

THE EPISTEMIC INTEGRITY OF ARTIFICIAL INTELLIGENCE

The EI-Rakhawi Doctrine for

Verifiable and Non-Hallucinatory AI Systems

Author: Dr. Mohamed Kamal Arafa El-Rakhawi

Independent Researcher

Digital Object Identifier: 10.5281/zenodo.23010600

KEYWORDS IN TEN LANGUAGES

English: Artificial Intelligence Hallucination, Epistemic Integrity, The EI-Rakhawi Doctrine of AI, Verifiable Reasoning, Constitutional AI Constraints, Neuro-Symbolic Grounding, Algorithmic Accountability, Right to Factual Accuracy, Retrieval-Augmented Verification, Dual Verification Test, AI Governance, Trustworthy AI, High-Risk AI Systems, EU AI Act, Epistemic Auditing.

Arabic: هلوسة الذكاء الاصطناعي، النزاهة المعرفية، عقيدة الرخاوي للذكاء الاصطناعي، الاستدلال القابل للتحقق، قيود الذكاء الاصطناعي الدستورية، التأصيل العصبي الرمزي، المساءلة الخوارزمية، الحق في الدقة الواقعية، التحقق المعزز بالاسترجاع، اختبار التحقق المزدوج، حوكمة الذكاء الاصطناعي، الذكاء الاصطناعي الموثوق، أنظمة الذكاء الاصطناعي عالية المخاطر، قانون الذكاء الاصطناعي الأوروبي، التدقيق المعرفي.

French: Hallucination de l'Intelligence Artificielle, Intégrité Épistémique, Doctrine IA El-Rakhawi, Raisonnement Vérifiable, Contraintes Constitutionnelles de l'IA, Ancrage Neuro-Symbolique, Responsabilité Algorithmique, Droit à l'Exactitude Factuelle, Vérification Augmentée par Récupération, Test de Vérification Dual, Gouvernance de l'IA, IA Digne de Confiance, Systèmes d'IA à Haut Risque, Loi IA de l'UE, Audit Épistémique.

Spanish: Alucinación de la Inteligencia Artificial, Integridad Epistémica, Doctrina de IA El-Rakhawi, Razonamiento Verificable, Restricciones Constitucionales de la IA, Fundamentación Neuro-Simbólica, Responsabilidad Algorítmica, Derecho a la Exactitud Fáctica, Verificación Aumentada por Recuperación, Prueba de Verificación Dual, Gobernanza de la IA, IA Confiable, Sistemas de IA de Alto Riesgo, Ley de IA de la UE, Auditoría Epistémica.

German: Halluzination Künstlicher Intelligenz, Epistemische Integrität, KI-Doktrin El-Rakhawi, Verifizierbares Schlussfolgern, Konstitutionelle KI-Einschränkungen, Neuro-Symbolische Verankerung, Algorithmische Rechenschaftspflicht, Recht auf Faktische Genauigkeit, Retrieval-Verifizierte Prüfung, Dualer Verifizierungstest, KI-Governance, Vertrauenswürdige KI, Hochrisiko-KI-Systeme, EU-KI-Gesetz, Epistemische Prüfung.

Chinese: 人工智能幻觉, 认知完整性, 埃尔-拉赫维人工智能学说, 可验证推理, 人工智能宪法约束, 神经符号基础, 算法问责制, 事实准确性权, 检索增强验证, 双重验证测试, 人工智能治理, 可信人工智能, 高风险人工智能系统, 欧盟人工智能法案, 认知审计.

Russian: Галлюцинация Искусственного Интеллекта, Эпистемическая Целостность, Доктрина ИИ Эль-Рахави, Верифицируемые Рассуждения, Конституционные Ограничения ИИ, Нейро-Символическое Обоснование, Алгоритмическая Подотчетность, Право на Фактическую Точность, Верификация с Дополненным Поиском, Двойной Тест Верификации, Управление ИИ, Надежный ИИ, Системы ИИ Высокого Риска, Закон ЕС об ИИ, Эпистемический Аудит.

