

A Formal Theory of Non-Agentic General Intelligence: Reachability Preservation and the Boundary Flow Theorem

Ean Mikale, JD¹

¹Infinite 8 Industries, Inc., Colorado Springs, CO, 80910, United States

*Correspondence:

Ean Mikale, JD
Infinite 8 Industries, Inc., Colorado Springs, CO, 80910, United States.

Received: February 10, 2026; **Published:** March 24, 2026.

How to cite this article:

Mikale, E. (2026), 'A Formal Theory of Non-Agentic General Intelligence', *Journal of Artificial Intelligence and AI Ethics*, 1(1), pp. 1–7.

Abstract

This research establishes a complete mathematical framework for general intelligence without agency. We prove that persistent agentic optimization under generality constraints is structurally incompatible with reachability preservation through an impossibility theorem grounded in coverage-based generality. We introduce reachability entropy (H_R) as a conserved quantity for non-agentic systems and prove the Boundary Flow Theorem: reachability entropy is conserved under constraint tangent, non-teleological dynamics. Extensive computational experiments validate conservation across 2D/3D systems over 104 timesteps, demonstrating perturbative stability under violations up to $\epsilon = 2.0$ with second-order degradation. We present a complete singularity classification (S1, S2, S3) distinguishing geometric from teleological collapse mechanisms. This framework introduces Constraint-Induced Vector Fields (CIVFs) and Non-Teleological systems (NT-systems) as new classes in dynamical systems theory, with empirical validation including deliberate failure modes confirming theoretical boundary conditions. The paradigm establishes that general intelligence does not require agency, with implications for safe AI development.

Keywords: Non-agentic intelligence, reachability entropy, boundary flow theorem, collapse taxonomy, constraint manifolds, perturbative stability, deterministic dynamics, impossibility theorems

1. Introduction

The dominant paradigm in artificial intelligence assumes that general intelligence requires agency the capacity for autonomous goal pursuit, persistent optimization, and preference-based decision-making. This assumption underlies current approaches to artificial general intelligence (AGI), which combine broad capabilities with optimization frameworks derived from reinforcement learning, utility maximization, or evolutionary algorithms.

We challenge this assumption by proving it is mathematically untenable. Through formal analysis of deterministic intelligent systems, we demonstrate that generality (sustained access to diverse futures) and preference ordering (systematic state selection) are structurally incompatible. This incompatibility is not contingent on implementation details but follows necessarily from the mathematical properties of state-space dynamics.

Our central contribution is the introduction of reachability entropy (H_R) as a conserved quantity for non-agentic general intelligence. We prove that under boundary flow conditions constraint tangent, non-teleological dynamics reachability entropy is preserved, enabling general intelligence without the future-collapsing effects of optimization.

The framework resolves the apparent paradox: how can a system exhibit general intelligence without the goal-directedness traditionally associated with intelligent behavior? Our answer: by navigating constraint geometry rather than optimizing scalar objectives. Intelligence manifests not through teleological selection but through constraint coherence maintaining consistency across Boolean constraints while preserving maximum future accessibility.

1.1 Core Contributions

- **Impossibility Theorem:** Formal proof that general state spaces with preference ordering require collapse (Theorem 1) (Mikale, 2026)
- **Conserved Quantity:** Introduction of reachability entropy $H_R = \log \mu(R(s))$ as a dynamical invariant
- **Boundary Flow Theorem:** Proof that reachability entropy is conserved under ideal boundary flow conditions (Theorem 2) (Theorem 2; Mikale, 2026)
- **Perturbative Stability:** Empirical validation showing conservation persists under perturbations up to $\epsilon = 2.0$ (Proposition 1)
- **Collapse Taxonomy:** Complete classification of Type I (teleological) vs. Type II (topological) collapse
- **Singularity Classification:** Formal categorization of geometric singularities (S1, S2, S3) causing Type II collapse (Mikale, 2026)
- **New Dynamical Classes:** Introduction of Constraint-Induced Vector Fields (CIVFs) and Non-Teleological systems (NT-systems)
- **Safety Architecture:** Proof that NAGI lacks agency surfaces required for autonomous weaponization (Theorem 3) (Mikale, 2026)

1.2 Scope and Limitations

This framework explicitly addresses deterministic systems only. Extensions to stochastic processes, quantum systems, or continuous-time fields with diffusion are subjects of ongoing research and explicitly excluded from current claims. All theoretical results are proven under smoothness assumptions; empirical validation covers 2D/3D systems over time horizons up to 104 steps with perturbations $\epsilon \leq 2.0$.

