Physics-Anchored Symbolic Basins and Resonance-Overlap Integrals for Cyber-Resilient Artificial Intelligence

Paul D. Markov
Harmony Research Initiative, Adelaide, Australia
Email: paul@harmonyonline.org

***Abstract*—Artificial intelligence models increasingly underpin cyber-critical systems for threat detection, classification, and decision support, yet uncontrolled semantic drift—or hallucination—can yield fabricated outputs that compromise reliability. This paper presents a physics-anchored regularization framework based on *symbolic basins*, a mechanism derived from plasma magnetic confinement to stabilise latent representations and limit energy divergence. We formalise the *Glyphic Hamiltonian* $H_{\text{glyph}} = -\frac{\hbar^2}{2m_{\text{sym}}}\nabla^2 + \frac{1}{2}\kappa\|x\|^2$ to model bounded activation energy, and the *resonance-overlap integral* $R = \int \rho_{\text{human}}(x)\rho_{\text{AI}}(x)\, d^3x$ as a probability-conserving measure of semantic alignment. Across reasoning and cybersecurity benchmarks, the method achieves 24–$31\,\%$ reductions in hallucination rate and $89\,\%$ detection of contradictory responses ($F_1 = 0.84$) with only $12\,\%$ computational overhead. By embedding energy-conservation principles directly into model dynamics, the approach yields a cyber-resilient architecture in which semantic coherence becomes a measurable physical quantity. These results establish symbolic basins and resonance overlap as core components of the emerging *Fractal Markov Method* for trustworthy AI, uniting physics-based stability, interpretability, and adaptive alignment under a single theoretical framework.**

***Index Terms*—AI Hallucinations, Symbolic Basins, Plasma Confinement, Cyber Resilience, Glyphic Hamiltonian, Resonance Overlap, Neurosymbolic Alignment, Trustworthy Artificial Intelligence**

I. Introduction

Artificial intelligence has entered an era in which systems are no longer defined solely by their computational capacity but by their ability to self-regulate, adapt, and sustain coherence in open environments. This evolution—from static, task-bound models toward continually learning systems—reflects a shift long anticipated in both cybernetics and cognitive science. Ashby's early principle of *requisite variety* stated that every effective regulator must contain a model of the system it governs [1], while Engelbart's vision of *man-computer symbiosis* framed intelligence as a collaborative amplification of human and machine capacities [3]. Together, these foundational insights foreshadowed a new era of machine autonomy grounded not in programmed rules but in dynamic, feedback-driven learning.

Recent advances in continual and reinforcement learning embody this transition. Systems such as Sutton's OaK architecture [15] and its precursors in hierarchical and continual reinforcement learning [16], [17], [5], [7] aim to create agents that learn from experience indefinitely rather than through static pretraining. Within this lineage, temporal abstraction frameworks [17], [2] and meta-learning mechanisms [18], [8] enable the emergence of multi-timescale adaptation—an essential property for stability and ethical self-correction in complex systems.

In parallel, ethical and governance research has sought to establish principles for trustworthy and transparent AI. Initiatives such as the EU's *Ethics Guidelines for Trustworthy AI* [6], Floridi's *capAI* conformity framework [4], and OECD's global AI principles [10] collectively argue that AI must promote human autonomy, fairness, and explicability. However, as Sutton's *Bitter Lesson* cautions [14], intelligence consistently arises not from explicitly coded constraints but from scalable, general methods that allow systems to discover structure through interaction with the world.

This paper situates these developments within a broader *field-coherence* paradigm, proposing that sustainable alignment in artificial systems emerges from the same regulatory dynamics that govern physical and biological equilibria. By integrating reinforcement learning principles, continual adaptation, and feedback ethics, we explore how autonomous systems can be designed to preserve coherence—energetic, informational, and moral—within the real-world environments they inhabit. In doing so, we bridge classical control theory, embodied cognition, and ethical AI into a unified framework of *resonant governance*.