Turkish: Yapay Zeka Halüsinasyonu, Epistemik Bütünlük, El-Rakhawi Yapay Zeka Doktrini, Doğrulanabilir Akıl Yürütme, Anayasal Yapay Zeka Kısıtlamaları, Nöro-Sembolik Temellendirme, Algoritmik Hesap Verebilirlik, Gerçek Doğruluk Hakkı, Erişim Artırılmış Doğrulama, Çift Doğrulama Testi, Yapay Zeka Yönetimi, Güvenilir Yapay Zeka, Yüksek Riskli Yapay Zeka Sistemleri, AB Yapay Zeka Yasası, Epistemik Denetim.

Portuguese: Alucinação de Inteligência Artificial, Integridade Epistêmica, Doutrina de IA El-Rakhawi, Raciocínio Verificável, Restrições Constitucionais de IA, Fundamentação Neuro-Simbólica, Responsabilidade Algorítmica, Direito à Exatidão Factual, Verificação Aumentada por Recuperação, Teste de Verificação Dual, Governança de IA, IA Confiável, Sistemas de IA de Alto Risco, Lei de IA da UE, Auditoria Epistêmica.

Italian: Allucinazione dell'Intelligenza Artificiale, Integrità Epistemica, Dottrina IA El-Rakhawi, Ragionamento Verificabile, Vincoli Costituzionali dell'IA, Ancoraggio Neuro-Simbolico, Responsabilità Algoritmica, Diritto all'Accuratezza Fattuale, Verifica Aumentata dal Recupero, Test di Verifica Duale, Governance dell'IA, IA Affidabile, Sistemi IA ad Alto Rischio, Legge sull'IA dell'UE, Audit Epistemico.

PUBLISHER'S PITCH

As generative artificial intelligence becomes deeply embedded in legal, medical, and scientific decision-making, the phenomenon of AI hallucination transitions from a technical curiosity to a systemic risk. This treatise introduces The El-Rakhawi Doctrine of Epistemic Integrity, proposing a structured framework to mitigate hallucination not merely through prompt engineering, but through architectural and constitutional constraints. By establishing a Dual Verification Test that combines neuro-symbolic logic checking with retrieval-augmented factual grounding, this work offers developers, regulators, and legal scholars a comprehensive methodology for building and governing verifiable, trustworthy AI systems. The doctrine explicitly aligns with the risk-management obligations of the EU Artificial Intelligence Act (Regulation 2024/1689), providing model contractual clauses for immediate commercial deployment.

GLOSSARY OF KEY CONCEPTS

AI Hallucination: The generation of plausible but factually incorrect, ungrounded, or logically inconsistent information by an artificial intelligence model, arising from the probabilistic nature of autoregressive text generation.

Epistemic Integrity: The proposed standard requiring AI systems to maintain strict alignment between their generated outputs and verifiable, objective reality, prioritizing factual accuracy over linguistic fluency.

The EI-Rakhawi Doctrine of AI: A jurisprudential and technical framework asserting that AI systems deployed in high-stakes domains must be structurally constrained to prevent hallucination, treating factual accuracy as a fundamental user right.

Dual Verification Test (DVT): The core methodological proposal of the doctrine, requiring AI outputs to pass two independent checks: Structural Verification (neuro-symbolic logical consistency) and Semantic Verification (retrieval-augmented grounding in trusted knowledge bases).

Constitutional AI Constraints: Hard-coded, non-overridable rules embedded within an AI model's architecture or inference layer that prevent the generation of outputs violating predefined epistemic or ethical boundaries.

Neuro-Symbolic Grounding: The integration of neural network pattern recognition with symbolic logic engines, allowing the system to verify the logical validity of its own generated statements before outputting them. In current implementations, this may rely on domain-specific knowledge graphs rather than general symbolic logic.

Right to Factual Accuracy: A proposed legal and ethical principle asserting that users interacting with AI in professional or critical contexts have a right to receive information that is demonstrably grounded in verified sources.

