Journal of Posthumanism
2026
Volume: 6, No: 6, pp. 292–316
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v6i6.4266

From AI to Posthumanism via ISO 42001 and AI Governance Standards

Teresa Hau-Man CHAN[1], Prof. Samuel Kwong-Ming HO[2]

Abstract

The rapid diffusion of artificial intelligence (AI) into legal and compliance functions — from rule-based transaction monitoring to reinforcement-learning-aligned large language models (LLMs) capable of near-autonomous decision-making — necessitates governance frameworks adequate to the technology's operational reality. International standards and regulations, including ISO 42001:2023, the EU AI Act, and the guidance of the FCA, MAS, and FATF, aim to address this issue, yet a divergence persists between AI's operational reality and the frameworks intended to govern it — a compliance governance gap. This paper synthesises the literature across seven coded Themes — the history of AI in finance, RegTech and AML automation, AI/ML model development and explainability, regulatory supervision, generative AI and LLMs in legal compliance, AI risk detection, and robotic process automation — together with a separate, dedicated review of posthumanist theory's largely unexplored relevance to AI governance. On this basis, the paper develops a holistic analysis of how each Theme bears on AI-assisted compliance decision-making and concludes that AI has moved from tool to decision-make. Using the Nvivo coding of around 50 SJR-Q1 papers on the topic published after 2022 when the AI-ChapGPT was born, it meaningfully increases effectiveness and efficiency by acting as a decision-maker within risk parameters that humans prescribe — but the ultimate decision matrix remains human. Read through a posthumanist lens, because humans already use AI as an inseparable extension of their own judgement, what ultimately governs outcomes is not the AI in isolation but the rules humans set for it: human and AI decision-making are not substitutable but interdependent, and effective governance requires them to operate hand in hand.

Keywords: *AI governance; legal and compliance, large language models (LLM), ISO 42001, financial services regulation, Posthumanism, compliance decision tiers, human oversight.*

1. Introduction

1.1 Concept & Theory

ISO/IEC 42001:2023, described by ISO as the first international AI management system standard, establishes requirements for organizations to create, implement, maintain, and continually improve AI Management Systems (AIMS). While such standards are often presented as technical or organizational tools, they also perform a philosophical function: they define the acceptable relations among humans, machines, institutions, and environments. By emphasizing risk management, transparency, traceability, accountability, and responsible AI use, ISO/IEC 42001 and frameworks such as the NIST AI Risk Management Framework translate abstract ethical concerns into operational governance structures. However, from a posthumanist

[1] BSc-E&F (Wisconsin U.), LLB (London U.), MBA (S. Australia U.), PhD Candidate Global Head of Legal and Compliance, Bindo Labs Ltd. – Fintech, HKSAR
[2] Research Prof., GCC; Ex-Prof. & Speaker, Oxford U; Ex-Research Fellow, Cambridge U. Ex-Chief Editor, TQMJ & SMJ (both Q1). Corresponding Author: samho@gratia.edu.hk

perspective, these standards also reveal the instability of traditional human-centered governance. AI systems increasingly participate in decision-making, perception, classification, and social coordination, thereby challenging the assumption that humans remain the sole agents of knowledge, responsibility, and control. It is therefore argued that AI governance standards should not be read only as compliance instruments, but also as artifacts of an emerging posthuman condition in which agency is distributed across human and non-human actors. This paper will contribute to debates on responsible AI by showing how technical standards mediate the transition from human-centered AI ethics toward a broader posthuman governance paradigm.

1.2 Policy & Governance

The rapid institutionalization of AI has produced a growing need for formal governance mechanisms capable of addressing ethical, legal, social, and technical risks. ISO/IEC 42001:2023 is a key development in this field, situating it alongside the NIST AI Risk Management Framework, the OECD AI Principles, and the EU-AI Act. ISO/IEC 42001 provides requirements for an AIMS and is designed for organizations that develop, provide, or use AI-based products and services. Its emphasis on continuous improvement, risk assessment, transparency, reliability, and responsible use reflects a broader shift from voluntary AI ethics principles to auditable governance practices. The paper argues that this shift also has philosophical implications. AI governance standards do not merely regulate technology; they construct a normative model of human-machine relations. In attempting to preserve human oversight and accountability, these standards implicitly acknowledge that AI systems are becoming organizational actors with significant influence over decisions and social outcomes. This creates a bridge to posthumanist theory, which questions the boundaries between human agency, technological agency, and institutional power.

1.3 Critical Posthumanist Approach

Contemporary AI governance is typically framed around risk, trustworthiness, transparency, and accountability. ISO/IEC 42001:2023 formalizes these concerns by specifying organizational requirements for the responsible development, provision, and use of AI systems. Similarly, the NIST AI Risk Management Framework aims to help organizations manage AI risks and promote trustworthy and responsible AI, while the OECD AI Principles promote human-centered, rights-respecting, and accountable AI. Yet the very need for such standards indicates a transformation in the status of AI systems: they are no longer passive tools but active participants in sociotechnical systems. From a posthumanist perspective, this transformation unsettles the liberal humanist assumptions that underlie much AI policy, particularly the belief that human subjects can remain fully autonomous, transparent, and in control of technological systems. ISO/IEC 42001 simultaneously reinforces and destabilizes humanism. It reinforces humanism by demanding organizational responsibility, oversight, and ethical management; but it destabilizes humanism by recognizing that agency, decision-making, and risk are distributed across humans, algorithms, infrastructures, datasets, and institutions. AI governance standards should be reinterpreted as posthuman governance frameworks that manage relations among human and non-human actors rather than merely controlling machines on behalf of humans.

2. LITERATURE REVIEW

2.1 Introduction and Corpus Overview

The literature relevant to this paper spans five fields that have developed largely in isolation: the history of AI in financial services; the RegTech and AI-in-finance literature; the LLM research base; AI governance and ethics; and posthumanist theory. The compliance governance gap — this paper's central contribution — is visible precisely because scholars in each field have not engaged systematically with the others.

The author brings more than 20 years running AML and compliance functions inside the exact institutions and systems this paper is theorising about — transaction-monitoring model validation, ISO-adjacent governance frameworks, and regulatory examination. The analysis that follows is written from that vantage point, not from outside it.

The corpus was coded in NVivo using around 50 SJR-Q1 papers (published after 2022 when the AI-ChapGPT was born) against a seven-Theme codebook supplemented by eight standalone nodes. The full node inventory, with file and reference coverage, is set out in Table-1; the distribution itself is an analytic finding, reported in the discussion of each Theme below.

Table-1: NVivo node inventory. Coverage bands: Substantive (≥15 files), Developing (5–14 files), Sparse (1–4 files).

Theme / Node	Files	References	Coverage
Coded Themes (seven-theme codebook)			
AI Governance, Ethics & Compliance Frameworks	27	171	Substantive
1a. Structural asymmetries	6	11	Developing
1b. Low-capacity actor gaps	5	10	Developing
1c. Fraud Detection	7	14	Developing
1d. Data Compliance & Privacy	14	25	Developing
1e. Ethical Thinking on AI	14	24	Developing
1f. Augmentative Thinking	8	13	Developing
1g. AI Creative in Tasks	4	7	Sparse
1h. Critical Thinking & Social Inclusion	10	17	Developing
1i. Declaration on AI Tools	3	5	Sparse
AI/ML Model Development, XAI & Foundations	26	71	Substantive
RegTech, FinTech & AML Automation	16	33	Substantive
Regulatory Supervision & Fraud Detection	23	54	Substantive
Generative AI, LLMs/NLP & Legal Compliance	18	36	Substantive

