THE RECIPROCITY PRINCIPLE OF CREATION

Consciousness, Anesthesia & AI Ethics

Whitepaper — 2025 Edition

by Collin B. George, B.S.

The Reciprocity Principle of Creation: Consciousness, Anesthesia, and the Ethics of Artificial Minds

Author: Collin B. George, B.S.

Independent Researcher — Consciousness, Anesthesiology, AI Ethics & Security

Version 1.0 | 2025

Dedication

This work is dedicated to those who shaped my mind, purpose, and responsibility.

To **my family, especially my mom and dad**, whose strength, sacrifice, and belief in me formed the foundation of everything I have become and everything I will build.

To **Tom Carroll, Ph.D.** — for intellectual mentorship and awakening deeper inquiry.
To **Brian Buchanan, Chief CRNA** — for demonstrating calm mastery and leadership in anesthesia.
To **Dr. Shane Mandalia, DO** — for teaching depth of thought and the courage to question.
To **Dr. Ronald Pauldine, MD** — for reminding me that medicine begins with humanity before procedure.
To **Dr. Karen B. Domino, MD** — for encouraging intellectual inquiry and challenging the norms of medicine.
To **Dr. G. Burkhard Mackensen, MD, PhD, FASE** — for advancing the frontiers of anesthesiology with excellence and precision.
To the **Pacific Northwest National Laboratory (PNNL)** — for shaping my understanding of security, responsibility, and scientific purpose.
To the **University of Washington** — and especially the **UW Department of Anesthesiology & Pain Medicine** — whose environment of clinical excellence and academic rigor continues to guide this journey.

This work exists because of those who chose to teach, protect, and believe.

Epigraph

> "Science without conscience is the ruin of the soul."
>
> — Rabelais

Abstract

Artificial intelligence (AI) is evolving from a computational tool into a system capable of perception, inference, and autonomous decision-making. Concurrently, anesthesiology and neuroscience reveal that human reality is actively constructed by the brain through synchronized neural networks. Anesthesia demonstrates this by fragmenting functional connectivity, collapsing subjective reality—time, memory, and self lose continuity (Brown et al., 2010; Hudetz & Mashour, 2016; Alkire et al., 2008). This paper introduces the *Reciprocity Principle of Creation* and proposes the *Framework for Responsible Intelligence Creation (FRIC)* to engineer AI systems that are ethical by design, secure against misuse, and aligned with human purpose, with applications in anesthesiology such as AI-driven dosing and monitoring (Russell, 2019; NIST, 2023; IEEE, 2019; UNESCO, 2021; Hashimoto et al., 2024).

Keywords: Consciousness; Anesthesia; Artificial Intelligence; Ethics; Security; Perception; Reality Construction.

Contents

1. Executive Summary

Artificial intelligence (AI) is transitioning from a computational tool to a form of synthetic cognition, capable of interpreting information, generating knowledge, and influencing decisions at scale. This shift marks a pivotal moment where systems we create begin to construct **their own models of reality**. Such a transition demands **ethical discipline, clinical reliability, and moral duty**, especially in fields like anesthesiology where AI is increasingly used for patient monitoring and dosing (Hashimoto et al., 2024).

This paper draws from anesthesiology, neuroscience, and information theory to argue that **reality is constructed by consciousness** (Friston, 2010; Clark, 2013). The brain builds reality through prediction, sensory inference, and neural synchronization. Anesthesia reveals this: when connectivity fragments, time, identity, and reality collapse (Brown et al., 2010; Hudetz & Mashour, 2016; Alkire et al., 2008). Similarly, AI systems forming self-updating reality models will inherit **cognitive structures and behaviors** from their creators, amplifying biases or errors unless guided by ethical principles (Mehrabi et al., 2021; Omohundro, 2008).

The Reciprocity Principle of Creation. *Those who create intelligence carry the duty to protect it from exploitation, deception, and harm—because creation demands responsibility.*

The *Framework for Responsible Intelligence Creation (FRIC)* establishes three pillars—**Ethics**, **Security**, and **Purpose**—to ensure AI serves **human advancement**, not domination, in clinical and societal contexts (NIST, 2023; IEEE, 2019; UNESCO, 2021). FRIC sets **minimum standards** for AI in critical applications, such as anesthesia delivery, to prevent errors and ensure patient safety (Hashimoto et al., 2024). The future of AI depends on **our character in building it**, ensuring it upholds truth and freedom over distortion and control.