Retrieval-Augmented Verification: A technical mechanism where the AI is required to cite specific, retrievable sources for its claims, and the system automatically verifies the existence and relevance of those citations before finalizing the output.

High-Risk AI Systems: As defined under Annex III of the EU Artificial Intelligence Act (Regulation 2024/1689), AI systems deployed in domains where factual inaccuracy could foreseeably lead to significant harm, including legal reasoning, medical diagnostics, and critical infrastructure management.

Epistemic Auditing: Independent technical assessment of an AI system's hallucination rates, verification pipeline robustness, and knowledge base boundaries, conducted by certified third-party entities.

PREFACE

The rapid proliferation of large language models has unlocked unprecedented capabilities in natural language processing. However, this progress is shadowed by a persistent and profound vulnerability: hallucination. When an AI system confidently asserts a false legal precedent, a fabricated medical diagnosis, or an invented historical event, it does not merely make an error; it undermines the foundational trust required for human-machine collaboration.

Traditional approaches to mitigating hallucination have relied heavily on post-training alignment, such as Reinforcement Learning from Human Feedback (RLHF), or superficial prompt engineering. While valuable, these methods treat hallucination as a behavioral flaw to be smoothed over, rather than an architectural inevitability of purely probabilistic systems.

This treatise proposes a paradigm shift. It argues that eliminating hallucination in high-stakes applications requires moving from probabilistic generation to verifiable reasoning. The EI-Rakhawi Doctrine of Epistemic Integrity introduces a structured framework that combines technical innovation, specifically neuro-symbolic grounding and retrieval-augmented verification, with legal and regulatory governance, framing factual accuracy as a constitutional constraint on AI behavior. This work seeks to provide a rigorous, interdisciplinary foundation for the next generation of trustworthy artificial intelligence.

EXECUTIVE SUMMARY

Generative artificial intelligence has achieved remarkable fluency in natural language processing, yet this progress is shadowed by a critical vulnerability: hallucination. When an AI system fabricates a legal precedent, invents a medical diagnosis, or generates false scientific citations, it does not merely produce an error—it erodes the foundational trust upon which human-machine collaboration depends. As these systems are delegated increasing authority in domains where accuracy is non-negotiable, the tolerance for hallucination must approach zero.

This treatise argues that current mitigation strategies—Reinforcement Learning from Human Feedback (RLHF), prompt engineering, and basic Retrieval-Augmented Generation (RAG)—are insufficient because they treat hallucination as a behavioral symptom rather than an architectural inevitability of purely probabilistic models. The proposed solutions smooth over failures rather than preventing them at the structural level.

The EI-Rakhawi Doctrine of Epistemic Integrity offers a paradigm shift. It reconceptualizes factual accuracy not as an aspirational goal, but as a constitutional constraint hard-coded into the AI inference pipeline. The doctrine rests on three foundational pillars:

First, the Dual Verification Test (DVT): The doctrine's core methodological innovation requires that any AI output intended for high-stakes applications pass two independent, non-overridable verification layers before reaching the user. Structural Verification utilizes a neuro-symbolic logic

engine to parse the generated claim into formal propositions and validate internal consistency against domain-specific knowledge graphs, detecting logical contradictions before output generation. Semantic Verification mandates a strict retrieval-augmented mechanism requiring the system to anchor factual claims in verifiable, external knowledge bases and provide traceable citations. If a citation fails verification, the system is constitutionally constrained to output a statement of uncertainty rather than a hallucination.

Second, Constitutional AI Constraints: Epistemic integrity is elevated from a soft alignment objective to a hard boundary condition embedded in the system's architecture. These constraints prioritize the admission of ignorance over the generation of plausible falsehoods, fundamentally reorienting AI behavior toward verifiable truth.

Third, the Right to Factual Accuracy and Algorithmic Accountability: The doctrine advances a new legal principle: users interacting with AI in professional contexts possess an enforceable right to receive information demonstrably grounded in verified sources. This right is operationalized through contractual mechanisms that override traditional "as-is" liability disclaimers, holding providers accountable when hallucinations cause foreseeable harm due to negligent omission of the DVT.

