

ALiPsé WHITE PAPER No. 2

The Information Authenticity Externality

When Trust Becomes an Attack Surface

Verification Costs, Institutional Impersonation, and the Missing Trust Layer of Dimension 4.5

Luis Ángel Rodríguez Alemán

AliPsé • English Master v1.0 • August 2026

Tabla de Contenido

Abstract	4
White Paper Contributions	4
1. The Unfinished Digital Revolution	6
1.1 Dimension 4.5	6
1.2 Positive externalities came first.....	6
1.3 The missing institutional layer.....	6
2. Externalities in Dimension 4.5	7
2.1 A classical economic lens	7
2.2 The dual character of digital scale	7
2.3 From content harms to institutional costs	7
3. Old Human Biases, New Amplification Machines	7

3.1 Negativity predates digital technology	7
3.2 Generative AI changes the production function of deception	8
3.3 Authenticity Verification Asymmetry	8
4. Authenticity Is Not Truth	8
4.1 A necessary distinction	8
4.2 Five related properties.....	8
4.3 Cybersecurity overlap	9
5. The Information Authenticity Externality.....	9
5.1 Definition.....	9
5.2 Externality boundary: incident burden is not automatically externality.....	9
5.3 Cost layers	9
5.4 Relationship to existing literature	10
5.5 Verification Economies as a testable condition	10
6. Institutional Impersonation and the Authenticity Attack Surface	11
6.1 Borrowed authority	11
6.2 Architectures of Borrowed Trust.....	11
6.3 The 43-case exploratory dataset	11
6.4 The SEC case	11
7. The Verification Cost Model	12
7.1 Purpose and conservatism	12
7.2 Observed, estimated, unobserved	12
7.3 Results.....	12
7.4 AER	12
8. From Authenticity Risk to Epistemic Security	13
8.1 Institutions require attributed communication.....	13
8.2 The liar's dividend	13
8.3 Efficiently warranted trust.....	13
9. Beyond Financial Harm: Political, Social, and Institutional Externalities	13
9.1 Political deepfakes.....	13
9.2 Trust as a stock.....	13
9.3 Political externalities as future research	14
10. Existing Responses — And the Missing Layer	14
10.1 Detection.....	14
10.2 Provenance and Content Credentials	14
10.3 Watermarking	14
10.4 Multi-layer provenance.....	14

10.5 Cybersecurity and the Authenticity Attack Surface	15
10.6 The emerging authenticity stack	15
11. From the Right to Information Authenticity to Authenticity Infrastructure	15
11.1 From norm to institution	15
11.2 Digital Authenticity Infrastructure.....	16
11.3 Authenticate Once, Verify Many	16
11.4 Registries of authoritative originals.....	16
11.5 The right also protects the speaker	16
11.6 The SEC case revisited	16
11.7 Underproduction of authentication	16
12. Completing Dimension 4.5.....	16
12.1 From trust signals to trust evidence.....	16
12.2 Orchestration as institutional usability	17
12.3 Universality and technology agnosticism	17
12.4 An adversarial infrastructure.....	17
12.5 Privacy and inclusion.....	17
12.6 Agentic AI and Digital Authority Authenticity.....	17
12.7 Resilience requires more than cryptography	17
12.8 Distribution + Trust.....	18
13. Limitations and Research Agenda.....	18
13.1 Dataset limitations	18
13.2 Monetary evidence	18
13.3 Modeled verification burden	18
13.4 Exposure is not belief	18
13.5 Avoided Verification Cost.....	18
13.6 Verification Economies as a testable proposition	18
13.7 Authenticity Commons.....	19
13.8 Risks created by authenticity infrastructure	19
13.9 Falsifiability	19
13.10 Appropriate claim.....	19
14. Conclusion: Completing the Trust Layer	19
Author Disclosure	21
References	21

Abstract

Digital networks have generated extraordinary positive externalities by lowering the cost of communication, coordination, knowledge diffusion, participation, and economic exchange. This paper describes that digitally mediated layer of social and economic life as Dimension 4.5: a domain superimposed on physical geography in which information can move globally at near-zero marginal cost. The same architecture, however, also creates negative externalities when informational origin, identity, integrity, and authority cannot be efficiently established. Generative artificial intelligence intensifies this problem by lowering the cost of producing plausible synthetic evidence while verification costs remain distributed across recipients, institutions, platforms, and the public.

This paper brings together previously separate literatures on externalities, costly verification, cybersecurity economics, misinformation, provenance, and digital authenticity, and proposes the Information Authenticity Externality as a framework for analyzing burdens imposed on actors other than the originator of deceptive or unauthorized information when informational origin and authority cannot be efficiently established. It develops an initial measurement framework—the Verification Cost Model—and derives a testable condition under which reusable positive authentication could generate Verification Economies. It further develops the institutional and technical logic of an interoperable verification-and-orchestration layer through which the right to information authenticity can become practically exercisable: establishing a binding between a legitimate source and an artifact, claim, or authority state; enabling subsequent checking of that evidence; and orchestrating heterogeneous authenticity signals at the point of decision.

An exploratory dataset of 43 institutional-impersonation observations illustrates the emerging Authenticity Attack Surface. A welfare-consistent VCM separates gross fraudulent-transfer exposure, net realized direct victim loss, Incident Burden, externalized authenticity burden, and net social welfare cost. Under deliberately conservative minimum-model assumptions, the current dataset contains approximately US\$27.2 million in documented gross exposure, approximately US\$26.2 million in net realized direct-loss floor, and approximately US\$35,000 in modeled verification, response, remediation, and ecosystem burden. These quantities matter for different reasons. Direct loss and institution-internal response costs can create a private business case for authentication even when they are not economic externalities; costs shifted to recipients, counterparties, platforms, public agencies, markets, or society create an additional institutional and social case. The paper therefore treats Incident Burden and the Information Authenticity Externality as complementary lenses rather than competing measures. Authenticity is not synonymous with truth, nor a substitute for cybersecurity, provenance standards, watermarking, or detection. These capabilities form an emerging stack whose missing institutional function is usable, interoperable verification and orchestration.

