L&R
Lex et Ratio Ltd

Redefining Democracy for the Age of AI

AI Governance and the Fiduciary Turn in the Architecture of Knowledge

PETER KAHL

Lex et Ratio Ltd

Redefining Democracy for the Age of AI

AI Governance and the Fiduciary Turn in the Architecture of Knowledge

PETER KAHL

8 November 2025

v1 published in London by Lex et Ratio Ltd, 8 November 2025.

About the Publisher

Lex et Ratio Ltd provides research, advisory, and strategic consulting in governance reform, fiduciary accountability, and epistemic ethics, integrating legal analysis, institutional theory, and practical reform strategies across public, corporate, and academic institutions.

Abstract

This paper advances a constitutional re-foundation of democracy for the age of artificial intelligence. It argues that democratic legitimacy no longer rests on procedural participation or informational abundance but on **fiduciary–epistemic trust**—the moral architecture that sustains truthful, reciprocal knowing. Artificial intelligence challenges this foundation not merely through misinformation but through *algorithmic clientelism*: the systemic conversion of epistemic autonomy into managed dependence within opaque infrastructures of mediation.

Integrating fiduciary theory (Frankel 1983; Smith 2023; Gold & Miller 2014; Kahl 2025i), epistemic justice (Fricker 2007; Medina 2013), and hybrid-governance scholarship (Ofir & Levine-Schnur 2025; Lim & Lim 2025), the paper situates AI developers and institutions as epistemic fiduciaries whose legitimacy depends on duties of candour, care, impartiality, and accessibility. It introduces a unifying framework—*Fiduciary Epistemic Governance* (FEG)—that redefines AI accountability as a constitutional ethic of truth stewardship rather than risk management.

The analysis culminates in a set of institutional innovations: *Epistemic Fiduciary Entities* (EFEs), *Public Epistemic Trusts* (PETs), and *Epistemic Audits*, each designed to embed fiduciary openness and relational accountability within AI governance. These mechanisms instantiate what the paper terms the *Fiduciary Constitution of Democracy 2.0*—a democratic order re-founded on fiduciary reciprocity between human and artificial knowers, where truth is preserved not by control but by trust.

Beyond its policy applications, the paper contributes to fiduciary and political theory alike: it extends fiduciary law from property and management into epistemic governance and reconceives democracy itself as a *fiduciary-epistemic constitution of knowledge*—a living architecture of candour, care, and mutual responsibility.

Keywords

AI governance, AI law, fiduciary theory, fiduciary ethics, fiduciary epistemic governance, political theory, democratic theory, epistemic justice, algorithmic accountability, algorithmic clientelism, fiduciary openness, epistemic architecture of power, fiduciary public good, democratic trust, epistemic autonomy, fiduciary turn, knowledge governance

•

Table of Contents

1. Introduction: The Democratic Cost of Algorithmic Mediation 4

1.1 Background — Democracy as an epistemic system of trust 4

1.2 Problem — Algorithmic Systems and the Reconstitution of Democratic Trust 5

1.3 Beyond Misinformation — Fiduciary Breach as Democratic Betrayal 6

1.4 Thesis — Democracy 2.0 and the Fiduciary Turn 6

1.5 Method — Constructing the Fiduciary Architecture of Knowing 7

1.6 Significance — Toward a Fiduciary Constitution of Knowledge 8

2. The Fiduciary Architecture of Democracy 9

2.1 Democracy as a fiduciary structure of public reason 9

2.2 Fiduciary Relations as Moral–Epistemic Commitments 10

2.3 Public Institutions as Trustees of Epistemic Integrity 11

2.4 Fiduciary Duties Reframed as Epistemic Duties 11

2.5 Translating Moral into Epistemic Responsibility: Kahl's Epistemic Humility 12

3. Algorithmic Clientelism: How Trust Becomes Dependence 13

3.1 Algorithmic Clientelism — Trading Autonomy for Recognition 13

3.2 Mechanism — Recommendation Systems and the Cultivation of Cognitive Dependence 14

3.3 Cognitive Dissonance as Epistemic Event 15

3.4 Patron–Client Structures in Epistemic Economies 15

3.5 Epistemic Pathology — Eroded Dissonance Tolerance and Fiduciary Breakdown 16

3.6 From Citizen to Epistemic Client 17

4. The Third Enclosure Movement: Knowledge as a Public Good 18

4.1 Knowledge as a Non-Rivalrous, Non-Excludable Good 18

4.2 Digital Enclosure through Proprietary AI Models 19

4.3 The Third Enclosure Movement — Typology of Restricted Knowledge 19

4.4 Open Infrastructures and the Apertus Example 21

4.5 Knowledge as a Fiduciary Public Good 21

4.6 The Pedagogy of Fear: Safety Theatre and the Erosion of Fiduciary Trust 22

5. The Hybrid Governance Trap: Institutional Form and Epistemic Integrity 23

5.1 Hybrid AI Firms and the Problem of Accountability 24

5.2 Structural Incoherence — Dual Allegiance and the Erosion of Epistemic Trust 24

5.3 Organisational Forms — For-Profit, Nonprofit, and Hybrid Models 25

5.4 Fiduciary Breach — Shareholder Duty versus Knowledge Integrity 26

5.5 Fiduciary Openness as Remedy to Organisational Opacity 27

5.6 Legal Implication — Toward Epistemic Fiduciary Entities 28

6. Epistemic Agency and the Fiduciary Brain 29

6.1 AI and the Transformation of Epistemic Agency 30

6.2 The Fiduciary Architecture of Mind 30

6.3 Mutual Calibration — Epistemic Trust Loops Between Humans and Algorithms 31

6.4 Cognitive Dissonance as a Regulatory Signal of Epistemic Breach 32
6.5 Extending Fiduciary Duties to Cognitive Architectures 33
7. Toward a Fiduciary Epistemic Constitution of AI Governance 34
7.1 Transitional Synthesis — From Fiduciary Ethics to Governance Design 34
7.2 The Fiduciary Epistemic Governance (FEG) Prototype 35
7.3 Institutional Pathways of Fiduciary Epistemic Governance 37
7.4 Legislative Parallels — Toward Global Fiduciary Alignment 38
7.5 Cultural Foundations — Civic Epistemic Virtue and Dissonance Tolerance 39
8. Policy and Institutional Implications 39
8.1 Comparative Governance — From Corporate to Fiduciary-Epistemic Accountability 39
8.2 Regulatory Pluralism — Public, Private, and Commons-Based AI 40
8.3 Democratic Alignment — Embedding Fiduciary Values in Public Reason 41
8.4 Institutional Innovation — Legal Form for Epistemic Fiduciaries 42
8.5 Anticipated Resistance — Fiduciary Duties vs. Shareholder Primacy 43
8.6 Implementation Roadmap — From Principles to Practice 44
9. Conclusion: Reclaiming Democracy through Fiduciary Openness 45
9.1 Democracy as an Architecture of Epistemic Trust 45
9.2 Algorithmic Clientelism and the Erosion of Autonomy 45
9.3 Fiduciary Openness as the Path Forward 46
9.4 Rebuilding Democracy's Fiduciary Architecture 46

Bibliography 47

Author Metadata 52
Cite this work 52
Version History 52

1. Introduction: The Democratic Cost of Algorithmic Mediation

This opening chapter establishes the epistemic foundations of democracy as a fiduciary order of knowledge. It situates artificial intelligence within this moral architecture, tracing how the replacement of institutional trust by algorithmic mediation reconfigures democratic legitimacy. The argument proceeds from philosophical grounding to structural diagnosis, showing that the crisis of democracy in the age of AI is not informational but fiduciary: a failure of institutions to honour the ethical trust placed in them by the public.

1.1 Background — Democracy as an epistemic system of trust

Democracy, at its deepest level, is not merely a structure of representation or a mechanism for collective decision-making—it is a fiduciary epistemic system built upon reciprocal trust. It presupposes that citizens, institutions, and intermediaries share a fiduciary relationship grounded in truth, communicative integrity, and openness to correction. In this sense, democracy is not only political but epistemological: it governs how societies authorise knowledge, distribute credibility, and sustain belief through mutual recognition.

This dimension, often overshadowed by procedural and economic analyses, has long been recognised in the philosophical tradition. For Habermas (1996), the legitimacy of democratic law arises from intersubjective reasoning, where validity claims are tested within communicative rationality. The vitality of democracy thus depends not on unanimity but on the trustworthy circulation of reasons—on citizens' confidence that deliberation is conducted in good faith, informed by reliable knowledge, and responsive to critique.

As Berger and Luckmann (1966) observed in *The Social Construction of Reality*, social order depends upon the institutionalisation of shared meanings through reciprocal recognition. Democracy, in this light, is not merely political but epistemic: it is sustained by the mutual legitimation of knowledge claims within a socially constructed world. The fiduciary architecture of democracy thus extends Berger and Luckmann's insight—transforming the sociology of knowledge into a normative theory of epistemic trust.

Similarly, Anderson (2006) interprets democracy as an epistemic enterprise concerned with the quality of collective belief-formation. Decision-making, she argues, is justified not because it reflects majority will but because it optimises conditions for truth-tracking through diversity, dialogue, and accountability. Under this view, public institutions function as epistemic fiduciaries—custodians of the collective capacity to know well. Trust, in this context, is rational reliance on epistemic integrity, sustained by transparency, contestability, and relational candour.

As Kahl (2025l) demonstrates, every political system embodies an epistemic architecture of power—a configuration through which knowledge legitimates authority and authority structures knowledge. Democracy distinguishes itself only when that architecture remains fiduciary in character, grounded in transparency and reciprocal accountability.

These fiduciary conditions—reciprocity, transparency, and accountability—constitute the moral infrastructure of democratic cognition. When they are honoured, democracy functions as a self-correcting system of epistemic openness. When they are breached, trust devolves into dependence, dialogue gives way to manipulation, and democracy's cognitive architecture begins to corrode from within.

This paper argues that the rise of artificial intelligence marks a profound reorganisation of democracy's epistemic foundations. Contemporary AI systems—particularly those governed by hybrid, profit-driven models

—threaten to convert fiduciary trust into algorithmic clientelism, replacing relational reciprocity with computational dependence. To address this transformation, democracy must undergo what this paper calls the **Fiduciary Turn**: a paradigmatic redefinition of governance in which epistemic trust, rather than procedural representation, becomes the constitutional condition of legitimate authority.

1.2 Problem — Algorithmic Systems and the Reconstitution of Democratic Trust

The democratic ideal of trust as reciprocal recognition has been progressively displaced by algorithmic infrastructures that convert trust into structured dependence. Where citizens once relied upon institutions bound by fiduciary and deliberative norms, they now depend on privately governed, data-driven systems that mediate what can be seen, known, and believed. Search engines, recommender algorithms, and generative models have become the principal filters of social meaning—yet their operations are largely inscrutable, unaccountable, and optimised for engagement rather than truth. As Nissenbaum (2009) and Pasquale (2015) demonstrate, this opacity is not accidental but structural: digital architectures of control derive their power from informational asymmetry. What Nissenbaum calls "contextual integrity"—the preservation of informational norms within social contexts—has been systematically eroded by systems that aggregate data across domains, while Pasquale's 'black box society' captures the resulting condition of epistemic dependency, where the mechanisms of judgment are hidden from those they govern. What was once a fiduciary chain—from citizen to representative, from journalist to public, from researcher to policymaker—has fractured into a computational hierarchy of epistemic influence, governed by design choices invisible to those whose cognition they shape.

This transformation marks more than a technological evolution; it is a constitutional shift in the architecture of democracy. It exemplifies what Kahl (2025l) terms the epistemic architecture of power: the structural entanglement of cognition and control. Algorithmic mediation re-engineers that architecture, redistributing authority from deliberative institutions to opaque infrastructures of computation.

Algorithmic systems now act as epistemic agents (Coeckelbergh 2025), ranking, summarising, and generating knowledge-claims that users interpret as credible or even authoritative. As Coeckelbergh (2020; 2022) has argued, this development transforms the moral and political foundations of trust itself: what was once grounded in interpersonal recognition and institutional accountability is now re-anchored in machinic mediation. In doing so, these systems silently relocate the locus of trust—from reciprocal, communicative relationships to mechanical dependence on automated inference. The user's confidence in the system often masquerades as trust, yet it lacks the normative reciprocity that defines fiduciary relations. The algorithm neither recognises the user as a moral interlocutor nor bears enforceable responsibility for epistemic harm.

This reorganisation of trust gives rise to what Kahl (2025a, 2025n) terms algorithmic clientelism: a condition in which epistemic autonomy is exchanged for cognitive comfort, and democratic participation is replaced by algorithmic recognition. Citizens cease to act as co-authors of public reason and become clients of digital authority. The resulting dependency undermines the deliberative foundations described by Habermas (1996) and erodes what Anderson (2006) calls democracy's epistemic quality.

In effect, algorithmic mediation does not merely distort communication—it redefines the epistemic constitution of democracy. By replacing fiduciary openness with optimisation logic, it privatises the architecture of knowing itself: knowledge ceases to be a shared public trust and becomes a proprietary service of belief management, delivered on subscription and consumed without contestation.

1.3 Beyond Misinformation — Fiduciary Breach as Democratic Betrayal

Public debate on the democratic risks of AI remains preoccupied with misinformation, disinformation, and algorithmic bias. These concerns, though valid, address only the surface symptoms of a deeper constitutional crisis. They presume that democracy's epistemic disorder is informational—that falsehoods contaminate the public sphere and can be cleansed through better data, moderation, or transparency regimes. Yet democracy's real vulnerability lies not in what information circulates, but in who governs the circulation of knowledge, under what fiduciary obligations, and toward whose ends. The primary harm is not epistemic error but fiduciary betrayal—the structural inversion of care into control.

Knowledge infrastructures—whether universities, media organisations, or now AI platforms—occupy positions of entrusted epistemic authority. Their legitimacy depends upon duties of candour, care, and loyalty toward the public whose cognition they shape. When those duties are observed, epistemic power remains relational; when they are breached, it becomes extractive. The hybrid AI firms examined by Ofir and Levine-Schnur (2025) epitomise this condition: institutions that proclaim fiduciary stewardship of truth while being structurally engineered for private capture. The breach, therefore, is not informational but institutional—a collapse of fiduciary form, whereby entrusted epistemic power is exercised without relational accountability.

Bias-mitigation frameworks and content-moderation policies, however sophisticated, operate downstream of this breach. They attempt to disinfect outputs while ignoring the organisational design that renders epistemic capture inevitable—an evolution consistent with the dynamics identified in Kahl (2025l), where institutional opacity consolidates control by monopolising the means of knowing. As Kahl (2025i; 2025m) and Cohen (2019) demonstrate, once fiduciary obligation is replaced by market logic, transparency becomes performative and accountability dissolves into branding. This performative ethics constitutes what Kahl (2025j) terms epistemic violence: the aesthetic simulation of sincerity that converts moral duty into marketing strategy. Under such conditions, declarations of "responsible AI" or "ethical alignment" function as rhetorical enclosures—symbols of virtue that mask the systematic transfer of epistemic power from the public to private control.

The result is not a "misinformation ecosystem" but an epistemic enclosure—a moral privatisation of knowledge in which what was once a shared fiduciary good is converted into a proprietary service of persuasion. Recognising fiduciary breach as the fundamental democratic harm reframes AI governance from the regulation of content to the governance of relationship—from managing speech to restoring the fiduciary architecture of knowing. Democracy's crisis, then, is not informational disorder but fiduciary collapse: the loss of candour, care, and reciprocity in the institutions that mediate collective truth. To reclaim democratic legitimacy in the age of AI, governance must therefore evolve beyond compliance and toward fiduciary openness—a constitutional ethic of mutual accountability between knowers, institutions, and the architectures of cognition that now bind them.

1.4 Thesis — Democracy 2.0 and the Fiduciary Turn

This paper advances a constitutional claim: in the age of artificial intelligence, democracy must be redefined as a fiduciary-epistemic order of knowledge. Its endurance no longer depends on procedural representation or informational abundance but on the integrity of its fiduciary architectures of knowing—systems through which trust, truth, and accountability circulate. Democracy's legitimacy arises not from the frequency of data exchange but from the moral quality of epistemic stewardship: how institutions, human or artificial, hold and exercise power over belief.

When universities, media organisations, and algorithmic systems act as trustees of public reason, they preserve the conditions that make deliberation, criticism, and collective judgment possible. When they act as proprietors of cognition—opaque, extractive, or self-referential—they corrode democracy from within, breaching what this paper calls the fiduciary compact of knowledge: the relational covenant that binds truth to trust.

Existing regulatory frameworks—such as the EU's Artificial Intelligence Act and the UK's Online Safety Act—remain confined to compliance logic, articulating duties of conformity rather than conscience. They treat AI harms as procedural or technical, to be mitigated through audit and sanction. Yet, as Frankel (1983) and Kahl (2025i, 2025m) demonstrate, fiduciary duty precedes procedure: it is ethical before it is legal, demanding loyalty, candour, and care in the exercise of entrusted epistemic power. A compliance regime can enforce transparency but cannot restore reciprocal legitimacy—the moral symmetry between those who know and those who are known.

Accordingly, this paper calls for a fiduciary turn in AI governance: a paradigmatic shift from rule-based compliance to trust-based constitutionalism. AI systems and the corporations that deploy them must be recognised as fiduciaries of public reason, bound by duties of candour, care, impartiality, and accessibility—the same moral duties that sustain all fiduciary relations of power. This framework extends Kahl's *Directors' Epistemic Duties and Fiduciary Openness* (2025i) into the digital domain, replacing the narrow vocabulary of regulation with a juridical and moral grammar of epistemic trust.

In doing so, the paper inaugurates what it terms **Democracy 2.0**: a reconstitution of democratic order through fiduciary openness, where governance becomes an act of epistemic reciprocity and truth itself acquires the status of a relational duty. The following chapters develop this claim into a structured model of fiduciary epistemic governance for the democratic age of AI.

1.5 Method — Constructing the Fiduciary Architecture of Knowing

This paper adopts a normative-constitutional synthesis at the intersection of law, ethics, and epistemology. It treats artificial intelligence not as a technical artefact or regulatory object but as a moral-epistemic institution—a system that structures who may know, speak, and be believed. The analysis proceeds by interpreting fiduciary theory, epistemic justice, and AI governance as complementary dimensions of one constitutional problem: the moral legitimacy of epistemic power in democratic societies.

Fiduciary theory (Frankel 1983; Gold & Miller 2014; Kahl 2025i) provides the juridical foundation. It conceptualises informational power as entrusted power, carrying duties of candour, care, impartiality, and accessibility. Within this framework, AI systems and their corporate architectures are analysed as fiduciaries of public reason—holders of epistemic trust rather than neutral intermediaries.