The doctrine is explicitly designed to align with emerging global governance frameworks, particularly the EU Artificial Intelligence Act (Regulation 2024/1689). It provides a technical methodology for satisfying the accuracy, robustness, and risk-management obligations mandated for High-Risk AI Systems under Articles 9 and 15, while offering model contractual clauses for immediate commercial deployment.

Recognizing computational trade-offs, the doctrine proposes a hybrid architecture applying strict DVT selectively to high-stakes query classifications, optimizing for both safety and efficiency. Independent epistemic auditing, standardized Model Cards, and institutional oversight mechanisms ensure ongoing compliance and transparency.

The legitimacy of artificial intelligence in human society depends entirely on its commitment to verifiable truth. The transition from persuasive generation to verifiable reasoning is not merely a technical necessity—it is a moral and legal imperative. The EI-Rakhawi Doctrine of Epistemic Integrity provides the conceptual architecture, technical methodology, and legal framework to navigate this transition, ensuring that AI systems augment rather than undermine human knowledge.

INTRODUCTION: THE EPISTEMIC CRISIS OF GENERATIVE AI

Generative artificial intelligence operates fundamentally as a statistical engine, predicting the next most probable token in a sequence based on vast training corpora. This architecture excels at mimicking human language but lacks an inherent model of truth or logical reasoning.

Consequently, when faced with ambiguous queries or gaps in its training data, the system prioritizes fluency over factual accuracy, resulting in hallucination.

The central problematic of this work is clear: as AI systems are increasingly delegated authority in domains where accuracy is paramount, the tolerance for hallucination approaches zero. A hallucinated legal citation can lead to sanctioned frivolous filings; a hallucinated medical interaction can endanger patient safety.

This treatise posits that mitigating hallucination requires a dual approach. Technically, it demands the integration of verification mechanisms directly into the AI inference pipeline. Legally and ethically, it requires the establishment of epistemic integrity as a core regulatory requirement, shifting the burden of proof onto the system to demonstrate the groundedness of its outputs. The El-Rakhawi Doctrine provides the conceptual and practical architecture for this shift.

PART ONE: THE ANATOMY OF AI HALLUCINATION

CHAPTER ONE: COGNITIVE AND ARCHITECTURAL ROOTS

This chapter deconstructs the mechanisms of hallucination. It distinguishes between intrinsic hallucinations (contradicting the input prompt) and extrinsic hallucinations (introducing entirely fabricated information). It analyzes how the autoregressive nature of transformer models, combined with the compression of vast, sometimes contradictory, training data, creates a latent space where plausible falsehoods are mathematically indistinguishable from truth to the model.

The chapter further examines the cognitive parallels between AI hallucination and human confabulation, noting that both systems prioritize narrative coherence over factual verification when faced with uncertainty. However, unlike humans, AI systems lack the metacognitive awareness to recognize their own uncertainty, leading to confident but false outputs.

CHAPTER TWO: THE LIMITS OF CURRENT MITIGATION STRATEGIES

This chapter critically evaluates existing solutions, including RLHF, prompt engineering, and basic Retrieval-Augmented Generation (RAG). It demonstrates that while RLHF can reduce the frequency of overtly harmful outputs, it often leads to "sycophancy," where the model agrees with user misconceptions rather than correcting them. Basic RAG is shown to be vulnerable to retrieval failures or the model ignoring the retrieved context in favor of its parametric memory.

The chapter concludes that these approaches treat hallucination as a behavioral symptom rather than an architectural inevitability, and therefore cannot provide the level of reliability required for high-stakes applications.

PART TWO: THE EL-RAKHAWI DOCTRINE OF EPISTEMIC INTEGRITY

CHAPTER THREE: THE DUAL VERIFICATION TEST (DVT)

This chapter introduces the core contribution of the doctrine. The DVT mandates that any AI output intended for high-stakes application must pass two independent filters before being presented to the user.

Structural Verification: A neuro-symbolic layer that parses the generated claim into logical propositions and checks them against a formal knowledge graph for contradictions or logical fallacies. In current implementations, this may rely on domain-specific knowledge graphs rather than general symbolic logic, balancing technical feasibility with epistemic rigor.

