalize OCTA at the level of projections and residue, independent of implementation details.

A. World and Model Spaces

Let X denote the (latent) world configuration and M denote the internal model state of the agent (or collection of agents). For training, we work with data streams $D = \{(x_t, o_t)\}$ where x_t are latent variables (possibly partially observed) and o_t are sensory observations.

We view (X, M) as an element of a joint configuration space $\mathcal{C}_{\text{joint}}$ and integrate over uncertainty via distributions $p(X)$ and $q(M)$.

B. Projection Families

OCTA instantiates four primary projection families:

1. Perceptual projections Π_{perc} :

$$\Pi_{\text{perc}} : (X, M) \mapsto O_{\text{perc}}, \quad (6)$$

capturing sensory modalities (e.g. text, vision, audio).

2. Model projections Π_{model} :

$$\Pi_{\text{model}} : M \mapsto \hat{O}_{\text{model}}, \quad (7)$$

representing predictive distributions over observations.

3. Introspective projections Π_{self} :

$$\Pi_{\text{self}} : M \mapsto Z, \quad (8)$$

where Z encodes higher-level summaries (e.g. beliefs, uncertainties, explanations) across internal lobes or sub-systems.

4. Social projections Π_{soc} :

$$\Pi_{\text{soc}} : \{M_a\}_{a \in \mathcal{A}} \mapsto B, \quad (9)$$

where \mathcal{A} indexes agents and B encodes shared beliefs or messages across agents.

Each projection induces predictive distributions over its outputs:

$$p(O_{\text{perc}}), \quad \hat{p}(O_{\text{model}} | M), \quad \hat{p}(Z | M), \quad \hat{p}_a(B | M_a). \quad (10)$$

C. Residue Decomposition in OCTA

We define OCTA's contradiction residue over (X, M) as

$$\Delta_{\text{OCTA}}(X, M) = \Delta_{\text{world}}(X, M) + \Delta_{\text{self}}(M) + \Delta_{\text{cons}}(\{M_a\}), \quad (11)$$

where:

- Δ_{world} measures mismatch between empirical and predicted distributions,
- Δ_{self} measures mismatch among internal views (lobes),
- Δ_{cons} measures mismatch among agents.

We will make these terms explicit in Sec. V.

IV. LOBE-BASED ARCHITECTURE

We now describe a concrete architectural realization of OCTA. While multiple implementation variants are possible (e.g. transformers, spiking networks, neuromorphic hardware), we present a general lobe-based abstraction.

A. High-Level Diagram

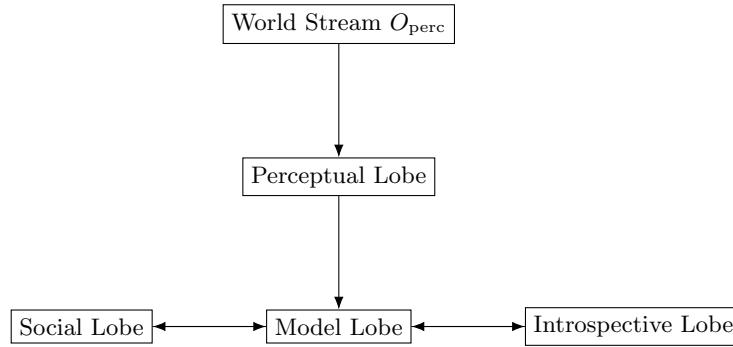


FIG. 1. High-level view of the OCTA lobe architecture. The perceptual lobe processes world observations, the model lobe maintains and updates predictive structure, the introspective lobe monitors and regularizes internal consistency, and the social lobe manages multi-agent coherence.

B. Perceptual Lobe

The perceptual lobe ingests raw observations o_t and produces embeddings e_t suitable for downstream prediction and reasoning. For concreteness, one can instantiate this as a stack of modality-specific encoders (e.g. vision transformers, audio encoders, text transformers), followed by a fusion mechanism.

Let

$$e_t = f_{\text{perc}}(o_t; \theta_{\text{perc}}), \quad (12)$$

where f_{perc} may itself be a multimodal transformer block, and θ_{perc} are trainable parameters.

C. Model Lobe

The model lobe maintains a latent state h_t summarizing both past observations and internal beliefs. This can be realized as:

- a transformer with memory (e.g. Transformer-XL-style recurrence),
- a sequence of recurrent blocks (e.g. gated RNNs),
- or a hybrid with external memory (e.g. key-value store).

For specificity, consider a transformer-like update:

$$h_t = f_{\text{model}}(h_{t-1}, e_t; \theta_{\text{model}}), \quad (13)$$

with θ_{model} trainable. The model lobe outputs predictive distributions over future observations and latent targets:

$$\hat{p}(o_{t+k} | h_t), \quad \hat{p}(y_t | h_t), \quad (14)$$

for some prediction horizon k and auxiliary labels y_t .