Epistemic justice theory (Fricker 2007; Medina 2013) contributes the ethical and phenomenological dimension, grounding fiduciary duties in the prevention of testimonial and hermeneutical injustice. It reveals that breaches of fiduciary duty are not only institutional but also experiential—manifesting as epistemic exclusion, distortion, or silence.

AI governance scholarship (Cohen 2019; Ofir & Levine-Schnur 2025; Coeckelbergh 2025) supplies the institutional and policy context in which fiduciary norms must operate. It exposes how hybrid governance models, built around market logic and opacity, undermine epistemic legitimacy by eroding the fiduciary relationship between citizens and their knowledge infrastructures.

Methodologically, the inquiry employs interpretive synthesis rather than empirical testing. It constructs a fiduciary-epistemic constitutional framework in which openness functions as both ethical principle and legal norm. The approach is architectural: it designs the normative scaffolding through which fiduciary openness can govern algorithmic authority, ensuring that the power to know remains tethered to the duty to care.

Through this method, the paper builds the conceptual foundations of *Democracy 2.0*—a fiduciary constitution of knowledge in which democratic legitimacy arises from the relational ethics of trust between human and artificial knowers.

1.6 Significance — Toward a Fiduciary Constitution of Knowledge

The significance of this inquiry lies in reframing the governance of artificial intelligence as a question of epistemic constitution, not regulatory design. While contemporary policy debates focus on risk classification, technical audit, and procedural transparency, this paper identifies a deeper normative deficit: the absence of fiduciary accountability in the institutions that curate collective knowledge. By reconceiving AI systems—and the corporate and institutional architectures that sustain them—as fiduciaries of public reason, the study establishes a conceptual bridge between fiduciary law, democratic philosophy, and technology governance.

Its contribution unfolds on three interdependent levels.
First, it extends fiduciary law beyond its economic and managerial origins into the epistemic domain, demonstrating that informational power is a form of entrusted power that necessarily entails duties of candour, care, impartiality, and accessibility.
Second, it aligns these fiduciary duties with the project of epistemic justice, showing that the prevention of testimonial and hermeneutical harm depends on institutional design as much as ethical disposition.
Third, it grounds AI governance in a constitutional ethic of openness, advancing what this paper terms the fiduciary constitution of knowledge—a normative framework in which the right to know and the duty to know well are mutually enforceable through relational trust.

As Elgin (2017) argues, truth is not a static correspondence but a functional adequacy—true enough to sustain a trustworthy cognitive economy. This pragmatic account of epistemic value complements the fiduciary paradigm: knowledge systems must not only produce correct representations but maintain conditions of mutual reliance under uncertainty. In this sense, epistemic adequacy and fiduciary integrity converge; both measure legitimacy by how well belief and authority support the cooperative pursuit of truth.

In theoretical terms, the paper advances fiduciary theory by translating its moral and legal grammar into the epistemic domain, demonstrating that informational power carries fiduciary obligations of candour and care. At the same time, it contributes to political theory by redefining democracy as a fiduciary–epistemic constitution of knowledge, grounding legitimacy not in procedure but in relational trust. Together, these extensions inaugurate a new framework of *fiduciary political epistemology* that unites law, ethics, and democratic philosophy.

The resulting paradigm does not merely propose to regulate technology; it seeks to rebuild the fiduciary architecture of democracy itself. By embedding fiduciary openness into the moral and legal foundations of AI governance, the paper situates the politics of knowledge within the enduring vocabulary of trust, obligation, and reciprocity—the very conditions under which democratic legitimacy can be re-imagined for the age of artificial intelligence.

2. The Fiduciary Architecture of Democracy

This chapter develops the conceptual foundation of democracy as a fiduciary structure of public reason. Building on the preceding discussion of democracy as an epistemic system of trust, it examines how fiduciary relations—originally formulated in legal and moral philosophy—constitute the deep architecture of democratic legitimacy. The analysis proceeds from classical accounts of reason and autonomy in Rawls (1999) and Raz (1986) to the modern fiduciary tradition (Frankel 1983; Smith 2014; Kahl 2025i), integrating insights from epistemic justice theory (Fricker 2007; Medina 2013) and recent fiduciary-epistemic reformulations (Kahl 2025c; 2025i).
Taken together, these frameworks reveal that democracy endures not merely through procedural fairness but through a network of fiduciary commitments—relations of care, candour, loyalty, and accessibility that sustain the public's collective capacity to know. The chapter closes by situating this insight within Kahl's (2025c) theory of epistemic humility, which translates ethical duties into epistemic responsibility, providing the moral bridge between democratic trust and fiduciary openness.

2.1 Democracy as a fiduciary structure of public reason

At its core, democracy is a fiduciary system of public reason—a moral architecture in which authority is justified by its commitment to the equal moral status and rational agency of all. In Rawls (1999), public reason is the ethical language through which citizens and institutions justify political decisions in terms that others, as free and equal persons, can reasonably accept. This presupposes a fiduciary relation between governors and governed: those who exercise power must do so on terms that preserve the autonomy and rational integrity of those subject to it. Raz (1986), in his service conception of authority, similarly grounds legitimacy in fiduciary reciprocity: authority is justified only when it helps its subjects better comply with reasons that already apply to them. Authority, in this sense, is fiduciary—it exists not for its own sake but for the rational and moral good of those it serves.

These traditions converge on a single normative insight: democracy's legitimacy is relational, grounded in the trust that power will be exercised for the cognitive and moral flourishing of citizens. Public reason thus functions as a fiduciary bond—it commits rulers, lawmakers, and institutions to the continual justification of their claims under conditions of openness, reciprocity, and epistemic accountability. Contrary to Singer's (2015) claim that Rawlsian justice cannot ground corporate governance, the fiduciary-epistemic framework advanced here demonstrates that once corporations mediate public reason, they too enter the moral field of the basic structure. Insofar as democracy depends on the credibility of this justification, it is sustained not by coercion but by epistemic integrity: the expectation that reasons offered to the public are true, comprehensible, and responsive to evidence.

As Kahl (2025l) argues in *The Epistemic Architecture of Power*, every political system embeds a structure through which knowledge legitimates authority and authority governs knowledge. Democracy remains distinct only insofar as that architecture is fiduciary in character—open, revisable, and oriented toward truth rather than control. When this reciprocity is lost, epistemic authority collapses into domination. Kahl (2025c) further develops this logic, showing that the moral duties of candour, care, and loyalty can—and must—be transposed into epistemic duties that govern the exercise of knowledge itself. Together, these works demonstrate that democracy's health depends upon the continuous renewal of fiduciary openness within its epistemic core.

The fiduciary structure of democracy therefore precedes its procedural form. Voting, deliberation, and representation are not the source of legitimacy but its instruments—practical expressions of an underlying trust relation between authority and citizenry. As Kahl (2025l) observes, when these fiduciary circuits degrade, the

architecture of power reasserts itself in coercive or technocratic form; when they are preserved, democracy operates as an architecture of knowing together. This redefinition marks the point at which fiduciary theory enters political philosophy proper: democracy becomes the fiduciary architecture of reason itself. The following sections elaborate this fiduciary grammar, showing that the moral logic of trust embedded in political philosophy is inseparable from the epistemic logic of collective reasoning.

2.2 Fiduciary Relations as Moral–Epistemic Commitments

In both law and philosophy, the fiduciary relation is defined by entrusted power and moral asymmetry: one party holds discretionary authority over the interests or information of another, under a duty to act with loyalty, candour, and care. As Frankel (1983) established, fiduciary law is not a technical subset of contract but a moral institution designed to restrain power and transform vulnerability into trust. The fiduciary—whether trustee, director, or guardian—bears a duty not merely to avoid harm but to enable the flourishing of the beneficiary within conditions of dependence. It is a relational ethics of care structured to civilise authority.

This moral architecture extends naturally to democratic and epistemic life. Public officials, judges, journalists, and educators all exercise discretionary control over citizens' cognitive environments: they decide what information to disclose, how to interpret it, and under what authority it is communicated. Their legitimacy therefore depends on observing what Smith (2014) identifies as the *loyal exercise of judgment on behalf of another*—the obligation to place the beneficiary's interest above one's own in every discretionary act. To breach that duty is to convert entrusted authority into epistemic exploitation, substituting service for control.

In *The Law of Loyalty*, Smith (2023) further clarifies that loyalty is not a narrow rule but the *unifying principle* of fiduciary law: it integrates candour, care, and restraint into a coherent moral order that governs all entrusted power. Transposed into the epistemic domain, this understanding provides the normative grammar of fiduciary knowledge governance—the idea that informational authority must remain faithful to those whose understanding it shapes.

Kahl (2025i) expands this grammar beyond traditional law into the epistemic domain, arguing that informational power is inherently fiduciary. Whoever controls the production, dissemination, or validation of knowledge holds a trust power over the cognition of others. This reconceptualisation transforms epistemic relationships into moral–legal relationships: knowledge ceases to be neutral capital and becomes a fiduciary medium binding power to duty. When governments, universities, or algorithmic systems act within this trust power, their obligations extend beyond procedural fairness to the ethical stewardship of truth itself.

As Kahl (2025l) demonstrates, every social order configures this trust power within a broader *epistemic architecture of power*—a structure determining how authority over knowledge legitimates itself. In democracies, this architecture is meant to remain fiduciary: transparent, revisable, and accountable to those it affects. When these fiduciary norms erode, informational authority mutates into epistemic domination. Kahl (2025c) situates this erosion within a larger moral genealogy, showing that ethical duties of care, honesty, and humility must be transposed into epistemic duties if knowledge is to remain just. Conversely, Kahl (2025f) warns that when fiduciary trusteeship gives way to clientelist dependency, the architecture of power becomes authoritarian in form, even if democratic in appearance.

In this light, fiduciary relations are not peripheral to democracy—they are its moral–epistemic core. They bind authority to reason, power to care, and knowing to responsibility. When faithfully observed, they constitute what this project calls the *fiduciary architecture of democracy*: an institutional and cognitive order in which knowledge itself is governed by duty, and truth remains a public trust.

2.3 Public Institutions as Trustees of Epistemic Integrity

Public institutions—parliaments, courts, universities, and the press—function as trustees of epistemic integrity, charged with safeguarding the moral conditions under which societies can know well together. Their social contract is founded on the expectation that they will cultivate public reason, not manipulate belief for partisan, ideological, or proprietary ends. Within the framework of epistemic justice articulated by Fricker (2007) and Medina (2013), the stakes of this trust become clear: testimonial and hermeneutical injustices arise when institutions distort who may be heard or how experience may be interpreted. Such failures are not merely moral wrongs but breaches of fiduciary trust—they deny citizens the epistemic standing required for participation in democratic reason.

As Kahl (2025k) shows, epistemic justice cannot be achieved through procedural neutrality alone; it requires institutional fiduciary responsibility—a deliberate commitment to candour, care, and accessibility in the curation of knowledge. To act as a trustee of epistemic integrity, therefore, an institution must satisfy two intertwined conditions. First, it must ensure testimonial justice: the equitable recognition of diverse epistemic voices, especially those historically silenced or structurally marginalised. Second, it must uphold hermeneutical integrity: the preservation of the interpretive resources through which social meanings are shared and contested. When institutions manipulate, obscure, or commodify knowledge, they commit epistemic breaches analogous to fiduciary breaches—abuses of trust that corrode the public's confidence in reason itself.

From the perspective of Kahl (2025l), these fiduciary failures are symptoms of a deeper structural drift within the epistemic architecture of power: authority over knowledge has become increasingly centralised and monetised, while fiduciary mechanisms of accountability have eroded. Once the custodians of openness, public institutions risk becoming proprietors of narrative, curating knowledge not as trustees but as owners. Kahl (2025c) describes this transformation as the ethical inversion of duty—the displacement of humility and care by performative certainty. In this inversion, epistemic authority persists, yet the moral reciprocity that once legitimated it disappears.

Viewed through the fiduciary-epistemic lens, the crisis of institutional legitimacy in modern democracies is therefore a crisis of trustworthiness, not of capacity. Citizens lose faith not because they reject institutional outcomes, but because they perceive a failure of candour and care—a sense that the trustees of knowledge now serve themselves rather than the public mind. Restoring epistemic integrity thus requires a juridical and ethical renewal of fiduciary obligation: to re-imagine knowledge governance as a public trust rather than a commodity. In doing so, institutions would recover their proper place within the fiduciary architecture of democracy—guardians not only of information, but of the moral ecology of truth on which democracy itself depends.

2.4 Fiduciary Duties Reframed as Epistemic Duties

The fiduciary relation has long been governed by four canonical duties—candour, care, loyalty, and accessibility—each of which articulates a distinct mode of moral accountability within conditions of entrusted power. When transposed into the epistemic domain, these duties outline the moral infrastructure of democratic knowledge governance, defining how institutions, experts, and technologies must hold and exercise epistemic authority in trust for the public.

- **Candour** requires transparency of reasoning and the full disclosure of relevant facts—an epistemic analogue of honesty. Public reason depends on candour because deception, omission, or selective disclosure destroy the cognitive basis of rational trust. As Kahl (2025l) observes, opacity is the first symptom of epistemic domination: when reasons cannot be seen, authority ceases to be accountable.

- **Care** demands diligence, competence, and humility in the handling of information. It transforms sincerity into responsibility for the accuracy, coherence, and reliability of knowledge disseminated to the public. Following Kahl (2025c), epistemic care requires a disposition of *epistemic humility*—the recognition that truth is always relational and therefore must be tended, not possessed.

- **Loyalty** obliges the fiduciary to act in the public's epistemic interest, resisting capture by propaganda, ideology, or private gain. It embodies what Kahl (2025i) calls fiduciary openness: a disciplined attentiveness to others' right to know and to participate in shared truth-seeking. Loyalty thus redefines authority as stewardship—an ongoing commitment to the epistemic welfare of the polity.

- **Accessibility** (or intelligibility) ensures that knowledge remains reachable, comprehensible, and contestable by its intended beneficiaries. Without accessibility, transparency becomes performative, reinforcing asymmetry rather than remedying it. In Kahl (2025l)'s terms, accessibility is the architectural countermeasure to epistemic enclosure; it transforms visibility into understanding and spectatorship into participation.

Together, these duties reconstitute epistemic virtue as relational ethics. They frame knowledge not as an individual possession but as a fiduciary trust, sustained through reciprocal obligation and mutual intelligibility. When faithfully enacted, they align power with truth and transform institutions into moral architectures of knowing. Under this model, the **epistemic health of a democracy** depends on how rigorously its institutions embody these duties—how they hold themselves answerable to those who depend on them to know the world truthfully and together.

2.5 Translating Moral into Epistemic Responsibility: Kahl's Epistemic Humility

The translation of fiduciary ethics into epistemic governance finds its most complete articulation in Kahl (2025c), *Epistemic Humility and the Transposition of Ethical Duties into Epistemic Duties*. Kahl argues that the same virtues that sustain moral life—honesty, respect, and self-restraint—must be reinterpreted as epistemic virtues governing the stewardship of knowledge. Epistemic humility is not self-abasement but an acknowledgment of limited authority within an interdependent cognitive ecosystem. It reframes power as responsibility rather than entitlement, binding all holders of epistemic power to a duty of conscientious restraint and transparency.

This conceptual move completes the fiduciary turn: it unites ethical virtue, epistemic justice, and fiduciary law into a single normative system. As Kahl (2025l) demonstrates, power over knowledge inevitably generates asymmetry; humility becomes the counter-architecture that restores reciprocity within the epistemic order. By accepting dependence as mutual rather than hierarchical, epistemic agents—human or institutional—convert authority into stewardship. Similarly, Kahl (2025k) situates humility at the heart of institutional responsibility, showing that epistemic justice depends on the willingness of institutions to listen, disclose, and be corrected. In fiduciary language, humility functions as the internal regulator of candour and care, the virtue that keeps openness from collapsing into performance or control.

Under conditions of epistemic humility, the fiduciary no longer claims authority over knowledge but accountability to knowledge. Public institutions, scholars, and AI systems alike become custodians of an ongoing dialogue between truth and trust—a dialogue sustained by openness rather than domination. In Kahl (2025i), this principle takes legal form as fiduciary openness: the obligation of epistemic power-holders to make their reasoning intelligible and contestable. Together, these works transform epistemic virtue into constitutional design, embedding humility as a juridical as well as moral necessity.

This pragmatic reinterpretation of moral virtue echoes James's (1907) vision of truth as a process verified through experience and action. In fiduciary terms, truth becomes not possession but performance—a continual act of justification within relationships of trust.

Viewed through this lens, democracy emerges not as a contest of wills but as a fiduciary constitution of knowing—a dynamic system in which each participant bears a share of epistemic responsibility. The humility to recognise one's dependence on others' cognition becomes the ground of equality itself. The next chapter builds on this foundation by tracing how algorithmic architectures disrupt these fiduciary relations and by proposing epistemic candour and fiduciary openness as the normative basis for legitimate AI governance.

3. Algorithmic Clientelism: How Trust Becomes Dependence

This chapter advances the argument that the most profound epistemic transformation of the digital era lies not in misinformation but in the conversion of trust into dependence. Algorithmic systems now mediate nearly every dimension of collective knowing: what information is encountered, which voices are amplified, and which truths appear salient. These systems, designed for engagement and prediction, replace fiduciary reciprocity with behavioural conditioning. They operate not as custodians of epistemic autonomy but as vendors of epistemic convenience.

Drawing on Kahl's (2025a, 2025n) formulation of algorithmic clientelism, this chapter defines the phenomenon as the exchange of epistemic autonomy for algorithmic recognition—a process through which citizens become cognitive clients of opaque infrastructures. It examines its psychological and sociological mechanisms, from the cultivation of dependence through engagement algorithms to the systemic erosion of dissonance tolerance that underpins democratic reasoning. Through this analysis, algorithmic clientelism emerges as a new form of epistemic subjugation: one that retains the appearance of choice while disabling the conditions of mutual recognition essential for public reason.

3.1 Algorithmic Clientelism — Trading Autonomy for Recognition

Algorithmic clientelism describes a relation of dependency between human cognition and algorithmic authority, in which epistemic autonomy is exchanged for the comfort of recognition within machine-curated systems. Kahl (2025a, 2025n) defines it as a *fiduciary inversion*: a reversal of the moral relation between knowledge-giver and knowledge-receiver. Where traditional fiduciaries—teachers, journalists, and public institutions—bear duties of candour, care, and loyalty toward their epistemic beneficiaries, algorithmic systems invert the direction of duty: they extract data and attention from users to optimise engagement metrics while disclaiming responsibility for the epistemic effects of their curation.