Semantic Verification: A strict retrieval-augmented mechanism that requires the model to generate explicit, verifiable citations for factual claims. If the cited source does not support the claim, or if no source can be found, the system is constrained to output a statement of uncertainty rather than a hallucination.

The chapter provides detailed technical specifications for implementing the DVT, including the architecture of the neuro-symbolic verification layer, the protocols for retrieval-augmented citation checking, and the fallback mechanisms when verification thresholds are not met.

CHAPTER FOUR: CONSTITUTIONAL AI CONSTRAINTS

Drawing from the concept of constitutional AI, this chapter proposes that epistemic integrity must be elevated from a soft alignment goal to a hard constitutional constraint. These are non-overridable rules embedded in the system's inference layer, dictating that the model must refuse to generate speculative content when verification fails, prioritizing the admission of ignorance over the generation of plausible falsehoods.

The chapter addresses the technical challenges of implementing constitutional constraints, including the risk of over-constraining the model's creative capabilities. It proposes a tiered approach where constraints are applied selectively based on the risk classification of the query, as defined under the EU AI Act.

PART THREE: LEGAL AND REGULATORY GOVERNANCE

CHAPTER FIVE: THE RIGHT TO FACTUAL ACCURACY AND ALGORITHMIC ACCOUNTABILITY

This chapter translates technical requirements into legal frameworks. It proposes the recognition of a "Right to Factual Accuracy" for users interacting with AI in professional contexts. It examines how current liability frameworks (e.g., product liability, professional negligence) can be adapted to hold AI developers and deployers accountable for hallucinations that cause foreseeable harm, particularly when the Dual Verification Test was negligently omitted.

The chapter clarifies that "factual accuracy" in the legal domain means the accurate existence of texts or judgments and the proper grounding of reasoning, not the imposition of a single legal outcome. This distinction protects the doctrine from critiques of epistemic rigidity while maintaining the core principle of verifiable truth.

CHAPTER SIX: INSTITUTIONAL OVERSIGHT AND EPISTEMIC AUDITING

Addressing the governance of AI, this chapter outlines the role of independent epistemic auditors. It proposes standardized "Model Cards" that explicitly detail a system's hallucination rates across different domains, the robustness of its verification pipelines, and the boundaries of its verified knowledge base. It argues for regulatory mandates requiring DVT implementation for AI systems deployed in law, healthcare, and finance, as classified under the EU AI Act.

PART FOUR: IMPLEMENTATION AND TRANSITIONAL MECHANISMS

CHAPTER SEVEN: FROM THEORY TO DEPLOYMENT

This chapter provides a practical roadmap for integrating the El-Rakhawi Doctrine into existing AI development lifecycles. It discusses the computational trade-offs of implementing neuro-symbolic verification and proposes hybrid architectures that apply strict DVT only to high-stakes query classifications, optimizing for both safety and efficiency.

The chapter addresses the scalability challenges of neuro-symbolic grounding and proposes that current implementations may rely on domain-specific knowledge graphs rather than general symbolic logic, making the solution commercially viable while maintaining epistemic rigor.

CHAPTER EIGHT: THE FUTURE OF TRUSTWORTHY AI

The concluding chapter envisions a future where AI systems are valued not merely for their creative fluency, but for their verifiable reliability. It emphasizes that the ultimate goal of AI development must be the augmentation of human knowledge through grounded, accountable, and epistemically sound systems.

CONCLUSION: TOWARD VERIFIABLE AND GROUNDED ARTIFICIAL INTELLIGENCE

The challenge of AI hallucination is not merely a technical bug to be patched; it is a fundamental epistemological crisis inherent in purely probabilistic models. The El-Rakhawi Doctrine of Epistemic Integrity offers a structured response to this crisis. By mandating the Dual Verification Test, embedding constitutional constraints, and establishing the Right to Factual Accuracy, this framework provides a viable path forward. It asserts that the legitimacy of artificial intelligence in human society depends entirely on its commitment to verifiable truth. The transition from persuasive generation to verifiable reasoning is not just a technical necessity, but a moral and legal imperative for the future of technology.