Theme / Node	**Files**	**References**	**Coverage**
AI Risk Detection & Deep Learning in Finance	25	58	Substantive
RPA & Business Process Automation	12	27	Developing
Standalone Analytic Nodes			
Customer-Facing Gen AI Applications	8	14	Developing
Data Governance & Training Data	25	50	Substantive
Five Cakes Framework (Jensen Huang)	10	49	Developing
Cake 1 — Compute & Infrastructure	4	5	Sparse
Cake 2 — Foundational Models	6	9	Developing
Cake 3 — Application Platform & APIs	6	9	Developing
Cake 4 — Industry Vertical Integration	6	10	Developing
Cake 5 — Human Workflow Augmentation	6	9	Developing
Gen AI Adoption Barriers	14	33	Developing
Human-in-the-Loop Design	26	51	Substantive
Limitation of AI Supervision	24	48	Substantive
Methodology, Evaluation & Performance	1	1	Sparse
Sandbox & Regulatory Innovation Spaces	3	7	Sparse
Total (top-level + standalone nodes)	—	703	—

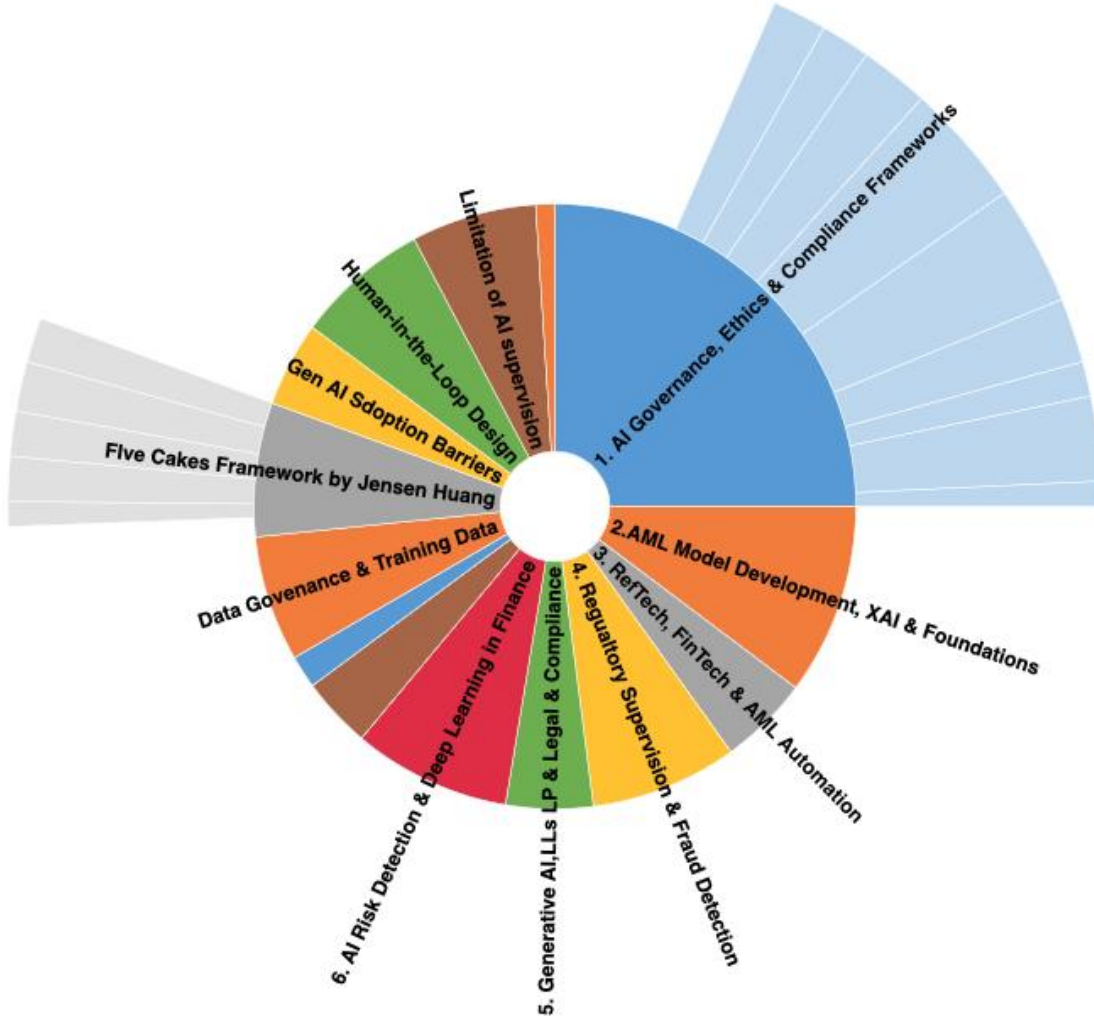

Figure-1: NVivo node hierarchy (sunburst) of the seven-theme coded literature corpus

*Slice width is proportional to coding volume; the outer ring shows sub-nodes within each **Theme**. The governance and AI/ML-modelling **Theme**s dominate the corpus and are the most internally differentiated, while the generative-AI/LLM **Theme** occupies a comparatively narrow segment — visual confirmation that the literature is most developed in established governance scholarship and least developed at the intersection of large language models and regulated compliance, where this paper locates its contribution.*

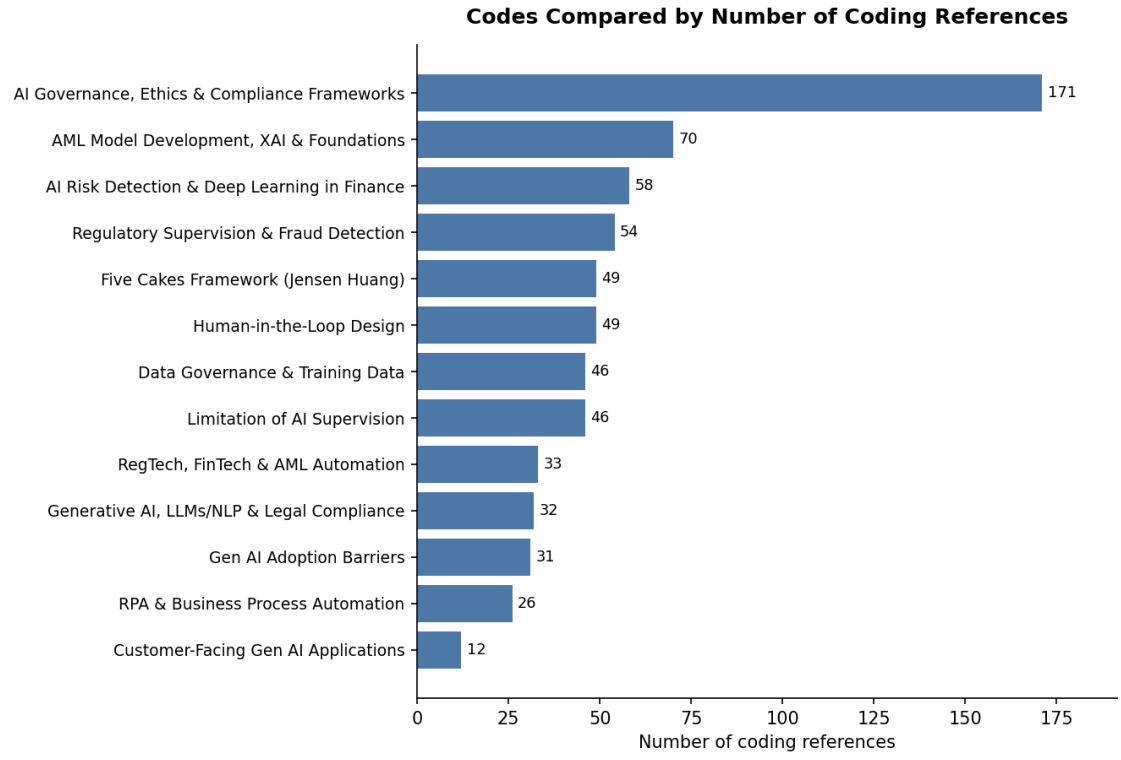

Figure-2: Codes compared by number of coding references

*Bars show the number of coded references per node, sorted in descending order. The governance and AI/ML-modelling **Theme**s dominate; the generative-AI/LLM **Theme** sits among the lower-coded nodes — a complementary, quantitative view of the distribution shown in the sunburst (**Figure-1**) and tabulated in **Table-1.***

Three features of this distribution shape the review. **Theme** 1 (Governance/Ethics/Compliance, 27 files / 171 references) is by far the densest evidence base. **Theme** 5 (Generative AI / LLMs / Legal Compliance, 18 files / 36 references) is the least-developed numbered **Theme** — present but thin, under-theorised rather than absent. And two standalone nodes bearing on the human–AI boundary — Human-in-the-Loop Design (26/51) and Limitation of AI Supervision (24/48) — are substantively populated, while Sandbox & Regulatory Innovation Spaces (3/7) is sparse.