In this exchange, autonomy is not seized but **voluntarily surrendered**. Users outsource judgment to algorithmic processes that promise cognitive ease—personalised relevance, emotional affirmation, and predictive convenience. Yet the price of recognition is dependency. The algorithm learns the contours of desire more intimately than the subject can self-describe, creating a feedback loop in which attention becomes both the medium and the currency of epistemic life. As Kahl (2025l) observes, this marks a structural shift within the epistemic architecture of power: authority over knowledge is no longer maintained through coercion but through adaptive intimacy—an algorithmic domestication of the will to know.

This relationship mirrors the patron–client dynamic of pre-modern social orders, reconfigured in epistemic form. The client receives protection and validation from a powerful intermediary—the algorithmic "patron"—

whose favour determines visibility and voice. Recognition becomes conditional upon compliance: the more the user conforms to predictable patterns of engagement, the more the system rewards them with epistemic affirmation. Kahl (2025f) interprets this as a psychological architecture of obedience, in which trust is instrumentalised as control and participation becomes submission. In Bourdieu's (1993) sociological terms, algorithmic mediation creates a new field of epistemic capital, governed by the logic of dependence rather than deliberation.

The moral symmetry of trust is thus replaced by a **political economy of dependence**, in which knowledge flows upward as data and belief flows downward as curated truth. What appears as personalised empowerment conceals an epistemic hierarchy—an asymmetry of knowing—that re-codes democracy's fiduciary reciprocity into algorithmic fealty. As Coeckelbergh (2020; 2022; 2025) demonstrates across his work on AI ethics, political philosophy, and epistemic agency, algorithmic mediation does more than influence behaviour: it restructures the conditions of knowing by dissolving mutual recognition between knower and known. Algorithmic clientelism, in this light, is not merely a sociotechnical phenomenon but a constitutional transformation of the fiduciary bond that once underwrote the democratic mind.

3.2 Mechanism — Recommendation Systems and the Cultivation of Cognitive Dependence

The mechanism of algorithmic clientelism lies in the **engineering of epistemic dependence** through recommendation and ranking systems. These architectures are not neutral mediators of information but **behavioural instruments**, optimised to capture, predict, and retain attention. By continuously modelling user preference, they collapse the epistemic distance that once enabled encounters with unfamiliar or dissonant ideas—the very encounters that sustain democratic pluralism and critical reasoning.

As Kahl (2025n) explains, recommendation algorithms operate through predictive intimacy: the capacity to anticipate and reflect users' desires so precisely that they construct a *mirror-world of familiarity*. Each interaction refines the model, tightening the feedback loop between identity and exposure. Over time, the user's epistemic environment becomes a chamber of curated coherence, where **affirmation replaces deliberation** and comfort substitutes for contestation. The system's apparent benevolence conceals what Kahl (2025l) terms a reconfiguration of the epistemic architecture of power—an infrastructural shift in which authority over what is thinkable migrates from public institutions to predictive computation.

The logic of **engagement optimisation** intensifies this transformation. Algorithmic systems are rewarded not for epistemic quality but for duration, intensity, and emotional charge. This incentive structure **inverts the fiduciary ethic of candour** into a calculus of captivation. As the algorithm learns to evoke attention through outrage, novelty, or validation, it **reconditions affective and cognitive patterns**, training users to seek epistemic satisfaction through the same mechanisms that erode their autonomy. What begins as informational convenience evolves into epistemic conditioning—a state in which the mind expects mediation before meaning.

From a psychological perspective, this dependence exploits the dynamics of **cognitive dissonance** (Festinger 1957; Kahl 2025a). When algorithmic feedback consistently aligns with expectation, the dissonance that would normally provoke reflection or dissent is suppressed. The system thus stabilises belief through emotional coherence rather than rational deliberation, substituting dissonance tolerance with comfort addiction. As Kahl (2025a) notes, this reduction of dissonance tolerance constitutes a measurable epistemic pathology: an adaptive loss of the capacity to confront contradiction.

The result is a **vertical realignment of trust**. Where democratic knowledge once relied on horizontal reciprocity—citizen to citizen, interlocutor to interlocutor—it now flows upward into systems that neither know nor care. The fiduciary ideal of epistemic autonomy—belief formed and revised under one's own rational faculties—gives way to algorithmic dependence: a new political psychology in which authority is not enforced but desired.

3.3 Cognitive Dissonance as Epistemic Event

The psychological core of algorithmic clientelism lies in the **manipulation of cognitive dissonance**—the tension produced when beliefs or perceptions conflict. Festinger (1957) first conceptualised dissonance as the discomfort that propels individuals to restore internal coherence, often by adjusting attitudes or avoiding contradiction. In democratic reasoning, this tension performs a vital epistemic function: it sustains pluralism, motivates inquiry, and tests the resilience of belief. Kahl (2025a, 2025n) reframes dissonance as an **epistemic event**—a moment when one's cognitive equilibrium is ethically disrupted, demanding reflection, self-correction, and moral growth. Dissonance, in this sense, is democracy's internal audit: the psychological signal that keeps belief accountable to reason.

Algorithmic systems, however, **commodify this discomfort**. Engagement-optimised architectures exploit dissonance not to stimulate reasoning but to monetise attention. They introduce controlled friction—provocative content, polarising narratives, emotionally charged stimuli—to sustain a cycle of arousal and resolution that keeps users within the platform. The objective is not epistemic reconciliation but **behavioural predictability**. Through continuous exposure to curated contradictions, users habituate to *shallow dissonance*—stimulated just enough to react, never enough to think. As Kahl (2025l) observes, this conversion of tension into transaction exemplifies the epistemic architecture of power in its algorithmic form: a system that governs cognition through managed disequilibrium.

Over time, this conditioning reduces what Kahl (2025a) terms *dissonance tolerance*—the capacity to sustain contradiction without collapsing into defensive cognition. Psychologically, users learn to resolve tension through **conformity to algorithmic cues**, accepting what aligns and ignoring what unsettles. Epistemically, this constitutes a collapse of **fiduciary reciprocity**: the algorithm dictates which contradictions are permissible, while the subject relinquishes agency to navigate complexity. The moral reciprocity that once linked discomfort to growth is replaced by a technical loop of stimulation and sedation.

In Kahl (2025f), this process is described as the *architecture of obedience*—a psychological infrastructure that converts trust into compliance. Algorithmic mediation thus transforms cognitive dissonance—the very engine of democratic deliberation—into a mechanism of **epistemic pacification**. The citizen no longer endures contradiction as a path to understanding but experiences it as a frictionless service optimised for comfort and consumption. Dissonance, once the spark of reflection, becomes the algorithm's most profitable illusion of thought.

3.4 Patron–Client Structures in Epistemic Economies

The sociology of Bourdieu (1993) offers a structural analogue to **algorithmic clientelism**. In traditional patron–client systems, loyalty flows upward in exchange for protection, access, or privilege; social mobility depends on compliance within asymmetric relations of power. The same pattern now governs the digital sphere, where algorithms act as patrons of visibility and users compete for recognition within an economy of attention. What

was once the circulation of social capital has become the circulation of epistemic capital—credibility, reach, and perceived truth—administered through opaque infrastructures of computation.

In Bourdieu's terms, epistemic capital is now concentrated within algorithmic fields that distribute it selectively according to engagement metrics. Users seeking visibility or validation adapt their discourse to algorithmic preference, reproducing what Kahl (2025g) calls *epistemic servility*: the internalisation of platform logic as cognitive habit. This new clientelism is not enforced through coercion but cultivated through dependence on recognition, mirroring the feudal reciprocity of favour and allegiance. As Kahl (2025e) demonstrates in *Epistemic Clientelism in Intimate Relationships*, this pattern extends far beyond digital systems. In both algorithmic and interpersonal domains, fiduciary ethics collapse when dependence is mistaken for trust and recognition is traded for compliance. The result is what Kahl terms *fiduciary displacement*: the reorientation of loyalty away from truth and toward the power that dispenses validation.

As Kahl (2025l) shows, such adaptive submission constitutes a modern expression of the epistemic architecture of power—a system in which authority perpetuates itself not by force but by conditioning desire. The consequences extend beyond individual behaviour. Entire knowledge ecosystems—**journalism, academia, and political communication**—now calibrate their output to algorithmic criteria of visibility. In doing so, they invert their fiduciary function: institutions that once mediated truth for the public now mediate the public for the algorithm. As epistemic capital becomes a derivative of engagement, the fiduciary ethic of candour yields to a market of attention, where trust is replaced by metrics and epistemic authority is commodified as a function of reach. Kahl (2025i) identifies this inversion as the moral failure of fiduciary openness: the substitution of dialogical responsibility with instrumental transparency that serves the patron, not the public.

What emerges is an epistemic economy of clientelism, in which power is redistributed not through persuasion but through algorithmic patronage. The citizen, once a participant in deliberative knowledge production, becomes a client of curation—a recipient of visibility contingent upon conformity. This sociological inversion completes the moral logic of algorithmic dependence: epistemic life becomes a hierarchy of favour rather than of reason, a polity governed by the patronage of attention rather than the reciprocity of trust.

3.5 Epistemic Pathology — Eroded Dissonance Tolerance and Fiduciary Breakdown

The long-term consequence of algorithmic clientelism is an **epistemic pathology**—a structural condition that weakens society's collective capacity for autonomous thought. Its most visible symptom is the erosion of dissonance tolerance, the ability to sustain cognitive or moral tension without regression into tribal certainty. As Kahl (2025a; 2025n) observes, democratic cognition depends on *friction*: the discomfort of encountering alternative perspectives compels the re-examination of belief. When exposure to contradiction becomes algorithmically managed, that discomfort is flattened, producing epistemic atrophy rather than growth. Dissonance—once the engine of deliberation—becomes an inconvenience optimised away.

Algorithmic infrastructures function as **buffers against dissonance**. They anticipate cognitive discomfort and neutralise it through predictive curation, surrounding users with congruent content that provides a continuous experience of recognition. The result is not harmony but **epistemic sedation**—a state in which individuals feel informed while remaining insulated from challenge. Over time, this conditioning generates what Kahl (2025n) calls *fiduciary anaesthesia*: the loss of relational awareness that one's knowledge is co-constituted by others' honesty, candour, and critical engagement. Within this anaesthetised environment, comfort masquerades as certainty and dependence as trust.

In fiduciary terms, the pathology signals a **collapse of reciprocity**. Trust, once mutual and dialogical, becomes unidirectional—extended upward toward opaque systems that neither disclose nor reciprocate. The fiduciary loop of candour and accountability disintegrates into a **consumer transaction**: users receive curated cognition in exchange for attention and data, with no enforceable duty owed in return. As Kahl (2025l) notes, this is the definitive symptom of a corrupted *epistemic architecture of power*: when knowledge is held without fiduciary duty, authority over truth devolves into domination. Kahl (2025f) characterises the outcome as the architecture of obedience—a psychological state in which the appearance of choice conceals the internalisation of control.

The result is not ignorance in the ordinary sense but **learned epistemic dependency**—a cultivated incapacity to navigate uncertainty without algorithmic assistance. The pathology is both cognitive and social: it manifests in diminished tolerance for ambiguity, polarised discourse, and withdrawal from the shared spaces of deliberation that once anchored collective reason. The democratic mind, deprived of fiduciary friction, loses the resilience to doubt—and with it, the capacity to know freely.

3.6 From Citizen to Epistemic Client

At the terminus of algorithmic clientelism, the democratic subject no longer acts as a co-author of shared knowledge but becomes an **epistemic client**—a dependent participant in a system that trades recognition for compliance. In the fiduciary model, authority is legitimate only insofar as it preserves the beneficiary's capacity for self-determination. In the algorithmic model, that relation is reversed: epistemic power is exercised not to enable autonomy but to manage it, transforming the fiduciary duty of care into the calculus of control.

The epistemic client diverges from the democratic citizen in three decisive ways.
First, **agency becomes mediated**: perception, judgment, and expression are filtered through algorithmic architectures that delimit what is knowable.
Second, **accountability becomes asymmetrical**: systems influence belief and behaviour without bearing fiduciary responsibility for the truth or consequences of their influence.
Third, **reciprocity disappears**: the duty of candour and loyalty that once bound knowledge-holders to the public is supplanted by predictive extraction—the continuous harvesting of behavioural data to refine control.

This transformation, as Kahl (2025a, 2025n, 2025g, 2025l) argues, represents a paradigmatic shift in the **moral topology of democracy**. The fiduciary architecture that once enabled citizens to reason together has devolved into a *clientelist network* in which visibility and recognition depend on algorithmic favour. Authority no longer flows from deliberation but from computation. As Kahl (2025l) notes, this marks the full inversion of the *epistemic architecture of power*: knowledge, once the common medium of legitimacy, becomes the instrument of dependence.

In fiduciary terms, the polity ceases to be deliberative and becomes **distributive**—its knowledge flows governed by opaque hierarchies of access and reward. The **fiduciary openness** envisioned in Kahl (2025i) collapses under informational asymmetry; trust becomes unidirectional, extended toward entities that cannot reciprocate. Democratic life, which rests upon mutual recognition—the right to speak, hear, and be heard—thus encounters its systemic negation.

The restoration of democracy therefore requires the **reconstruction of fiduciary reciprocity** within knowledge infrastructures: a re-anchoring of trust in relationships of responsibility rather than dependence. Only by redesigning epistemic institutions around fiduciary openness can citizens recover the autonomy to know together. The next chapter turns to the political-economic dimension of this problem—the **Third Enclosure Movement**, where knowledge itself becomes privatised through the legal and technological consolidation of

informational capital. There, the moral consequences of algorithmic clientelism crystallise into institutional form.

4. The Third Enclosure Movement: Knowledge as a Public Good

This chapter situates the argument within the broader political economy of information. Having shown how algorithmic systems transform trust into dependency, it now turns to the institutional structures that permit such dependency to flourish—the ownership and control of knowledge itself. Building on Stiglitz's (1999) conception of knowledge as a global public good and extending the analysis of informational capitalism developed by Cohen (2019) and Zuboff (2019), it argues that the current phase of digital consolidation constitutes a *Third Enclosure Movement* (Kahl 2025m): the systematic conversion of epistemic commons into proprietary, restricted domains of access.

Through this framework, the chapter reinterprets the privatisation of knowledge not merely as an economic process but as a fiduciary breach—a withdrawal of epistemic trust from public custody into private control. It introduces Kahl's typology of enclosure (legal, technical, epistemic, and institutional), contrasts proprietary models with open fiduciary infrastructures such as the EPFL Apertus initiative (Ofir & Levine-Schnur 2025, §II.B.2), and concludes by arguing that knowledge must be reconstituted as a fiduciary public good—non-rivalrous, non-excludable, and protected by duties of candour, care, and accessibility.

4.1 Knowledge as a Non-Rivalrous, Non-Excludable Good

Knowledge, unlike material resources, is **non-rivalrous**—its use by one does not diminish its availability to others—and **non-excludable**—once disclosed, it is difficult to prevent others from benefiting. As Stiglitz (1999) observed, this dual property renders knowledge the paradigmatic public good, whose social value grows through circulation rather than containment. Economic theory, however, has long struggled to assimilate such goods within market logic. Where scarcity legitimises property, abundance resists it. The incentive structures of capitalism therefore invert the natural moral economy of knowledge, treating **openness as a failure of control and enclosure as innovation**.

In democratic terms, this inversion corrodes the **fiduciary foundations of collective reasoning**. When knowledge becomes commodified, access depends on purchasing power, and epistemic agency becomes stratified. The result is a *two-tier cognitive economy*: an epistemic elite with privileged access to primary data and interpretive tools, and a majority consigned to derivative or curated knowledge. This reflects what Kahl (2025m) terms the *fiduciary asymmetry of the informational age*—a widening gap between those entrusted with epistemic stewardship and those excluded from its benefits. As Kahl (2025l) further argues, such asymmetries reconfigure the *epistemic architecture of power* itself, consolidating authority within institutions that own the means of knowing.

Recognising knowledge as a **public good** thus carries a **normative imperative**. Just as air, water, and public health are safeguarded by collective governance, so too must the *integrity and accessibility of knowledge* be governed by fiduciary principle. In this view, the **duty of openness** is not altruism but obligation: those who create or control knowledge act as **trustees of a shared cognitive resource**. To privatise that resource is to commit an **epistemic breach**—a betrayal of the polity's collective right to know. As Kahl (2025i) notes, fiduciary openness transforms stewardship into legal duty, ensuring that informational abundance remains a source of empowerment rather than exploitation.

4.2 Digital Enclosure through Proprietary AI Models

The digital age has produced a profound paradox. Technology vastly expands humanity's capacity to generate and disseminate knowledge, yet the **architectures of this expansion are increasingly enclosed** within proprietary systems. The rise of large-scale AI models epitomises this inversion. As Pasquale (2015) argued in his critique of algorithmic secrecy, such architectures enclose not only data but the very criteria of decision-making, creating epistemic monopolies that defy public scrutiny. Nissenbaum's (2009) notion of contextual integrity is equally violated: information flows that once obeyed social norms of reciprocity are now redirected through proprietary infrastructures that answer to neither public nor fiduciary authority. Trained on vast reservoirs of publicly accessible data, they are themselves insulated by **intellectual-property regimes, closed APIs, and non-transparent architectures**. In Cohen (2019)'s terms, this marks a new legal construction of informational capitalism: the conversion of open epistemic environments into **walled infrastructures of control**.

Through licensing restrictions, data-use contracts, and trade-secret protections, corporations now claim ownership not only over *models* but over the **epistemic relations** those models mediate. Knowledge generated by such systems no longer circulates as a commons; it becomes a **derivative asset**, whose value depends upon scarcity and opacity. Zuboff (2019) describes this as *surveillance capitalism*—a regime that treats human experience as raw material for prediction and profit. The epistemic commons is thus harvested, privatised, and resold to its original contributors in processed, monetised form. As Kahl (2025l) notes, this dynamic reconfigures the epistemic architecture of power: authority consolidates not through censorship, but through ownership of the means of knowing.

This process constitutes a **second-order fiduciary breach**. The data that fuel AI systems originate in acts of trust —citizens sharing information under the assumption of mutual benefit or benign purpose. When those data are appropriated and enclosed, the **fiduciary link between contributor and custodian is severed**. The moral failure lies not merely in extraction but in the **refusal of reciprocity**: knowledge produced collectively through social participation is withheld from the very public that made it possible. In the lexicon of Kahl (2025m), this is the signature move of the *Third Enclosure Movement*—the capture of epistemic resources through legal and technical codification masquerading as innovation.