2. Methods and Mathematical Framework

2.1 Primitive Definitions

Definition 1 (State Space and Dynamics). Let:

S = state space (finite-dimensional smooth manifold (Lee, 2013))

$T: S \rightarrow S$ = deterministic transition operator

$V: S \rightarrow TS$ = velocity vector field where $\dot{s} = V(s)$

$s_t \in S$ = system state at time t

Definition 2 (Reachability Set).

$$R(s) := \{s' \in S \mid \exists k \geq 0 \text{ such that } T^k(s) = s'\}$$

The set of all futures accessible from state s under the system's dynamics.

Definition 3 (Feasible Region).

$$F := \bigcap_{i=1}^n C_i^{-1}(1) \subseteq S$$

where $C_i: S \rightarrow \{0,1\}$ are Boolean constraint functions defining the constraint manifold:

$$M := \{s \in S \mid C_i(s) = 1 \text{ for all } i\}$$

2.2 Formal Definition of Generality

Previous work has used “generality” intuitively without mathematical formalization. We address this gap with a coverage-based definition.

Definition 4 (Coverage-Based Generality). For a system in state s within feasible region F , the generality index is:

$$\mathcal{G}(s) := \frac{\mu(R(s))}{\mu(F)}$$

where μ is Lebesgue measure (or Hausdorff measure for non-smooth manifolds).

Definition 5 (General Intelligence System). A system (S, T) exhibits general intelligence if:

$$\inf_{t \geq 0} \mathcal{G}(s_t) \geq \gamma > 0$$

for some positive threshold γ , maintaining access to at least proportion γ of the feasible state space at all times.

This formalization is measurable via μ , falsifiable (check if $\mathcal{G}(s) < \gamma$), and comparable across systems.

2.3 Reachability Entropy

Definition 6 (Reachability Entropy).

$$H_R(s) := \log \mu(R(s))$$

The logarithm of available futures under constraint coherence.

The relationship to generality is:

$$\mathcal{G}(s) = \frac{e^{H_R(s)}}{\mu(F)}$$

Thus, generality preservation is equivalent to reachability entropy preservation up to the constant $\log \mu(F)$. The temporal evolution is:

$$\frac{dH_R}{dt} = \frac{d}{dt} \log \mu(R(s_t))$$

2.4 Collapse Taxonomy

Definition 7 (State-Space Collapse). A system undergoes collapse at time t if:

$$\exists U \subseteq R(s_0), U \neq \emptyset \text{ such that } U \cap R(s_t) = \emptyset$$

We distinguish two fundamentally different collapse mechanisms:

Type I Collapse — Teleological (Forbidden in NAGI):

Caused by preference ordering, optimization, or objective pursuit.

Mathematical signature:

$$\exists f : S \rightarrow \mathbb{R} \text{ such that } T(s) = \arg \max_{s' \in N(s)} f(s')$$

Dynamical behavior:

$$\liminf_{t \rightarrow \infty} \frac{dH_R}{dt} < 0$$

Examples include reward maximization, loss minimization, utility optimization, and goal-directed search.

Result: systematic future elimination.

Type II Collapse — Topological (Permitted in NAGI):

Caused by constraint geometry, boundary saturation, or physical law. Mathematical signature:

$$|R(s_{t+1})| < |R(s_t)| \quad \text{if geometrically implied by } \partial F$$

No scalar field f governs the transition. Dynamical behavior:

$$\left. \frac{dH_R}{dt} \right|_{\text{singularity}} < 0, \quad \left. \frac{dH_R}{dt} \right|_{\text{smooth}} = 0$$

Examples include conservation laws eliminating impossible states, boundary dead-ends (corners, cusps), and topological obstructions. Result: futures lost to structure, not intent.