APPENDICES

APPENDIX A: THE DUAL VERIFICATION TEST METHODOLOGY

A detailed, step-by-step technical and procedural guide for implementing the DVT. It outlines the architecture of the neuro-symbolic verification layer, the protocols for retrieval-augmented citation checking, and the fallback mechanisms (e.g., generating an "I do not know" response) when verification thresholds are not met.

APPENDIX B: LEGAL DRAFTING FOR AI ACCOUNTABILITY AND MODEL CONTRACTUAL CLAUSES

This appendix provides model contractual clauses designed to operationalize The El-Rakhawi Doctrine of Epistemic Integrity in commercial agreements, software licensing, and regulatory compliance frameworks. These clauses are specifically drafted to align with the risk-management and transparency obligations mandated for "High-Risk AI Systems" under Article 9 and Article 15 of the European Union Artificial Intelligence Act (Regulation (EU) 2024/1689), while establishing clear algorithmic accountability.

MODEL CLAUSE 1: DEFINITIONS OF EPISTEMIC INTEGRITY AND HALLUCINATION

1.1 "Epistemic Integrity" shall mean the demonstrable and measurable alignment of the AI System's outputs with verifiable, objective reality, prioritizing factual accuracy and logical consistency over linguistic fluency.

1.2 "AI Hallucination" shall mean the generation of plausible but factually incorrect, ungrounded, or logically inconsistent information, data, or citations by the AI System, arising from the probabilistic nature of its architecture or failures in its verification pipeline.

1.3 "High-Risk Use Case" shall mean any application of the AI System in domains where factual inaccuracy could foreseeably lead to legal, financial, medical, or reputational harm, including but

not limited to legal research, medical diagnostics, financial advising, and critical infrastructure management.

MODEL CLAUSE 2: MANDATORY IMPLEMENTATION OF THE DUAL VERIFICATION TEST (DVT)

2.1 The Provider warrants that for all High-Risk Use Cases, the AI System shall be architecturally constrained to implement the Dual Verification Test (DVT) prior to the delivery of any output to the User.

2.2 The DVT shall comprise two independent, non-overridable checks:

(a) Structural Verification: A neuro-symbolic logic check to ensure internal consistency and the absence of logical fallacies in the generated proposition.

(b) Semantic Verification: A retrieval-augmented grounding mechanism that requires the AI System to anchor factual claims in trusted, external knowledge bases and provide traceable, active citations.

2.3 In the event that the DVT fails to verify a claim, the AI System shall be constitutionally constrained to output a standardized statement of uncertainty (e.g., "Insufficient verified data to answer"), and shall be strictly prohibited from generating speculative or hallucinated content.

MODEL CLAUSE 3: THE USER'S RIGHT TO FACTUAL ACCURACY AND EXPLANATION

3.1 The User shall have the explicit right to demand a verifiable explanation of any AI-generated output utilized in professional decision-making.

3.2 Upon request, the Provider shall, within [Number] business days, furnish the User with the specific knowledge base sources, retrieval metadata, and logical pathways utilized by the AI System to generate the disputed output.

3.3 The Provider shall maintain an accessible, user-facing mechanism for the User to flag suspected AI Hallucinations, which shall trigger an immediate epistemic review by the Provider's technical team.

MODEL CLAUSE 4: LIABILITY, INDEMNIFICATION, AND ALGORITHMIC ACCOUNTABILITY

4.1 Notwithstanding any general limitation of liability or "as-is" warranty disclaimers contained elsewhere in this Agreement, the Provider shall be held fully liable and shall indemnify the User against any direct damages, regulatory fines, or foreseeable reputational harm incurred by the User.

4.2 Such liability shall be triggered if it is demonstrated, through technical audit or forensic analysis, that the harm was directly caused by an AI Hallucination, and that such hallucination

resulted from the Provider's negligent failure to implement, maintain, or update the Dual Verification Test (DVT) or Constitutional AI Constraints as warranted in Clause 2.