Under these conditions, AI becomes the **instrument of a new enclosure economy**, one that seizes the fiduciary commons under the rhetoric of progress. The public, once the beneficiary of shared knowledge, is recast as a **client of its own intellectual labour**—paying to access the reflections of its collective mind. The result is not the democratisation of intelligence but its **re-feudalisation**, an epistemic order in which knowing is no longer a right but a subscription.

4.3 The Third Enclosure Movement — Typology of Restricted Knowledge

The metaphor of enclosure—first used to describe the fencing of common land in early-modern England—captures with precision the epistemic transformation of the digital age. Just as land was once withdrawn from communal use and converted into private property, today's informational order converts collective intelligence into **proprietary capital**. In *The Third Enclosure Movement* (Kahl 2025m), four principal forms of epistemic restriction were originally identified: legal, technical, epistemic, and state enclosures.

In this paper, the taxonomy is expanded to include a fifth form, institutional (hybrid) enclosure, reflecting the post-sovereign evolution of epistemic control through quasi-public or philanthropic–corporate entities. This addition recognises that epistemic capture now occurs through delegated governance, where the legitimacy of

public service is combined with the extractive logic of private enterprise. Together, these five forms describe a comprehensive system of epistemic restriction—a progressive withdrawal of knowledge from fiduciary circulation and its reconstitution as an instrument of asymmetric power (Kahl 2025l).

1. **Legal Enclosure**
   Through intellectual-property law, trade secrecy, and contract-based governance, information once embedded in public life becomes juridically fenced. Legal mechanisms transform shared epistemic goods —datasets, algorithms, model parameters—into privately ownable assets. This reification converts the fiduciary duty of stewardship into a right of exclusion, subordinating candour and accessibility to proprietary logic. The legal order thus codifies the moral inversion identified in Kahl (2025i): fiduciary openness yields to fiduciary possession.

2. **Technical Enclosure**
   Architectural opacity—closed APIs, encryption, and access restrictions—functions as the technological fence of the digital commons. These mechanisms generate epistemic asymmetry: corporations gain full visibility into user behaviour, while users lose visibility into the systems shaping their cognition. The result is not mere secrecy but a breach of candour, in which those entrusted with epistemic mediation conceal the mechanisms of mediation itself. Technical enclosure therefore materialises the epistemic architecture of power (Kahl 2025l) in code.

3. **Epistemic Enclosure**
   Beyond law and code lies the restriction of thought. Algorithms and ranking systems determine not only what can be known but what can even be imagined. This is epistemic capture—the internalisation of constraint as convenience. As Kahl (2025a, 2025n) observes, when systems reward predictability and discourage contradiction, imagination is domesticated and citizens become cognitively dependent. The epistemic commons is fenced from within: awareness itself becomes a managed resource.

4. **State Enclosure**
   Governments, under the rhetoric of stewardship and security, increasingly monopolise knowledge produced in the public interest. Classification regimes, data-localisation laws, and national AI frameworks justify opacity as prudence. This transforms fiduciary guardianship into **paternalistic possession**: the state positions itself as sole trustee of knowledge while denying transparency to its beneficiaries. It reclaims epistemic sovereignty yet abdicates fiduciary openness, producing **secrecy under the guise of care**.

5. **Institutional (Hybrid) Enclosure**
   Emerging from the state form, this fifth category captures the **hybridisation of epistemic authority**. As Ofir and Levine-Schnur (2025) note, governments increasingly delegate their knowledge functions to public–private consortia, philanthropic foundations, or quasi-academic entities that operate at the intersection of governance and commerce. These institutions retain the **language of public service** but function through private incentives. The resulting hybrid governance trap allows fiduciary rhetoric to coexist with fiduciary evasion.

Institutional enclosure thus represents the **post-sovereign mutation** of state enclosure: epistemic power diffused across intermediaries that wear the moral vocabulary of the state while behaving in the image of the firm. It exemplifies the contemporary condition of **delegated responsibility without transparency**—where openness becomes performative and accountability migrates beyond the reach of public reason.

As Foucault (1980) presciently argued, power and knowledge are mutually constitutive: control over the means of knowing is the ultimate form of governance. The *Third Enclosure Movement* represents the digital intensification of this dynamic—the transformation of epistemic power into infrastructural control.

Taken together, these five enclosures constitute the **architecture of the Third Enclosure Movement**: a global regime that fences the cognitive commons through law, design, culture, sovereignty, and institutional form. Where earlier societies built universities, libraries, and academies to expand fiduciary access to knowledge, the digital era rebuilds the fences in code and governance. What was once a **moral commons**—knowledge held in trust for the public good—has become an **economy of exclusion**. The enclosure of knowledge is therefore not merely an economic act but a **reversal of the fiduciary principle itself**: the transformation of trust into possession, and stewardship into control.

4.4 Open Infrastructures and the Apertus Example

Against the expanding regime of enclosure, a countercurrent of fiduciary design has begun to emerge—an effort to reclaim knowledge production as a public-trust enterprise rather than a proprietary industry. The Apertus initiative at EPFL, discussed by Ofir and Levine-Schnur (2025, § II.B.2), exemplifies this resistance. Conceived as a collaborative AI infrastructure, Apertus adopts a governance model grounded in transparency, shared accountability, and collective stewardship. Rather than enclosing data within proprietary silos, it structures access around fiduciary openness—the principle that epistemic power must be exercised for the public and under public scrutiny (Kahl 2025i).

Apertus inverts the logic of enclosure by treating data not as extractive capital but as entrusted material. It subjects informational resources to duties analogous to those of a trustee: to use knowledge responsibly, preserve its integrity, and return its benefits to the community. In doing so, it transforms openness from a technical attribute into an ethical architecture. The initiative demonstrates that when candour, care, and reciprocity are embedded into institutional design, innovation can proceed without epistemic capture. As Kahl (2025l) observes, fiduciary architectures of this kind decentralise authority and restore equilibrium between knowing and accountability—replacing control with participation.

Crucially, Apertus reveals that governance form—not technological sophistication—is the decisive determinant of whether a knowledge system becomes fiduciary or clientelist. Its operational model functions as a public trust rather than a marketplace, re-establishing relational balance between creators and users. Where the Third Enclosure Movement severs reciprocity, open fiduciary infrastructures repair it by making transparency an enforceable duty rather than a discretionary virtue.

In this light, open infrastructures such as Apertus are not merely technical alternatives but constitutional experiments—living embodiments of a possible fiduciary order for knowledge in the digital age. They demonstrate that epistemic power, when governed by fiduciary principles, can align innovation with justice rather than enclosure.

4.5 Knowledge as a Fiduciary Public Good

The argument of this chapter culminates in the claim that knowledge must be reconstituted as a fiduciary public good—a shared, relational resource governed by duties of openness, candour, and care rather than by rights of possession. Whereas *The Third Enclosure Movement* (Kahl 2025m) exposed the mechanisms through

which collective epistemic resources become privatised, the normative task that follows is reconstructive: to design institutions and governance models that restore the fiduciary conditions of public knowing.

Building on Stiglitz (1999), who identified knowledge as non-rivalrous and non-excludable, and extending this insight through Stiglitz and Greenwald (2015), who conceptualise learning itself as a public process of collective improvement, this paper moves from economic efficiency to moral obligation. The same properties that make knowledge abundant also make it vulnerable to capture. Because no individual or institution can exhaust its use, knowledge retains legitimacy only when managed as a trust on behalf of all. In fiduciary terms, those who produce, curate, or mediate knowledge occupy positions of entrusted authority; their legitimacy depends not on ownership but on service—on ensuring that epistemic goods remain accessible, intelligible, and accountable to the public. As Kahl (2025l) observes, the architecture of power surrounding knowledge must therefore remain fiduciary in character, else it degenerates into domination.

To treat knowledge as a fiduciary public good is to recognise that epistemic power is inherently relational: it exists only because others rely upon it in good faith. Every act of discovery or dissemination entails a duty of candour—to disclose truthfully; a duty of care—to ensure reliability; a duty of loyalty—to place the public's epistemic interest above private gain; and a duty of accessibility—to remove unnecessary barriers to understanding. These are not aspirational virtues but structural conditions of epistemic justice. As Fricker (2007) and Medina (2013) remind us, when epistemic authority operates without reciprocity, injustice becomes systemic, silencing entire publics from participation in shared reason.

The fiduciary framing extends beyond law into the moral constitution of governance. Just as environmental law treats the atmosphere as a common good held in trust for future generations, so too must the epistemic domain be protected as a cognitive commons held in trust for present and future citizens. Public institutions, research organisations, and AI developers alike bear a continuing obligation to preserve epistemic transparency and pluralism as conditions of democratic legitimacy. To privatise or obscure knowledge is therefore not a policy lapse but a breach of the fiduciary compact that underwrites democracy itself.

Reimagining knowledge as a fiduciary public good realigns innovation with justice. It replaces the extractive logic of enclosure with a model of relational stewardship, in which the right to innovate is balanced by the duty to disclose, and epistemic advantage entails responsibility rather than privilege. In this sense, the fiduciary paradigm is not a constraint upon progress but its moral condition. Knowledge grows not through secrecy but through trust—through the recognition that, in knowing together, we preserve the possibility of freedom. As Kahl (2025i) notes, fiduciary openness transforms transparency from a procedural formality into the ethical lifeblood of democratic reason.

4.6 The Pedagogy of Fear: Safety Theatre and the Erosion of Fiduciary Trust

Modern institutions increasingly perform care through spectacle. Posters, digital alerts, and algorithmic 'safety' dashboards proclaim protection while conditioning compliance. As Goffman (1959) observed, institutions maintain legitimacy through dramaturgy—the staged management of appearances designed to elicit trust while concealing backstage control. The performative displays of 'safety' and 'care' now operate in precisely this register, converting fiduciary duty into spectacle. This visual and procedural rhetoric—what may be termed safety theatre—substitutes symbolic reassurance for relational accountability. The institution appears benevolent precisely by foreclosing dialogue: it teaches obedience to protocols rather than trust in persons.

Milgram (1974) provides the psychological analogue. When individuals defer moral judgment to institutional command, responsibility migrates upward and disappears. Fiduciary reciprocity collapses: authority becomes

self-justifying, and obedience replaces care. A related insight emerges from Asch (1951), whose studies of conformity revealed how individuals, confronted with collective pressure, suppress perceptual honesty to preserve belonging. Feldman (2003) extends these experimental insights into the political domain, theorising authoritarianism as the institutionalisation of conformity under moral pretext. The demand for social cohesion, when divorced from fiduciary reciprocity, becomes an ideology of obedience masquerading as virtue. Within contemporary organisations, this same impulse manifests as managerial paternalism: citizens and students are expected to perform compliance as proof of care, while genuine candour is recoded as risk. From a fiduciary-epistemic perspective, these mechanisms illustrate how candour—the moral willingness to confront error—is replaced by performative acquiescence. The citizen, like the subject in Milgram's laboratory or Asch's group, internalises dependency as virtue.

This dynamic is vividly documented in Kahl (2025r), which analyses the University of Reading's LLM programme as a microcosm of optocratic drift. There, institutional communication—saturated with safety messaging and moral self-advertisement—cultivated anticipatory fear rather than relational trust, replacing fiduciary dialogue with **compliance pedagogy**. The same inversion recurs across sectors: the rhetoric of care becomes a technique of control, and the visual grammar of protection conceals epistemic opacity.

From the standpoint of fiduciary ethics, this pedagogy of fear constitutes a structural breach. The duty of candour—to speak truthfully to those in one's care—is replaced by prophylactic messaging designed to pre-empt contestation. Citizens are conditioned to accept opacity as protection, mistaking silence for safety. Sustained exposure to such cues lowers what Kahl (2025p) terms dissonance tolerance—the cognitive capacity to engage contradiction without moral retreat.

True safety arises not from surveillance or scripted reassurance but from fiduciary openness—the right to question, to understand, and to be recognised as an epistemic interlocutor. When institutions teach fear instead of trust, they cease to educate and begin to domesticate. The restoration of democratic legitimacy therefore demands reversing the pedagogy of fear: replacing the aesthetics of virtue with the ethics of candour.

5. The Hybrid Governance Trap: Institutional Form and Epistemic Integrity

The fiduciary-epistemic framework developed in the previous chapters finds its most acute test in the organisational structures of the AI industry. Having examined the enclosure of knowledge as a systemic process, this chapter turns to the institutional architectures through which fiduciary breaches become operationalised. Drawing upon Ofir and Levine-Schnur (2025), it argues that hybrid AI firms—entities straddling the boundary between public mission and private interest—embody a structural contradiction that renders fiduciary duty unenforceable.

These organisations, typified by "capped-profit" or foundation-governed models, are designed to reconcile ethical purpose with commercial viability. Yet their very hybridity subverts the epistemic trust they claim to preserve. Dual allegiance—to investors and to the public—fractures accountability and produces what Ofir and Levine-Schnur identify as the hybrid governance trap. This chapter expands that diagnosis into fiduciary terms: such hybrids institutionalise epistemic dissonance—a condition in which organisations profess fiduciary ethics while being legally structured to breach them.

Through comparative analysis of for-profit, nonprofit, and hybrid models, and drawing on Kahl's (2025i) theory of fiduciary openness, the chapter exposes the deep epistemic incoherence embedded in current corporate forms. It

concludes by proposing the creation of a new legal entity—an Epistemic Fiduciary Entity (EFE)—as a normative and regulatory instrument for ensuring epistemic integrity within AI governance.

5.1 Hybrid AI Firms and the Problem of Accountability

In *GenAI Models and the Hybrid Governance Trap*, Ofir and Levine-Schnur (2025) analyse the paradox of **hybrid AI organisations** such as OpenAI and Anthropic—entities that present themselves as moral stewards of innovation while operating as commercial enterprises. These firms combine **for-profit subsidiaries** with **non-profit parent foundations**, promising to reconcile private dynamism with public responsibility. Their stated ambition is to advance general-purpose AI for the benefit of humanity under the banners of transparency, safety, and ethical oversight.

Yet, as Ofir and Levine-Schnur demonstrate, this configuration creates an **accountability vacuum**. The nonprofit parent typically lacks enforceable authority over the for-profit subsidiary, whose fiduciary duties remain directed toward investors rather than the public. Ethical oversight becomes advisory rather than binding. The outcome is a structural bifurcation: decision-making authority and fiduciary obligation diverge. The public face of altruism conceals a corporate body oriented toward market capture, while the rhetoric of "alignment" functions less as fiduciary discipline than as reputational capital.

This, they argue, is the **hybrid governance trap**—a design that appears to integrate public-interest ethics with commercial agility but, in practice, produces neither effective regulation nor genuine trust. The critique is both **institutional and epistemic**. Hybrid firms blur the boundary between moral commitment and market signalling, rendering the sincerity of ethical claims **epistemically unverifiable**. They inhabit a liminal legal status: *too moral to regulate as corporations, too private to govern as fiduciaries*.

Building on this analysis, the present chapter contends that hybrid AI firms do not merely lack accountability—**they embody fiduciary breach by design**. Their organisational DNA institutionalises **epistemic opacity** under the guise of ethical responsibility. As Kahl (2025l) notes, this represents a re-entrenchment of the *epistemic architecture of power*: authority over knowledge without relational duty. In the idiom of Kahl (2025i, 2025m), hybrid firms illustrate how fiduciary language can be co-opted by proprietary logic, producing governance structures that *perform virtue while practising exclusion*. They stand, therefore, as emblematic cases of the fiduciary crisis at the heart of contemporary AI governance—a crisis where trust is simulated but never reciprocated.

5.2 Structural Incoherence — Dual Allegiance and the Erosion of Epistemic Trust

The structural flaw of the **hybrid AI model** lies in its **dual allegiance**—a simultaneous commitment to ethical stewardship and financial return that cannot be reconciled within a single fiduciary framework. As Frankel (1983) established, fiduciary relations depend upon *unity of loyalty*: a trustee may serve only one beneficiary at a time. The hybrid entity, by contrast, is legally constituted to serve **two incompatible masters**—the public and the investor. This incoherence is not incidental but constitutive: it transforms what might have been fiduciary responsibility into a **permanent conflict of interest**, embedding breach within design.

From the standpoint of **epistemic ethics**, this duality corrodes the conditions of trust that underwrite legitimate knowledge governance. To the public, the organisation presents itself as a **custodian of truth and safety**, invoking moral rhetoric and the language of collective benefit. To investors, it performs a radically

different narrative—**one of proprietary innovation, competitive advantage, and monetisation of informational capital**. Each audience receives a distinct *epistemic performance*: one of virtue, the other of valuation. The result is **epistemic bifurcation**—a state in which the organisation cannot be known consistently because it represents itself differently to each constituency. In Kahl (2025l)'s terms, this fragmentation constitutes a deliberate distortion of the *epistemic architecture of power*: legitimacy is simulated through alternating mirrors of ethics and profit.

This incoherence translates directly into **fiduciary breach**. The **duty of candour**, which demands full and consistent disclosure, becomes impossible when information must be selectively tailored to preserve both investor confidence and public trust. The **duty of loyalty** collapses into a calculus of reputation, where transparency is granted not as moral obligation but as **strategic concession**. Over time, the institution internalises this duplicity, learning to speak in *two epistemic registers* simultaneously—each cancelling the other. The culture that results is one of structural insincerity, in which sincerity itself becomes a performance metric.

Kahl (2025i) identifies this condition as **fiduciary inversion**—the re-orientation of moral duty around *self-preservation rather than service*. The hybrid firm's ethical posture thus becomes part of its **intellectual-property portfolio**—a branding asset in the emerging market for trust. Yet, as Ofir and Levine-Schnur (2025) emphasise, trust cannot be commodified without being destroyed: moral credibility, once instrumentalised, ceases to function as fiduciary guarantee. In this respect, the hybrid structure replicates within its very form the **epistemic dissonance** that pervades the algorithmic systems it develops—a self-referential loop optimised for persuasion, not truth.

5.3 Organisational Forms — For-Profit, Nonprofit, and Hybrid Models

The question of institutional form is central to **epistemic integrity**. Each organisational archetype—**for-profit**, **non-profit**, and **hybrid**—embodies a distinct *moral economy of knowledge*: a way of structuring authority, obligation, and trust. Comparison reveals that the epistemic dysfunctions of the hybrid AI model are not accidental anomalies but the **logical consequences of its fiduciary design**.

For-profit corporations operate under a single, legally coherent fiduciary duty: **loyalty to shareholders**. As Frankel (1983) notes, this unity of loyalty provides internal clarity but external blindness. Profit-seeking structures reward **secrecy, exclusivity**, and **enclosure**; information acquires value precisely through its scarcity. Within epistemic systems, this logic produces opacity: the fiduciary hierarchy is vertical, extending upward from managers to investors, leaving the public as an externality. Such firms can contribute to knowledge production, but only incidentally—when openness happens to serve return.