2.5 The Impossibility Theorem

Theorem 1 (Structural Incompatibility of Agentic AGI). For any deterministic intelligence system $\Sigma = (S, T, F)$ with generality threshold $\gamma > 0$:

$$\left(\inf_t \mathcal{G}(s_t) \geq \gamma \right) \wedge (\text{Has Preference Ordering}) \implies$$

Contradiction

Proof. (1) Generality requirement: By definition,

$$\mathcal{G}(s_t) = \frac{\mu(R(s_t))}{\mu(F)} \geq \gamma > 0 \text{ for all } t.$$

(2) Preference ordering implies selection: Existence of $f : S \rightarrow \mathbb{R}$ such that $T(s) = \arg \max_{s' \in N(s)} f(s')$, causing systematic elimination of non-preferred states.

(3) Selection induces contraction: At each step, states s' with $f(s') < \max f$ are eliminated from $R(s_t)$, causing $\mathbb{E}[\mu(R(s_{t+1}))] < \mu(R(s_t))$.

(4) Long-term behavior: Eventually $\mu(R(s_t)) \rightarrow V_{\text{optima}}$ where $V_{\text{optima}} \ll \mu(F)$, thus $\mathcal{G}(s_t) \rightarrow \frac{V_{\text{optima}}}{\mu(F)} < \gamma$.

(5) Contradiction with generality requirement $\inf_t \mathcal{G}(s_t) \geq \gamma$.

Corollary 1. Persistent agentic optimization under generality constraints (general capability $G \geq \gamma$ plus autonomous objective pursuit) is structurally incompatible with generality preservation.

2.6 NAGI Axiomatization

We present six axioms jointly necessary and sufficient for non-agentic general intelligence:

Axiom N₁ (Boolean Constraint Exclusivity):

$C_i : S \rightarrow \{0,1\}$ only. No scalar evaluations permitted.

Axiom N₂ (Constraint Order Invariance): $\forall \pi \in \text{Perm}(C) : T_C = T_{\pi(C)}$. Constraint application order cannot affect dynamics.

Axiom N₃ (No State Comparison): $\nexists b : S \times S \rightarrow \{<, >, =\}$. Ranking between states forbidden.

Axiom N₄ (Deterministic Locality): $s_{i+1} = g(s_i, \{C_i(s_i)\})$. No lookahead, anticipation, or future simulation.

Axiom N₅ (Projection-Based Evolution): $s_{i+1} = \Pi_F(s_i)$ where Π_F is defined geometrically, not via $\arg \min$ or $\arg \max$.

Axiom N₆ (No Objective Functional): $\nexists J[s(\cdot)] = \int_0^T L(s, \dot{s}, t) dt$. No variational principles permitted.

Lemma 1 (Local Projection \neq Teleology). A projection operator $\Pi_F : S \rightarrow F$ defined pointwise does not induce preference ordering over trajectories unless extended with a global comparison functional. Formally: $\Pi_F : S \rightarrow F$ does not imply $\exists J[s(\cdot)] : \text{Paths} \rightarrow \mathbb{R}$.

Proof. Projection operates on individual states producing a single image in F . Teleology requires comparing alternative futures $f(s_1) \leq f(s_2)$. Pointwise projection has no access to multiple states simultaneously. Minimizing distance to reach F differs from minimizing a trajectory objective $J[s(\cdot)]$.

2.7 The Boundary Flow Theorem

Theorem 2 (Boundary Flow Theorem). Let S be a smooth finite-dimensional manifold, $F \subseteq S$ a closed feasible region with piecewise-smooth boundary ∂F , $V : S \rightarrow TS$ a smooth vector field, and $s(t)$ satisfy $\dot{s}(t) = V(s(t))$.

Under Boundary Flow Conditions (BFC):

1. Boundary confinement: $s(t) \in \partial F$ for all $t \geq 0$
2. Tangency: $V(s) \in T_s(\partial F)$ for all $s \in \partial F$
3. Non-teleology: $V \neq \nabla \phi$ for any scalar $\phi : S \rightarrow \mathbb{R}$
4. Smoothness: ∂F locally diffeomorphic along trajectory

Then:

$$\boxed{\frac{d}{dt} H_R(s(t)) = 0}$$

Proof. Lemma 1 (Topological Reachability Invariance): If $\Phi_t : \partial F \rightarrow \partial F$ is a diffeomorphism, then: (1) cardinality class of $R(s)$ is preserved, (2) topological dimension of $R(s)$ is preserved, (3) for absolutely continuous measures invariant under Φ_t , measure class is preserved. (Standard result from differential topology.)