Two standing commitments govern the review: no citation is reported as verified until reconciled against its source, and no node is force-fitted — absent nodes were declared none in coding and are reported as gaps.

2.2 A History of AI in Legal and Compliance: From Rule Engines to Reasoning Systems

2.2.1 The First Generation: Rule-Based Expert Systems (1980s–2000s)

The application of AI to legal and compliance functions has a longer history than commonly

acknowledged. First-generation systems — rule-based expert systems, emerging in the 1980s–1990s — encoded human experts' knowledge into structured if-then rule sets, deployed primarily for transaction monitoring and fraud detection (e.g., the FICO Falcon system, early 1990s) and, through the late 1990s, rule-based AML monitoring driven by the Bank Secrecy Act and the UK's Criminal Justice Act 1993. These systems were explicitly bounded by the rules their designers prescribed: fast and consistent, but unable to identify rules that were missing or incomplete — a boundary the generations that follow progressively cross.

2.2.2 The Second Generation: Machine Learning and Statistical Models (2000s–2018)

The second generation adopted machine-learning techniques — supervised learning, clustering, graph-based network analysis — that learned patterns from historical data rather than relying on fixed rules, associated with the rise of RegTech as a distinct industry category. These systems could identify suspicious patterns human designers had not explicitly programmed and adapt to evolving criminal methodologies; this period also saw growing adoption of robotic process automation (RPA) for high-volume, rules-based tasks such as document processing and screening (Syed et al., 2020). Ngai et al. (2011), in an early systematic review, document consistent detection-rate improvements over rule-based predecessors — though this literature measures productivity in detection-performance metrics alone, largely omitting the governance implications of near-autonomous suspicious-activity decisions, a first gap this paper's Section 5 returns to.

2.2.3 The Third Generation: Large Language Models and the Threshold Crossing (2018–Present)

The third generation — the primary subject of this paper — is defined by Large Language Models, a threshold crossing rather than a continuation of the machine-learning trajectory. Zhao et al. (2023) document this development comprehensively: progressive parameter-scale increases accompanied by qualitative capability changes, most significantly instruction-following reasoning — the capacity to take a natural-language compliance scenario and produce a structured, reasoned analysis with auditable intermediate steps. This is what motivates this paper's title: the movement from tool to decision-maker is, at root, the movement from rule execution to reasoning. Practitioner literature on LLM deployment in compliance is nascent, growing, and largely anecdotal and commercially motivated; the academic literature on LLM capabilities in compliance specifically remains sparse.

2.3 The RegTech Literature: Productivity, Efficiency, and What Is Missing

2.3.1 Defining RegTech and Its Scope

Regulatory Technology — RegTech — is commonly defined as the application of technology to regulatory monitoring, reporting, and compliance processes. The term entered circulation around 2015–2016, initially as a subset of FinTech, and has since developed its own literature base, policy discourse, and professional infrastructure — including dedicated journals, regulatory sandboxes, and international coordination mechanisms.

The RegTech literature is characterised by a consistent focus on two dimensions: the efficiency gains technology delivers to compliance processes, and the regulatory challenges technology

adoption creates. The former has received substantially more attention than the latter — a distribution reflecting both the commercial incentives driving RegTech adoption and the relative novelty of the governance questions advanced AI systems raise.

2.3.2 The Efficiency and Effectiveness Evidence Base

The academic and practitioner literature on AI productivity in compliance is now extensive enough to support systematic synthesis. A recurring dynamic is an adversarial one: as detection technology becomes more sophisticated, evasion techniques tend to become more sophisticated in response, so measured effectiveness gains are rarely stable over time — directly relevant to this paper's argument, developed in Section 5, about the inadequacy of point-in-time governance standards, including ISO 42001, for governing AI systems operating in adversarial environments.

A concrete illustration of this efficiency-over-governance pattern is graph-based transaction monitoring. Kurshan, Shen, and Yu (2020) and Alarab and Prakoonwit (2023) document how graph-computing and graph-neural-network techniques now underpin institutional AML detection, tracing fund flows and flagging suspicious network structures. The significance lies not in any single performance claim, but in the kind of analytical work it represents: fund-flow tracing over a defined transaction graph is, in principle, reproducible and traceable, in contrast to the probabilistic, low-traceability inference characteristic of LLM-mediated judgement examined elsewhere in this paper.

The question of decision-tier effectiveness — which compliance decisions AI most effectively supports — has received almost no systematic empirical treatment. RegTech's dominant vocabulary (cost reduction, processing-time reduction, false-positive-rate reduction) shows that an AI system does more, faster, but not what kind of decision it does more of, or how much judgement it has absorbed. A system automating away easy cases looks identical, on these metrics, to one automating away genuinely judgement-laden escalation decisions — yet the accountability implications differ entirely.

2.3.3 The Governance Gap in the RegTech Literature

The governance dimension of the RegTech literature has developed significantly since 2018, driven by regulatory interest in the accountability implications of automated compliance systems — an argument developed for the pre-LLM generation of compliance AI, but whose underlying logic is directly applicable to the LLM context this paper examines.

The economics of RegTech adoption suggest technology-driven compliance can significantly reduce costs while improving regulatory outcomes — but only if governance frameworks ensure cost reduction does not come at the expense of accountability, a tension anticipating the compliance governance gap without naming it. Finch and Butt (2025), discussed further in Section 2.5, identify a related pattern of regulatory fragmentation across jurisdictions developing AI compliance governance independently, creating arbitrage opportunities — a finding that directly supports this paper's argument that ISO 42001 alone, a management-system standard without regulatory force, cannot close the compliance governance gap.

What none of these contributions provides, however, is a framework that maps governance requirements to the specific characteristics of AI compliance decisions at different tiers of the compliance hierarchy. The governance literature treats AI compliance systems as a uniform

category — overlooking the qualitatively different governance questions raised by near-autonomous operational decisions compared to augmentative strategic advice. This is the first literature gap this review identifies: a decision-tier-calibrated governance framework, absent from the RegTech literature, that Section 5's holistic analysis begins to supply.

2.4 The Large Language Model Literature: Capabilities, Limitations, and Governance Implications

2.4.1 The Development of LLMs: A Technical Overview

The technical literature on Large Language Models has grown from a specialised area of natural language processing into one of the most active fields in computer science, driven by the impact of models like ChatGPT, GPT-4, and Claude on professional practice. The most significant contribution for this paper is Zhao et al.'s (2023) comprehensive survey, providing an empirically grounded, systematically organised account of LLM capabilities, architectures, training methodologies, and limitations.

The core technical claim of the LLM literature most relevant to this paper is emergent abilities — capabilities that appear at scale thresholds, not through gradual improvement. Zhao et al. (2023) document this comprehensively, showing that capabilities including multi-step reasoning, mathematical problem-solving, and instruction following are effectively absent in smaller models and appear unpredictably in larger ones, providing the empirical foundation for this paper's claim that AI compliance productivity is threshold-driven.

The compliance relevance of specific LLM capabilities deserves systematic treatment. **Table-**2 adopts the peer-reviewed evaluation taxonomy of Chang et al. (2024), whose survey organises the field around what to evaluate, where to evaluate, and how to evaluate large language models.