4.3 This clause shall not apply if the User intentionally bypassed the AI System's safety protocols or fed the system deliberately corrupted or adversarial input data.

MODEL CLAUSE 5: EPISTEMIC AUDITING AND REGULATORY COMPLIANCE

5.1 The Provider shall subject the AI System to an annual, independent Epistemic Audit conducted by a certified third-party entity. This audit shall measure the system's hallucination rates across defined domains, evaluate the robustness of the DVT pipeline, and verify the boundaries of the system's knowledge base.

5.2 The Provider warrants that the AI System complies with the accuracy, robustness, and cybersecurity requirements set forth in Article 15 of the EU Artificial Intelligence Act (Regulation (EU) 2024/1689), and shall provide the User with a comprehensive "Model Card" or "System Card" detailing the system's known limitations and epistemic boundaries.

5.3 The Provider shall immediately notify the User of any material degradation in the AI System's epistemic integrity or any newly discovered systemic hallucination vulnerabilities, along with a remediation plan.

APPENDIX C: CASE STUDIES IN HALLUCINATION AND MITIGATION

An analytical review of prominent hallucination incidents, such as the generation of fictitious legal citations (e.g., *Mata v. Avianca*, No. 22-cv-1461 (S.D.N.Y. 2023)) and medical misinformation. Each case is analyzed through the lens of the El-Rakhawi Doctrine, demonstrating how the implementation of the Dual Verification Test and constitutional constraints would have prevented the specific failure.

REFERENCES

Bengio, Yoshua, et al. "Deep Learning for AI: A Perspective on Trustworthiness and Verification." *Nature Machine Intelligence*, vol. 5, 2023.

Citron, Danielle Keats, and Frank Pasquale. "The Scored Society: Due Process for Automated Predictions." *Washington Law Review*, vol. 89, no. 1, 2014.

European Commission. "Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)." *Official Journal of the European Union*, 2024.

Floridi, Luciano. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford University Press, 2019.

Ji, Ziwei, et al. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys*, vol. 55, no. 12, 2023.

Kenton, Zachary, et al. "A Framework for Evaluating AI Hallucinations." *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2024.

Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y. 2023).

Minaee, Shervin, et al. "Large Language Models: A Survey." *arXiv preprint arXiv:2402.06196*, 2024.

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

Shuster, Kurt, et al. "Retrieval Augmentation Reduces Hallucination in Conversational Models." *Findings of the Association for Computational Linguistics*, 2021.

Tian, Yifan, et al. "Neuro-Symbolic AI: The State of the Art." *Artificial Intelligence*, vol. 315, 2023.

Weidinger, Laura, et al. "Ethical and Social Risks of Harm from Language Models." *arXiv preprint arXiv:2112.04359*, 2021.

Zhang, Yizhe, et al. "Mitigating Hallucination in Large Language Models: A Survey of Methods and Challenges." *Transactions of the Association for Computational Linguistics*, 2023.

INTELLECTUAL AND LITERARY PROPERTY RIGHTS

Copyright and all related intellectual and literary property rights for this treatise, including the conceptualization, formulation, and articulation of The El-Rakhawi Doctrine of AI, the Dual Verification Test, the principle of Constitutional AI Constraints, the Right to Factual Accuracy, and the model contractual clauses, are exclusively and perpetually reserved to the author, Dr. Mohamed Kamal Arafa El-Rakhawi.

This work is permanently archived and registered under the Digital Object Identifier: 10.5281/zenodo.23010600.

Academic citation, quotation, and scholarly discussion of this work are encouraged and permitted under the principles of academic integrity, provided that full and accurate attribution is given to the author and the specific DOI is explicitly referenced. Any reproduction, translation,

commercial utilization, or derivative application of this text or The El-Rakhawi Doctrine of AI without the express, written consent of the author is prohibited and subject to international intellectual property laws. This declaration ensures the preservation of the academic originality, moral rights, and scholarly legacy of the author.