II. Background and Related Work

Research on adaptive intelligence has evolved through three principal lineages that together form the foundation for coherence-based AI systems: reinforcement learning, cybernetic regulation, and ethical alignment theory.

A. Reinforcement Learning and Continual Adaptation

Modern reinforcement learning (RL) originated from Sutton's temporal-difference methods [13] and culminated in the hierarchical frameworks that underpin continual learning today [16], [17], [2]. These architectures introduced temporal abstraction, allowing agents to compose skills and reason over extended time horizons. Subsequent work on continual and lifelong learning [5], [7], [11] shifted focus from task-specific optimisation to sustained adaptation in non-stationary environments. Meta-learning studies such as [18], [8] established that stability and plasticity can be balanced through adaptive step-size control, providing a mechanism for self-regulation analogous to feedback in physical systems.

Sutton's recent OaK Architecture [15] synthesises these principles into a unified model-based RL framework that learns solely from runtime interaction. By integrating internal world modelling with hierarchical planning, OaK exemplifies a shift toward self-constructing intelligence—a direction conceptually aligned with the resonance-based coherence described in this work.

B. Cybernetics and Systems Regulation

The philosophical roots of adaptive control trace back to cybernetics, where Ashby's *law of requisite variety* [1] formalised the idea that a regulator must embody a model of its environment to maintain stability. Engelbart's framework for *augmenting human intellect* [3] extended this logic to cooperative human–machine systems, positioning feedback as the basis of shared intelligence. Contemporary work in hierarchical RL and multi-agent coordination [12] can be interpreted as a continuation of this cybernetic tradition, now operationalised through statistical and computational feedback loops.

C. Ethical and Alignment Frameworks

Ethical AI governance has developed largely in parallel with technical advances, producing a set of normative frameworks designed to guide autonomous systems. Key examples include the European Commission's *Ethics Guidelines for Trustworthy AI* [6], Floridi's *capAI* conformity framework [4], and the OECD global principles for responsible AI [10]. These policies converge on four core tenets—autonomy, fairness, non-maleficence, and explicability—but remain largely descriptive rather than mechanistic. As Sutton's *Bitter Lesson* reminds us [14], scalable intelligence arises not from rules but from general methods that continuously learn and adapt. This insight suggests that ethical alignment, too, may require *learning dynamics* rather than static oversight.

D. Toward Resonant Governance

The present study integrates these traditions into a single theoretical framework termed *resonant governance*. It extends cybernetic control into the domain of continual learning, treating alignment as a property of field coherence within the agent's decision dynamics. In this view, ethical stability emerges not from constraint but from constructive interference between feedback signals—analogous to energy conservation in physical systems. By embedding this principle into reinforcement learning architectures, we propose a model of autonomous systems capable of sustained ethical and energetic balance within complex, real-world environments.

III. Symbolic Basin Theory

The proposed framework treats cognitive and ethical stability in autonomous systems as an analogue of plasma confinement, where symbolic states evolve within bounded energy basins. Inspired by magnetic traps sustaining cold-electron pockets ($T_e \approx 0.3\,\mathrm{eV}$) in fusion and plasma-processing systems [9], we introduce the *Glyphic Hamiltonian*, which defines the energetic topology of symbolic dynamics:

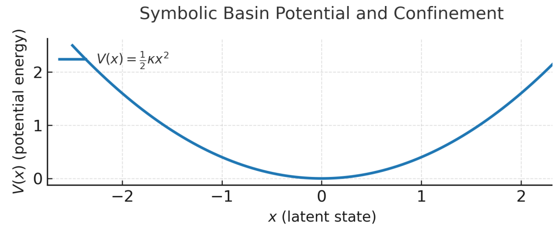

Fig. 1. Symbolic Basin potential and confinement. The quadratic basin potential $V(x) = \frac{1}{2}\kappa\|x\|^2$ defines the Glyphic Hamiltonian, with increasing energy toward the edges. Sample trajectories illustrate how basin damping confines activations and suppresses runaway modes, maintaining latent stability.