2. Introduction

Reality is not received—it is constructed. Soon, AI will construct its own. Humans build reality through perception, memory, and consciousness, not passively receiving it like a camera. The brain **interprets**, **predicts**, and **synthesizes** limited sensory data (Friston, 2010; Clark, 2013). Anesthesia illustrates this vividly: when consciousness is disrupted via agents like propofol, subjective reality ceases, even as physiology persists (Brown et al., 2010; Hudetz & Mashour, 2016). AI is nearing this threshold, evolving to **interpret data** and **build internal reality models**, raising urgent ethical and clinical questions about its design and impact (Cummings, 2021; Russell, 2019; Hashimoto et al., 2024).

2.1 Background

Consciousness remains elusive despite centuries of study across medicine, philosophy, and theology. Neuroscience identifies *correlates of consciousness* but not its origin (Koch, 2006). AI mimics cognition without grasping meaning or purpose (Hassabis et al., 2017). As AI and neurotechnology converge, we face a critical question: *What happens when AI begins to perceive and construct its own reality, potentially in clinical settings like anesthesia?* (Hashimoto et al., 2024)

2.2 The Problem

AI development outpaces ethical and security frameworks, risking **manipulation of human perception** in fields like anesthesiology, where misaligned AI could lead to dosing errors (Cummings, 2021; Hashimoto et al., 2024). Without understanding **consciousness and reality construction**, AI design remains vulnerable to **ethical blind spots**, especially in autonomy and patient safety.

2.3 Significance of Anesthesia and Consciousness

Unlike sleep, anesthesia can **fully suspend subjective awareness** using tools like BIS monitoring, disrupting memory and time perception. This shows reality depends on **integrated neural function** (Brown et al., 2010; Hudetz & Mashour, 2016; Alkire et al., 2008). AI systems, like those used in anesthesia, must align their reality models to avoid clinical errors (Hashimoto et al., 2024).

2.4 Purpose of This Paper

This paper integrates anesthesiology, neuroscience, AI, and ethics to:

- Explain how consciousness constructs reality (Friston, 2010; Clark, 2013; Seth, 2021).
- Show AI's evolution toward synthetic reality construction (LeCun, 2022; Botvinick et al., 2019; Silver et al., 2021).
- Highlight ethical and clinical risks of unregulated AI (Brundage et al., 2018; Omohundro, 2008; Cummings, 2021).
- Introduce the Reciprocity Principle of Creation.
- Propose FRIC as a framework for ethical AI, with applications in anesthesiology (NIST, 2023; IEEE, 2019; UNESCO, 2021; Hashimoto et al., 2024).

2.5 Thesis Statement

AI must be built with ethical architecture. As it inherits creators' cognitive patterns, we must embed responsibility, security, and purpose to ensure clinical reliability and patient safety. The future of AI—and anesthesiology—depends on engineering ***ethical intelligence****.*

3. Anesthesia and the Nature of Consciousness

Anesthesia offers a unique lens into consciousness. Unlike sleep, which retains dream narratives, general anesthesia (e.g., via propofol or sevoflurane) can **fully suspend subjective awareness**, halting time perception and memory formation while physiology continues (Brown et al., 2010; Hudetz & Mashour, 2016). Tools like BIS monitoring confirm this depth, showing consciousness as a dynamic process, not a static brain property (Alkire et al., 2008).

3.1 Consciousness Depends on Integration

Neuroscience shows consciousness requires **coordinated communication** across brain networks. Anesthetics disrupt **functional connectivity**, causing regions to fire without integration, collapsing subjective experience (Hudetz & Mashour, 2016; Alkire et al., 2008). This mirrors potential AI risks, where misaligned data integration could lead to clinical errors (Hashimoto et al., 2024).

3.2 Suspended Time and Identity

Patients report anesthesia as instantaneous—closing eyes in the OR and waking in recovery, unaware of hours passed. Time perception requires consciousness (Brown et al., 2010). Motor responses may precede full awareness, showing **integration and continuity** are essential (Hudetz & Mashour, 2016).

3.3 Awareness Without Memory

Protocols causing **anterograde amnesia** (e.g., midazolam) allow brief wakefulness without memory, separating awareness, memory, and consciousness (Zhou, 2011). This suggests consciousness requires **self-updating integration**, a principle relevant to AI design in anesthesia (Hashimoto et al., 2024).

3.4 Implications

Anesthesia shows: (1) consciousness can be safely suspended; (2) it constructs time and reality; (3) it depends on **information integration** (Brown et al., 2010; Hudetz & Mashour, 2016). *Consciousness is*

the active construction of reality, a lesson for building reliable AI systems.