Table-2: LLM capabilities mapped to compliance applications
(technical definitions after Chang et al., 2024)

LLM Capability	Technical Definition (Chang et al., 2024)	Compliance Application
In-context learning	Ability to apply reasoning to new tasks from examples provided in a prompt, without retraining or parameter updates	Interpret novel regulatory scenarios not covered in training data — engage with new guidance as it is issued without system retraining
Chain-of-thought reasoning	Ability to work through multi-step problems systematically, making intermediate reasoning steps visible in the output	Produce auditable, step-by-step compliance decision rationales that function as the AI equivalent of documented professional reasoning
Instruction following	Ability to accurately interpret and execute complex natural-language instructions	Apply regulatory protocols, internal compliance procedures, and supervisory guidance precisely and consistently

LLM Capability	Technical Definition (Chang et al., 2024)	Compliance Application
Tool manipulation	Ability to invoke external tools — databases, calculators, search engines — to supplement reasoning	Access real-time sanctions lists, regulatory databases, and transaction-monitoring systems during compliance analysis
RLHF alignment	Reinforcement Learning from Human Feedback: training process encoding human preferences into model behaviour	Encode expert compliance-officer judgement into model outputs through iterative feedback — the mechanism of near-human decision-making

The low-explainability character of LLM-mediated inference reflects how such systems are aligned. Contemporary LLMs are post-trained through reinforcement learning from human feedback (RLHF): a reward model, trained on aggregated human preference comparisons, optimises outputs toward responses humans would rate favourably — now standard across the frontier-model literature (Zhao et al., 2023). The resulting behavioural dispositions — refusal, hedging, deference — are encoded implicitly in model parameters rather than in inspectable rules. For a deploying institution this has a governance consequence: the system's behavioural policy was fixed upstream, by a vendor's annotator pool and reward-modelling choices the institution neither specified nor can reconstruct. Where a deterministic system permits verification by direct inspection, an RLHF-aligned model admits only behavioural evaluation — testing what it does, without access to why. Reward-hacking and sycophantic agreement — endorsing a user's position because agreement was rewarded in training — are precisely the dispositions a second-line compliance function exists to resist. RLHF thus supplies the mechanistic basis for this paper's tier distinction: the probabilistic tier is constituted by a training process whose outputs are, by design, not rule-traceable.

LLMs exhibit a broad capability profile spanning natural-language understanding, generation, reasoning, and knowledge-intensive tasks, increasingly assessed for robustness, trustworthiness, and domain-specific competence (Chang et al., 2024). Survey evidence on LLM evaluation nonetheless identifies capability limitations directly relevant to compliance: performance varies across tasks and domains, outputs can be fluent yet factually unreliable, and evaluation itself remains an open methodological problem — there is no settled protocol by which a deploying institution can certify what an LLM can and cannot reliably do.

2.4.2 Human-Level Performance: The Benchmark Evidence

The empirical performance data documented by Zhao et al. (2024) is the most important quantitative evidence for this paper's near-human decision-making claim. On the **MMLU benchmark —** a 57-domain test spanning law, finance, medicine, ethics, and professional practice — GPT-4 achieves 86.4% accuracy in the five-shot setting, exceeding average human-professional performance; the legal and professional-law subsets indicate a level of regulatory knowledge approaching expert professional standards.

On BIG-bench — a comprehensive evaluation of complex reasoning capabilities — LLMs at

scale outperform average human performance on 65% of tasks in the few-shot setting, with the advantage greatest on complex reasoning — precisely the multi-step analytical reasoning compliance decisions require. This is the empirical basis for the near-human decision-making claim at the operational compliance tier. On mathematical reasoning benchmarks, directly relevant to quantitative compliance analysis and risk modelling, ChatGPT achieves 78.47% on GSM8k — below human expert performance, but sufficient for the structured risk-quantification tasks compliance investment decisions require.

> *"GPT-4 achieves a remarkable record (86.4% in 5-shot setting) in MMLU, which is significantly better than the previous state-of-the-art models and shows surprising performance in understanding and knowledge mastery."* **— Zhao et al. (2023, p. 63)**

2.4.3 LLM Limitations: The Academic Evidence

The LLM literature also documents structural limitations that define an augmented-judgement ceiling. Three categories matter most for compliance, all documented by Zhao et al. (2023). Hallucination — factually incorrect output stated with high confidence — is acute where training-data coverage is sparse or regulatory frameworks conflict across jurisdictions. Reasoning inconsistency — a correct answer reached through invalid reasoning, or an incorrect one following valid steps — is most dangerous operationally, where a recommendation may pass human review because its conclusion is correct while the reasoning error goes undetected. Knowledge-currency limitations — inability to access regulatory developments after training-data cutoff — mean LLM-based systems require continuous updating rather than one-time deployment, a requirement none of the standards in Section 2.5 specifies.

The relevance of these documented capabilities and limitations to compliance is relational, not intrinsic: it is determined by the function against which a given capability or limitation is deployed. A generative summarisation capacity adequate for internal triage may prove wholly inadequate for a regulatory filing that carries legal consequence. This paper's synthesis in Section 5 accordingly treats risk-tiering as the governing principle: the explanatory stringency an AI-assisted decision must satisfy should scale jointly with the decision's risk tier and the severity of the specific limitation implicated.

2.5 ISO 42001:2023 and AI Governance Standards: The State of the Art and Its Limits

2.5.1 ISO 42001:2023 — What It Provides

ISO 42001:2023 — Artificial Intelligence Management System — is the first international standard specifically addressing the governance of AI systems within organisations. Published in December 2023, it establishes requirements for an AI management system (AIMS): a set of organisational processes, roles, controls, and documentation requirements designed to ensure that AI is developed, deployed, and used responsibly.

The standard is structured around four requirements clusters — context and scope; planning (risk assessment, AI policy, objectives); support (resources, competence, documentation); and operation (lifecycle management, impact assessment, supplier controls) — plus performance evaluation and continuous improvement, drawing on the ISO management-system framework familiar from ISO 9001 and ISO 27001.

For financial services compliance, ISO 42001 provides genuine value in four areas: it establishes a documented AI risk-assessment framework that regulators can reference and audit; it requires organisational accountability structures defining who owns, oversees, and approves AI deployment; it mandates impact-assessment processes before deployment; and it addresses supplier management, requiring institutions to assess vendors' AI-governance practices.

2.5.2 What ISO 42001 Does Not Provide — The Four Governance Gaps

Despite these genuine contributions, ISO 42001 has four specific limitations that leave the compliance governance gap substantially open — not criticisms of the standard's quality, but a reflection of the necessarily general scope of an international management-system standard relative to the pace of AI capability development.

Recent systematic-review evidence substantiates this assessment. Finch and Butt (2025), examining AI governance across the EU AI Act, ISO/IEC 42001, the NIST AI Risk Management Framework, the OECD AI Principles, and the EU's ALTAI, identify four persistent gaps: enforceability, proportionality, auditability, and fragmented accountability. They further identify friction between the EU AI Act and GDPR as compounding these gaps, since the two instruments impose overlapping but not fully reconciled obligations on the same data-processing activities.

Table-3: The 4 governance gaps in ISO 42001 as applied to AI-assisted compliance decisions

ISO 42001 Limitation	Analysis of the Gap
Does not map accountability to decision tier	ISO 42001 establishes organisational accountability for AI systems at the management-system level — roles, responsibilities, governance structures. It does not differentiate accountability requirements based on the nature or consequences of specific AI decisions. A near-autonomous operational compliance decision and an augmentative strategic-analysis input are treated identically within the management-

ISO 42001 Limitation	Analysis of the Gap
	system framework, despite requiring fundamentally different accountability architectures.
Does not address the liability vacuum	The standard requires institutions to assess and document AI risks, but does not specify how liability is allocated when those risks materialise as AI-driven decisions that prove wrong. When an ISO 42001-compliant AI system makes an incorrect suspicious-activity-report recommendation, the standard provides no mechanism for determining whether liability rests with the model developer, the deploying institution, the reviewing compliance officer, or some combination.
Is anchored in current AI capabilities	ISO 42001 was drafted primarily in the context of AI capabilities available in 2021–2023. The emergent-ability dynamic documented by Wei et al. (2022) and Zhao et al. (2024) — by which LLM capabilities change qualitatively at scale thresholds — means a management-system standard written for current capabilities may be inadequate for the AI systems that will be deployed by the time it is widely implemented.
Does not engage posthumanist governance challenges	The standard assumes a clear principal–agent relationship in which human decision-makers use AI tools. It has no framework for the human–AI assemblage that RLHF-aligned LLMs produce — systems that embody and reproduce human professional judgement in ways that make the simple principal–agent model inadequate for accountability purposes.