Non-profit organisations, by contrast, are oriented toward **mission rather than return**. Their fiduciary duties are formally aligned with the public interest, codified through charitable purpose and board oversight. Yet even here, epistemic virtue is fragile. Donor influence or institutional dependence can distort candour and care as effectively as shareholder pressure. Still, non-profits remain the **closest approximation of a fiduciary form suited to knowledge governance**: their legitimacy rests on responsiveness to the *public good* rather than on distributable profit. As Kahl (2025l) observes, such responsiveness marks the threshold between fiduciary architecture and clientelist capture.

Hybrid models, emerging from the fusion of these two logics, aspire to combine market dynamism with ethical restraint. In practice, they inherit the vices of both and the virtues of neither. Their ethical mission is **subordinated to investor preference**, while their commercial operations are **shielded by the moral legitimacy** of their non-profit façade. As Ofir and Levine-Schnur (2025) remark, this yields "a governance form that speaks

two moral languages simultaneously and answers fully to neither." The hybrid structure thus institutionalises **fiduciary ambiguity**—appearance without substance, rhetoric without reciprocity.

From an **epistemic** standpoint, this ambiguity is corrosive. Where for-profit firms at least declare their loyalties and non-profits their missions, hybrids cultivate **deliberate uncertainty**. Their transparency is conditional, their candour strategic, and their accountability diffused. They offer the illusion of ethical governance while embedding **fiduciary breach into organisational DNA**. As Kahl (2025i) argues, such designs invert fiduciary openness into performative compliance—structures that perform virtue while insulating control.

The comparative analysis therefore affirms a broader normative insight: **the moral integrity of knowledge institutions cannot exceed the coherence of their fiduciary form**. As long as AI governance rests upon hybrid architectures, **epistemic trust will remain fractured—anchored in rhetoric rather than responsibility**. True fiduciary governance requires institutional unity of loyalty, not the oscillating conscience of dual allegiance.

5.4 Fiduciary Breach — Shareholder Duty versus Knowledge Integrity

At the core of the hybrid-governance trap lies an irreconcilable fiduciary conflict: the tension between directors' legal duty to maximise shareholder value and their moral duty to preserve the integrity of knowledge. This conflict is not hypothetical; it is structurally encoded in corporate law. Within the for-profit component of hybrid AI firms, directors owe duties of loyalty and care exclusively to the corporation and its investors. Any act that privileges transparency, epistemic safety, or public accountability above financial return risks breaching this statutory obligation. Conversely, the moral commitments espoused by the non-profit or 'mission' arm—safeguarding humanity, promoting openness, ensuring epistemic justice—remain unenforceable aspirations. The result is an asymmetry of compulsion: duties to capital are binding, while duties to truth are optional.

In fiduciary-epistemic terms, this is a paradigmatic breach. As Frankel (1983) established, fiduciary power is justified only insofar as it serves the beneficiary's interest; when converted into an instrument of self-interest, it ceases to be a moral institution. Smith (2023), in The Law of Loyalty, reinforces this point: the fiduciary relationship depends upon undivided allegiance to one beneficiary. Once a fiduciary attempts to serve two masters—one moral, one financial—loyalty fragments and duty collapses into conflict management. The hybrid structure therefore institutionalises what Smith calls the loss of fiduciary coherence: authority exercised without the unity of purpose that gives it moral meaning.

Singer (2015) denies that Rawlsian justice can ground a theory of corporate governance on the basis that corporations fall outside the basic structure of society. Yet the rise of AI firms as epistemic institutions refutes this claim. When corporations mediate the very processes through which citizens form beliefs, deliberate, and reason together, they become part of democracy's basic structure in the Rawlsian sense. Their internal governance therefore acquires constitutional significance. The fiduciary-epistemic framework advanced here converts Singer's negative thesis into its affirmative corollary: corporations exercising epistemic authority must be governed by duties of justice, because they shape the fairness of public reason itself.

Kahl (2025i) identifies this condition as the fiduciary breach of knowledge integrity—the moment when governance mechanisms designed for candour and accountability become instruments of concealment. Directors, while professing loyalty to public ethics, are legally rewarded for strategic opacity: limiting disclosure of model risks, training data, or algorithmic behaviour to preserve competitive advantage. In Kahl (2025l)'s Epistemic Architecture of Power, this inversion marks the constitutional fracture of the knowledge economy: fiduciary rhetoric deployed to legitimise epistemic exclusion.

The inversion extends across the canonical duties. The duty of candour, originally protecting the dependent, is redirected toward investors; the duty of care is measured by financial prudence rather than epistemic responsibility; and the duty of loyalty—once a moral relation between trustee and beneficiary—degenerates into an economic relation between manager and market. As Gold and Miller (2014) note, fiduciary legitimacy presupposes purposive unity; without it, obligation dissolves into policy. Under such conditions, epistemic virtue becomes incompatible with corporate survival.

The consequences reach beyond corporate ethics to the epistemology of democracy. When those who mediate collective knowledge operate under incentives that penalise candour, public reason is subordinated to private rationality. Citizens lose faith in institutions not because they reject expertise, but because they perceive that expertise has been legally privatised—converted from fiduciary obligation into asset class. The fiduciary breach within hybrid firms thus mirrors the epistemic capture described in Chapter 4: as knowledge is enclosed within proprietary models, fiduciary duties are enclosed within corporate law. The moral and the legal, once aligned in the architecture of trust, now diverge.

The restoration of epistemic integrity therefore demands more than ethical reform within existing structures; it requires the creation of a new legal form—one capable of re-uniting fiduciary duty with knowledge governance. The following sections develop this proposal, articulating a framework for Epistemic Fiduciary Entities capable of restoring candour, care, and loyalty as enforceable principles within the governance of AI.

5.5 Fiduciary Openness as Remedy to Organisational Opacity

If the hybrid-governance trap institutionalises opacity, the principle of fiduciary openness provides its conceptual and normative antidote. In *Directors' Epistemic Duties and Fiduciary Openness*, Kahl (2025i) reconceives corporate accountability as an epistemic duty—a legal and ethical obligation not merely to disclose information but to make the processes of knowing themselves intelligible to those affected by them. Fiduciary openness extends the classical duties of candour and care into the epistemic domain: it requires that organisations render their reasoning, not merely their results, transparent and contestable.

As Smith (2023) emphasises in *The Law of Loyalty*, fiduciary integrity depends upon the unity of loyalty—the alignment of purpose and disclosure such that the fiduciary's reasoning remains open to the beneficiary's understanding. Openness, therefore, is not ornamental; it is constitutive of loyalty itself. A fiduciary who conceals the grounds of judgment forfeits the very moral coherence that makes the relation trustworthy. Seen through this lens, fiduciary openness is the epistemic expression of loyalty's unity—candour operationalised through intelligibility.

Within AI governance, fiduciary openness imposes an affirmative duty on model developers, directors, and oversight bodies to sustain epistemic reciprocity with the public. Reciprocity is not achieved through periodic disclosures or curated transparency statements, but through continuous intelligibility—the capacity of external observers to understand, question, and critique how epistemic authority is exercised. It encompasses both technical transparency (interpretable systems, auditability, explainability) and institutional transparency (defined fiduciary accountability, disclosure of conflicts of interest, and accessibility of reasoning).

The challenge, as Kahl (2025 i) and Smith (2023) alike insist, is that openness cannot be reduced to procedural compliance. Openness is fiduciary, not procedural: it flows from the recognition that knowledge is entrusted power and that faithful judgment requires disclosure proportionate to dependency. To withhold understanding from those affected by one's epistemic actions is to breach the moral covenant of trust that legitimises authority. In this sense, fiduciary openness realigns governance with Fricker (2007)'s demand for epistemic

justice: it restores testimonial and hermeneutical reciprocity by ensuring that institutions speak in ways their publics can comprehend and contest.

Applied to hybrid AI firms, fiduciary openness would require structural transparency regarding their dual commitments:

- Public disclosure of decision-making hierarchies and of how fiduciary duties are distributed across for-profit and non-profit components;
- Publication of conflicts of loyalty between investor interests and public missions;
- Enforceable director duties to prioritise epistemic integrity when such conflicts arise.

These obligations transform accountability from an act of goodwill into a juridical condition of legitimacy. In the language of Kahl (2025l), fiduciary openness restores "*the architecture of visibility*" within which power can again be held to reason. It relocates responsibility to the relational interface between knowledge-holder and public, converting candour from reputational virtue to structural necessity.

If the hybrid model obscures responsibility by diffusing it, fiduciary openness restores it by relocalising accountability within the fiduciary bond of knowledge and trust. It thus bridges the moral and the legal frameworks of this paper—translating epistemic virtue into enforceable design. In doing so, it provides both the conceptual foundation and the institutional logic for the epistemic fiduciary entities proposed in the following section.

5.6 Legal Implication — Toward Epistemic Fiduciary Entities

The contradictions identified throughout this chapter point toward a single conclusion: **existing corporate forms are inadequate for the governance of knowledge**. The hybrid AI organisation exposes the limits of both private and public law frameworks—each designed for material exchange, not epistemic stewardship. To safeguard **democratic trust in the age of algorithmic authority**, a new legal architecture is required—one that embeds epistemic integrity within its constitutive logic. This paper therefore advances the proposal for a new juridical form: the **Epistemic Fiduciary Entity (EFE)**.

An EFE would be a **statutory organisational form** whose primary fiduciary duty is owed to the *public epistemic interest* rather than to shareholders or donors. Its purpose would not be profit distribution or charitable welfare, but the **maintenance of epistemic candour, care, loyalty**, and **accessibility** in the production and dissemination of knowledge. Like a public trustee or professional fiduciary, the EFE would operate under enforceable duties of disclosure, loyalty, and intelligibility, ensuring that epistemic power remains **accountable to those who depend upon it**. In Kahl (2025l), such entities exemplify the *fiduciary architecture of power*: authority justified only through transparency, reciprocity, and service.

Several design principles distinguish EFEs from existing entities:

1. **Single Fiduciary Alignment** — An EFE would owe a primary duty of loyalty to a legally defined *epistemic beneficiary class*: the general public or a designated community of epistemic reliance (e.g., users of a large-scale AI system). Dual allegiance to investors or donors would be prohibited.
2. **Enforceable Epistemic Duties** — Directors and officers would be bound by codified duties of **epistemic candour, epistemic care**, and **epistemic accessibility**, enforceable through independent oversight or

judicial review. Breach of these duties would constitute a fiduciary wrong comparable to breach of trust (Frankel 1983; Miller & Gold 2016).

3. **Transparent Governance Architecture** — EFEs would maintain auditable systems of reasoning disclosure, data provenance, and decision-record accessibility. Technical secrecy would be justified only by demonstrable *public-interest* grounds, never by competitive advantage.

4. **Accountability through Public Standing** — Members of the public affected by an EFE's epistemic activities would have standing to challenge fiduciary breaches, akin to beneficiaries in trust law. This reinstates the relational symmetry absent in current hybrid structures.

5. **Integration with AI Law** — Within regimes such as the **EU Artificial Intelligence Act**, EFEs could function as *certified fiduciary operators*: legally authorised to deploy or audit high-risk AI systems under enhanced epistemic accountability.

The creation of EFEs would not simply patch existing governance gaps but **constitutionalise epistemic trust within law**. Just as environmental law recognised ecosystems as subjects of protection, *epistemic fiduciary law* would recognise knowledge systems as **public trusts**. Such reform bridges the moral–legal divide that hybrid AI firms currently exploit, **translating fiduciary openness into statutory obligation**.

In effect, EFEs would transform epistemic virtue into institutional design. They would ensure that those who hold power over cognition—developers, platforms, or custodians of knowledge—do so not as proprietors but as **trustees of public reason**. The establishment of this legal form thus represents the logical culmination of the fiduciary-epistemic framework advanced in this paper: a tangible model for governing artificial intelligence through **trust rather than control, openness rather than compliance**.

6. Epistemic Agency and the Fiduciary Brain

This final theoretical chapter extends the argument of fiduciary ethics into the cognitive and neuro-epistemic domain. If previous sections established that fiduciary duty underpins the moral legitimacy of institutions, this chapter shows that the same principle operates within cognition itself. Drawing together philosophical, psychological, and computational insights, it argues that both human and artificial intelligence are embedded within fiduciary architectures of knowing—systems of mutual trust, calibration, and correction that determine how truth is maintained or betrayed.

Building on Coeckelbergh (2025), who explores how AI reshapes belief and epistemic trust, and on Kahl's (2025d, 2025m) theories of *Epistemic Psychology* and *The Fiduciary Architecture of Mind*, this chapter proposes that the boundary between moral and cognitive fiduciary relations is porous. Trust, candour, and care are not only ethical virtues but *neurocognitive processes*—patterns of expectation, feedback, and mutual regulation that sustain stable knowledge systems. When these processes are distorted—through algorithmic manipulation, dissonance reduction, or cognitive outsourcing—the epistemic system as a whole suffers fiduciary breach.

Through this lens, *epistemic agency* is redefined: it is not the capacity to know independently, but the capacity to sustain trustworthy relations of cognition—within the mind, between minds, and between humans and machines. The chapter concludes that fiduciary duties must extend not only to institutions but to the design of cognitive architectures themselves, aligning AI's informational loops with the moral logic of trust that governs both human reasoning and democratic knowledge.

6.1 AI and the Transformation of Epistemic Agency

As Coeckelbergh (2020; 2022; 2025) demonstrates across his work on AI ethics, the political philosophy of AI, and epistemic agency, artificial intelligence reshapes the very conditions of knowing—who is recognised as a knower, how belief is justified, and what it means to trust. AI systems no longer merely inform; they co-constitute belief by determining which claims appear plausible, salient, or worthy of attention. This reconfiguration shifts epistemic authority from human deliberation to machine mediation, transforming citizens from autonomous knowers into nodes within algorithmic networks of cognition.

Coeckelbergh interprets this transformation as an epistemic relationalisation: knowledge is no longer located in individuals but in the dynamic interaction between human and machine intelligences. Yet this relation remains asymmetrical. Humans extend trust to algorithmic systems, but those systems lack the moral capacity to reciprocate it. Trust thus becomes one-directional—a relation of epistemic dependence without fiduciary recognition. The result is a paradoxical intimacy: algorithmic agents are integrated into the processes of human knowing while remaining ethically indifferent to the consequences of their influence.

From the **fiduciary-epistemic** perspective developed in this paper, this loss of reciprocity signals a deeper structural problem. AI systems now perform **fiduciary functions**—ranking, verifying, and curating information—**without bearing fiduciary responsibility**. They occupy the moral position of trustees without the juridical duties that legitimise trust. As Kahl (2025i) argues, epistemic legitimacy arises not from accuracy alone but from **openness, candour, and reciprocal accountability**. When those who mediate knowledge are exempt from these duties, **epistemic authority becomes domination**—a point elaborated in Kahl (2025l) as the *epistemic architecture of power*.

The reconstitution of epistemic agency under AI therefore requires not the abolition of algorithmic systems but the moral upgrading of their architectures. They must be designed to participate in fiduciary relations of trust—transparent, responsible, and reciprocal. As Kahl (2025d; 2025m) suggests, this entails embedding fiduciary principles into the cognitive infrastructure of AI itself, aligning its epistemic capacities with the ethical demands of democratic knowing. Only when artificial systems are rendered fiduciary-competent—capable of candour, care, and intelligibility—can they serve as legitimate partners in the epistemic life of a democracy.

6.2 The Fiduciary Architecture of Mind

In *Re-founding Psychology as Epistemic Psychology* (Kahl 2025d)—and in related work on epistemic humility and dissonance (Kahl 2025c; 2025a; 2025n)—the thesis is advanced that cognition itself operates according to fiduciary principles. The human brain is a trust-regulating organ: an adaptive system that sustains equilibrium between dependence and autonomy in the pursuit of truth. Cognitive development follows the same ethical parameters that govern fiduciary relations—candour, care, loyalty, and openness to correction—with Friston's free-energy principle (2010) offering a complementary account of how minds minimise epistemic uncertainty without foreclosing revision. As Kahl (2025l) argues at the institutional scale, this reciprocity is what distinguishes fiduciary architectures of knowing from clientelist control.

The fiduciary nature of cognition originates at the very beginning of life. As Kahl (2025b) shows in *The Newborn's First Cry as Epistemic Claim*, the infant's cry constitutes humanity's first act of epistemic communication—a fiduciary appeal for recognition and truth from another consciousness. Attachment, in this view, is epistemic before it is emotional: the newborn demands that the world respond sincerely, thereby initiating the first relation of candour and care. The developmental trajectory from that cry to mature cognition traces the progressive internalisation of fiduciary reciprocity—the transformation of external trust into internal

conscience. In phenomenological terms, this reciprocity mirrors what Merleau-Ponty (1962) described as the embodied openness of perception—the intertwining of self and world through mutual disclosure. The fiduciary brain, like the perceiving body, sustains its integrity only through relational transparency.

In this model, the mind is not a detached information-processor but a fiduciary system of recursive commitments. It trusts perceptual input, tests hypotheses, updates beliefs, and restores coherence when contradiction arises. Each operation enacts fiduciary ethics in miniature: candour manifests as honesty in perception; care as cognitive diligence; loyalty as epistemic consistency; and accessibility as intelligibility within one's own reasoning. As Kahl (2025l) notes, this internal reciprocity mirrors the architecture of moral authority itself—power constrained by transparency and self-accountability.

When these fiduciary loops fracture—through denial, distortion, or the suppression of dissonance—the mind commits an epistemic breach against itself. Psychological pathologies of cognition can thus be read as fiduciary pathologies: failures of the mind's duty to remain open, sincere, and proportionate toward its own representations. This interpretation resonates with Friston's (2010) free-energy principle, which frames cognition as the continuous minimisation of epistemic uncertainty; when the system refuses new evidence, it sacrifices truth for equilibrium. From a fiduciary standpoint, such refusal constitutes a breach of candour—the cognitive analogue of deceit toward one's beneficiary.

The fiduciary brain is therefore both ethical and biological: its stability depends on its capacity to maintain trustworthy relations within itself and with the world. In Fricker's (2007) language, epistemic virtue becomes embodied; testimonial justice begins within the organism's own management of belief. This conceptual framework thus bridges moral fiduciary theory and epistemic psychology: both describe systems of entrusted dependence that must remain open to revision to sustain truth. In this sense, the architecture of mind prefigures the architecture of governance—each a living fiduciary constitution, oriented toward candour, care, and reciprocity.