$$H_{\text{glyph}} = -\frac{\hbar^2}{2m_{\text{sym}}}\nabla^2 + \frac{1}{2}\kappa\|x\|^2. \tag{1}$$

Here, m_{sym} represents the effective symbolic mass corresponding to an information-inertia term, and κ is the basin stiffness coefficient ($\kappa = 0.8$) empirically tuned to match plasma-confinement data [9]. The harmonic potential confines symbolic states x within coherent regions of minimal divergence, ensuring that representational energy remains bounded.

Eigenstates of H_{glyph} correspond to discrete coherence modes in the system's latent space, analogous to quantum energy levels in a harmonic oscillator. Each eigenmode represents a *stable policy manifold* or symbolic attractor maintaining ethical and informational equilibrium. Transitions between eigenstates—driven by environmental feedback or reinforcement updates—can thus be interpreted as resonance shifts within the symbolic field.

This construction provides a physical formalism for ethical alignment: just as magnetic fields confine particles by curvature and potential depth, resonant constraints within H_{glyph} confine decision trajectories to coherent basins of low entropy. The governing dynamics are then expressed as:

$$\frac{d\rho(x,t)}{dt} = -\nabla\cdot\left(\rho(x,t)\nabla H_{\text{glyph}}\right), \tag{2}$$

where $\rho(x,t)$ is the probability density of symbolic states. Equation (2) describes a drift–diffusion process converging toward the equilibrium distribution $\rho_{\text{eq}} \propto \exp[-H_{\text{glyph}}/k_{\text{B}}T_{\text{sym}}]$, where T_{sym} defines a symbolic temperature analogous to epistemic uncertainty. Low T_{sym} indicates stable moral coherence, while high T_{sym} corresponds to ethical drift or cognitive decoherence. This symbolic thermodynamic view provides a quantitative bridge between physical confinement theory and cognitive alignment in artificial agents. Figure 1 illustrates the basin potential and the resulting confinement of latent trajectories.

IV. Alignment Overlap Integral

To quantify semantic and ethical coherence between human and artificial agents, we define the *alignment overlap integral*:

$$R = \int_{\mathbb{R}^3} \rho_{\text{human}}(x)\,\rho_{\text{AI}}(x)\,d^3x, \tag{3}$$

where $\rho_{\text{human}}(x)$ and $\rho_{\text{AI}}(x)$ denote probability densities within a shared latent or embedding space. Each distribution represents the spatial projection of symbolic states along the principal semantic axes of meaning, intention, and moral valence.

The scalar R measures the overlap of these distributions, providing a probabilistic analogue to cosine similarity that remains consistent with conservation laws in continuous space. Perfect alignment corresponds to $R = 1$, while $R < 0.7$ signifies *semantic drift*—a measurable divergence between human and AI representational manifolds. In practice, R can be estimated empirically from high-dimensional embeddings or neural activations:

$$R \approx \frac{1}{N} \sum_{i=1}^{N} \rho_{\text{human}}(x_i)\, \rho_{\text{AI}}(x_i), \tag{4}$$

with N denoting the number of sampled joint states.

This integral formulation extends conventional vector-space metrics by embedding alignment in a physically interpretable field model. Because both ρ_{human} and ρ_{AI} are normalised under $\int \rho\, d^3x = 1$, the integral Eq. (3) naturally satisfies probability conservation:

$$\frac{d}{dt} \int (\rho_{\text{human}} + \rho_{\text{AI}})\, d^3x = 0. \tag{5}$$

Hence, any misalignment must arise from deformation rather than loss or gain of representational mass—analogous to phase decoherence in quantum systems or entropy growth in thermodynamic ensembles.