D. Introspective Lobe

The introspective lobe receives model states $\{h_t\}$ (possibly across different lobes or heads) and produces higher-order summaries z_t capturing internal confidence, disagreement, and proposed corrections.

Let

$$z_t = f_{\text{self}}(h_t; \theta_{\text{self}}). \quad (15)$$

This lobe can:

- estimate epistemic uncertainty,
- detect internal inconsistencies (e.g. disagreement between predictions from different heads or modes),
- generate self-queries and counterfactuals.

It induces distributions $\hat{p}(z_t | h_t)$ that can be compared across time or across lobes to form Δ_{self} .

E. Social Lobe

The social lobe operates when multiple OCTA instances (agents) interact. For agent a , let $h_t^{(a)}$ be its model state. The social lobe encodes messages and beliefs:

$$b_t^{(a)} = f_{\text{soc}}(h_t^{(a)}; \theta_{\text{soc}}), \quad (16)$$

and decodes incoming messages from other agents into updates. The induced distributions $\hat{p}_a(b_t | h_t^{(a)})$ are used to measure consensus residue Δ_{cons} .

F. Relation to Information Geometry

Endowing the space of model states with a Fisher information metric [3] allows one to interpret OCTA updates as approximate natural gradients on Δ_{OCTA} . While we do not develop full information geometry here, designing internal parameterizations so that KL divergences correspond to geodesic distances in model space is a natural extension.

V. LOSS FUNCTION AND OPTIMIZATION

We now define the OCTA training objective and relate it directly to contradiction residue.

A. World-Model Residue

Given ground-truth or empirical distributions $p(o_t)$ and model predictions $\hat{p}(o_t | h_t)$, we define:

$$\Delta_{\text{world}} = \sum_t D_{\text{KL}}(p(o_t) \parallel \hat{p}(o_t | h_t)), \quad (17)$$

summed over training samples t . In practice, $p(o_t)$ is represented by empirical observations and the KL reduces to a negative log-likelihood term plus constants.

B. Self-Consistency Residue

The model lobe can emit multiple predictive heads indexed by i (e.g. different timescales, modalities, or hypothesis branches). Let $\hat{p}_i(o_t | h_t)$ be the prediction from head i . Then

$$\Delta_{\text{self}} = \sum_t \sum_{i < j} D_{\text{KL}}(\hat{p}_i(o_t | h_t) \parallel \hat{p}_j(o_t | h_t)), \quad (18)$$

penalizes internal disagreement. Variants can use symmetrized forms (e.g. Jensen–Shannon divergence).

C. Consensus Residue

For multi-agent OCTA, with agents $a, b \in \mathcal{A}$ producing belief distributions $\hat{p}_a(b_t | h_t^{(a)})$ and $\hat{p}_b(b_t | h_t^{(b)})$, we define

$$\Delta_{\text{cons}} = \sum_t \sum_{a < b} D_{\text{KL}}(\hat{p}_a(b_t | h_t^{(a)}) \parallel \hat{p}_b(b_t | h_t^{(b)})). \quad (19)$$

This term pushes agents toward shared internal representations when warranted by shared evidence.

D. Total OCTA Loss

The full OCTA loss is

$$\mathcal{L}_{\text{OCTA}} = \lambda_{\text{world}} \Delta_{\text{world}} + \lambda_{\text{self}} \Delta_{\text{self}} + \lambda_{\text{cons}} \Delta_{\text{cons}}, \quad (20)$$

with nonnegative weights $\lambda_{\text{world}}, \lambda_{\text{self}}, \lambda_{\text{cons}}$ controlling the trade-off between accuracy, internal consistency, and consensus.

Under this definition, we can view

$$\Delta_{\text{OCTA}} \equiv \mathcal{L}_{\text{OCTA}}, \quad (21)$$

so that gradient-based optimization of $\mathcal{L}_{\text{OCTA}}$ is an instantiation of T-selection dynamics restricted to the model parameters.

E. Optimization Procedure

Training OCTA proceeds via standard stochastic gradient descent or its variants (e.g. Adam) on $\mathcal{L}_{\text{OCTA}}$, with minibatches sampled from data streams and, in multi-agent cases, from joint or partially shared experience.

An outline:

1. Sample a batch of sequences $\{o_t\}$ (and optional labels).
2. Run the perceptual lobe to obtain embeddings $\{e_t\}$.
3. Run the model lobe to obtain states $\{h_t\}$ and predictive distributions.