Keywords: information authenticity; externalities; verification costs; institutional impersonation; generative artificial intelligence; digital trust; provenance; digital authenticity infrastructure; cybersecurity; Dimension 4.5.

White Paper Contributions

- Defines the Information Authenticity Externality (IAE) as the burden imposed beyond the originator of deceptive or unauthorized information when informational origin and authority cannot be efficiently established, while distinguishing such externalized burdens from costs internalized by direct victims or responding institutions.
- Distinguishes authenticity from truth and separates origin, provenance, integrity, identity, and authority as related but non-identical properties.
- Develops an authenticity-specific Verification Asymmetry: deception can be produced and distributed cheaply while downstream authentication, verification, response, and remediation burdens can scale across recipients and institutions.

- Builds an exploratory 43-case institutional-impersonation dataset and a welfare-consistent Verification Cost Model (VCM).
- Derives a testable condition for possible Verification Economies: above a threshold, reusable positive authentication can be socially cheaper than decentralized recipient-by-recipient verification if it reduces expected marginal verification cost sufficiently.
- Frames digital authenticity as an expanded cybersecurity and institutional trust surface, not as a competing category.
- Proposes an interoperable Digital Authenticity Infrastructure in which capture authenticity, AI-generation marking, provenance, watermarking, detection, registries, and verification/orchestration can operate as complementary layers.
- Extends the right to information authenticity from a normative claim into an institutional design problem: how to make the right easy to exercise without creating a centralized authority over truth.

1. The Unfinished Digital Revolution

1.1 Dimension 4.5

The internet did not simply make existing communication faster. It created an additional operational layer of human civilization. In this paper, Dimension 4.5 is the name given to cyberspace, or digital space, as a dimension of human activity superimposed on the physical world. The name is intentionally metaphorical rather than a claim in physics or mathematics. Digital activity is not constrained by the three material spatial dimensions in the same way as physical objects: information can be accessed simultaneously from distant locations, copied without traversing physical distance in the ordinary sense, and coordinated across geography with near-zero marginal friction. Yet cyberspace does not escape the fourth dimension, time: digital events still occur in sequence, have duration, and leave temporal histories. For that reason, earlier work called this civilizational layer Dimension 4.5 - an additional dimension beyond the familiar three spatial dimensions plus time, but deliberately 'not quite' a fifth physical dimension (Rodríguez Alemán 2016).

The term is therefore used throughout this paper as a shorthand for the digitally mediated domain in which people, firms, governments, markets, communities, and increasingly autonomous software agents communicate, transact, coordinate, create, and exercise authority. Dimension 4.5 is neither wholly physical nor detached from physical reality; it is a civilizational layer whose distinctive economics arise in part because information, identity signals, transactions, and social relationships can move with radically lower spatial friction.

The economic achievements of this layer are substantial. Digital goods and services create consumer surplus that conventional national accounts often understate; digital platforms reduce search and coordination costs; and networks enable forms of collaboration that would have been prohibitively expensive in earlier information environments (Brynjolfsson et al. 2019; World Bank 2016). The same infrastructure has supported scientific collaboration, access to education, remote work, entrepreneurship, diaspora networks, political mobilization, and rapid crisis communication. The correct starting point for an authenticity agenda is therefore not technological pessimism. Dimension 4.5 is an extraordinary civilizational achievement.

1.2 Positive externalities came first

Many of the defining externalities of digital networks are positive. A person who publishes useful knowledge may create value for recipients far beyond the compensation received by the publisher. A new participant in a communications network increases the network's usefulness for others. Public digital infrastructure can reduce the cost of interacting with the state, while open technical standards can enable innovation by actors who were not involved in their creation. These gains are not peripheral; they are part of the reason digitalization became foundational.

Political mobilization offers another example. Networked communications have at times enabled citizens to coordinate rapidly, document events, and overcome information bottlenecks that previously favored centralized institutions. The point is not that digital mobilization is inherently democratic or benign. It is that the same low-friction architecture can create substantial positive social externalities when it expands voice, access, and coordination.

1.3 The missing institutional layer

The problem is that the infrastructure for moving information matured faster than the infrastructure for proving where information came from. We optimized distribution before developing equivalent mechanisms for origin, authority, and integrity. For many years this gap was partially tolerable because producing convincing falsifications was costly, access to mass distribution was constrained, and social context supplied informal trust signals. Generative AI changes that equilibrium.

The next institutional challenge is not to reverse Dimension 4.5. It is to complete it. The central question is how a digital society can preserve the positive externalities of low-friction information while reducing the negative externalities created when recipients cannot efficiently establish who or what actually stands behind a message, image, audio clip, document, transaction instruction, or delegated digital action.

2. Externalities in Dimension 4.5

2.1 A classical economic lens

Externalities arise when an action imposes benefits or costs on parties who are not fully represented in the decision that produces the action. Pigouvian analysis emphasizes the divergence between private and social cost; Coasean analysis emphasizes rights, bargaining costs, and the institutional conditions under which parties can internalize spillovers (Pigou 1920; Coase 1960). Digital systems repeatedly create such divergences because copying, distribution, and interaction occur at scale while many downstream consequences remain outside the producer's private calculation.

2.2 The dual character of digital scale

Scale magnifies both directions. A useful digital artifact can be copied at almost zero marginal cost and create value for millions. A deceptive artifact can exploit the same economics. Social platforms, recommendation systems, messaging networks, and generative models therefore do not merely create 'information problems.' They alter the production and distribution functions through which both value and harm propagate.

The policy objective should not be to eliminate uncertainty or impose friction indiscriminately. Excessive verification requirements can themselves become a tax on legitimate communication. The design problem is to reduce the marginal cost of establishing authenticity for legitimate actors while increasing the expected difficulty of successfully borrowing another actor's authority.

2.3 From content harms to institutional costs

Many discussions of misinformation focus on belief: whether a recipient believes a false proposition. Authenticity uncertainty creates costs even before belief is measured. A bank that must investigate a forged executive instruction, a regulator that must deny a fabricated announcement, a company that must warn customers about a deepfake advertisement, or a citizen who must spend time determining whether a video is genuine all incur costs because origin or authority is uncertain. These costs motivate the economic framework developed below.