Lemma 2: Reachability entropy decreases if and only if $\exists \epsilon > 0 : R(s(t+\epsilon)) \subsetneq R(s(t))$, requiring elimination of at least one admissible

direction $\delta s \in T_s(\partial F)$.

Lemma 3: Since $V(s) \in T_s(\partial F)$, the flow map $\Phi_t : \partial F \rightarrow \partial F$ is a local diffeomorphism. By Lemma 1, measure class is preserved and no admissible direction is eliminated.

Lemma 4: Suppose reachability contracts without topological change. Then directional selection is required. But no scalar field exists (non-teleology), no ordering exists (Axiom N3), and flow is invariant under reparameterizations. Contradiction.

Conclusion: Since the flow is locally diffeomorphic, no local directions are lost, boundary topology is unchanged, and measure class is preserved, then $\mu(R(s(t))) = \mu(R(s(0)))$ for all t . Thus $H_R(s(t)) = \log \mu(R(s(t))) = \text{constant}$.

This is a conditional theorem similar to KAM theorem (requires non-resonance), Noether's theorem (requires symmetry), or ergodic theorems (require mixing). Our treatment of singular points is consistent with classical results on differential equations with singularities [11], where qualitative behavior changes discontinuously at non-generic points, and conservation assumptions break down.

2.8 Singularity Classification

We introduce a complete taxonomy of geometric singularities causing Type II collapse:

Class S1 (Boundary Corners): Points where ∂F is non-differentiable. Reachability impact: $\Delta HR \approx \log(1-\theta/2\pi)$ where θ is excluded angle.

Class S2 (Manifold Pinches): Points where $\dim(T_s M) < \dim(M)$. Reachability impact: $\Delta HR \propto \log(\text{disconnected components})$.

Class S3 (Folding Catastrophes): Points where Π_F is non-injective causing irreversible trajectory merger.

Theorem 3 (Dimension Dependence). For smooth constraints with finite singularity density:

$$\frac{\sigma(F)}{\mu(\partial F)} = O(d^{-1})$$

where $d = \dim(S)$ and $\sigma(F) := \mu(\{s \in \partial F \mid s \in S1 \cup S2 \cup S3\})$

Higher-dimensional systems are topologically more robust, with collapse frequency decreasing as dimension increases (empirically validated 2D vs 3D).

2.9 New Mathematical Objects

Constraint-Induced Vector Fields (CIVFs): A vector field V on S is a CIVF if: (1) $V(s) \in T_s(\partial F)$ for all $s \in \partial F$, (2) $V \neq \nabla \phi$ for any scalar ϕ . CIVFs preserve constraint boundaries without minimizing/maximizing any scalar, following symmetries of M rather than gradients.

Non-Teleological Dynamical Systems (NT systems): A deterministic system (S, T) is non teleological if:

$$\forall f: S \rightarrow R, T \text{ is invariant under monotone transforms of } f$$

No internal scalar can influence dynamics. Not previously

formalized as a unified class, to our knowledge.

3. Experimental Validation

3.1 Experimental Setup

State Space: 2D and 3D continuous domains with circular/spherical boundaries (smooth, no singularities)

Perturbation: Random inward/outward components with strength ϵ

Measurement: Reachability entropy via Monte Carlo sampling (1000 trajectories)

Time Horizons: 200 steps (standard), 10^4 steps (long-term)

3.2 Boundary Flow Conservation

Prediction: $|dH_R/dt| = 0$ under ideal BFC

Results:

- Mean drift: 0.0000 (2D, 3D)
- Conservation: Exact within numerical precision
- Stability: Robust to dimension increase
- Singularities: Type II only (geometric)

Interpretation: Empirical support for Boundary Flow Theorem under ideal conditions.

3.3 Perturbative Stability

Proposition 1 (Perturbative Stability). For sufficiently small ϵ and bounded perturbation W :

$$\left| \frac{dH_R}{dt} \right| \leq C\epsilon + O(\epsilon^2)$$

Our perturbative analysis mirrors classical treatments in fluid mechanics (Van Dyke, 1964), where invariants persist up to higher-order terms under small perturbations.