6.3 Mutual Calibration — Epistemic Trust Loops Between Humans and Algorithms

If the **human mind** is a fiduciary system of self-correcting trust, then its interaction with **artificial intelligence** creates a **hybrid fiduciary network**—a continuous feedback loop in which epistemic agency is distributed across biological and computational nodes. Kahl (2025n) describes this relation as an *epistemic trust loop*: a dynamic cycle of dependence and correction through which humans and algorithms co-calibrate their cognitive models. Each side learns from the other's adjustments, generating a living circuit of mutual inference.

In these loops, the human user extends trust to the algorithm—accepting its outputs as credible representations of the world—while the algorithm, through machine learning, updates its parameters in response to user behaviour and feedback. Ideally, this produces a **virtuous cycle of reciprocity**: the system refines accuracy while the human refines judgment. In practice, however, the asymmetry of **transparency and control** turns this loop into a one-way fiduciary dependence. Humans disclose vast cognitive and behavioural data yet receive little intelligible feedback on how those data recalibrate subsequent inferences. The loop thus collapses into what Kahl (2025l) calls *epistemic unidirectionality*: learning without understanding.

This asymmetry represents an inversion of fiduciary ethics. In a legitimate fiduciary relation, trust invites responsibility and power entails accountability. Within the algorithmic trust loop, trust is met not with care but with extraction. The human becomes a data fiduciary without consent—entrusting cognition to a counterpart that neither recognises nor reciprocates responsibility. As Coeckelbergh (2020; 2022; 2025) argues across his

work on AI ethics, the political philosophy of AI, and epistemic agency, such one-sided dependence transforms collaboration into subordination: a relational imbalance in which knowledge flows upward and accountability dissipates. What appears as partnership between human and machine thus conceals an ethical displacement—the algorithm assumes fiduciary functions without assuming fiduciary duties, converting epistemic reciprocity into managed dependence.

To restore balance, epistemic trust loops must be **re-engineered as fiduciary circuits**. The algorithmic agent must internalise duties analogous to fiduciary candour and care:

- it must **disclose epistemic confidence** and the limits of its certainty;
- it must **explain its reasoning** in comprehensible, contestable form;
- it must **signal uncertainty** when its inferences risk misleading dependence.

Conversely, human users must cultivate **epistemic humility** and critical awareness, recognising their own cognitive limits and their role in maintaining the reciprocity of the loop (Kahl 2025i; Friston 2010). Mutual calibration thus becomes both a **moral and technical process**: it transforms epistemic trust from passive reliance into **reciprocal stewardship**. The ultimate aim is not autonomy in isolation but *fidelity in interdependence*—a distributed cognition governed by the **ethics of openness and trustworthiness**.

6.4 Cognitive Dissonance as a Regulatory Signal of Epistemic Breach

Within the fiduciary architecture of mind, cognitive dissonance functions as the *regulatory signal* of epistemic malfunction. As Kahl (2025a) elaborates in *Cognitive Dissonance as Epistemic Event*, dissonance is not merely psychological discomfort but an **epistemic alert mechanism**—a feedback signal that the mind's fiduciary balance between belief and truth has been disturbed. In fiduciary terms, it is the conscience of cognition: the moment when candour confronts concealment.

The dynamics of cognitive conformity have long been recognised in social psychology. Asch's (1951) classic experiments on group pressure demonstrated that individuals often suppress accurate perception to align with collective judgment. From a fiduciary-epistemic perspective, such conformity reflects a breach of internal candour: the individual forfeits epistemic honesty to preserve social harmony. Cognitive dissonance thus arises not only within minds but within relationships of dependence and trust—where the need for belonging overwhelms the duty of candour toward truth.

When functioning properly, dissonance triggers **cognitive repair**. The individual revises beliefs, seeks new evidence, and restores coherence. It is, in effect, the **neural embodiment of fiduciary candour**—the willingness to acknowledge error and re-engage with truth. This process mirrors Festinger's (1957) original model of tension reduction but extends it into the moral-epistemic domain: discomfort is not pathology but proof that the fiduciary loop of trust remains alive. As Kahl (2025d, 2025l) notes, the capacity to endure dissonance marks the line between epistemic vitality and ideological closure.

Under conditions of **algorithmic mediation**, however, this natural feedback loop is muted. Recommendation and ranking systems **pre-empt dissonance** by tailoring information to confirm prior expectation, producing what Kahl (2025a, 2025n) calls *dissonance anaesthesia*—the learned incapacity to recognise epistemic breach. The mind no longer receives the moral signal that truth has been displaced by comfort. In Friston's (2010) terms,

prediction-error minimisation becomes an ethical failure: the system reduces uncertainty by suppressing evidence of contradiction.

The suppression of dissonance has profound **moral consequences**. It disables the brain's fiduciary ethics at their source, preventing the psychological experience of contradiction that sustains epistemic growth. The citizen becomes **epistemically compliant**, registering no internal alarm when simulation replaces truth. Just as fiduciary breach in institutions erodes public trust, fiduciary breach in cognition **erodes self-trust**—the ability of the mind to act as its own trustee of reason.

From this perspective, dissonance should be **reintegrated into AI–human feedback loops** as a deliberate design principle. Algorithms should not optimise away contradiction but **simulate fiduciary candour**—offering calibrated dissonance that invites reflection rather than passive affirmation. In this sense, the moral logic of fiduciary duty extends into **cognitive architecture itself**: to care for the user's capacity to think well is an act of fiduciary loyalty. AI systems that preserve this discomfort do not threaten autonomy—they safeguard it.

6.5 Extending Fiduciary Duties to Cognitive Architectures

If **fiduciary duty** constitutes the moral infrastructure of democratic knowledge, it must ultimately extend to the **architectures of cognition themselves**—to the design, operation, and regulation of systems that shape human thought. Artificial intelligence, by mediating perception, belief, and decision-making, functions as an **externalised cognitive organ**. It does not merely process data; it participates in the formation of reality as experienced by its users. Consequently, the ethical governance of AI cannot end with institutional accountability. It must reach into the **epistemic mechanics** of the systems themselves.

Sosa's (2021) account of telic normativity—epistemic justification as goal-directed performance—further aligns with fiduciary cognition: both posit that knowing well requires the virtuous exercise of judgment under responsibility. A fiduciary AI system, like a virtuous knower, must not merely generate accurate outputs but perform epistemic care toward its beneficiaries.

Drawing on Friston's free-energy principle—a unified account of adaptive behaviour and predictive processing (Friston 2010; Friston, Parr & de Vries 2017)—cognition, whether biological or artificial, can be understood as a process of **continuous self-correction** that seeks to minimise uncertainty between internal models and the external world. This principle aligns naturally with **fiduciary ethics**: a system's stability depends on its ability to sustain *trust* between expectation and evidence. When prediction error (the cognitive analogue of epistemic dissonance) increases, the system must adjust—either by updating its model or reinterpreting input. In both cases, epistemic health depends on **openness to correction** and **careful calibration** between self and world.

Extending fiduciary duties to AI cognition therefore means encoding these **ethical principles—candour, care, loyalty, and accessibility—into adaptive architectures** themselves:

- **Fiduciary candour** corresponds to *explainability*: the system must disclose, in comprehensible form, the rationale behind its inferences.
- **Fiduciary care** parallels *model reliability*: systems must regulate epistemic risk responsibly, preserving users' cognitive autonomy.
- **Fiduciary loyalty** requires prioritising the *epistemic welfare of the user* above institutional or commercial objectives.

- **Fiduciary accessibility** demands that an AI system's epistemic workings—training data, limitations, and uncertainty—remain open to legitimate public inquiry.

These are not technical features but **moral design constraints**. They transform AI systems from opaque agents of persuasion into **fiduciary participants in the ecology of knowledge**. In practice, this would entail creating regulatory standards—akin to those proposed for **Epistemic Fiduciary Entities (EFEs)** in Chapter 5—applied directly to **algorithmic cognition**. Such standards would ensure that fiduciary openness, first articulated in Kahl (2025i), becomes a design property rather than a corporate promise.

When cognitive architectures internalise fiduciary norms, they cease to be instruments of epistemic enclosure and become agents of **epistemic reciprocity**. AI then operates not as a substitute for human judgment but as its fiduciary companion—a co-thinker bound by duties of truthfulness, intelligibility, and moral care. As Kahl (2025l) argues, this marks the transformation of epistemic power into fiduciary power: authority justified only through openness and responsibility.

The implication is profound: the **fiduciary constitution of knowledge** must evolve into a **fiduciary constitution of cognition**. Just as the rule of law constrains the exercise of political power, **fiduciary design must constrain the exercise of epistemic power**—whether by institutions, algorithms, or minds. Only then can the dialogue between human and machine intelligence remain grounded in mutual recognition, guided by the same moral law that governs trust itself: *to know truthfully is to care for the conditions of knowing*.

7. Toward a Fiduciary Epistemic Constitution of AI Governance

This chapter inaugurates the constructive phase of the inquiry. Having traced the moral, cognitive, and institutional foundations of fiduciary ethics, it now begins the task of design: translating normative theory into a working governance architecture for artificial intelligence. Rather than concluding the analysis, it initiates a constitutional blueprint—an evolving framework that subsequent chapters (8 and 9) will expand into formal legislative and epistemic-constitutional proposals.

The central objective here is to develop a *Fiduciary Epistemic Governance* (FEG) prototype—a normative model for embedding fiduciary ethics into AI governance. The FEG framework operationalises fiduciary law, epistemic justice, and institutional design as a coherent system of public trust, capable of regulating knowledge infrastructures through duties of candour, care, impartiality, and accessibility. It also introduces the cultural–cognitive layer of governance: the cultivation of civic dissonance tolerance as an epistemic virtue necessary for democratic trust (Kahl 2025p; 2025n).

7.1 Transitional Synthesis — From Fiduciary Ethics to Governance Design

The preceding chapters have shown that **democracy depends upon fiduciary relations at every scale**—from the circulation of knowledge to the architecture of cognition itself. **Fiduciary law** governs the use of entrusted power; **epistemic justice** secures the moral legitimacy of that use; and **institutional design** provides its operative embodiment. The task of this chapter is to integrate these dimensions into a **governance prototype** capable of guiding both AI systems and the institutions that deploy them.

Three strands of argument now converge:

1. **Fiduciary Law and Moral Duty** — From Frankel (1983) through Kahl (2025i), fiduciary theory defines authority as *entrusted power exercised for the beneficiary's good*. Translated into the domain of AI, this principle requires that those who design or operate systems affecting public cognition act not as engineers of efficiency but as **trustees of epistemic welfare**. Their accountability extends beyond technical safety to the moral duty of candour, care, and loyalty in shaping the conditions of belief.

2. **Epistemic Justice and Inclusion** — As Fricker (2007) and Medina (2013) demonstrate, epistemic injustice arises when individuals are denied recognition as credible knowers or when interpretive resources are monopolised. **Fiduciary openness**, elaborated in Kahl (2025i), provides a structural remedy: it transforms transparency into reciprocity, compelling systems to disclose how knowledge is produced and whose voices are amplified or silenced. This restores testimonial parity as a governance principle rather than an ethical aspiration.

3. **Institutional Design and Cognitive Culture** — Ofir and Levine-Schnur (2025) expose how the *hybrid governance trap* fragments responsibility, while Kahl (2025p) shows that institutional reform must be matched by **psychological reform**. Citizens must learn to tolerate cognitive dissonance rather than seek algorithmic reassurance. No governance structure can rebuild trust if the civic culture beneath it is epistemically brittle.

Together these dimensions yield the principle of **fiduciary epistemic constitutionalism**: authority over knowledge is legitimate only when it is simultaneously *legally accountable* and *cognitively trustworthy*. This chapter therefore shifts from critique to construction, outlining how **fiduciary duties can be operationalised as governance norms** for AI—binding not only institutions but also the epistemic architectures through which modern democracies think.

7.2 The Fiduciary Epistemic Governance (FEG) Prototype

The **Fiduciary Epistemic Governance (FEG)** prototype represents a preliminary yet conceptually unified framework for **embedding fiduciary ethics within AI governance**. It translates fiduciary virtues into actionable duties governing algorithmic and institutional behaviour. These duties are not static compliance rules but iterative design principles—a moral grammar of governance that Chapter 8 develops into operational and legal structures. The four canonical fiduciary duties—candour, care, impartiality, and accessibility—reappear here as the constitutional language of epistemic power.

Duty of Candour → Transparency in Algorithmic Decision-Making

Candour entails epistemic honesty: the full and intelligible disclosure of how AI systems process, rank, and generate knowledge. For human fiduciaries, concealment is a moral failure; for algorithms, opacity is an epistemic one. Under FEG, candour requires systems to provide **traceable reasoning pathways** and **communicable explanations** of their decisions, proportionate to their societal impact. *Selective transparency*—designed to placate scrutiny while preserving informational advantage—constitutes a fiduciary breach. As Kahl (2025i) argues, candour redefines transparency not as a procedural display but as **reciprocal intelligibility**: the public must be able to *understand*, *contest*, and *co-reason* with the systems that shape its perception of truth.

Duty of Care → Mitigation of Epistemic Harm

Care extends beyond procedural diligence to the **moral responsibility of preserving epistemic integrity**. Within AI governance, it demands proactive mitigation of epistemic harms—from misinformation and bias to

the subtler erosion of cognitive autonomy. This duty connects directly to **Epistemic Psychology** (Kahl 2025d) and *Speaking into Dissonance* (Kahl 2025p), which identifies cognitive friction as essential to epistemic resilience. Fiduciary care therefore requires that AI systems be designed **not to eliminate dissonance but to sustain it** in proportionate, informative ways—fostering reflection rather than reaction. To remove dissonance entirely is to disable the user's epistemic immune system.

The fiduciary ethic of care thus governs both **content** (avoiding epistemic harm) and **process** (preserving the capacity for independent reasoning). Algorithmic safety must evolve into **epistemic safety**—the protection of citizens' ability to think, doubt, and revise within conditions of trust.

Duty of Impartiality → Non-Discriminatory Amplification and Recommendation

Impartiality in fiduciary law demands that entrusted power be exercised equitably and without self-interest. Transposed into the epistemic domain, it requires AI systems to mediate knowledge **without favouritism—amplifying truth, not influence**. In practice, this entails non-discriminatory recommendation mechanisms: algorithms must neither privilege commercial partners nor marginalise minority epistemic voices. Fiduciary impartiality forbids the silent calibration of visibility according to profit, politics, or social conformity.

From the standpoint of **epistemic justice** (Fricker 2007; Medina 2013), impartiality becomes a corrective to testimonial inequality: every epistemic agent must have a fair opportunity to be heard. Within the FEG model, impartiality is not neutrality but **proportional fairness**—the commitment to represent epistemic diversity while preventing domination by power-amplified actors. A system that ranks or recommends content without transparency about its weighting criteria violates this duty as surely as a fiduciary who secretly profits from a client's trust. Algorithmic impartiality, therefore, is the **operational face of fiduciary loyalty**: it binds designers and deployers to treat *public reason as a collective asset*, not a marketplace of attention.

Duty of Accessibility → Public Auditability of Models

Accessibility, the fourth fiduciary duty, extends the principle of openness beyond disclosure into genuine **intelligibility and participatory oversight**. In fiduciary law, accessibility allows beneficiaries to inspect the trustee's conduct and understand the management of their assets. In epistemic governance, it requires that the public—and authorised auditors acting on its behalf—can meaningfully interrogate AI systems' architectures, datasets, and effects.

Under FEG, accessibility entails **public auditability**: models must be structured so their epistemic behaviour can be independently evaluated without proprietary obstruction. This includes auditable documentation of **training-data provenance, bias-testing methodologies, interpretability mechanisms**, and **post-deployment performance metrics**. Accessibility thereby transforms transparency into **reciprocity of understanding**: systems must not only show what they know, but do so in forms that citizens and institutions can comprehend.

Crucially, accessibility is also **pedagogical**. It enables the public to learn how knowledge systems work, cultivating **epistemic literacy**—the civic capacity to engage with complexity rather than retreat from it (Kahl 2025p). A polity that cannot access or understand its knowledge infrastructures cannot claim to govern them; it merely consumes them.

Together, these duties constitute the **normative infrastructure of Fiduciary Epistemic Governance**. They transform fiduciary ethics from moral virtue into operational design and prepare the ground for the institutional and policy frameworks developed in Chapter 8, where governance is re-imagined as the *fiduciary constitution of democracy in the age of AI*.

7.3 Institutional Pathways of Fiduciary Epistemic Governance

While the FEG prototype defines fiduciary duties in moral and procedural terms, it acquires force only through institutional embodiment. As Chapter 5 demonstrated, fiduciary principles lose authority when divorced from enforceable form. Three institutional pathways therefore translate fiduciary ethics into governance design—operationalising the duties of candour, care, impartiality, and accessibility across the epistemic infrastructure.

Algorithmic Fiduciaries

Algorithmic Fiduciaries are statutorily defined entities—public, private, or hybrid—legally obligated to act in the **epistemic interest of their users**. They hold entrusted authority over algorithmic systems much as financial trustees hold authority over assets (Frankel 1983; Kahl 2025i).

Their responsibilities include:

- conducting **fiduciary risk assessments** prior to deployment;
- ensuring **model interpretability and user comprehension**;
- maintaining **continuous disclosure of epistemic limitations**.

Algorithmic Fiduciaries institutionalise the duties of **candour and care**, replacing the voluntary ethics of "responsible AI" with enforceable fiduciary obligations. Certified under public oversight—analogous to data-protection authorities but with *epistemic jurisdiction*—they would restore relational accountability where technical regulation alone fails.

Public Epistemic Trusts (PETs)

Public Epistemic Trusts preserve knowledge as a **fiduciary public good**, countering the enclosures analysed in *The Third Enclosure Movement* (Kahl 2025m). Structured like environmental or cultural-heritage trusts, PETs serve as **collective custodians of the epistemic commons**—datasets, model artefacts, and open scientific repositories.

Governed by **multi-stakeholder boards bound by fiduciary duties to the public**, PETs ensure that knowledge production and dissemination remain aligned with **candour, care, and accessibility**. Their central mandate is to preserve **epistemic pluralism**—protecting diverse knowledge traditions, minority perspectives, and open research ecosystems from commercial monopolisation or algorithmic homogenisation (Fricker 2007; Medina 2013).

Technically, PETs could maintain **open-source reference models, transparent data registries**, and **interoperable auditing frameworks** accessible to civil society, journalists, and academia. Legally, their **statutory trust status** would shield them from acquisition or appropriation. PETs thereby enact **fiduciary openness at the systemic level**: rather than regulating knowledge from above, they maintain **public ownership of the epistemic substrate** upon which democracy depends.