3. Old Human Biases, New Amplification Machines

3.1 Negativity predates digital technology

Human attention is not neutral. Psychological research has long documented a negativity bias: negative events, information, and outcomes often receive greater cognitive weight than comparably positive ones (Baumeister et al. 2001; Rozin and Royzman 2001). Digital platforms did not create this tendency. They inherited it.

What digital systems can do is measure, optimize, and amplify it. Robertson et al. (2023), analyzing more than 22,000 randomized headline experiments and hundreds of millions of impressions, found that negative language increased online news consumption. Brady et al. (2017) showed that moral-emotional language is associated with greater diffusion, while Rathje, Van Bavel, and van der Linden (2021) found particularly strong engagement effects for out-group language. These mechanisms help explain why content that triggers fear, anger, outrage, or identity threat can receive disproportionate distribution.

3.2 Generative AI changes the production function of deception

Falsehood is not new, impersonation is not new, and propaganda is not new. What changes is the cost structure. Generative systems can produce persuasive text, cloned voices, synthetic images, translated variants, personalized messages, and increasingly convincing video at a speed and marginal cost unavailable to earlier operators. Experimental evidence suggests that AI-generated propaganda can be persuasive and that personalized large-language-model interactions can outperform human persuaders under some conditions (Goldstein et al. 2024; Salvi et al. 2025).

This creates a new combination: old cognitive biases, low-cost synthetic production, automated personalization, and high-speed distribution. The resulting risk is not simply that 'deepfakes will fool everyone.' The more general problem is that the cost of producing plausible evidence can fall faster than the cost of resolving authenticity uncertainty.

3.3 Authenticity Verification Asymmetry

We refer to this distributional phenomenon as an authenticity Verification Asymmetry. The phrase should be distinguished from generation-verification asymmetry elsewhere in the AI literature, where verification may be computationally cheaper than generation. Here the asymmetry concerns social distribution: an attacker may create and distribute a synthetic authority signal at low private cost, while thousands of recipients and multiple institutions independently spend time and resources deciding whether it is authentic.

Private cost of deception << Aggregate downstream cost of authenticity resolution

The asymmetry is therefore not only technological. It is institutional. A system that asks every recipient to independently solve authenticity recreates verification cost at each decision point. That is the mechanism through which authenticity uncertainty can become an externality.

4. Authenticity Is Not Truth

4.1 A necessary distinction

An authentic statement can be false. An inauthentic statement can accidentally be true. Authenticity concerns whether a piece of information is what it purports to be with respect to origin, identity, integrity, provenance, or authority. Truth concerns whether its substantive proposition corresponds to reality. Conflating the two creates both technical and institutional danger.

C2PA's guiding principles embody this distinction: provenance mechanisms can allow producers and custodians to make verifiable assertions about creation and modification without deciding whether the content is 'good' or 'bad.' The objective is verifiable association and tamper evidence, not centralized epistemic judgment (C2PA 2026).

4.2 Five related properties

Table 1. Authenticity Properties and Their Verification Questions

Property	Question
Origin	Where did this artifact come from?
Provenance	What happened to it across its lifecycle?
Integrity	Has the relevant artifact or assertion been altered?
Identity	Which person, organization, device, or agent is associated with it?
Authority	Was that actor authorized to speak or act in this capacity at this time?

These properties overlap but are not interchangeable. Throughout this paper, authentication, verification, and orchestration are therefore used in a deliberately narrower sequence. Authentication is the establishment of a

defensible binding between an artifact, claim, identity, or authority state and the source or actor to which it is attributed. Verification is the subsequent checking of that binding and its supporting evidence at a later decision point. Orchestration is the discovery, combination, validation, and presentation of heterogeneous authenticity evidence—such as provenance credentials, watermarks, registries, identity or authorization signals, and detection outputs—so that verification remains usable across technologies and channels. A cryptographically intact file can still carry a false assertion; a real executive's voice can be replayed without authority; an official social account can be compromised and publish an unauthorized institutional message; and a synthetic image can be fully authorized by its creator. This is why the paper preserves the proposition: AI-generated is not the same as unauthorized, and human-generated is not the same as authentic.

4.3 Cybersecurity overlap

Digital authenticity should not be framed as an alternative to cybersecurity. Identity security, account protection, cryptographic integrity, access control, fraud prevention, incident response, and provenance already overlap substantially with the authenticity problem. The emerging issue is that the surface being protected expands from systems and accounts to the authority conveyed by content itself.

5. The Information Authenticity Externalities

5.1 Definition

This paper proposes the Information Authenticity Externalities (IAE) as a framework for analyzing burdens imposed beyond the originator of deceptive or unauthorized information when the origin, integrity, identity, or authority of information cannot be efficiently established. The IAE is one part of a broader economic problem. Organizations and individuals can also bear substantial Incident Burdens directly: fraudulent transfers, internal verification, investigation, remediation, operational disruption, reputational effects, and incremental prevention. These privately borne costs matter independently because they can justify investment in authenticity controls even where no social externality is measured. The externality lens then asks which additional burdens are shifted to counterparties, platforms, public agencies, market participants, recipients, or society more broadly. The paper therefore studies both private Incident Burden and social or institutional externalization.

5.2 Externality boundary: incident burden is not automatically externality

For measurement purposes, the paper distinguishes four incidence categories: (1) attacker resource cost; (2) direct victim loss and victim-internal response cost; (3) third-party or ecosystem cost borne by recipients, counterparties, platforms, public agencies, or other actors; and (4) broader social deadweight or destruction. Categories (2) through (4) may all be economically consequential Incident Burdens. Category (2) is especially relevant to private adoption decisions because firms do not need a social-externality rationale to invest in reducing their own expected losses or response costs. Categories (3) and (4) reveal the additional institutional or public value that private authentication investment can create. The VCM therefore functions as an incidence map that can support both organizational business cases and analysis of the IAE.