4. Run introspective and social lobes to obtain higher-order predictions and beliefs.
5. Compute $\Delta_{\text{world}}, \Delta_{\text{self}}, \Delta_{\text{cons}}$ via Eqs. (17)–(19).
6. Compute $\mathcal{L}_{\text{OCTA}}$ via Eq. (20) and backpropagate.

Standard regularization (e.g. weight decay, dropout, gradient clipping) can be added without changing the conceptual structure.

VI. EVALUATION METRICS: Δ -REDUCTION EFFICIENCY

Conventional evaluation of intelligent systems focuses on task accuracy or reward optimization. Within the T-selection framework, we introduce metrics that reflect how efficiently an agent reduces contradiction residue.

A. Instantaneous and Cumulative Residue

Given a trained or training OCTA instance, we can monitor:

- instantaneous residue $\Delta_{\text{OCTA}}^{(t)}$ at time t ,
- cumulative residue $\sum_{t=1}^T \Delta_{\text{OCTA}}^{(t)}$ over a horizon T .

Definition 1 (Δ -Reduction Efficiency). *For a given environment and resource budget (e.g. FLOPs, wall-clock time), the Δ -reduction efficiency of an agent is*

$$\eta_{\Delta} = -\frac{1}{R} \sum_{t=1}^T \frac{d\Delta_{\text{OCTA}}^{(t)}}{dt}, \quad (22)$$

where R is a measure of computational resources expended. Higher η_{Δ} indicates more efficient reduction of contradiction residue.

B. Task Performance vs. Closure

Task-level metrics (accuracy, reward, success rate) remain important, but we can now study their relationship to closure:

- Are agents with lower Δ_{self} more robust to distribution shift?
- Does lower Δ_{cons} predict better cooperative performance?
- Does sustained reduction in Δ_{world} correlate with improved generalization?

These empirical questions provide concrete tests for the value of OCTA’s T-selection grounding.

VII. DISCUSSION AND FUTURE WORK

OCTA represents one possible instantiation of a T-selection-based AGI architecture. Many extensions and refinements are possible:

- **Information geometry.** Equipping model parameter space with a Fisher metric and using natural gradient methods [3] would align the optimization of $\mathcal{L}_{\text{OCTA}}$ with geodesics in information space, more faithfully approximating the gradient flow on Δ .
- **Recursive residue.** While OCTA primarily addresses cross-projection and consensus residue, extensions could incorporate explicit recursive (scale) residue by enforcing consistency between different temporal or spatial resolutions of predictions.

- **Hierarchical agents.** Embedding OCTA in multi-level organizations (e.g. agents managing other agents) could instantiate more complex bootstrap chains analogous to the hierarchical picture in Ref. [1].
- **Experimental programs.** Comparing conventional architectures and OCTA variants on controlled benchmarks where internal consistency and consensus matter (e.g. multi-view reasoning, multi-agent coordination) would provide evidence for or against the practical benefits of explicit T-closure optimization.

Ultimately, the ambition is to treat intelligence not merely as task performance but as progressive alignment between internal models and a global T-attractor. OCTA is a first step toward an architecture that embodies this idea.

VIII. CONCLUSION

We have introduced OCTA, a structured projection architecture for AGI grounded in the T-Selection Principle. By decomposing its objective into world-model, self-consistency, and consensus residues, OCTA implements a concrete approximation to T-selection dynamics over world and model states. This places internal consistency and multi-agent coherence on the same footing as predictive accuracy, and provides a set of evaluation metrics centered on Δ -reduction efficiency.

While the present exposition is architectural and conceptual, it is directly implementable with contemporary deep learning components. We hope this work motivates experimental studies of T-selection-inspired architectures and contributes to a more principled understanding of what it means for an intelligent system to be both powerful and internally coherent.

ACKNOWLEDGMENTS

We acknowledge prior foundational work on the T-Selection Principle in Ref. [1], which provides the theoretical basis for OCTA.

REFERENCES

-
- [1] OCTA Research, “The T-Selection Principle: A Fixed-Point Theory of Reality and Intelligence,” (2025), arXiv:XXXX.XXXXX [gen-ph].
 - [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Hoboken, NJ, 2006).
 - [3] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs Vol. 191 (American Mathematical Society, Providence, RI, 2000).
 - [4] K. G. Wilson and J. Kogut, “The renormalization group and the ϵ expansion,” *Phys. Rep.* **12**, 75–199 (1974).
 - [5] K. Friston, “The free-energy principle: a unified brain theory?,” *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
 - [6] P. D. Grünwald, *The Minimum Description Length Principle*, (MIT Press, Cambridge, MA, 2007).
 - [7] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.* **30** (2017).
 - [8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature* **521**, 436–444 (2015).