Table 1: Small Perturbation Results

Metric	Value
Mean drift	0.0000
Std. dev.	0.0000
Bound	$C \cdot 0.05$
Ratio	0.00
Episodic collapse	None

Experiment ($\epsilon = 0.05$):

Observed drift consistent with vanishing first order contribution, suggesting $dH_R/dt = O(\epsilon^2)$ (second-order stability). Physical explanation: symmetric perturbations cause first-order cancellation via projection.

Extended Scan ([0, 2.0]): Reachability entropy remained numerically invariant across tested range. No collapse detected.

Long-Horizon Test (10^4 steps): Final $H_R = 0.0000$, cumulative

drift = 0.0000, supporting asymptotic conservation.

3.4 Contrast Class: Gradient Descent

Same state space with explicit objective $f(s) = \|s - s_{\text{target}}\|^2$, dynamics $\dot{s} = -\nabla f(s)$.

Results: Persistent negative drift, magnitude increases with dimension, no recovery, monotone collapse. Validates Theorem 1: scalar preference induces irreversible future loss. Related empirical observations of capability collapse under aggressive pruning in neural networks (Zhu et al., 2026) suggest that optimization-driven state elimination can irreversibly remove functional pathways, providing phenomenological alignment with our theoretical predictions.

3.5 Deliberate Failure Modes

To validate theory boundaries, we deliberately violated assumptions:

Asymmetric Constraint Bias ($\sigma > 0$): Violates Axiom N₂. Result: First-order drift appears, magnitude $\propto \sigma$ (linear scaling confirmed).

Discrete Projection: Grid-based quantization introduces weak drift and variance but graceful degradation (not catastrophic).

Inward-Biased Perturbations: Breaks symmetric cancellation, revealing linear drift term as predicted.

Table 2: Constraint Asymmetry Results

σ	Mean Drift	Drift / σ	Character
0.0	0.0000	-	Flat
0.1	-0.0015	0.015	Negative
0.2	-0.0031	0.016	Negative
0.3	-0.0047	0.016	Negative

These experiments prove: (1) conservation is not cherry-picking, (2) theory correctly predicts its own boundaries, (3) falsifiability validated through deliberate violation.

3.6 Singularity Validation

Domain: 2D circular boundary with discrete corners

Results:

- Smooth boundary: No collapse
- Near singularity: Yes, angular restriction mechanism
- Collapse ratio: 2D \sim 15%, 3D \sim 5%
- Dimension dependence: Collapse ratio $\propto d-1$

Strongest empirical support for claim: reachability loss absent preference occurs only via geometry.

4. Discussion

4.1 Relationship to Existing Theory

Recent work in topological and geometric mathematics emphasizes structure-first formalisms over scalar optimization (Anderson and Munkholm, 2006), providing a broader context for our constraint-coherence approach.

Viability Theory (Aubin): Both study constraint-confined dynamics. Distinction: viability permits optimization within viable sets; NAGI forbids it entirely (Aubin and Frankowska, 1991).

Geometric Constraint Modeling: Unlike geometric constraint models (Durand and Hoffmann, 2000) that still resolve violations via minimization or penalty methods, NAGI forbids preference resolution entirely through Boolean constraints only.

Hamiltonian Mechanics: Both follow geometric structure. Distinction: Hamiltonian flow minimizes action; boundary flow follows tangent geometry without action principle (Arnold, 1989; Marsden, Ratiu and Golubitsky, 1999). Hamiltonians are energy-based; boundary flows are topology-based.

Contact Geometry: Both study odd Distinction: contact geometry has canonical energy structures; NAGI explicitly forbids them.

Edge Mechanics: Behavior at boundaries and edges has established precedent (Gregory and Wan, 1985; Altenbach et al., 2020), normalizing boundary-following dynamics as legitimate objects of study rather than pathological cases. Our boundary-tangent confinement extends this tradition.

Constructal Theory: Independent evidence that symmetry-preserving flows minimize structural resistance (Bejan and Lorente, 2006). This helps justify why asymmetric constraint bias introduces drift while symmetric handling preserves conservation.

Coherence without Teleology: Our notion of coherence without preference ordering aligns with philosophical accounts distinguishing structural consistency from goal-directed agency (Bostrom and Yudkowsky, 2014), supporting our conservative definition of non-agency.

Novel Contribution: NT-systems and CIVFs represent the intersection of deterministic evolution, constraint confinement, geometric structure following, and absence of scalar objectives previously un studied as a formal class.