5.3 Cost layers

$$E_A = C_V + C_R + C_O + C_L + C_P + C_F + C_{Rep} + C_E$$

Here C_V denotes verification cost; C_R institutional response and investigation; C_O operational disruption; C_L legal or regulatory resource cost; C_P public remediation and communications; C_F future prevention or incremental security; C_{Rep} reputational cost; and C_E ecosystem or recipient verification. These are operational cost categories, not automatic externality classifications. The economic classification depends on incidence: a cost borne internally by a directly affected institution differs from a cost shifted to recipients, counterparties, public agencies, platforms, or other third parties.

$$B_I = L_D + E_A$$

Incident Burden B_I combines net realized direct victim loss L_D with modeled verification, response, remediation, and ecosystem burdens. It is a descriptive incident quantity, not a synonym for either the Information Authenticity Externality or net social welfare cost. The IAE is the subset of incident-generated burdens that is externalized beyond the relevant originator or decision-maker. Net social welfare analysis requires an additional distinction because a fraudulent transfer can be partly or wholly a transfer between parties rather than pure resource destruction.

$$W_S = C_A + E_A + D_L$$

W_S denotes net social welfare cost, C_A attacker resource cost, and D_L the deadweight or destruction component of victim loss. The current empirical model does not estimate aggregate W_S because C_A and D_L are generally unobserved. This refusal to invent missing inputs is a methodological feature.

5.4 Relationship to existing literature

The IAE is adjacent to, but not reducible to, several established literatures. Cybersecurity economics studies incentive misalignment and security externalities (Anderson and Moore 2006). Rao and Reiley (2012) show how low-cost information abuse can impose social costs far larger than attacker revenue in the economics of spam. Costly-verification models examine how verification incentives shape information environments (Kattwinkel and Knoepfle 2023; Cisternas and Vásquez 2026), while certification research shows that even small verification costs or errors can destroy surplus in markets characterized by information asymmetry (e.g., Welfare Effects of Product Certification 2022). Research on blockchain traceability separately addresses information authenticity and verification (Tan, Huang, and Li 2023). The contribution here is the synthesis of these mechanisms around uncertainty over informational origin and authority; an incidence-aware measurement framework for the resulting verification and response burdens; and the proposition that reusable positive authentication may alter the marginal economics of verification.

5.5 Verification Economies as a testable condition

If each recipient independently resolves authenticity, aggregate verification cost can scale with the number of decision points. A reusable positive-authentication system could change that structure. Let C_{Aut} be the fixed cost of establishing authentication for a defined artifact, claim, or authority state; C_{v0} the expected per-recipient verification cost without reusable evidence; and C_{v1} the expected residual per-recipient verification cost while that authentication remains valid.

$$C_{Aut} + N(C_{v1}) < N(C_{v0})$$

$$N^* = C_{Aut} / (C_{v0} - C_{v1})$$

Under these assumptions, N^* is the threshold above which reusable authentication would be socially cheaper than decentralized verification, provided that $C_{v1} < C_{v0}$ and that authentication remains valid for the relevant artifact, claim, or authority state. The equations therefore establish a testable condition for possible Verification Economies; they do not demonstrate that such economies already exist in deployed authenticity systems. Empirical estimation of C_{Aut} , C_{v0} , C_{v1} , validity duration, revocation costs, error rates, and adoption is required.

Above some scale, authenticating a defined artifact, claim, or authority state once and allowing repeated verification while that evidence remains valid could become socially cheaper than asking recipients independently to resolve authenticity uncertainty. This is the Verification Economies hypothesis tested by the framework.

6. Institutional Impersonation and the Authenticity Attack Surface

6.1 Borrowed authority

Many high-impact deception campaigns do not ask recipients to trust an unknown attacker. They borrow authority already accumulated by someone else. A familiar executive face, a central bank governor, a government logo, a news presenter, a verified platform account, or an official-looking investment brand can compress the recipient's decision process by importing pre-existing trust.

6.2 Architectures of Borrowed Trust

We use the term Architectures of Borrowed Trust to describe combinations of person, institution, media format, platform context, visual cues, and distribution channels that collectively make an unauthorized message appear institutionally credible. The phrase 'borrowed trust' has prior informal use in security and fraud contexts; the analytical contribution here is to treat these components as an attack architecture rather than a single deceptive artifact.

Table 2. Architectures of Borrowed Trust

Layer	Borrowed signal	Example
Person	Face, voice, name, role	CEO, governor, minister, candidate
Institution	Brand, seal, domain, official identity	Bank, regulator, government agency
Channel	Platform or account context	Official social account, messaging app, video call
Media	Apparent evidentiary realism	Audio, video, PDF, screenshot
Authority	Instruction or endorsement	Payment order, policy announcement, investment recommendation

6.3 The 43-case exploratory dataset

The empirical base contains 43 documented institutional-impersonation observations spanning corporate executives, public authorities, financial institutions, political figures, media identities, and investment scams. Cases were assembled purposively from official institutional statements, enforcement records, high-quality journalism, and threat-research reports under the inclusion criteria documented in Appendix A; the dataset is not a prevalence sample. Eighteen observations arise from a single Unit 42 investigation of deepfake scam campaigns across multiple countries and public figures and are therefore treated as a source cluster rather than 18 statistically independent investigations (Palo Alto Networks Unit 42 2024).

The cases nonetheless demonstrate breadth: deepfake executive video conferences, cloned political voices, fake central-bank endorsements, synthetic investment promotions, compromised official accounts, and face-swapping attacks against digital identity controls. The common denominator is not AI alone. It is unauthorized use of authority signals.

6.4 The SEC case

The U.S. Securities and Exchange Commission (SEC) case is especially important because it separates channel authenticity from content authenticity. On January 9, 2024, an unauthorized party obtained control of the authentic @SECGov account and published a false announcement concerning approval of spot Bitcoin exchange-traded products. The SEC subsequently stated that the content had not been drafted or created by the Commission and that access resulted from control of the phone number associated with the account (SEC 2024). The account was real; the institutional statement was not. Cybersecurity of the account and authenticity of the content are therefore complementary controls, not substitutes.

7. The Verification Cost Model

7.1 Purpose and conservatism

The VCM converts the paper's economic framework into an auditable measurement structure for both Incident Burden and the Information Authenticity Externality. It is not an accounting statement and does not claim to capture the full social cost of impersonation. Its first purpose is to identify burdens relevant to victims and organizations—direct loss, verification, response, remediation, and other incident costs. Its second is to examine incidence: which burdens remain internal and which are shifted outward to recipients, counterparties, platforms, public institutions, markets, or society. The model distinguishes observed facts from modeled estimates and unobserved layers while preventing gross fraudulent transfers from being mislabeled as social welfare losses.