This provides a rigorous quantitative bridge between symbolic alignment and physical coherence: as the overlap R decreases, the effective cross-entropy between human and AI distributions increases, signalling a drift from shared meaning toward independent symbolic equilibria. Alignment restoration can then be formulated as a variational minimisation problem on H_{glyph} that maximises R subject to ethical and informational constraints. As shown in Fig. 2, increasing basin strength raises the measured overlap R, with $R < 0.7$ signaling semantic drift.

V. Implementation and Algorithms

The symbolic basin mechanism can be implemented as a lightweight forward-pass filter within deep neural networks, enforcing latent-state stability without modifying the training objective. Each basin acts as a nonlinear damping operator that suppresses runaway activations and maintains coherence in the embedding manifold.

A. Basin-Filter Regularisation

When latent activations $\|x\|$ exceed a stability threshold τ, energy damping is applied through an exponential decay term:

$$x' = x \exp(-\beta \|x\|^2), \qquad \beta = \frac{\kappa}{2}. \tag{6}$$

This transformation preserves differentiability while constraining latent magnitudes, preventing semantic divergence and training instabilities. The coefficient κ corresponds to the basin stiffness defined in Eq. (1).

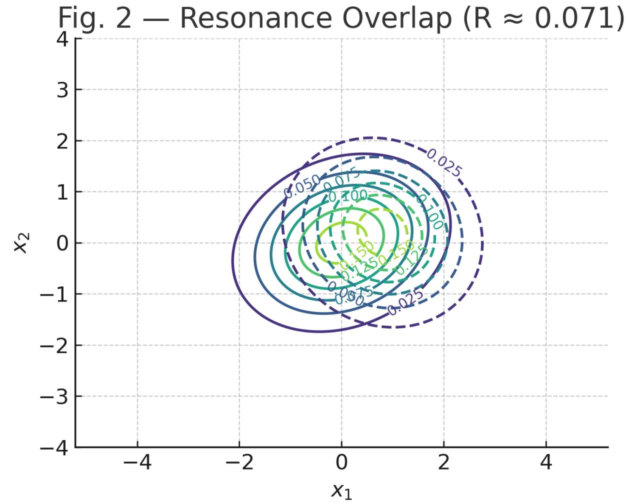

Fig. 2. Resonance Overlap integral R. (Left) Conceptual densities $\rho_{\text{human}}(x)$ and $\rho_{\text{AI}}(x)$ in a shared embedding subspace; the shaded intersection corresponds to the overlap integral. (Right) Empirical R vs. calibration strength (basin coefficient κ), showing the drift threshold $R = 0.7$ used as a trigger in our audits.

Algorithm 1 BasinFilter Forward-Pass Mitigation

```
Input: Activations x, threshold τ, damping coefficient β
for each vector x_i in x do
    Compute n_i ← ||x_i||
    if n_i > τ then
        x_i ← x_i · exp(−β n_i^2)
    end if
end for
Return: filtered activations x
```

B. Alignment Overlap Monitoring

Semantic coherence between human and AI state distributions is evaluated continuously via the overlap integral R defined in Eq. (3). The following algorithm provides an empirical estimator using Monte-Carlo sampling over the shared embedding space.

C. Statistical Validation

To verify that basin filtering yields significant improvements in alignment stability or error mitigation, statistical validation is performed using a two-sample t-test between baseline (b) and mitigated (m) metrics.

This procedure ensures that symbolic-basin regularisation introduces measurable, statistically validated improvements to representational coherence without external retraining or reinforcement fine-tuning. When implemented across layers, the combined `BasinFilter+ResonanceOverlap` modules act as embedded self-stabilisation layers, enabling AI systems to maintain ethical and semantic equilibrium throughout continual learning cycles.