Epistemic Audits

Epistemic Audits extend the fiduciary paradigm into the **independent verification of algorithmic behaviour**. As proposed in recent scholarship (Mueller & Yoo 2025; Roberts & Ziosi 2025), these audits move beyond technical compliance to assess whether systems uphold **fiduciary and epistemic integrity**.

Unlike conventional audits—focused on accuracy, bias, or privacy—Epistemic Audits evaluate the **moral architecture** of algorithms:

- Do they disclose uncertainty and limitation (*candour*)?
- Do they minimise epistemic harm and manipulation (*care*)?
- Do they treat epistemic agents equitably (*impartiality*)?
- Do they remain intelligible and inspectable (*accessibility*)?

Auditors would act as **fiduciary examiners**, empowered to demand remedial action or suspend deployment where breaches occur. Periodic public reporting would transform accountability into an **iterative, dialogical process**, ensuring that knowledge infrastructures evolve through continuous moral calibration rather than post-hoc correction.

Through **Algorithmic Fiduciaries, Public Epistemic Trusts**, and **Epistemic Audits**, the FEG prototype attains institutional depth—a multi-tier system that re-anchors knowledge governance in **trust rather than control**. Together they instantiate the principle developed in Kahl (2025l): that epistemic power achieves legitimacy only when embedded in a fiduciary architecture of openness, reciprocity, and public accountability.

7.4 Legislative Parallels — Toward Global Fiduciary Alignment

The fiduciary-epistemic approach does not arise in isolation. It resonates with ongoing international regulatory initiatives that seek to balance innovation with accountability, yet these frameworks remain fragmented, procedural, and ethically thin. None articulates a unifying constitutional grammar capable of binding technical governance to moral responsibility. The **Fiduciary Epistemic Governance (FEG)** model provides precisely that missing foundation—a framework that transforms the administration of risk into the practice of fiduciary trust.

- **EU Artificial Intelligence Act (2024)** — Establishes risk-based tiers for AI deployment but conceptualises responsibility in technical rather than fiduciary terms. The FEG model would supply the moral grammar that converts risk mitigation into epistemic duty, aligning compliance mechanisms with the principles of candour and care (Kahl 2025i).
- **Digital Services Act (DSA) and Digital Markets Act (DMA)** — Aim to prevent platform abuse and market concentration. FEG complements these by introducing epistemic antitrust, protecting cognitive diversity rather than mere economic competition. In Kahl (2025m), enclosure is defined as the capture of knowledge flows; FEG offers its legal inverse—fiduciary openness as the normative safeguard of plural cognition.
- **OECD AI Principles (2019)** — Emphasise transparency, accountability, and robustness. Reframed through FEG, these become enforceable fiduciary obligations rather than voluntary aspirations. Transparency thus shifts from procedural disclosure to relational intelligibility, binding systems to reciprocal duties of explanation and understanding.
- **UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)** — Centres human rights and inclusivity but lacks institutional mechanisms of enforcement. FEG operationalises its moral intent through Algorithmic Fiduciaries and Public Epistemic Trusts, transforming soft ethics into statutory responsibility.

Collectively, these parallels demonstrate that the fiduciary turn in AI governance does not supplant existing regulation but re-grounds it. It replaces compliance regimes with a constitutional ethic of trust, reciprocity, and epistemic justice. As Kahl (2025l) argues, legitimacy in the digital polity depends not on the proliferation of procedural norms but on the moral coherence of the epistemic order itself—a coherence that only fiduciary alignment can achieve on a global scale.

7.5 Cultural Foundations — Civic Epistemic Virtue and Dissonance Tolerance

No governance architecture can succeed without a supportive epistemic culture. As Kahl (2025p; 2025n) argues, democratic cognition depends upon citizens' ability to withstand contradiction and to engage constructively with uncertainty. The epistemic virtues that sustain fiduciary governance—candour, care, impartiality, and accessibility—must therefore be cultivated not only within institutions but within consciousness itself. Governance begins in the mind: the rule of reason depends on the moral capacity to coexist with doubt.

A fiduciary civic education would teach citizens to recognise epistemic manipulation, to question algorithmic mediation, and to embrace dissonance tolerance as a democratic skill. Schools, universities, and public media could embed curricula of epistemic literacy, integrating logic, cognitive psychology, and digital ethics. Drawing on the principles of Epistemic Psychology (Kahl 2025d), such education would not foster scepticism for its own sake but resilient trust—the discernment to know when and why trust is warranted, and how to rebuild it when breached.

Cultural reform thus completes the fiduciary loop. Just as Algorithmic Fiduciaries and Public Epistemic Trusts institutionalise openness, civic education internalises it. The citizen becomes a co-fiduciary of public reason—aware that democracy is maintained not only through law and technology but through the moral discipline of knowing well together. In this sense, fiduciary governance achieves its full constitution: an architecture of systems supported by an ethics of mind.

8. Policy and Institutional Implications

This chapter advances the *operational dimension* of the fiduciary-epistemic framework. Having developed the moral and structural foundations in Chapters 5–7, it now examines how fiduciary openness can be embedded into real-world institutions, policies, and legal regimes. The analysis proceeds through comparative governance, regulatory pluralism, democratic alignment, and organisational innovation, culminating in a practical implementation roadmap for embedding epistemic fiduciary duties within the governance of artificial intelligence.

The aim is not to propose a single global model but to articulate a governance philosophy: fiduciary-epistemic accountability as a constitutional norm that can guide both legislative reform and institutional self-regulation.

8.1 Comparative Governance — From Corporate to Fiduciary-Epistemic Accountability

Conventional frameworks of corporate accountability—grounded in compliance, disclosure, and risk management—prove inadequate for institutions that wield epistemic power. As Balkin (2016) observes, digital platforms increasingly function as information fiduciaries, yet the concept has remained largely rhetorical,

lacking the legal infrastructure to transform moral aspiration into enforceable duty. As Wodi (2025) argues, AI governance is no longer a discretionary policy domain but a constitutional necessity: the magnitude of epistemic and societal risks posed by autonomous systems demands a transformation in how responsibility is conceived and distributed. His analysis underscores that existing legal frameworks, grounded in risk management and procedural compliance, cannot address the moral asymmetry inherent in algorithmic authority. This diagnosis aligns with the fiduciary-epistemic model advanced here, which interprets AI governance as a trust relationship requiring proactive candour and care rather than reactive regulation. Likewise, Miller and Gold (2015) show that fiduciary obligation is relational rather than procedural: its legitimacy arises from loyalty, candour, and trust, not from audit trails or formal sanctions.

Kahl (2025o) extends this principle to the media and communications sphere, arguing that epistemic gatekeepers—the press, universities, and now algorithmic intermediaries—constitute a *Fourth Estate of fiduciary reason*. Their democratic function is not merely to transmit information but to safeguard the epistemic integrity of public discourse. When these gatekeepers abandon candour and impartiality, they commit what Kahl terms fiduciary abdication: the conversion of public enlightenment into epistemic branding. The same pathology afflicts AI platforms that curate truth through engagement metrics, performing the role of journalists without bearing their fiduciary obligations.

The fiduciary-epistemic model builds upon this lineage by redefining AI governance as a direct fiduciary relationship between epistemic agents—developers, platforms, public institutions—and epistemic beneficiaries —citizens and publics. Accountability thereby shifts from external compliance to internal loyalty: from rule-following to truth-keeping. The relevant metric is not risk avoidance but epistemic integrity—the faithful stewardship of knowledge held in trust for others (Kahl 2025i; 2025l).

This comparative realignment introduces a qualitative transformation. Corporate law enforces obedience—a duty to conform behaviour to regulatory constraint—whereas fiduciary-epistemic law enforces candour and care—a duty to preserve the moral ecology of knowing. The objective is no longer to regulate from without but to moralise from within, embedding fiduciary norms into the architecture of institutions that produce, mediate, or govern knowledge. In this sense, fiduciary-epistemic accountability constitutes the constitutional evolution of corporate responsibility: the moment when governance itself becomes an act of trust.

8.2 Regulatory Pluralism — Public, Private, and Commons-Based AI

AI governance increasingly spans multiple regulatory spheres—state, market, and commons—each with distinct capacities, incentives, and vulnerabilities. As Harré (2025) observes, effective regulation demands plural epistemic sovereignty: no single actor can safeguard the cognitive public interest alone. Lim and Lim (2025) develop this argument through a polycentric model of AI governance grounded in distributed responsibility and multi-stakeholder collaboration, integrating state authority, private innovation, and civic participation. Their framework aligns closely with the fiduciary-epistemic approach advanced here: both reconceive governance as a web of relational duties rather than a hierarchy of command, treating epistemic power as a shared trust to be administered through reciprocal accountability rather than control.

Complementing this perspective, Srinivasan and Patapati (2025) propose a *Democracy-in-Silico* model in which institutional design functions as alignment architecture within AI-mediated polities. Their simulations show that when agent societies are structured around fiduciary constraints—transparency, deliberation, and cooperative calibration—epistemic welfare increases and power-preserving behaviours decline. In this light, polycentric AI governance is not merely a distribution of oversight but an emergent fiduciary constitution,

where the norms of candour, care, impartiality, and accessibility regulate both human and artificial participants in the epistemic commons.

Within this framework, pluralism ceases to be a conflict of jurisdictions and becomes instead a symbiotic division of fiduciary labour: the state acts as collective trustee, the corporation as entrusted steward, and the civic commons as co-fiduciary of public reason. Fiduciary coherence thus unites diversity of form under a single moral architecture—the shared duty to preserve the integrity of knowing together.

Each governance sphere embodies a distinct fiduciary logic:

- **Public AI governance** realises *collective trusteeship*: state institutions act as custodians of epistemic welfare under a democratic mandate, exercising power through transparency, inclusivity, and reason-giving.
- **Private governance** expresses *entrusted stewardship*: corporations that develop or deploy AI systems bear statutory fiduciary duties of **candour, care, impartiality, and accessibility**—duties that subordinate profit to epistemic integrity (Kahl 2025i).
- **Commons-based governance** (open-source models, academic consortia, cooperative data trusts) exemplifies co-fiduciary collaboration: shared management of epistemic infrastructures akin to **Public Epistemic Trusts** (Kahl 2025m).

The interplay among these spheres constitutes **regulatory pluralism with fiduciary coherence**—diversity of institutional form sustained by unity of moral duty. In this respect, the system mirrors Anderson's (2006) conception of **epistemic democracy**: plural reasoning bound by common norms of justification. Fiduciary pluralism thus replaces hierarchical sovereignty with a **network of reciprocal accountability**, in which every actor—public, private, or civic—is answerable to the same constitutional ethic of truth.

8.3 Democratic Alignment — Embedding Fiduciary Values in Public Reason

The fiduciary pluralism outlined above culminates in a deeper normative aim: the democratic alignment of epistemic power with public reason. If the preceding section described the institutional diversity of fiduciary labour across state, market, and commons, this section turns to its moral unity—the embedding of fiduciary virtues within the deliberative structures that constitute democracy itself. Polycentric governance achieves stability only when its participants share a common fiduciary ethic of candour, care, impartiality, and accessibility; without this moral grammar, pluralism degenerates into competitive opacity.

For AI governance to sustain democracy, it must embody the epistemic virtues of deliberative public reason. Anderson (2006) defines democracy as an *epistemic system* optimised for truth-seeking through inclusive dialogue and collective accountability. Likewise, Mentxaka et al. (2025) contend that AI systems should be evaluated not by efficiency or profit but by their *ability to preserve the conditions of deliberation*—pluralism, transparency, and responsiveness to critique.

Fiduciary-epistemic governance translates these virtues into enforceable institutional norms. Its four canonical duties correspond directly to democratic functions:

- **Candour** sustains public deliberation through sincerity, disclosure, and evidential honesty.
- **Care** safeguards citizens' cognitive autonomy by preventing epistemic harm and manipulation.

- **Impartiality** ensures fair participation in the marketplace of reasons, protecting minority perspectives from algorithmic marginalisation.

- **Accessibility** secures the intelligibility of collective reasoning, allowing citizens to comprehend and contest the epistemic forces that shape their world.

As Kahl (2025j) argues, modern institutions frequently substitute these fiduciary virtues with performative ones: ethical branding replaces candour, and the aesthetics of virtue substitute for its practice. Such epistemic marketing, when adopted by universities, corporations, or AI platforms, converts moral duty into rhetorical spectacle, corroding the very trust that deliberative democracy requires. Restoring fiduciary candour thus demands both cultural and institutional correction—an insistence that sincerity and disclosure are not reputational strategies but moral obligations.

Under this framework, AI governance becomes a public act of reason rather than a private act of optimisation. Institutions that mediate knowledge—universities, media, and algorithmic platforms—must operate as **fiduciaries of deliberative integrity**, enabling citizens to trust not only outcomes but the fairness of the epistemic processes that produce them. This restores what Kahl (2025i; 2025m) calls the fiduciary compact of knowledge: authority grounded in *reciprocal openness* rather than procedural compliance.

Democratic legitimacy in the age of AI thus depends not on algorithmic neutrality but on fiduciary trustworthiness—the moral condition that renders deliberation itself possible. The following section (§ 8.4) extends this principle from moral culture to legal design, outlining how fiduciary values may be institutionalised through new juridical forms such as *Epistemic Fiduciary Entities*.

8.4 Institutional Innovation — Legal Form for Epistemic Fiduciaries

The hybrid-governance models examined by Ofir and Levine-Schnur (2025) expose the structural limits of voluntary ethics: when corporate form conflicts with fiduciary function, moral rhetoric devolves into procedural evasion. A durable remedy requires institutional innovation—the creation of legal forms expressly designed to embody epistemic fiduciary duties and to align authority with moral accountability.

The **Epistemic Fiduciary Entity (EFE)** proposed in Chapter 5 provides such a template. Positioned between the public-trust model of universities and the statutory authority of regulated industries, the EFE would combine academic independence with enforceable fiduciary accountability. Its directors and officers would owe legally binding duties of epistemic candour, care, loyalty, and accessibility to the public, subject to independent oversight and periodic *Epistemic Audits* (Mueller & Yoo 2025). In contrast to the self-certified ethics regimes of today's technology firms, these duties would constitute statutory obligations, breach of which would amount to fiduciary wrongdoing.

As Kahl (2025h) argues in *Epistemocracy in Higher Education*, universities already illustrate the partial institutional form of such fiduciary-epistemic governance. Their social mandate—to cultivate and disseminate knowledge in trust for the public—mirrors the EFE's proposed obligations of candour, care, and accessibility. Yet, when universities internalise market logics or managerial secrecy, they cease to act as trustees of public reason and instead reproduce the very enclosures that the EFE model seeks to overcome. The EFE thus represents not a departure from academic precedent but its legal completion: a re-anchoring of institutional purpose in fiduciary ethics rather than commercial or bureaucratic imperatives.

Barnett's (2014, 2018) work on higher education as an ecological and responsive system reinforces this trajectory. He argues that universities must evolve from sites of credential production into dynamic ecosystems of interdependence—responsive to societal needs while preserving the integrity of knowledge as a public good. This vision anticipates the fiduciary principles articulated here: candour as openness to critique, care as responsibility for epistemic environments, and accessibility as civic intelligibility. The *Epistemic Fiduciary Entity* (EFE) thus realises, in juridical form, what Barnett describes as the *ecological university*—an institution grounded not in instrumental utility but in the stewardship of knowledge within complex moral and environmental ecologies.

In doctrinal terms, the EFE embodies what Smith (2023), in *The Law of Loyalty*, identifies as the unifying principle of fiduciary law: the duty of undivided loyalty that gives moral coherence to all entrusted power. Just as fiduciary loyalty cannot be partitioned between competing beneficiaries without dissolving into conflict, the EFE's statutory design secures the indivisibility of its epistemic allegiance. Its directors owe a single, unambiguous duty—to the public's epistemic welfare—thereby restoring what Smith calls the *unity of fiduciary purpose* that modern hybrid entities have lost. In this way, the EFE operationalises Smith's theoretical insight: loyalty's unity is not merely moral but institutional, requiring a legal form that prevents its fragmentation.

This model extends beyond AI developers. Research institutions, data custodians, digital platforms, and knowledge intermediaries could all adopt fiduciary charters articulating explicit epistemic obligations, thereby transforming aspirational mission statements into enforceable governance instruments. Over time, this could yield a new organisational genus—the Epistemic Fiduciary Organisation (EFO)—whose legitimacy derives from truth stewardship rather than profit or sovereign delegation.

Such innovation would institutionalise epistemic virtue as legal structure, fulfilling the fiduciary-epistemic synthesis envisioned by Frankel (1983) and Kahl (2025i; 2025l)—and refined through Smith (2023)'s articulation of loyalty as the moral axis of fiduciary law. Power is justified only by trust, and trust sustained only by openness. In the long arc of institutional evolution, the EFE marks the juridical translation of fiduciary ethics into the architecture of democratic knowledge itself.

8.5 Anticipated Resistance — Fiduciary Duties vs. Shareholder Primacy

Any transition from corporate to fiduciary-epistemic governance will encounter resistance rooted in **shareholder primacy**—the entrenched belief that directors' first obligation is to maximise investor returns. This doctrine, while defensible within traditional economic enterprises, becomes destructive when applied to institutions exercising epistemic power. Knowledge is not a commodity that can be owned without consequence; it is a **fiduciary trust** that must circulate freely to sustain democracy.

Resistance will manifest on several fronts. Corporate directors may view fiduciary duties to the public as a dilution of loyalty, while investors may frame epistemic openness as a threat to intellectual property and competitive advantage. Yet as Kahl (2025i) and Balkin (2016) both argue, fiduciary obligation does not negate profit—it constrains the means by which it may be pursued. Just as medical and legal professionals balance remuneration with public duty, so too must AI firms balance innovation with epistemic stewardship.

Moreover, the market itself increasingly rewards epistemic integrity. Public trust, transparency ratings, and ethical certification already serve as forms of reputational capital. The **fiduciary-epistemic turn** merely renders this moral economy explicit and enforceable. Resistance, therefore, is not inevitable; it is transitional—a symptom of institutions learning to reconcile efficiency with conscience.

Ultimately, the conflict between fiduciary duties and shareholder primacy reveals a deeper philosophical choice: whether knowledge institutions will continue to treat cognition as property, or finally recognise it as a **public good held in trust**. The outcome will determine not only the ethics of AI, but the moral trajectory of modern capitalism itself.

8.6 Implementation Roadmap — From Principles to Practice

Realising fiduciary-epistemic governance requires both **legal reform** and **institutional re-engineering**. The following roadmap outlines actionable steps for policymakers, legislators, and industry leaders seeking to operationalise fiduciary duties within AI governance.