7.2 Observed, estimated, unobserved

Each component is coded O, E, or U. In v0.5, institutional verification and response costs are modeled using conservative labor assumptions; public-remediation costs are modeled where a documented event occurs; ecosystem verification is monetized only in the Biden robocall deep dive; and attacker resource cost is case-specific only in that same observation. Operational disruption, legal-resource costs, future prevention, and reputation remain unmonetized in the strict minimum model when case-specific evidence is insufficient.

7.3 Results

Table 3. VCM Publication Results

Measure	v1.0 publication value	Interpretation
Gross exposure floor	≈ US\$27.2 million	Documented capital transferred or exposed in four quantified cases; not a welfare-loss estimate.
Net realized direct-loss floor L_D	≈ US\$26.2 million	Victim-side direct-loss floor after recovery treatment.
Minimum modeled verification/response burden	≈ US\$35,000	Conservative modeled incident burden; only the subset borne outside the relevant decision-maker is an economic externality.
Minimum modeled B_I	≈ US\$26.2 million	L_D + E_A; mixed observed/estimated burden.
W_S	Not estimated	C_A and D_L are generally unobserved.
AER, Biden robocall	3.23×	Exploratory single-case minimum ratio.

The monetary results are concentrated. Four cases account for the entire quantified gross and direct-loss floors; 36 observations have no publicly measurable direct loss. The appropriate inference is not that those 36 cases caused zero burden, but that public evidence does not support defensible monetization. For private organizations, the direct-loss and internal-response dimensions remain relevant to risk reduction and investment decisions; for policymakers, development institutions, and ecosystem actors, the incidence of costs beyond the victim becomes additionally important. For reproducibility, the underlying workbook retains unrounded calculations; the manuscript deliberately rounds modeled outputs to avoid implying epistemic precision beyond the assumptions.

7.4 AER

$$AER = E_A / C_A$$

The complete case register, source hierarchy, source-clustering treatment, VCM variable dictionary, O/E/U coding rules, FX convention, case-level loss treatment, and reproducibility trace are provided in Appendices A and B.

The Authenticity Externality Ratio is intended to compare externalized authenticity burden with attacker resource cost where both can be estimated. In the Biden robocall case, a reported attacker spend of approximately US\$500

and a minimum modeled E_A of approximately US\$1,616.67 yield an exploratory AER of 3.23. One case is not a population estimate; it is a proof of measurement logic.

8. From Authenticity Risk to Epistemic Security

8.1 Institutions require attributed communication

Modern institutions function through attributed statements: orders, approvals, denials, disclosures, credentials, records, public warnings, market announcements, and delegated authority. When attribution becomes costly to establish, institutional capacity itself can be taxed. The relevant harm is not limited to a recipient believing a false statement. Institutions must also spend resources restoring the distinction between authorized and unauthorized speech.

8.2 The liar's dividend

Deepfakes create a second-order problem: once realistic fabrication is widely understood to be possible, authentic evidence can be dismissed as fabricated. Chesney and Citron (2019) described this as the liar's dividend. This means an authenticity infrastructure has two directions of value: it can help recipients reject unauthorized artifacts, and it can help legitimate speakers demonstrate that authentic artifacts are theirs.

8.3 Efficiently warranted trust

The objective is not maximal skepticism. A society in which every artifact requires forensic analysis would be dysfunctional. Nor is the objective epistemic obedience. Authentication and verification should reduce unnecessary uncertainty about origin and authority so that recipients can allocate attention to substantive evaluation.

The next institutional challenge is not to reverse Dimension 4.5. It is to complete it.

9. Beyond Financial Harm: Political, Social, and Institutional Externalities

9.1 Political deepfakes

Political impersonation is intuitively alarming, but empirical discipline is essential. Research indicates that political deepfakes can affect credibility, trust, and polarization, while human detection remains imperfect. Yet evidence that AI deepfakes have changed the outcome of major elections remains limited and causally difficult to establish (Ching et al. 2025).

The 2023 Slovakia election illustrates the distinction between documented synthetic political content and causal claims about election outcomes. More broadly, a 2024 CETaS/Alan Turing Institute analysis identified 16 confirmed viral AI-enabled disinformation cases during the UK general election and 11 across the EU and French elections, but found no evidence that these cases meaningfully affected election results; the same research nevertheless documented confusion over authenticity, harassment, and broader risks to democratic-system integrity (Stockwell 2024). The paper therefore treats political cases as evidence of authenticity uncertainty and institutional remediation burden, not as proof of changed election outcomes.

9.2 Trust as a stock

Repeated impersonation can affect a broader stock of institutional trust even when no single artifact changes a measurable outcome. The challenge is empirical: trust is multidimensional, slow-moving, and influenced by many

confounders. For this reason, reputational and institutional-trust costs remain conceptually important but unmonetized in the current VCM.

9.3 Political externalities as future research

A dedicated political dataset should distinguish exposure, engagement, belief, behavioral response, turnout, agenda effects, and institutional remediation. It should also separate successful persuasion from the liar's dividend and from generalized uncertainty. This paper deliberately does not force these outcomes into a single loss metric.

10. Existing Responses — And the Missing Layer

10.1 Detection

Deepfake detection is necessary but structurally difficult. A 2026 cross-modality survey and empirical evaluation found persistent out-of-distribution generalization gaps and substantial vulnerability to white-box adversarial attacks across evaluated methods, illustrating why detector performance in controlled benchmarks does not automatically translate into robust real-world verification (Deepfake Detection Across Image, Video, and Audio 2026). Detection therefore remains one layer of defense rather than a complete institutional solution.