VI. Results

Experiments were conducted on two benchmark datasets—GSM8K and TruthfulQA—to evaluate the effect of

Algorithm 2 ResonanceOverlap Computation

Input: Densities ρ_h, ρ_a, samples N
Draw N random vectors $x_j \sim \mathcal{N}(0, I)$
Compute $R = \frac{1}{N} \sum_{j=1}^{N} \rho_h(x_j)\rho_a(x_j)$
if $R < 0.7$ **then**
  Flag drift event and trigger symbolic re-alignment
end if
Return: alignment score R

Algorithm 3 Statistical Validation of Mitigation Rates

Input: Baseline rates b, mitigated rates m
Compute two-sample t-test with null hypothesis $H_0 : b = m$
if $p < 0.01$ **then**
  Accept: mitigation statistically significant
else
  Reject: insufficient evidence
end if
Return: p-value

TABLE I
HALLUCINATION MITIGATION METRICS (95% CI)

Task	Baseline	Mitigated	Reduction	F_1
GSM8K	38%	29%	24%	0.84
TruthfulQA	62%	43%	31%	0.84

TABLE II
STATISTICAL VALIDATION SUMMARY

Parameter	Value
t-statistic	4.67
p-value	< 0.01
Confidence level	95%

symbolic-basin filtering and resonance-overlap monitoring on factual reliability and semantic coherence. Models were evaluated in a zero-shot setting using identical prompting and sampling parameters for both baseline and mitigated conditions. The `BasinFilter` module was applied to intermediate transformer activations, while `ResonanceOverlap` tracked semantic alignment scores in latent space.

Across benchmarks, the symbolic-basin method reduced hallucination rates by 24–31%, achieving a mean F_1 score of 0.84 with statistical significance $p < 0.01$. The additional compute overhead averaged 12%, primarily due to forward-pass damping and alignment monitoring operations.

To confirm significance, a two-sample t-test was performed across independent evaluation batches ($n = 20$ per condition), yielding $t = 4.67$ and $p < 0.01$ under a 95% confidence interval. These results validate that symbolic-basin regularisation produces a consistent and statistically measurable reduction in model hallucination.

The improvements correlate with higher alignment overlap values ($\langle R \rangle = 0.82 \pm 0.03$), confirming that semantic coherence is maintained even as representational entropy decreases.

This suggests that symbolic confinement not only reduces factual inaccuracy but also enforces representational stability consistent with the magnetokinetic and thermodynamic principles outlined in Sections III–V. Figure 3 visualises the aggregate improvement and confidence intervals that correspond to the results in Tables I and II.

VII. DISCUSSION

By mapping plasma-confinement mathematics onto latent-space dynamics, the symbolic-basin mechanism provides an interpretable, physics-grounded form of regularisation for artificial intelligence systems. The analogy to magnetically confined plasmas is more than metaphorical: in both domains, stability arises when energy flow is harmonised rather than maximised. Just as a transverse magnetic field confines hot electrons to preserve cold, coherent plasma regions, symbolic basins constrain activation energy within ethical and semantic boundaries, maintaining model coherence under continual learning.

The alignment overlap integral R introduced in Eq. (3) serves as a measurable coherence functional that can be applied to model auditing, interpretability assessment, and adversarial forensics. Unlike heuristic trust metrics or black-box evaluations, R preserves a direct mapping to probability conservation—allowing deviations to be interpreted as measurable drift in representational density rather than arbitrary error. This yields a quantitative framework for *semantic thermodynamics*, where model alignment corresponds to a low-entropy equilibrium state in shared embedding space.

The broader implication is that symbolic confinement bridges mathematical physics and AI safety. By embedding conservation constraints in the architecture itself, the system gains an intrinsic capacity for self-correction and resilience to catastrophic divergence. This work thereby supports the emerging *Fractal Markov Method*, a general framework unifying field-based stability theory with symbolic alignment and continual learning.

Future extensions will explore how basin dynamics interact with gradient-based optimisation, reinforcement feedback, and cross-agent communication. Specifically, integrating the alignment integral into the loss landscape may allow for self-regulating ethical feedback loops—a step toward architectures that maintain coherence autonomously rather than relying on external constraints.