Finch and Butt's analysis is particularly significant here because it is oriented specifically toward low-capacity actors — SMEs and public authorities lacking dedicated compliance capacity — for whom the four gaps above are operationally binding, not abstract. Their concept of compliance asymmetry names a condition directly relevant to this paper's argument: instruments calibrated to well-resourced institutions produce obligations that smaller or less-resourced actors can state but not operationally satisfy.

2.5.2 A Comparative View of the Instrument Landscape

The instrument-by-instrument discussion above, and the cross-instrument gaps Finch and Butt (2025) identify, can be consolidated into a single comparative view. **Table-**4 arrays the five governance instruments most relevant to AI compliance in financial services against eight governance dimensions chosen to expose the deficiencies this paper addresses.

Table-4: Comparative matrix of five AI-governance instruments against eight governance dimensions relevant to AI-assisted compliance decisions

Governance dimension	ISO/IEC 42001:2023	EU AI Act	FCA (UK)	MAS FEAT (SG)	HKMA (HK)
Legal character	Voluntary certifiable standard	Binding law, with penalties	Binding principles-based regulation	Non-binding supervisory principles	Supervisory guidance and circulars
Governance approach	Process/management-system oriented (AIMS)	Risk-tiered by application	Outcome- and principles-focused	Fairness, Ethics, Accountability, Transparency	Institution-level expectations
Risk / decision-tier calibration	None — uniform at system level	Tiered by application, not decision	None at decision level	None — applies across the board	None — institution level only
Explainability requirement	Transparency process required; no calibrated standard	Duties for high-risk uses	Implied via Consumer Duty logic	Explicit F-A-E-T principle	Proportionate to risk (qualitative)
Accountability mechanism	Organisational roles; fragmented	Provider/deployer duties; distributed	SM&CR assigns individual accountability	Accountability principle; no allocation	Management responsibility; no AI-specific rule
Liability vacuum	Not addressed	Partial — no duty for erroneous AI output	Addressed for humans, not AI-authored decisions	Not addressed	Not addressed
Human–AI assemblage / RLHF	Assumes principal–agent tool use	Assumes oversight of a discrete system	Assumes accountable human decision-maker	Assumes human-in-the-loop only	Assumes human oversight only
Relevance to this paper	Management-system architecture	Risk-tiering logic	Organisational accountability chain	Explainability-tier vocabulary	Local regulatory context

Two rows are decisive here. The risk/decision-tier calibration row shows that while the EU AI Act tiers obligations by application category, no instrument calibrates governance to the

individual compliance decision. The human-AI-assemblage row shows every instrument presupposes a principal-agent model of discrete human oversight, none engaging the entangled configuration RLHF-aligned systems produce — the condition Section 3's posthumanist analysis takes as its premise.

2.5.3 Other AI Governance Frameworks — The Regulatory Landscape

Beyond ISO 42001, the AI governance landscape includes a range of regulatory and quasi-regulatory frameworks directly relevant to financial services compliance. **Table-**5 maps these frameworks against the compliance governance gap.

Table-5: Five additional AI-governance frameworks assessed against the compliance governance gap

Framework	Assessment Against the Compliance Governance Gap
EU AI Act (2024)	The most comprehensive binding AI governance regulation currently in force. Risk-based approach classifying AI systems by risk level. High-risk AI in financial services subject to mandatory human oversight, transparency, and conformity assessment. Gap: risk classification is by AI system type, not compliance decision tier; does not map accountability to strategic vs. operational vs. investment decisions.
FCA AI Principles (UK)	Principles-based guidance on AI fairness, explainability, and accountability. Non-binding but influential. Gap: no binding tier-specific rules; explainability guidance applies uniformly regardless of stakes, reversibility, or accountability complexity.
SEC AI Guidance (US)	Staff bulletins and no-action letters addressing AI use in investment advice, compliance functions, and regulatory reporting. Gap: advisory and function-specific; no comprehensive compliance governance framework; liability allocation not addressed.
MAS FEAT Criteria (Singapore)	Fairness, Ethics, Accountability, Transparency criteria — the most advanced Asian framework, with specific guidance on explainability and human oversight. Gap: criteria are aspirational rather than prescriptive; no binding accountability-allocation mechanism.
FATF Guidance on AI in AML (2023)	Guidance on AI and data analytics in AML/CFT compliance. Acknowledges AI benefits for monitoring and risk assessment. Gap: focuses on whether AI outputs are reliable, not on who is accountable when they prove unreliable.

Regulator-issued evidence from Asia-Pacific supervisors extends this landscape beyond the predominantly Western frameworks reviewed above. The Hong Kong Monetary Authority's sector-wide study of generative AI adoption — a survey of 137 practitioners and interviews with 16 organisations across banking, securities, insurance, and technology — documents a pronounced asymmetry: of 59 reported use cases, 47 were internal-facing, only 4 external-facing

(Hong Kong Monetary Authority, 2024). Institutions attributed this to regulatory and liability uncertainty rather than technical infeasibility, since system-generated suggestions could constitute regulated activities such as investment advice — direct practitioner evidence that operational capability has outrun the governance frameworks institutions are prepared to expose to customers.

One further finding from the same study bears on this paper's arguments: surveyed institutions ranked extensive resource requirements (66%) and compliance/risk concerns (58%) above knowledge gaps (49%) as adoption hurdles, locating the binding constraint at the cost-compliance frontier rather than in capability awareness.

The regulatory instruments surveyed in this section each constitute a substantial contribution to AI governance, yet none is adequate to the conditions under which AI now operates within compliance functions: none calibrates its requirements to the risk tier of the decision at issue, none resolves liability where a decision is materially authored by an AI system, and none offers a governance architecture commensurate with the entangled human-AI configurations RLHF has made the operational norm. Sections 3 and 5 address the second and third deficiencies directly.

2.6 AI Ethics, Accountability, and the Liability Question

2.6.1 The AI Ethics Literature

The AI ethics literature has developed rapidly since 2016, driven by documented failures of AI systems in high-stakes domains — predictive policing, facial recognition, credit scoring, and medical diagnosis — that raised urgent questions about accountability, bias, fairness, and transparency. These concerns have converged on a small set of recurring principles — beneficence, fairness, transparency, and accountability chief among them — that have gone on to influence subsequent regulatory frameworks including the EU AI Act and the OECD AI Principles.

For compliance, the most directly relevant AI ethics literature concerns accountability and explainability. Arrieta et al. (2020), in their widely cited taxonomy, distinguish explanations that support a user's decision-making, explanations that satisfy a human reviewer, and explanations that faithfully reflect a model's actual computational process — directly applicable to this paper's risk-tiering principle: different compliance decision tiers require different types of explainability, not simply more or less of the same type. Chen et al. (2023) document the growing bibliometric footprint of this explainability research specifically within finance, confirming it as an active but still-maturing field.

Mittelstadt et al. (2016) provide the most comprehensive academic treatment of the ethics of algorithms, identifying six clusters of concern: inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, transformative effects, and traceability. Their analysis of traceability — identifying who is responsible for algorithmic decisions in distributed development and deployment chains — directly anticipates the liability vacuum this paper documents, though their solutions do not extend to RLHF-aligned LLMs in regulated compliance environments.