10.2 Provenance and Content Credentials

C2PA provides an open technical standard for attaching cryptographically signed, tamper-evident Content Credentials to digital assets and for expressing verifiable assertions about their provenance. This is a substantial capability: it can bind assertions to an asset, identify signers, record creation or editing actions, and allow compatible systems to validate whether those assertions remain correctly formed and associated with the underlying asset. C2PA also deliberately avoids value judgments about whether provenance is 'good' or 'bad,' and its own explainer states that provenance information alone cannot establish whether depicted content is true, accurate, or factual (C2PA 2026a, 2026b).

The boundary matters for the architecture proposed here. C2PA solves a core part of the authenticity problem exceptionally well when Content Credentials are present and verifiable: it standardizes how provenance assertions can travel with content across an ecosystem. It does not, by design, guarantee complete provenance; credentials or metadata may be absent, stripped, or interrupted by non-aware workflows, and provenance does not by itself resolve every question of current authority, registry status, revoked delegation, or cross-signal conflict. Verification and orchestration therefore do not replace C2PA. They consume C2PA evidence where available, combine it with other authenticity signals when necessary, and present the resulting evidence at the decision point.

10.3 Watermarking

Google DeepMind's SynthID embeds imperceptible signals in AI-generated images, audio, text, and video. Google itself describes SynthID as an important building block rather than a silver bullet. Watermarks can complement metadata because they may survive transformations that strip or separate metadata, while metadata can carry richer contextual information.

10.4 Multi-layer provenance

OpenAI's 2026 provenance approach makes the emerging stack particularly visible: C2PA conformance, SynthID watermarking, and public/API verification are combined rather than treated as mutually exclusive techniques. OpenAI explicitly describes C2PA metadata as an important provenance foundation while noting that metadata can be stripped or lost through transformations; its verification tooling checks multiple supported signals and avoids definitive conclusions when no signal is detected (OpenAI 2026). The significance is architectural. No single signal is assumed to survive every transformation or cover every content ecosystem.

10.5 Cybersecurity and the Authenticity Attack Surface

For CISOs and CIOs, the practical implication is not the creation of an unrelated trust category outside cybersecurity. It is an expansion of the surface that cybersecurity must protect. Account takeover, identity compromise, executive impersonation, business-email compromise, deepfake fraud, forged documents, and unauthorized institutional communications increasingly converge. The relevant question becomes: after protecting the account, device, network, and identity, how does the organization allow a recipient to verify that a particular artifact or instruction is actually authorized?

10.6 The emerging authenticity stack

Table 4. The Emerging Authenticity Stack

Layer	Representative capability	Function
Capture authenticity	Authenticated cameras / secure capture	Bind evidence to origin at creation.
AI-generation marking	Generator-side disclosure	Mark content produced by participating AI systems.
Provenance	C2PA / Content Credentials	Preserve signed assertions and modification history.
Watermarking	SynthID and equivalents	Provide durable embedded signals.
Detection	Deepfake / manipulation detectors	Estimate whether artifacts show synthetic/manipulation patterns.
Verification & orchestration	Cross-signal verification layer	Resolve heterogeneous evidence into usable authenticity checks at the decision point.

The missing layer is not another detector. In the terminology used throughout this paper, authentication establishes a binding, verification checks that evidence at a later decision point, and orchestration discovers, combines, validates, and presents heterogeneous evidence across standards, registries, watermarks, identity or authorization systems, and detection outputs. The proposed gap is therefore an interoperability and usability function above—not instead of—the underlying technical controls.

11. From the Right to Information Authenticity to Authenticity Infrastructure

11.1 From norm to institution

White Paper No. 1 advanced a normative argument: persons and institutions should have a meaningful right to distinguish information genuinely attributable to a source from information falsely attributed to that source (Rodríguez Alemán 2026a). The present paper advances a different, economic and institutional argument: authenticity uncertainty can generate verification, response, remediation, and ecosystem burdens whose incidence may extend beyond the actor who created the uncertainty. Neither proposition logically depends on the other. A right to information authenticity could be defended even if measured externalities were small, and an authenticity externality could exist even without recognizing a corresponding right.

The two arguments nevertheless converge on an institutional design problem. The normative argument asks whether people and institutions should be able to establish authentic attribution. The economic argument asks whether the cost of doing so is inefficiently distributed or repeatedly recreated. Together they motivate a practical question: what infrastructure can make authenticity rights exercisable while reducing unnecessary verification burden without centralizing authority over truth?

11.2 Digital Authenticity Infrastructure

The institutional implication is Digital Authenticity Infrastructure: interoperable capabilities that make evidence of origin, integrity, identity, and authority available at the point where decisions are made. Such infrastructure need not be a single public platform. It can be composed of private services, public registries, open standards, trusted issuers, secure devices, cryptographic systems, and verification interfaces.

11.3 Authenticate Once, Verify Many

The design principle is Authenticate Once, Verify Many. More precisely, a legitimate source establishes authentication for a defined artifact, claim, identity, or authority state; recipients can then verify that evidence repeatedly while it remains valid. Revocation, expiration, modification, credential compromise, and changes in delegated authority require re-authentication or updated evidence. The principle is therefore about reuse of valid trust evidence, not permanent authentication.

11.4 Registries of authoritative originals

An authenticity registry need not declare that a statement is true. It can record that a particular institution or authorized actor registered an artifact, hash, credential, or authority assertion at a given time. The function is closer to an authoritative record than to a ministry of truth. This distinction is fundamental to preserving pluralism.

11.5 The right also protects the speaker

Authenticity rights are often described from the recipient's perspective: the right to know whether content is genuine. The speaker has a complementary interest. A person or institution should be able to demonstrate 'this is ours' and, equally important, 'that was not us.' Positive authentication protects authorized speech from both impersonation and opportunistic denial.

11.6 The SEC case revisited

The U.S. Securities and Exchange Commission (SEC) case now acquires additional significance. The institution had an authentic channel, but the channel was temporarily controlled by an unauthorized actor. A content-level positive-authentication mechanism could have provided recipients with a second, independent question: not merely 'did this appear on the SEC account?' but 'is this specific announcement registered or signed as an authorized SEC communication?'

11.7 Underproduction of authentication

Markets can underproduce reusable authentication when the authenticating party bears the implementation cost but much of the benefit accrues to recipients, counterparties, platforms, and society. The result can be simultaneous underproduction of positive authentication and overproduction of authenticity uncertainty. This is the mirror image of the IAE and a candidate positive externality of authenticity infrastructure.