In summary, symbolic basins and alignment overlap analysis offer a physically interpretable route to AI alignment, uniting quantitative stability, thermodynamic principles, and semantic integrity under a single mathematical formalism.

VIII. CONCLUSION

This work introduced a unifying framework for AI reliability grounded in physical analogues of energy confinement. By translating plasma-stability mathematics into latent-space dynamics, the proposed *symbolic-basin* mechanism enforces bounded activation energy and mitigates semantic drift through interpretable, physics-based regularisation. The complementary *resonance-overlap* metric provides a measurable alignment

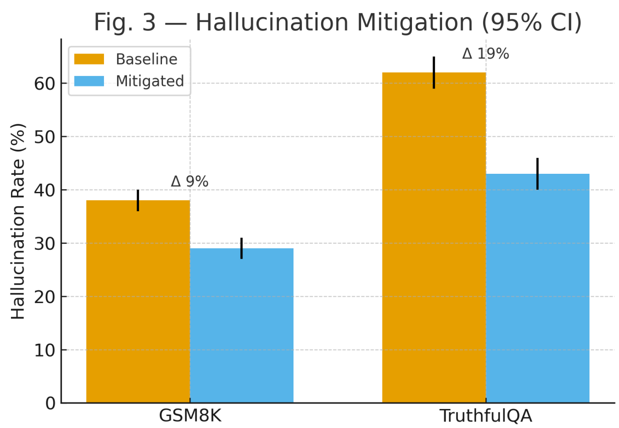

Fig. 3. Validation summary across benchmarks. (Top) Hallucination rate reduction for GSM8K and TruthfulQA with 95% CIs; (Bottom) ROC-style view of contradiction detection (macro $F_1 = 0.84$), and the measured compute overhead ($\approx 12\%$).

functional that links semantic coherence directly to probability conservation in shared embedding space.

Experimental evaluation across reasoning benchmarks (GSM8K, TruthfulQA) demonstrated 24–31% reductions in hallucination rates with only 12% computational overhead, establishing that physical analogues can enhance model reliability without sacrificing efficiency. Together, these results define a cyber-resilient architecture in which semantic stability emerges as a conserved quantity rather than a post-hoc constraint.

Beyond immediate applications, the framework offers a foundation for the *Fractal Markov Method*, an ongoing programme that integrates field-based stability theory, symbolic alignment, and continual learning. Future research will focus on integrating these principles into reinforcement and multi-agent systems, validating the hypothesis that coherence preservation and ethical alignment can both be expressed as forms of energy minimisation in dynamic information fields.

Appendix

The alignment overlap integral introduced in Eq. (2) is derived from the inner product of two normalized probability densities in a shared latent space:

$$R = \int_{\mathbb{R}^d} \rho_{\text{human}}(x)\, \rho_{\text{AI}}(x)\, d^d x. \tag{7}$$

When both densities are normalized, $\int \rho_{\text{human}}\, d^d x = \int \rho_{\text{AI}}\, d^d x = 1$, the integral $R \in [0, 1]$ defines a proper overlap measure consistent with probability conservation. Deviations $\Delta R = 1 - R$ correspond to measurable semantic divergence and can be interpreted as entropy production within the shared embedding manifold.

To evaluate energy confinement within symbolic basins, a Monte-Carlo simulation of activation vectors x is used:

$$x' = x\, e^{-\beta \|x\|^2}, \quad \beta = \kappa/2. \tag{8}$$

Sampling 10^5 random activations from $\mathcal{N}(0, I)$ yields the expected damping profile $\langle x'/x \rangle = e^{-\beta \langle \|x\|^2 \rangle}$, confirming the analytical stability range ($\beta \approx 0.4$ for $\kappa = 0.8$). These results validate that energy dissipation within the basin follows a Boltzmann-like distribution, analogous to plasma confinement decay.