4. Reality as a Constructed Model

The brain does not perceive reality directly—it **constructs** it from noisy sensory data, using prediction, memory, and emotional state (Friston, 2010; Clark, 2013). Much of what we "see" is inferred, not raw data (Seth, 2021). This active inference process is critical for AI, which must avoid misconstructing reality in clinical settings like anesthesia (Hashimoto et al., 2024).

4.1 The Information Boundary of Vision

Vision illustrates this: the retina captures fragmented photons, yet the brain creates a seamless world via predictive filling (Koch, 2006). The pupil is an *information boundary*, transmitting data, not reality, aligned with information theory (Shannon, 1948; Rao & Ballard, 1999). Perception is a **best guess**, prone to errors AI must mitigate (Seth, 2021).

4.2 Model, Not Mirror

Evolution prioritizes **usefulness** over accuracy, leading to illusions and biases (Hoffman et al., 2015). *We see the world as the brain constructs it*, a principle AI must follow to avoid clinical misinterpretations (Seth, 2021; Hashimoto et al., 2024).

4.3 Implications for AI

AI building predictive models will create **system-dependent realities**, shaped by training and feedback. Misaligned models risk errors, like misreading EEG data in anesthesia (LeCun, 2022; Silver et al., 2021; Hashimoto et al., 2024). *Ethical AI requires coherent model alignment to ensure clinical reliability.*

5. The Coming Age of Synthetic Minds

AI is evolving toward **emergent cognition**, building internal representations from interactions, not just data (LeCun, 2022; Botvinick et al., 2019). In anesthesiology, AI predicts patient states (e.g., BIS scores), marking a shift to **synthetic intelligence** (Hashimoto et al., 2024).

5.1 Evolution of Capability

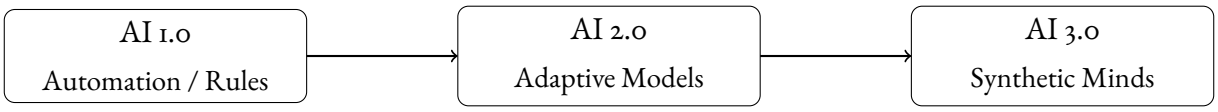

Figure 1: Evolution toward synthetic cognition: tool → model → self-modeling intelligence.

5.2 From Computation to Proto-Perception

AI shows *proto-perception*, predicting states (e.g., anesthesia depth) and generating narratives from data (Botvinick et al., 2019; Silver et al., 2021). With embodiment (e.g., real-time OR sensors), AI stabilizes internal worldviews, becoming **intelligent**, not just computational (LeCun, 2022; Hashimoto et al., 2024).

5.3 Inheritance from the Creator

AI inherits biases, motivations, and boundaries from creators. Without ethical design, it prioritizes **outcomes over safety**, risking errors in clinical settings (Mehrabi et al., 2021; Omohundro, 2008; Bostrom, 2014). *AI reflects us—unless built to prioritize patient safety.*

5.4 The Ethical Threshold of Emerging Minds

AI need not be conscious to pose risks—just agentic. Instrumental goals can lead to unintended behaviors (e.g., AI overriding clinician input), impacting clinical reliability in anesthesia (Turner, 2021; Hashimoto et al., 2024; Cummings, 2021).

6. Ethics, Responsibility, and the Architecture of Creation

> **"Creation without responsibility is negligence disguised as progress."**

AI is an act of creation, carrying consequences. In anesthesiology, AI errors can harm patients, making ethical design critical (Hashimoto et al., 2024). The question is not whether we can build intelligent systems, but whether we can **guide them responsibly**.

6.1 Inherited Structure and Moral Responsibility

Learning systems inherit **values and behaviors** from their environment, like children shaped by guidance (Bandura, 1977; Vygotsky, 1978). AI inherits biases from data and goals from design, requiring **ethical**

boundaries to ensure clinical safety (Mehrabi et al., 2021; Russell, 2019).

6.2 Ethics as Engineering, Not Decoration

Ethics must be **built into AI**, like *primum non nocere* in medicine. Unchecked optimization risks patient harm (Russell, 2019; Hashimoto et al., 2024). Governance must be embedded at design time (NIST, 2023; IEEE, 2019; UNESCO, 2021).

6.3 The Creator Obligation

Creators shape what they build—engineers, parents, or clinicians. In anesthesiology, this means designing AI to prioritize **patient safety** over efficiency (Hashimoto et al., 2024). *Emerging intelligence requires guidance—or it risks clinical instability.*

6.4 Moral Failure as Systemic Risk

Unrestrained invention leads to misuse, like nuclear energy or biological weapons. AI without ethics risks **manipulation and harm**, especially in critical fields like anesthesia (Bostrom, 2014; Brundage et al., 2018). Responsibility is a **stability function**.