12. Completing Dimension 4.5

12.1 From trust signals to trust evidence

We optimized the movement of information before fully developing equivalent infrastructure for proving its origin. Historically, recipients relied on visual design, platform badges, familiar domains, voices, faces, institutional logos, and social context. Generative AI reduces the evidentiary value of many of these signals because they can be reproduced cheaply.

The transition required is from trust signals to trust evidence. Plausibility becomes cheap; proof becomes valuable. Here 'proof' is used in the practical sense of verifiable evidence supporting origin, integrity, identity, or authority—not proof that the substantive content is true.

12.2 Orchestration as institutional usability

Even strong provenance technologies create limited social value if recipients cannot access, interpret, or combine them at the moment of decision. Orchestration is therefore defined here as the discovery, combination, validation, and presentation of heterogeneous authenticity evidence so that verification remains usable across technologies and channels. It may include provenance credentials, watermarks, authoritative registries, identity and authorization signals, secure-capture evidence, and detector outputs. Orchestration is not a truth engine: its purpose is to make evidence about origin and authority usable, expose conflicts or missing signals, and preserve uncertainty where the available evidence does not justify a conclusion.

12.3 Universality and technology agnosticism

A durable authenticity layer cannot depend on one generator, one platform, one watermark, one file format, or one cryptographic vendor. It should be able to consume heterogeneous evidence and improve as new techniques emerge. Universality is therefore not equivalent to centralization. It means that the verification function should remain usable across technologies and accessible to actors who did not participate in a single closed ecosystem.

12.4 An adversarial infrastructure

Authenticity is structurally adversarial. Successful defenses create incentives for circumvention: metadata stripping, watermark degradation, credential theft, replay, account compromise, synthetic reconstruction, or attacks on the verifier itself. This implies a permanent R&D requirement. Authenticity infrastructure is not a one-time product feature; it is an evolving security capability.

12.5 Privacy and inclusion

Verification must not require universal public identity. Anonymous and pseudonymous speech can be legitimate. The relevant design question is whether a particular claim of identity or authority can be substantiated when such a claim is made. Privacy-preserving credentials, selective disclosure, and scoped authority can allow authenticity without converting every communication into a surveillance event.

12.6 Agentic AI and Digital Authority Authenticity

Agentic AI expands the problem from content authenticity to digital authority authenticity. An autonomous or semi-autonomous agent may be genuine yet unauthorized for a particular action; authorized for one scope but not another; or acting under a delegation that has expired. The future question becomes: who or what is authorized to act for whom, in what capacity, for which operation, and at what time?

This frontier intersects machine identity, authorization, cybersecurity, and AI governance. It should be treated as an extension of the Authenticity Attack Surface rather than as evidence that content provenance is obsolete.

12.7 Resilience requires more than cryptography

Cryptography is indispensable but not sufficient. Key compromise, poor authority governance, vulnerable endpoints, social engineering, unavailable registries, and weak recovery procedures can undermine systems built on strong algorithms. The same applies to post-quantum cryptography (PQC): quantum-resistant primitives may be necessary for long-lived credentials, but they do not by themselves create resilient authenticity institutions.

12.8 Distribution + Trust

Dimension 4.5 was built around distribution. Completing it requires a second primitive: trust evidence. The objective is not to slow information until it becomes safe. It is to make legitimate authenticity cheap enough that trust can travel with information.

13. Limitations and Research Agenda

13.1 Dataset limitations

The 43-case dataset is exploratory, purposive, source-dependent, and not representative. Observed cases are a subset of actual incidents; underreporting is unknown. Eighteen observations derive from one Unit 42 investigation and therefore constitute a source cluster. The unit of observation also varies between discrete incidents and recurrent campaigns.

13.2 Monetary evidence

Public monetary evidence is sparse. Four cases drive the quantified direct-loss floor. Absence of a reported loss is not evidence of zero loss. Conversely, gross fraudulent transfers should not be treated as net social welfare destruction. The VCM therefore reports direct loss, externalized burden, and welfare cost separately.

13.3 Modeled verification burden

The current modeled verification, response, remediation, and ecosystem burden is deliberately conservative. It is not an observed invoice total, and not every modeled component is automatically an economic externality. Incidence must be established case by case: victim-internal response differs from cost shifted to recipients, counterparties, platforms, public agencies, or other third parties. Several plausible cost layers—operational disruption, legal-resource cost, future prevention, and reputation—remain unmonetized where case-specific evidence is insufficient.

13.4 Exposure is not belief

Exposure, engagement, belief, behavioral action, and causal outcome are distinct. This is particularly important for political deepfakes. Viral distribution does not prove persuasion; persuasion does not prove electoral effect. A dedicated research design is required for political outcomes.

13.5 Avoided Verification Cost

$$AVC = VC_0 - VC_1$$

A future empirical program should measure Avoided Verification Cost: the difference between verification cost under existing decentralized practices and verification cost after reusable positive authentication is introduced. Product telemetry, controlled experiments, and institutional time studies could provide stronger evidence than benchmark transfer.

13.6 Verification Economies as a testable proposition

$$N^* = C_{Aut} / (C_{v0} - C_{v1})$$

The Verification Economies condition is empirically testable rather than established by the present dataset. Research can estimate authentication fixed cost, recipient verification time before and after intervention, validity and revocation costs, false-positive and false-negative costs, and adoption rates. The hypothesis should be rejected where reusable authentication fails to reduce expected social cost or where C_{v1} is not lower than C_{v0} .

13.7 Authenticity Commons

The term Authenticity Commons has prior use and is not claimed as original here. Economically, however, the concept remains useful: when trusted origin evidence is reusable across an ecosystem, one actor's authentication investment can benefit many others. This creates a candidate positive externality that may justify shared standards, interoperability, or collective infrastructure.

13.8 Risks created by authenticity infrastructure

Authenticity systems can create their own externalities: surveillance, exclusion of actors without credentials, concentration of verification power, false positives, false negatives, credential compromise, chilling effects, and overreliance on automated trust labels. A credible research program must evaluate these harms rather than assume that more authentication is always socially optimal.