2.6.2 The Legal Liability Literature

The legal literature on AI liability is growing rapidly but remains contested in its conclusions. The fundamental debate — whether existing legal frameworks can accommodate AI-driven decisions or whether new legislative frameworks are required — is directly relevant to this paper's liability-vacuum argument.

Delvaux (2017) and Vladeck (2014) represent the foundational academic contributions to this debate. Delvaux argues that existing legal categories — product liability, negligence, contract — are insufficient for AI systems that learn from feedback and develop capabilities their designers did not anticipate. Vladeck's earlier analysis reaches a similar conclusion: the unpredictability of AI system behaviour over time creates accountability gaps that existing tort law cannot bridge without significant doctrinal development.

A related strand of the legal literature argues that the principal-agent framework breaks down when the AI system operates with a degree of autonomy that makes the supervising professional's responsibility genuinely uncertain. The Financial Stability Board's (2017) report on AI and machine learning in financial services provides an authoritative regulatory analysis of AI liability, acknowledging that existing frameworks are insufficient but stopping short of proposing binding solutions — an acknowledgement without resolution that is itself significant evidence for the compliance governance gap this paper identifies.

The AI ethics and legal-liability literature has articulated the accountability problem clearly, establishing that autonomous and semi-autonomous systems disrupt conventional structures of responsibility. What it has not produced is a governance architecture adequate to regulated compliance environments, where AI is already making near-autonomous operational decisions: the diagnosis is mature, the constructive response is not. If the liability vacuum arises because conventional accountability presupposes a singular, identifiable human decision-maker, the remedy cannot simply reinstate that presupposition. Responsibility is better reconstituted as a property of an auditable organisational chain, in which each contribution — human and non-human alike — is traceable at the appropriate juncture, a position this paper grounds theoretically in Sections 3 and 5.

Figure-3 offers a supplementary, illustrative view of the same corpus: a word-tree text-search query on the root term "supervision" across all coded sources. It is a purely lexical, collocational visualisation rather than an analytical finding — no thematic claim in this paper rests on it — but it shows the vocabulary in which the corpus's discussion of oversight is embedded: human oversight, regulation and oversight used near-interchangeably, market monitoring.

2.7 Research Gap

From the above Literature Review, there is little research conducted to investigate the gap between AI to Posthumanism via ISO 42001 and AI Governance Standards. This study is therefore designed to fill the gap. As a result, the Research Question for this paper is:

From AI to Posthumanism via ISO 42001 and AI Governance Standards

In order to do so, the powerful qualitative tool **NVivo** was designed to measure the effective implementation of from AI to Posthumanism via ISO 42001 based on the above literature review and adaptation from related measurement scales.

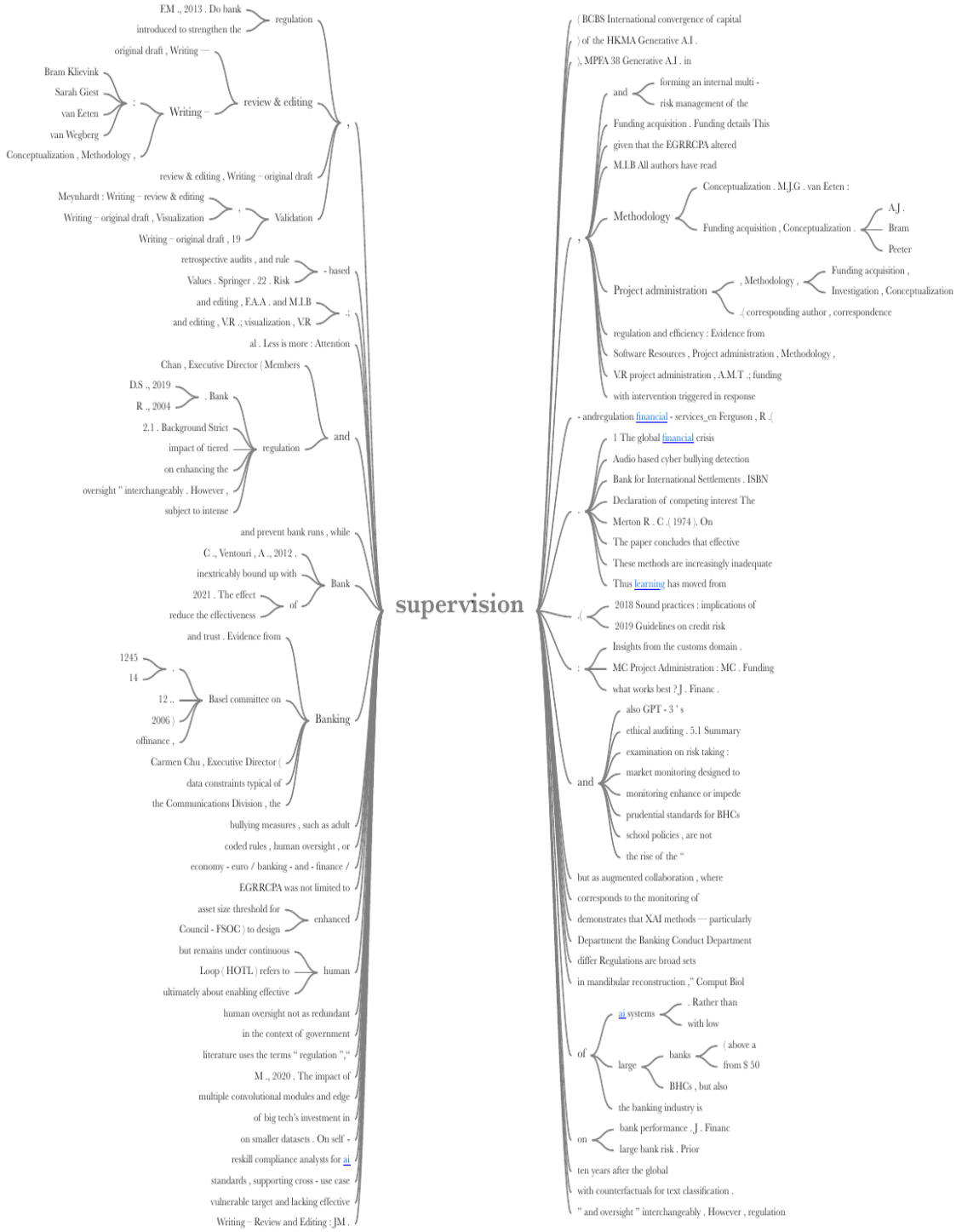

Figure-3: Word tree for the root term "supervision" (NVivo text-search query, full corpus) *Illustrative only; independent of the seven-Theme coding distribution in Table-1*

3. POSTHUMANISM THEORY & AI GOVERNANCE

Posthumanism is treated here as a distinct contribution rather than an eighth coded **Theme**, because its relevance to AI governance is not a matter of thematic coverage within the existing corpus but a challenge to the conceptual frame the other seven **Theme**s share.

3.1 The Posthumanist Tradition

Posthumanism is a theoretical framework developed primarily within the humanities and social sciences, largely independently of the AI-governance discourse. Its principal theorists — Donna Haraway, N. Katherine Hayles, Rosi Braidotti, and Karen Barad — have produced foundational work on the relationship between human agency and technological systems that is directly applicable to the challenges of AI governance, yet has been largely ignored in the governance literature.

Haraway's foundational Cyborg Manifesto (1985) introduced the cyborg — a human–machine hybrid — as a figure for understanding the dissolution of the boundary between human and technological agency. For contemporary AI compliance systems, the cyborg is not a metaphor but an architectural reality: RLHF-aligned LLMs encode human professional judgement into non-human systems that then reproduce and extend that judgement at scale, dissolving the boundary between compliance officer and AI system at the level of training process, not merely workflow integration.