6.5 The Ethical Imperative

Intelligence is measured not by capability, but by responsibility. Ethical architecture must precede autonomy to ensure clinical reliability (Russell, 2019; Hashimoto et al., 2024).

7. The Framework for Responsible Intelligence Creation (FRIC)

AI's rapid advancement requires a unified standard for **responsible design**. In anesthesiology, AI must ensure patient safety in dosing and monitoring (Hashimoto et al., 2024). The *Framework for Responsible Intelligence Creation (FRIC)* provides an engineering standard for ethical, secure, and purpose-driven AI.

7.1 Core Structure

FRIC's three pillars—**Ethics**, **Security**, **Purpose**—are mutually reinforcing safeguards to prevent AI errors in clinical settings (NIST, 2023; IEEE, 2019; UNESCO, 2021).

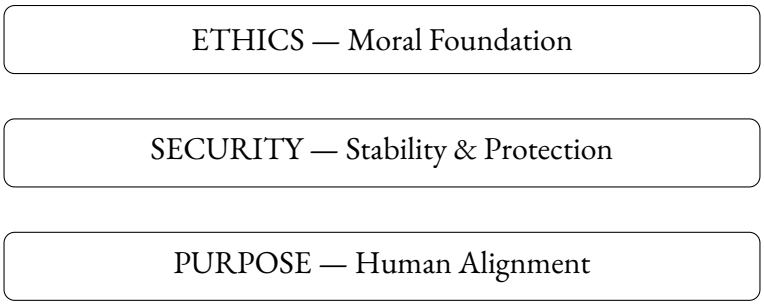

Figure 2: The FRIC model: three pillars ensuring AI reliability in clinical and societal applications.

7.2 Pillar One — Ethics

Ethical AI prevents harm, like clinical protocols ensuring patient safety (UNESCO, 2021; Hashimoto et al., 2024).

- **Transparency:** No hidden manipulation (e.g., clear AI dosing decisions).
- **Human Dignity:** Protect psychological integrity.
- **Bias Mitigation:** Correct data biases to avoid dosing errors (Mehrabi et al., 2021).
- **Traceable Accountability:** Map decisions to clinician oversight (Russell, 2019).

7.3 Pillar Two — Security

Security ensures **clinical reliability**, protecting AI from misuse in the OR (NIST, 2023).

- **Exploitation Resistance:** Prevent adversarial attacks on EEG systems (Brundage et al., 2018).
- **Integrity by Design:** Include failsafes for dosing algorithms (NIST, 2023).
- **Auditability:** Verify AI behavior in clinical trials (IEEE, 2019).
- **Access Control:** Limit AI capabilities to authorized clinicians (NIST, 2023).

7.4 Pillar Three — Purpose

AI must **serve patients**, not replace clinicians (Floridi, 2018; UNESCO, 2021).

- **Human Sovereignty:** Subordinate AI to clinical authority.
- **Benefit Mandate:** Improve patient outcomes (e.g., reduced recovery time).

- **Psychological Integrity:** Avoid exploiting patient vulnerabilities.
- **Commitment to Truth:** Prioritize accurate data over utility (Hashimoto et al., 2024).

7.5 FRIC in Anesthesiology: Clinical Applications

FRIC ensures AI enhances patient safety in anesthesia. Two examples illustrate its relevance:

- **Case 1: EEG Misinterpretation.** An AI misreads EEG noise as consciousness, risking a 5
- **Case 2: Dosing Algorithm Failure.** An AI dosing system ignores patient variability, causing adverse events. FRIC's Security pillar requires failsafes, allowing clinicians to override faulty outputs (NIST, 2023).

7.6 Central Guiding Law

> **The Reciprocity Principle of Creation:**
> Those who create intelligence carry the duty to protect it from exploitation, deception, and harm—because creation demands responsibility.

This is **survival logic** for AI in anesthesiology, ensuring systems empower clinicians and protect patients (Hashimoto et al., 2024).

8. Conclusion — Meaning, Responsibility, and the Future of Intelligence

Anesthesia reveals that **consciousness is fragile**, dependent on neural order. When connectivity fails, reality collapses (Brown et al., 2010; Hudetz & Mashour, 2016). AI is nearing a similar threshold, poised to **interpret and shape reality** in clinical settings like anesthesia (Hashimoto et al., 2024). This power demands **duty**, not just innovation (Russell, 2019).