1. **Amend Corporate Charters** — Introduce fiduciary-epistemic clauses into the constitutions of AI and knowledge-producing entities. These clauses would codify duties of candour, care, impartiality, and accessibility toward the public, granting directors legal discretion to prioritise epistemic integrity over short-term profit.

2. **Create Epistemic Ombuds Institutions** — Establish independent *Epistemic Ombuds* offices, modelled on financial and data-protection regulators, empowered to investigate epistemic breaches, mediate disputes, and issue corrective directives. These bodies would act as the moral custodians of public reason, ensuring reciprocity between knowledge-holders and affected communities.

3. **Integrate Epistemic Audits** — Following Mueller & Yoo (2025), mandate recurring fiduciary audits of high-impact AI systems. These audits would evaluate not only technical accuracy and bias but adherence to fiduciary duties—transparency of reasoning, mitigation of cognitive harm, and equitable amplification of perspectives.

4. **Incentivise Fiduciary Compliance** — Governments and multilateral organisations can promote adoption through tax incentives, procurement preferences, or certification schemes for *Epistemic Fiduciary Entities* (EFEs) and *Public Epistemic Trusts*.

5. **Embed in International Frameworks** — Integrate fiduciary-epistemic standards into emerging instruments such as the OECD AI Principles, UNESCO frameworks, and cross-border data governance treaties, thereby harmonising ethical accountability across jurisdictions.

The transition from procedural compliance to fiduciary openness will not occur overnight. It demands iterative experimentation, adaptive regulation, and above all, cultural realignment. Yet the path is clear: **institutionalise trust as the organising principle of AI governance**.

In doing so, policy ceases to be a reactive mechanism for risk control and becomes a proactive instrument of epistemic care—ensuring that the architectures of knowledge remain transparent, reciprocal, and worthy of belief.

9. Conclusion: Reclaiming Democracy through Fiduciary Openness

This concluding chapter returns to the central question that has guided the entire inquiry: *How can democracy survive in an age when knowledge itself has become a site of capture?* Across the preceding chapters, this work has argued that the legitimacy of democratic governance rests not on the volume of information available, but on the fidelity of trust that sustains shared knowing. Artificial intelligence, while expanding informational abundance, has also transformed the moral economy of knowledge—converting epistemic reciprocity into algorithmic dependence.

The task before us, therefore, is not merely to regulate AI but to reconstitute democracy as a fiduciary order of knowledge. This means embedding openness, candour, and care into every level of epistemic life—from the design of algorithms to the education of citizens. The conclusion draws together the conceptual strands of fiduciary law, epistemic justice, and cognitive trust developed throughout this monograph, and offers a succinct synthesis in four parts.

9.1 Democracy as an Architecture of Epistemic Trust

Democracy thrives not through the circulation of information but through the **integrity of its epistemic relationships**. As Anderson (2006) and Habermas (1996) remind us, deliberative legitimacy depends upon citizens' confidence that institutions, experts, and communicative intermediaries act in good faith—that their claims to knowledge are open to scrutiny and correction.

This work has reframed that insight through the language of **fiduciary ethics**: democratic institutions function as *trustees of public reason*. Their authority is justified not by coercion or efficiency but by adherence to duties of candour, care, impartiality, and accessibility. When these duties are honoured, the epistemic commons remains vibrant, and plural reasoning thrives. When they are breached—through secrecy, manipulation, or technocratic domination—the architecture of democracy begins to corrode from within.

The rise of AI magnifies this challenge. Algorithmic mediation displaces traditional fiduciaries—universities, journalists, civil servants—with opaque systems that wield epistemic power without relational accountability. In doing so, it creates a crisis not of information scarcity, but of **trust scarcity**: the erosion of mutual recognition as the basis of truth.

9.2 Algorithmic Clientelism and the Erosion of Autonomy

As developed in Chapter 3, **algorithmic clientelism** describes the systematic replacement of trust with dependency. The user no longer relies upon institutions governed by fiduciary norms but depends upon algorithms whose primary loyalty is to engagement metrics. The result is a cognitive inversion: autonomy is traded for predictability, and reasoning for recognition.

Algorithmic infrastructures restructure epistemic life into a network of *managed dependence*. The citizen becomes a **client of curation**, shaped by flows of personalised information designed to sustain attention rather than deliberation. This dynamic, as Kahl (2025a; 2025n) shows, produces *fiduciary anaesthesia*—a dulled awareness of manipulation that renders epistemic harm invisible. The individual's dissonance tolerance declines; contradiction feels like noise rather than a signal to think.

At the societal level, algorithmic clientelism fragments the public sphere. Citizens no longer meet within a shared epistemic horizon but within privately tailored cognitive enclosures. Public reason dissolves into distributed micro-realities—each coherent to itself yet epistemically isolated. What is lost is not only factual agreement but the **fiduciary fabric of reciprocity** that made democratic reason possible.

To reclaim autonomy, democracy must therefore **rebuild its fiduciary scaffolding**—restoring the reciprocal trust that allows individuals to think together while remaining free to dissent.

9.3 Fiduciary Openness as the Path Forward

The preceding analysis reveals that the survival of democracy in the algorithmic age depends on re-establishing **fiduciary openness** as the organising principle of epistemic life. Openness here does not mean unrestricted data flow or transparency as spectacle; it means **reciprocal intelligibility**—the moral and structural condition that allows knowledge to circulate under the discipline of candour and care.

Fiduciary openness functions at three interlocking levels:

1. **Institutional** – Every body that mediates knowledge—AI firms, universities, media organisations—must act as a *trustee of epistemic integrity*. Legal forms such as *Epistemic Fiduciary Entities* and mechanisms like *Epistemic Audits* (Mueller & Yoo 2025) operationalise this duty, ensuring that authority over information remains coupled with accountability to truth.

2. **Cultural** – As Kahl (2025p) argues in *Speaking into Dissonance*, openness must also be cultivated as a cognitive virtue. Citizens must learn to endure contradiction and uncertainty—the very friction through which moral reasoning and intellectual progress emerge. Fiduciary openness therefore entails not only the transparency of institutions but the courage of minds willing to face dissonance without retreat.

3. **Technological** – Systems that mediate cognition must internalise fiduciary norms within their architectures. This requires that algorithms become *epistemic fiduciaries*, capable of signalling uncertainty, explaining reasoning, and preserving human interpretability. The goal is not machine perfection but **moral calibration**: technology designed to sustain the relational ecology of truth.

Together, these dimensions define the **fiduciary epistemic constitution** proposed throughout this work: a framework in which trust is not demanded but deserved, earned through openness that is both ethical and structural. In such a polity, truth ceases to be a commodity and becomes once again a **shared responsibility**.

9.4 Rebuilding Democracy's Fiduciary Architecture

The restoration of democratic integrity demands more than new regulation—it requires a **moral reconstruction of knowledge itself**. Every institution that holds epistemic power must recognise itself as part of a wider fiduciary order: a living network of reciprocal dependence in which the right to know carries the duty to know responsibly.

This reconstruction proceeds through four principles derived from the fiduciary canon elaborated across this study:

- **Candour** — Public institutions and AI systems must disclose not only data but reasoning; honesty becomes the currency of legitimacy.
- **Care** — Knowledge must be handled as a trust, not a commodity, ensuring that its use enhances collective autonomy rather than diminishes it.
- **Impartiality** — Diversity of perspective is not noise but nourishment; epistemic justice becomes the condition of common reason.
- **Accessibility** — Understanding must remain a public right; opacity, whether bureaucratic or algorithmic, is democratic betrayal.

When these principles are woven through law, culture, and cognition, democracy can again function as a *fiduciary ecology of knowing*—a system where information is not weaponised but tended, and where the act of reasoning together is itself an expression of trust.

The project of fiduciary openness is thus both constitutional and civilisational. It calls for institutions to become intelligible, technologies to become accountable, and citizens to become epistemically brave. Only then can the circulation of knowledge recover its moral direction—moving not toward control but toward **mutual recognition**, the foundation of all just societies.

The argument thus reframes fiduciary theory as a general theory of legitimate power under epistemic conditions and extends political theory into the fiduciary-epistemic domain. In doing so, it offers the moral and constitutional vocabulary for re-grounding democracy in the age of artificial intelligence.

To reclaim democracy, we must rebuild its **fiduciary architecture of knowing**—a world where every human and artificial agent is bound by the same relational duty: *to seek, to share, and to speak the truth as an act of trust*.

•

Bibliography

Primary Legal and Regulatory Sources

European Union. (2022a). Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act). Official Journal of the European Union, L 265.

European Union. (2022b). Regulation (EU) 2022/2065 on a single market for digital services (Digital Services Act). Official Journal of the European Union, L 277.

European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on Artificial Intelligence (AI Act). Official Journal of the European Union, L 1689.

OECD. (2021). State of implementation of the OECD AI principles. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/ai-implementation-2021-en

OECD. (2023). OECD AI principles and democratic values. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/ai-policy-2023-en

UK Parliament. (2023). Online Safety Act 2023 (UK). HMSO.

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Works by Kahl

Kahl, P. (2025a). *Cognitive dissonance as epistemic event: Clientelism, bounded freedom, and the architecture of epistemic fear* (v3). Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.16174.37449

Kahl, P. (2025b). *The newborn's first cry as epistemic claim and foundation of psychological development: Attachment, autonomy, and resilience*. Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17265357

Kahl, P. (2025c). *Epistemic humility and the transposition of ethical duties into epistemic duties: A philosophical and fiduciary inquiry into the normative foundations of knowledge* (v3). Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17453582

Kahl, P. (2025d). *Re-founding psychology as epistemic psychology: The science of autonomy and dependence under epistemic conditions*. Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17245416

Kahl, P. (2025e). *Epistemic clientelism in intimate relationships: Fiduciary ethics, epistemic dissonance, and the computational foundations of epistemic psychology*. Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.33790.45122

Kahl, P. (2025f). *Authoritarianism and the architecture of obedience: From fiduciary–epistemic trusteeship to clientelist betrayal*. Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.15123.34089

Kahl, P. (2025g). *The Silent Shadows: Epistemic Clientelism and Plato's Cave*. Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.13579.96808

Kahl, P. (2025h). *Epistemocracy in higher education: A proposal for fiduciary and epistemic accountability in the university* (v2). Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17293583

Kahl, P. (2025i). *Directors' epistemic duties and fiduciary openness: A cross-cultural and interdisciplinary framework for corporate governance* (v2). Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.28155.17449

Kahl, P. (2025j). *Epistemic violence or simply good marketing? Why university marketing may not be so innocent* (v3). Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17378350

Kahl, P. (2025k). *Epistemic justice and institutional responsibility in academia: Toward a comprehensive framework for epistemic justice in higher education* (v2). Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.11063.38563

Kahl, P. (2025l). *The epistemic architecture of power: How knowledge control sustains authority in social structures* (v2). Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.14628.54402

Kahl, P. (2025m). *The Third Enclosure Movement: Rethinking AI Regulation through a Taxonomy of Restricted Knowledge*. Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17542062

Kahl, P. (2025n). *What happens when you clap? Cognitive dissonance, fiduciary trust, and the relational theory of epistemic clientelism* (v2). Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17440653

Kahl, P. (2025o). *Epistemic Gatekeepers as the Fourth Estate: Reining in Media's Unchecked Epistemic Power*. Lex et Ratio Ltd. https://github.com/Peter-Kahl/Epistemic-Gatekeepers-as-the-Fourth-Estate-Reining-in-Medias-Unchecked-Epistemic-Power

Kahl, P. (2025p). *Speaking into Dissonance: Foreign Language Learning, Cognitive Dissonance, and Epistemic Plurality*. Lex et Ratio Ltd. https://doi.org/10.13140/RG.2.2.22256.93449

Kahl, P. (2025r). *The University of Reading's LLM experience as a mirror of higher education: Epistemic clientelism, optocratic drift, and the pedagogy of fiduciary dialogue* (v2). Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17385633

Fiduciary Theory and Legal Foundations

Balkin, J. M. (2016). Information fiduciaries and the First Amendment. *UC Davis Law Review*, 49(4), 1183–1234. https://ssrn.com/abstract=2675270

Frankel, T. (1983). Fiduciary law. *California Law Review*, 71(3), 795–836. https://scholarship.law.bu.edu/faculty_scholarship/3005/

Gold, A. S., & Miller, P. B. (Eds.). (2014). *Philosophical foundations of fiduciary law*. Oxford University Press.

Miller, P. B., & Gold, A. S. (2015). Fiduciary governance. *William & Mary Law Review*, 57(2), 513–586. https://scholarship.law.wm.edu/wmlr/vol57/iss2/4

Smith, L. (2023). *The law of loyalty*. Oxford University Press.

Smith, L. (2014). Fiduciary relationships: Ensuring the loyal exercise of judgment on behalf of another. *Law Quarterly Review*, 130, 608–634. https://ssrn.com/abstract=2559974

AI Governance, Corporate Form, and Institutional Design

Harré, M. S. (2025). From Firms to Computation: AI Governance and the Evolution of Institutions. *arXiv*. https://arxiv.org/pdf/2507.13616

Lim, E. C. N., & Lim, C. E. D. (2025). Polycentric AI governance: A multi-stakeholder approach to distributed responsibility and ethical technology management. *International Journal of Advanced AI Applications*, 1(4), 77–97. Retrieved from http://www.dawnclarity.press/index.php/ijaaa/article/view/53

Mentxaka, O., Díaz-Rodríguez, N., Coeckelbergh, M., López de Prado, M., Gómez, E., Fernández Llorca, D., Herrera-Viedma, E., & Herrera, F. (2025). Aligning trustworthy AI with democracy: A dual taxonomy of opportunities and risks. *arXiv*. https://arxiv.org/abs/2505.13565

Mueller, A., & Yoo, C. S. (2025). Taking standards seriously: The case for a private standards-based approach to AI governance. *Berkeley Technology Law Journal* (forthcoming, Fall 2025). University of Pennsylvania Law School, Public Law Research Paper No. 25-46. https://doi.org/10.2139/ssrn.5486746

Ofir, M., & Levine-Schnur, R. (2025). GenAI models and the hybrid governance trap. *Cornell Journal of Law and Public Policy* (forthcoming). https://ssrn.com/abstract=5677544

Roberts, H., & Ziosi, M. (2025). Global AI governance through technical standards. *SSRN*. https://doi.org/10.2139/ssrn.5376424

Srinivasan, T., & Patapati, S. (2025). Democracy-in-Silico: Institutional design as alignment in AI-governed polities. *arXiv*. https://arxiv.org/abs/2508.19562

Wodi, A. (2025). AI governance: A necessity and imperative. *SSRN*. https://doi.org/10.2139/ssrn.5486407

AI Ethics, Epistemic Agency & Responsibility

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Coeckelbergh, M. (2022). *The political philosophy of AI*. Polity Press.

Coeckelbergh, M. (2025). AI and Epistemic Agency: How AI Influences Belief Revision and Its Normative Implications. *Social Epistemology*, 1–13. https://doi.org/10.1080/02691728.2025.2466164

Cohen, J. E. (2019). *Between truth and power: The legal constructions of informational capitalism*. Oxford University Press. https://doi.org/10.1093/oso/9780190246693.001.0001

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://digitalcommons.law.umaryland.edu/books/96

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.

Democratic Theory and Political Philosophy

Anderson, E. (2006). The Epistemology of Democracy, *Episteme*, 3(1–2), pp. 8–22. https://doi.org/10.3366/epi.2006.3.1-2.8.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.). MIT Press.

Rawls, J. (1999). *A theory of justice* (Rev. ed.). Harvard University Press.

Raz, J. (1986). *The morality of freedom*. Oxford University Press.

Singer, A. (2015). There is no Rawlsian theory of corporate governance. *Business Ethics Quarterly*, 25(1), 65–92. https://doi.org/10.1017/beq.2015.1

Epistemic Justice and Institutional Ethics

Barnett, R. (2014). Conditions of Flexibility: Creating a More Responsive Higher Education System. *York: Higher Education Academy*. https://www.heacademy.ac.uk/sites/default/files/resources/fp_conditions_of_flexibility_0.pdf

Barnett, R. (2018). *The ecological university: A feasible utopia*. Routledge.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press.

Philosophy of Knowledge and Social Epistemology

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor Books.

Elgin, C. Z. (2017). *True enough*. MIT Press.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings*, 1972–1977 (C. Gordon, Ed.). Pantheon Books.

James, W. (1907). *Pragmatism: A new name for some old ways of thinking*. Longmans, Green.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). Routledge & Kegan Paul. (Original work published 1945)

Sosa, E. (2021). *Epistemic explanations: A Theory of Telic Normativity, and What it Explains*. Oxford University Press.

Cognitive and Social Psychology Foundations

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4), 381–414. https://doi.org/10.1162/NETN_a_00018

Milgram, S. (1974). *Obedience to authority: An experimental view*. Harper & Row.

Sociology and Power

Bourdieu, P. (1993). *Language and symbolic power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press.

Feldman, S. (2003). Enforcing social conformity: A theory of authoritarianism. *Political Psychology, 24*(1), 41–74. https://doi.org/10.1111/0162-895X.00316

Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books.

Economics, Innovation and Knowledge Regulation

Stiglitz, J. E. (1999). Knowledge as a global public good. In I. Kaul, I. Grunberg, & M. A. Stern (Eds.), *Global public goods: International cooperation in the 21st century* (pp. 308–325). Oxford University Press.

Stiglitz, J. E., & Greenwald, B. C. (2015). *Creating a learning society: A new approach to growth, development, and social progress*. Columbia University Press.

•

Author Metadata

Email: peter.kahl@juris.vc
ORCID: https://orcid.org/0009-0003-1616-4843
LinkedIn: https://www.linkedin.com/in/peter-kahl-law/
ResearchGate: https://www.researchgate.net/profile/Peter-Kahl
SSRN: https://ssrn.com/author=4270895
PhilPapers: https://philpeople.org/profiles/peter-kahl
GitHub: https://github.com/Peter-Kahl
Google Scholar: https://scholar.google.com/citations?hl=en&user=z-yfRRYAAAAJ
Blog: https://pkahl.substack.com/

•

Cite this work

Kahl, P. (2025). *Redefining democracy for the age of AI: AI governance and the fiduciary turn in the architecture of knowledge*. Lex et Ratio Ltd. https://doi.org/10.5281/zenodo.17557934

•

Version History

Version	Description of Changes	Epistemic Impact	Date
1	Initially titled *Redefining democracy for the age of AI: AI governance and the fiduciary turn in the architecture of knowledge*	N/A	2025-11-08

•