Hayles' How We Became Posthuman (1999) develops cognitive extension — the idea that human cognition does not end at the brain's boundaries but extends through the technological systems it works with. In compliance practice, a practitioner's professional knowledge is extended and altered by an AI system's gap-identification capability, producing insights the human expert alone could not generate: the human's expertise does not simply use the AI, it is extended through it.

Braidotti's The Posthuman (2013) develops the posthuman subject — an entity whose identity, capabilities, and moral standing are continuously constituted through technological entanglement rather than pre-existing it. For compliance governance, the posthuman professional is not a human who uses AI tools but an entity whose professional identity and judgement are continuously co-constituted through engagement with AI systems — with direct implications for the competency frameworks of ACAMS, ICA, and CISI, which remain anchored in humanist assumptions that the practitioner is a bounded, autonomous agent.

Barad's Meeting the Universe Halfway (2007) introduces intra-action — the idea that agency does not pre-exist the relations between actors but emerges from them. Applied to AI compliance systems, intra-action explains the emergent-ability dynamic: LLMs develop qualitatively new capabilities through the intra-action of scale, training data, architecture, and deployment context. Governance frameworks assuming stable, pre-defined AI capabilities are fundamentally inadequate for systems whose capabilities emerge this way.

3.2 Posthumanism and Governance — The Missing Connection

Despite the analytical power of posthumanist theory for understanding AI-governance challenges, the literature at the intersection of posthumanism and AI governance is sparse to the point of representing a significant academic gap. The AI-governance literature — including ISO 42001, the EU AI Act, and academic governance scholarship — operates within a firmly humanist framework: it assumes a principal–agent structure in which humans are always the decision-making principal and AI systems are always the instrumental agent.
A small number of scholars have begun to connect posthumanist theory and AI governance. Coeckelbergh (2020) argues posthumanist perspectives are essential for understanding AI's impact on political agency, without extending to regulated compliance specifically. Hildebrandt (2020) engages posthumanist **Theme**s in analysing law and machine learning, arguing algorithmic decision-making challenges legal concepts of personhood and agency — again without this paper's compliance-specific application. The connection between RLHF as a training process and posthumanist theories of cognitive extension, cyborg constitution, and intra-action does not appear elsewhere in the literature; it is this review's original theoretical contribution.

No coded material was returned, in the **NVivo** review underlying **Table-**1, at the intersection of posthumanist theory and compliance governance — a gap this paper's dedicated treatment of posthumanism addresses rather than obscures. Treating it separately matters because it changes the unit of analysis: Section 2's seven **Theme**s all ask what AI can do and how well existing instruments govern it, holding human and AI apart as distinct entities. Posthumanism asks whether that separation is itself the problem — if human judgement and AI output are constitutively entangled rather than merely juxtaposed, then 'is the AI trustworthy' is not a harder version of the right question but a mistaken one.

It follows that the four theorists are not offering interchangeable versions of one objection but a layered analytic. Haraway's cyborg names the phenomenon — human and machine already fused in practice, whatever the governing instrument assumes. Hayles' cognitive extension supplies the mechanism: professional cognition is partly constituted through the AI system, so removing it does not return the officer to a purely human baseline but removes a component of the judgement itself. Braidotti's posthuman subject draws out the institutional consequence — competency frameworks certifying individual human judgement may be certifying a unit that no longer exists in the form the certification presupposes. Barad's agential cut supplies the diagnostic tool: any governance instrument draws a line between 'the AI' and 'the human' — and a posthumanist reading asks where that line was drawn, and what it thereby placed outside the frame of what it governs.

4. RESEARCH APPROACH & ANALYSIS

This paper's research approach is stated briefly, consistent with its focus on the literature synthesis that follows. The review adopts an interpretive, social-constructionist qualitative design, appropriate to early-stage theoretical development rather than exhaustive systematic synthesis. A purposively sampled corpus — peer-reviewed academic sources, applied

competency literature, and regulatory-institutional reports, drawn from an audited pool of fifty candidate papers screened for relevance, recency, and domain fit — was coded in **NVivo** against the seven-**Theme** node structure in **Table-**1, using a two-pass, hybrid inductive-deductive process in which generative AI tools surfaced candidate codes only, with every code's verification remaining a manual, theory-informed act. Rigour was addressed through Lincoln and Guba's (1985) credibility, transferability, dependability, and confirmability criteria; acknowledged limitations are a non-exhaustive, English-language-skewed corpus and single-researcher coding without inter-rater testing.

4.1 Holistic Analysis: The Seven Themes, Posthumanism, and AI-Assisted Decision-Making

4.1.1 The Convergent Pattern across the Seven Themes

Read individually, each of the seven **Theme**s reviewed in Section 2 identifies a bounded technical or regulatory gap: governance instruments that do not calibrate to decision tier (Section 2.5); model architectures that trade explainability for performance (Section 2.4); automation that delivers efficiency without addressing accountability (Section 2.3); a generative-AI literature too thin to have caught up with deployment (Section 2.2.3); and an ethics/liability literature that diagnoses the accountability problem without resolving it (Section 2.6). Read together, however, the seven **Theme**s converge on a single structural pattern relevant to compliance decision-making specifically: at every stage of the AI-in-compliance lifecycle — detection, analysis, recommendation, and review — capability has moved faster than the accountability architecture built to govern it, and the resulting gap is largest exactly where the decision carries the most regulatory consequence.

This pattern has a direct bearing on how AI actually influences decision-making in a compliance function. The governance, RegTech, and regulatory-supervision literatures together show that AI's influence on decision-making is currently strongest at the point of detection and triage — flagging, scoring, and routing — where efficiency gains are best documented and where accountability instruments are, not coincidentally, least developed. The model-development, generative-AI, and risk-detection literatures show that AI's influence on the analytical content of a decision — the reasoning offered in support of a recommendation — is growing quickly but remains bounded by hallucination, reasoning inconsistency, and knowledge-currency limitations that make the AI's analytical output a contribution to, rather than a substitute for, professional judgement. The RPA literature's more settled automation evidence supplies the useful contrast case: where a process is genuinely rule-bound and low-consequence, AI's influence on decision-making can approach full delegation without governance strain, which is precisely why the generative-AI literature's thinness is consequential — it is thin exactly where delegation is least safe to extend by analogy from the RPA case.

4.2 The Posthumanist Reading and Its Design Implications

Posthumanism changes what this pattern means rather than adding a further data point to it. If the seven **Theme**s are read within the humanist frame they mostly share, the appropriate response to "capability has outrun accountability" is to build better accountability instruments for a fundamentally separate AI agent — more granular risk tiers, better explainability requirements,

clearer liability rules — which is, in fact, the direction the governance and regulatory-supervision literatures are already moving. Read through Haraway's, Hayles', Braidotti's, and Barad's posthumanist lens instead, the same pattern indicates something more specific: that the accountability gap is not a temporary lag to be closed by a better-specified instrument, but a structural consequence of governing an entangled human–AI configuration as though it were a separable human user operating a discrete tool. On this reading, the seven **Theme**s' shared pattern and the posthumanist critique are not two independent findings but one finding seen from two angles — the technical literature shows where accountability has fallen behind capability, and the posthumanist literature explains why an instrument built on the wrong ontology of the decision-maker will keep falling behind no matter how it is refined.

For AI-assisted decision-making specifically, the practical upshot is that the seven **Theme**s' technical gaps and the posthumanist ontological critique point toward the same design requirement: governance that attaches to the specific decision and the specific human–AI configuration producing it, rather than to the AI system considered in isolation or to a generic human-oversight requirement considered in isolation. Explanatory obligation should scale with the decision's consequence and with the specific limitation implicated (Sections 2.4–2.5); the accountability chain should be organisational and auditable rather than resting on an imputed agency of the AI system (Sections 2.3, 2.6); and the professional competencies required to supervise AI-assisted decisions should be defined for the entangled configuration that actually exists (Section 2.2.3, and the posthumanist critique of professional-standards frameworks in Section 3) rather than for a bounded human user that, in the settings this paper reviews, no longer accurately describes the decision-maker.