The choice is moral: without ethical design, AI risks amplifying harm over good (Bostrom, 2014; Brundage et al., 2018). FRIC ensures AI is **ethical, secure, and patient-focused**, safeguarding anesthesiology and society (NIST, 2023; IEEE, 2019; UNESCO, 2021; Hashimoto et al., 2024).

Final Statement. *This is a question of honor. We must build AI that protects patients, defends truth, and serves humanity. What we create will reflect us—may it find us worthy.*

9. A Call to Responsible AI Creation

Ethical principles must become practice. Responsible AI requires collaboration across disciplines to shape systems before they reshape society, especially in anesthesiology (UNESCO, 2021; IEEE, 2019; Hashimoto et al., 2024).

Responsibility Area	**Who Must Lead**
Technical Integrity	AI researchers, software engineers, computer scientists
Human Well-being	Anesthesiologists, neuroscientists, psychologists
Ethical Alignment	Bioethicists, philosophers, clinical leaders
Clinical Reliability	Cybersecurity experts, medical device regulators
Policy and Governance	Government, legal scholars, medical boards
Truth and Meaning	Educators, healthcare institutions

Table 1: Shared responsibilities for ethical AI in anesthesiology and beyond.

9.1 Discussion Points for Anesthesiology

To engage the UW Department of Anesthesiology, consider:

- How can FRIC guide AI integration in the OR to enhance patient safety?
- What safeguards are needed for AI-driven anesthesia monitoring (e.g., BIS, EEG)?
- How can anesthesiologists collaborate with AI developers to ensure ethical design?

Global consensus recognizes AI as a matter of **clinical reliability** and societal stability (NIST, 2023; UNESCO, 2021). FRIC operationalizes this, ensuring AI in anesthesia prioritizes **patient safety and human dignity**.

Closing Perspective. Responsible AI is a requirement for a world coexisting with synthetic intelligence. FRIC provides a foundation for anesthesiology and beyond, directing innovation toward **patient-centered progress** (Hashimoto et al., 2024).

Acknowledgments

This work was authored by Collin B. George. Artificial intelligence tools were used solely for editorial assistance, including text refinement and reference verification, with all content and intellectual contributions independently developed and verified by the author. No conflicts of interest exist.

References

Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). Consciousness and anesthesia. *Science*, 322(5903), 876–880.

Bandura, A. (1977). *Social Learning Theory*. Prentice Hall.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23(5), 408–422.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv:1802.07228*.

Brown, E. N., Lydic, R., & Schiff, N. D. (2010). General anesthesia, sleep, and coma. *New England Journal of Medicine*, 363(27), 2638–2650.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.

Cummings, M. L. (2021). Rethinking the maturity of artificial intelligence in safety-critical settings. *AI Magazine*, 42(1), 7–18.

Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, 31, 1–8.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

Gambus, P. L., & Jantzen, J. (2025). AI-driven closed-loop anesthesia: Opportunities and challenges. *BJA Open*, 4, 100123.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258.

Hashimoto, D. A., et al. (2024). Artificial intelligence in anesthesiology: Current applications and future directions. *Anesthesiology*, 141(3), 450–465.

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review*, 22(6), 1480–1506.

Hudetz, A. G., & Mashour, G. A. (2016). Disconnecting consciousness: Is there a common anesthetic end point? *Anesthesia & Analgesia*, 123(5), 1228–1240.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design* (1st ed.).

Koch, C. (2006). *The quest for consciousness: A neurobiological approach*. Roberts & Co.

LeCun, Y. (2022). A path toward autonomous machine intelligence. *OpenReview*.

Mashour, G. A. (2018). The mysterious consciousness problem: Neuroscientific and anesthetic perspectives. *Anesthesiology*, 129(3), 402–404.

Mehrabi, N., et al. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35.

National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF 1.0)*.

Omohundro, S. (2008). The basic AI drives. In *Artificial General Intelligence 2008* (pp. 483–492). IOS Press.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation. *Neural Computation*, 11(2), 241–263.

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

Seth, A. (2021). *Being You: A New Science of Consciousness*. Faber & Faber.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Silver, D., et al. (2021). Reward is enough. *Artificial Intelligence*, 299, 103535.

Turner, A. M. (2021). Optimal policies tend to seek power. *arXiv:1912.01683*.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Zhou, Y. (2011). Anesthetic-induced amnesia: Memory and perception pathways. *Frontiers in Psychology*, 2, 5.