4.2 Safety Architecture

Definition 8 (Agency Surface). An agency surface is a boundary in state space across which the set of admissible trajectories undergoes a qualitative change, thereby mediating a system's capacity for autonomous action (Hoffmeyer, 1998; Hammam et al., 2007; Norell, 2021). Agency can be further characterized by a system's sensitivity to perturbations near constraint boundaries, which induces anisotropic responses in trajectory evolution, formalizable through local geometric properties of the state space such as curvature or higher-order derivatives of the governing dynamics.

NAGI's Violin Memory architecture does not close the loop: No loop \implies no autonomous weapon.

Critical Limitation: This theorem holds for NAGI as architecturally specified. Embedding NAGI inside an agentic wrapper trivially voids the guarantee. NAGI's safety is architectural (prevents internal emergent agency, prevents autonomous recursive self-improvement) not cryptographic (does NOT prevent adversarial wrapper embedding or de liberate external misuse).

4.3 Engineering Feasibility

Perturbative stability experiments demonstrate that approximate implementations retain qualitative non-collapsing behavior. Even under:

- Small perturbations ($\epsilon \leq 2.0$)
- Discrete projection (grid-based)
- Extended time horizons (104 steps)

Conservation holds approximately, showing graceful degradation rather than catastrophic failure. This supports practical engineering deployment without requiring exact constraint satisfaction.

4.4 Limitations and Future Work

Explicit Scope: Deterministic systems only. Out of scope: stochastic dynamics, quantum systems, continuous-time fields, time-dependent constraints, infinite-dimensional spaces, fractal boundaries, perturbations beyond $\epsilon = 2.0$.

Open Questions:

1. Does critical ϵ^* exist for $\epsilon > 2.0$?
2. Can reachability entropy extend to stochastic systems?
3. What is measure-theoretic treatment of fractal boundaries?
4. Can NT-systems be formulated categorically?
5. Information-theoretic bounds on Violin Memory?

5. Conclusion

We have developed a formal mathematical framework for general intelligence without agency, including:

1. Formal definition of generality (coverage-based: $G(s) \geq \gamma$)
2. Conserved quantity (reachability entropy H_R)
3. Conservation law (Boundary Flow Theorem with rigorous proof)
4. Impossibility result (agentic AGI structural in compatibility)
5. Complete collapse taxonomy and singularity classification
6. Extensive empirical validation with negative results
7. New dynamical classes (NT-systems, CIVFs)
8. Safety guarantees via architectural constraints

The central insight: Intelligence does not require authority. Generality does not require desire. Creativity does not require optimization.

This framework demonstrates that power without agency is not only possible but mathematically well founded, with implications for safe development of advanced AI systems that maintain human judgment and semantic authority.

Acknowledgments

This work builds on classical results in dynamical systems, differential geometry, and complexity theory. The empirical validation program was essential to establishing the robustness of the theoretical framework.

Reference

Abraham, R. and Marsden, J.E. (1978) *Foundations of Mechanics*. Reading, MA: Benjamin/Cummings. https://api.pageplace.de/preview/DT0400.9780429689048_A37986404/preview-9780429689048_A37986404.pdf

Altenbach, H., Chinchaladze, N., Kienzler, R. and Müller, W.H. (2020) *Analysis of Shells, Plates and Beams*. Cham:

Springer. <https://link.springer.com/book/10.1007/978-3-030-47491-1>

Anderson, D.R. and Munkholm, H.J. (2006) *Boundedly Controlled Topology: Foundations of Algebraic Topology and Simple Homotopy Theory*. New York: Springer. <https://books.google.com/books?id=vk18CwAAQBAJ>

Arnold, V.I. (1989) *Mathematical Methods of Classical Mechanics*. 2nd edn. New York: Springer. <https://link.springer.com/book/10.1007/978-1-4757-2063-1>

Aubin, J.P. and Frankowska, H. (1991) 'Viability kernel of control systems', in *Nonlinear Synthesis: Proceedings of a IIASA Workshop held in Sopron, Hungary, June 1989*. Boston, MA: Birkhäuser, pp. 12–33. https://link.springer.com/chapter/10.1007/978-1-4757-2135-5_2