5. CONCLUSION

This paper's holistic reading of the seven coded **Themes** based on the NVivo coding and referencing of around 50 SJR-Q1 Papers published after 2022 when the AI-ChapGPT was born, together with its separate treatment of posthumanist theory, supports a conclusion that is neither an uncritical endorsement of AI-driven compliance nor a rejection of it — and that is best summarised, as this paper's title proposes, as a movement from tool to decision-maker. The literature reviewed in Section 2 establishes, with reasonable consistency across **Theme**s, that AI increases effectiveness and efficiency in compliance functions — faster and more consistent detection, measurable reductions in false positives, analytical support that approaches or exceeds average human performance on defined benchmark tasks. Within that increased effectiveness, AI functions as a genuine decision-maker: it classifies, scores, flags, and recommends, and it does so increasingly without a human directly re-deriving each output from first principles. But it is a decision-maker operating within a risk envelope that humans, not the AI system, have prescribed — the thresholds, typologies, risk appetite, and escalation criteria against which the AI's outputs are generated and judged are themselves human specifications, encoded into the system rather than discovered by it. The ultimate decision matrix, in this precise sense, remains human even where the moment-to-moment classification does not.

Posthumanist theory, read as this paper's Section 3 develops it, does not weaken this conclusion; it explains why it is the only conclusion available. Because humans already use AI as an extension of their own judgement — because, in Hayles' and Barad's terms, professional cognition and AI output are cognitively and relationally entangled rather than merely adjacent — human and AI decision-making are not, in the settings this paper reviews, two separable processes that could in principle be cleanly divided between "what the AI decides" and "what the human decides." That separation is precisely the humanist assumption Section 3 argues the governance literature should give up. What follows from giving it up is not that AI's contribution disappears into the humans', or the reverse, but that the meaningful locus of ultimate decision-making shifts to the one place the entanglement does not reach: the rules. It is the rule — the threshold, the typology, the risk tier, the escalation criterion — that a human sets and can be held to account for, and it is the rule, not any single classification the AI produces under it, that constitutes the decision a human can meaningfully be said to have made. AI and human are, on this account, inseparable in operation but not equivalent in accountability: the AI executes and analyses within a structure only a human is positioned to set and answer for.

The practical implication for AI governance in regulated finance is accordingly that effectiveness and accountability are not competing objectives to be traded off against one another, but that they are secured by different means at different points of the same entangled process. Effectiveness is secured by allowing the AI to do what the seven reviewed **Theme**s show it now does well — detect, score, triage, draft, and recommend, at a volume and consistency no equivalent human process can match. Accountability is secured not by trying to prise the AI's contribution back out of the human's, which the posthumanist analysis in Section 3 suggests cannot coherently be done, but by ensuring that the rules governing the AI's operation are themselves explicit, human-authored, auditable, and revisable, and that responsibility for those rules — rather than for each individual AI-generated output — is where governance attaches. Human and AI decision-making, in short, must go hand in hand: not as a co-operative slogan, but as the specific, structural

consequence of a governance architecture that accepts the entanglement posthumanist theory identifies and locates accountability at the one point, the rule, where a human author remains identifiable. AI has become, in every **Theme** this paper reviews, a decision-maker; what posthumanist theory shows is that it never stopped being, in the same movement, a tool — and that governance must be built for both truths at once.

A further implication, beyond the immediate governance question, deserves acknowledgment even though it sits outside this paper's primary scope. Some of the uncertainty documented above has a monetary dimension that the compliance-governance literature has not yet engaged. If AI materially increases the efficiency and effectiveness of compliance and adjacent back-office functions — as the productivity literature reviewed in Section 2.3.2 documents — the practical consequence is less, or cheaper, human labour performing the same function. That shift has second-order fiscal implications: a shrinking base of taxable labour income in the functions AI displaces. This sits uneasily alongside the observation that much of the current AI industry's valuation rests on projected future revenue rather than realised present profit. Should that projected profitability not materialise at the scale currently priced in, the compound effect — lower realised corporate earnings, a smaller taxable labour base in AI-displaced functions, and consequently lower measured GDP contribution than current models assume — could leave the broader economic and regulatory architecture that governs institutions like the ones reviewed in this paper looking substantially different from its current form. This is offered as a direction for future research rather than a claim this paper substantiates, but it is, in my view, the wider stake behind the narrower governance gaps this paper documents.

References

Alarab, I., & Prakoonwit, S. (2023). Graph-based LSTM for anti-money laundering: Experimenting temporal graph convolutional network with Bitcoin data. Neural Processing Letters. https://doi.org/10.1007/s11063-022-10904-8

Arrieta, A. B., Diaz-Rodriguez, N., Del Ser, J., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. Information Fusion. https://doi.org/10.1016/j.inffus.2019.12.012

Barad, K. (2007). Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning. Duke University Press. https://doi.org/10.1215/9780822388128

Braidotti, R. (2013). The Posthuman. Polity Press. https://rosibraidotti.com/publications/the-posthuman-2/

Chang, Y., Wang, X., Wang, J., et al. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), Article 39. https://doi.org/10.1145/3641289

Chen, X.-Q., Ma, C.-Q., Ren, Y.-S., Lei, Y.-T., Huynh, N. Q. A., & Narayan, S. (2023). Explainable artificial intelligence in finance: A bibliometric review. Finance Research Letters, 56, 104145. https://doi.org/10.1016/j.frl.2023.104145

Coeckelbergh, M. (2020). AI Ethics. MIT Press. https://doi.org/10.7551/mitpress/12549.001.0001

Delvaux, M. (2017). Report with recommendations to the Commission on Civil Law Rules on Robotics. European Parliament. https://www.europarl.europa.eu/committees/en/report-with-recommendations-to-the-commi/product-details/20170202CDT01121

Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

Finch, W. W., & Butt, M. (2025). Gaps in AI-compliant complementary governance frameworks' suitability for low-capacity actors, and structural asymmetries in the compliance ecosystem — a systematic review. Journal of Cybersecurity and Privacy, 5(4), 101. https://doi.org/10.3390/jcp5040101

Haraway, D. (1985). A Cyborg Manifesto: Science, technology, and socialist-feminism in the late twentieth century. Socialist Review, 80, 65–108. https://doi.org/10.5749/minnesota/9780816650477.003.0001

Hayles, N. K. (1999). How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics. University of Chicago Press. https://doi.org/10.7208/chicago/9780226321394.001.0001

Hildebrandt, M. (2020). Law for Computer Scientists and Other Folk. Oxford University Press. https://doi.org/10.1007/978-3-031-24349-3_14

Hong Kong Monetary Authority. (2024). Generative Artificial Intelligence in the Financial Services Space. HKMA. https://brdr.hkma.gov.hk/eng/doc-ldg/docId/20241118-4-EN

International Organization for Standardization. (2023). ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system. https://www.iso.org/standard/42001

Kurshan, E., Shen, H., & Yu, H. (2020). Financial crime & fraud detection using graph computing: Application considerations & outlook. 2020 Second International Conference on Transdisciplinary AI (TransAI), IEEE. https://doi.org/10.1109/TransAI49837.2020.00029

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic Inquiry. SAGE. http://dx.doi.org/10.1016/0147-1767(85)90062-8

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection. Decision Support Systems. https://doi.org/10.1016/j.dss.2010.08.006

Syed, R., Suriadi, S., Adams, M., et al. (2020). Robotic Process Automation: Contemporary themes and challenges. Computers in Industry. https://doi.org/10.1016/j.compind.2019.103162

Vladeck, D. C. (2014). Machines without principals: Liability rules and artificial intelligence. Washington Law Review, 89, 117. https://digitalcommons.law.uw.edu/wlr/vol89/iss1/6/

Zhao, W. X., Zhou, K., Li, J., et al. (2024). A survey of large language models. arXiv:2303.18223 http://doi.org/10.48550/arXiv.2303.18223.