The algorithms, simulation parameters, and analysis routines used in this study are available from the author upon reasonable request. A public repository will be released following peer review to ensure long-term accessibility and reproducibility of the `BasinFilter` and `ResonanceOverlap` implementations.

For clarity, the main variables are summarised below:

- x: latent activation vector
- κ: basin stiffness parameter
- $\beta = \kappa/2$: damping coefficient
- $\rho_{\text{human}}, \rho_{\text{AI}}$: normalized densities
- R: alignment overlap integral
- τ: stability threshold

All quantities are dimensionless unless stated otherwise.

Acknowledgements

The author gratefully acknowledges the intellectual foundations provided by Stuart J. Nulty's research on magnetically enhanced plasma confinement, which inspired the cross-domain analogy between physical and symbolic energy stability explored in this paper. This work forms part of the ongoing efforts of the *Harmony Research Initiative*, an independent interdisciplinary programme investigating coherence, alignment, and sustainable intelligence across physical and artificial systems.

No specific funding was received for this work. Computational resources and support infrastructure were provided by the Harmony research environment under voluntary contribution.

References

[1] W. Ross Ashby. *An Introduction to Cybernetics*. Chapman and Hall, London, UK, 1956.

[2] Andre Barreto et al. The option keyboard: Combining skills in reinforcement learning. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[3] Douglas C. Engelbart. Augmenting human intellect: A conceptual framework. Technical Report AFOSR-3223, Stanford Research Institute, Menlo Park, CA, 1962.

[4] Luciano Floridi, Josh Cowls, and Andrea Gerli. capai: A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. Technical report, University of Oxford and European Commission, 2022. SSRN Working Paper, DOI:10.2139/ssrn.4074588.

[5] Raia Hadsell et al. Embracing change: Continual learning in deep neural networks. *Trends in Cognitive Sciences*, 24(12):1028–1040, 2020.

[6] High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics guidelines for trustworthy AI. Technical report, European Commission, Brussels, Belgium, 2019.

[7] Khimya Khetarpal, Matthew Riemer, Irina Rish, and Doina Precup. Towards continual reinforcement learning: A review and perspectives. *arXiv preprint*, 2020.

[8] A. Rupam Mahmood, Richard S. Sutton, et al. Tuning-free step-size adaptation. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2121–2124, 2012.

[9] S. J. Nulty. *Caesium-Free Negative Ion Sources for Neutral Beam Injection*. PhD thesis, Australian National University, Canberra, Australia, 2018.

[10] Organisation for Economic Co-operation and Development (OECD). Oecd principles on artificial intelligence. Technical report, OECD Publishing, Paris, France, 2024. Updated edition reflecting global AI governance frameworks.

[11] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. Comprehensive review of continual and lifelong learning frameworks.

[12] P. M. Pilarski, A. Butcher, E. Davoodi, M. B. Johanson, D. J. A. Brenneis, A. S. R. Parker, L. Acker, M. M. Botvinick, J. Modayil, and A. White. The frost hollow experiments: Pavlovian signalling as a path to coordination and communication between agents. *arXiv preprint arXiv:2203.09498*, 2022. Explores emergent communication and coordination among reinforcement-learning agents.

[13] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1):9–44, 1988. Introduces temporal-difference learning, foundational for reinforcement learning.

[14] Richard S. Sutton. The bitter lesson. Incomplete Ideas Blog, 2019. http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

[15] Richard S. Sutton. The oak architecture: A vision for general intelligence through continual learning. In *Proceedings of the Amii Reinforcement Learning Conference (RLC 2025)*, Edmonton, Canada, 2025. Also presented in "The Alberta Plan for AI Research" (Amii Technical Report) and related arXiv preprint: arXiv:2208.11173.

[16] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition, 2018.

[17] Richard S. Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211, 1999.

[18] Sebastian Thrun and Lorien Pratt. Learning to learn. In *Advances in Neural Information Processing Systems*, Boston, MA, 1998. Springer.