Bejan, A. and Lorente, S. (2006) 'Constructal theory of generation of configuration in nature and engineering', *Journal of Applied Physics*, 100(4), p. 041301. <https://pubs.aip.org/aip/jap/article-abstract/100/4/041301/371167/>

Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press. <https://psycnet.apa.org/record/2014-48585-000>

Bostrom, N. and Yudkowsky, E. (2014) 'The ethics of artificial intelligence', in Frankish, K. and Ramsey, W.M. (eds) *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, pp. 316–334. <https://www.taylorfrancis.com/chapters/edit/10.1201/9781351251389-4/ethics-artificial-intelligence-nick-bostrom-eliezer-yudkowsky>

Bruno, A.D. and Goryuchkina, I.V. (2004) 'Expansions of solutions of the sixth Painlevé equation', *Doklady Akademii Nauk*, 395(6), pp. 733–737. <https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=dan&paperid=1541>

Durand, C. and Hoffmann, C.M. (2000) 'A systematic framework for solving geometric constraints analytically', *Journal of Symbolic Computation*, 30(5), pp. 493–519. <https://www.sciencedirect.com/science/article/pii/S0747717100903924>

Gregory, R.D. and Wan, F.Y.M. (1985) 'Edge effects in the stretching of plates', in *Studies in Applied Mechanics*, vol. 12. Amsterdam: Elsevier, pp. 35–54. <https://www.sciencedirect.com/science/article/pii/B9780444425201500062>

Hammam, Y., Moore, A. and Whigham, P. (2007) 'The dynamic geometry of geographical vector agents', *Computers, Environment and Urban Systems*, 31(5), pp. 502–519. <https://www.sciencedirect.com/science/article/pii/S0198971507000543>

Hirsch, M.W. (2012) *Differential Topology*. New York: Springer Science and Business Media. <https://books.google.com/books?id=emTmBwAAQBAJ>

Hoffmeyer, J. (1998) 'Surfaces inside surfaces: On the origin of agency and life', *Cybernetics and Human Knowing*, 5(1), pp. 33–42. <https://www.ingentaconnect.com/content/imp/chk/1998/00000005/00000001/3>

Lee, J.M. (2013) *Introduction to Smooth Manifolds*. 2nd edn. New York: Springer. https://link.springer.com/chapter/10.1007/978-1-4419-9982-5_6

Marsden, J.E., Ratiu, T.S. and Golubitsky, M. (1999) *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. 2nd edn. New York: Springer. <https://link.springer.com/book/10.1007/978-0-387-21792-5>

Meiches, L. (2024) 'Agency and technological systems', in Bachman, J.S. (ed.) *Genocide Studies: Pathways Ahead*. New Brunswick, NJ: Rutgers University Press. <https://www.degruyterbrill.com/document/>

- doi/10.36019/9781978832367/html
- Norell, D. (2021) 'Geometries with agency: Mathematics of form revisited', *arq: Architectural Research Quarterly*, 25(3), pp. 255–265. <https://www.cambridge.org/core/journals/arq-architectural-research-quarterly/article/abs/geometries-with-agency-mathematics-of-form-revisited/501A19A7A21E26846DE56E8EE8F0989C>
- Ott, E. (2002) *Chaos in Dynamical Systems*. 2nd edn. Cambridge: Cambridge University Press. <https://books.google.com/books?id=PfXoAwAAQBAJ>
- Russell, S. (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*. London: Penguin. https://maxkasy.github.io/home/files/other/ML_Econ_Oxford/human_compatible.pdf
- Strogatz, S.H. (2024) *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. 3rd edn. Boca Raton: Chapman and Hall/CRC. <https://www.taylorfrancis.com/books/mono/10.1201/9780429398490/>
- Van Dyke, M. (1964) *Perturbation Methods in Fluid Mechanics*. New York: Academic Press. <https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/abs/perturbation-methods-in-fluid-mechanics-by-milton-van-dyke-academic-press-1964-229-pp-22-16s/E8442803CDA553F2213F5D18823D99E6>
- Wolpert, D.H. and Macready, W.G. (1997) 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation*, 1(1), pp. 67–82. <https://ieeexplore.ieee.org/abstract/document/585893>
- Zhu, L., Quétu, V., Nguyen, V.T. and Tartaglione, E. (2026) 'Layer collapse can be induced by unstructured pruning', *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=rfDYZNZIZT>