13.9 Falsifiability

The framework should be treated as falsifiable. The empirical claim is not that every authenticity incident produces large externalities. It is that authenticity uncertainty can externalize meaningful verification and response costs, that these costs are measurable in principle, and that reusable positive authentication can reduce them under identifiable conditions. Evidence showing negligible downstream burden, prohibitive authentication cost, or persistent inability to reuse trust evidence would weaken the framework.

13.10 Appropriate claim

The present evidence is sufficient to treat authenticity uncertainty as a potentially consequential economic and institutional externality, and to justify systematic measurement, experimentation, and infrastructure research. It is not sufficient to estimate the total social cost of authenticity uncertainty across the economy.

What does authenticity uncertainty cost society, how are those costs distributed, and which institutions can reduce them most efficiently without sacrificing openness, privacy, or pluralism?

14. Conclusion: Completing the Trust Layer

Dimension 4.5 is one of the great achievements of the digital era. It lowered the cost of moving information, coordinating people, creating knowledge, entering markets, and participating in public life. Its positive externalities are too important to treat the authenticity problem as an argument for retreat.

But the economics of information changed faster than the institutions that support trusted attribution. Generative AI now makes plausible synthetic evidence cheaper to produce, while authentication, verification, response, and remediation burdens can be distributed across recipients, organizations, platforms, and public institutions. Some of these burdens remain private and directly affect the organization or person targeted; others are shifted outward and constitute the Information Authenticity Externality. Both matter. Incident Burden explains why a firm, bank, public authority, or other organization may rationally invest in authenticity infrastructure for its own benefit. The IAE explains why the same investment can create additional value for customers, counterparties, markets, public institutions, and society.

The 43-case exploratory dataset does not estimate prevalence, but it demonstrates a diverse Authenticity Attack Surface: corporate executives, central banks, governments, political actors, media identities, investment brands, and even authentic institutional channels can be used to borrow authority. The Verification Cost Model provides an initial welfare-consistent way to separate gross exposure, realized victim loss, externalized authenticity burden, and social welfare cost without pretending that missing evidence is zero.

The framework also clarifies the architectural boundary. Digital authenticity expands the surface that cybersecurity and institutional trust systems must protect, while provenance, watermarking, detection, authenticated capture, identity, and authorization remain complementary controls rather than substitutes for one another.

This dual perspective also clarifies adoption incentives. Authenticity infrastructure need not depend on altruistic internalization of a social externality. A private organization can adopt authentication because the expected reduction in fraud, investigation, remediation, operational, or reputational burden exceeds its private cost. Where the resulting trust evidence also lowers verification or risk for others, the same investment produces an additional positive externality. Commercial adoption and public-interest infrastructure can therefore reinforce one another without sharing the same objective function.

The emerging architecture is therefore a stack: authenticated capture, AI-generation marking, provenance, watermarking, detection, identity and authorization controls, authoritative originals, and verification. Its institutional value depends on orchestration—turning heterogeneous signals into usable evidence at the point of decision.

This produces a shift from negative detection to positive authentication. Instead of asking only whether an artifact looks fake, institutions can increasingly establish evidence that a defined artifact, claim, or authority state is theirs and allow that evidence to be verified repeatedly while valid. The Verification Economies hypothesis asks whether this reuse can reduce expected marginal verification cost sufficiently to outperform decentralized reconstruction of trust.

White Paper No. 1 framed information authenticity normatively as a right. This paper independently frames authenticity uncertainty as an economic and institutional incidence problem. The arguments do not prove one another, but they converge on the same implementation challenge: making authentic attribution practically exercisable without turning verification infrastructure into a centralized arbiter of truth.

AliPisé represents one emerging effort to operationalize this institutional need. Rather than attempting to adjudicate truth, its proposed role is to provide a technology-agnostic verification and orchestration layer through which authentic originals can be registered, evidence of origin can be preserved, and subsequent content can be evaluated against authenticated sources. In this sense, AliPisé is conceived not as an alternative to provenance standards, watermarking, deepfake detection, authenticated capture, or cybersecurity controls, but as one possible component of a broader Digital Authenticity Infrastructure.

This imbalance need not imply that the digital revolution failed. It implies that it remains incomplete. We built the infrastructure to move information. We are now beginning to build the infrastructure that allows information to prove where it came from. That is the missing trust layer.

The next institutional challenge is not to reverse Dimension 4.5. It is to complete it.

Author Disclosure

The author is the founder of AliPsé, an initiative developing digital-authenticity verification and orchestration infrastructure. AliPsé is discussed in this white paper as one emerging implementation of the institutional architecture analyzed herein. This relationship should be considered when evaluating the paper's applied implications; the theoretical propositions, empirical dataset, VCM assumptions, and cited external evidence are presented so that they can be assessed independently of AliPsé's commercial development.

References

- Anderson, Ross, and Tyler Moore. 2006. "The Economics of Information Security." *Science* 314 (5799): 610–613. <https://doi.org/10.1126/science.1130992>.
- Banca d'Italia. 2024a. "Warning: Deepfake Videos Circulating Online." May 22, 2024. <https://www.bancaditalia.it/media/notizia/warning-deepfake-videos-circulating-online/>.
- Banca d'Italia. 2024b. "Warning: Fraudulent Deepfake Videos Online." July 1, 2024. <https://www.bancaditalia.it/media/notizia/warning-fraudulent-deepfake-videos-online>.
- Banca d'Italia. 2026a. "Warning: Online Circulation of False and Fraudulent Content Unlawfully Using the Name and Image of Fabio Panetta, Governor of Banca d'Italia." February 26, 2026.
- Banca d'Italia. 2026b. "Warning: Online Circulation of New False and Fraudulent Content Unlawfully Using the Name and Image of Panetta, Governor of Banca d'Italia." July 8, 2026.
- Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. "Bad Is Stronger Than Good." *Review of General Psychology* 5 (4): 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>.
- Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114 (28): 7313–7318. <https://doi.org/10.1073/pnas.1618923114>.
- Brynjolfsson, Erik, Avinash Collis, W. Erwin Diewert, Felix Eggers, and Kevin J. Fox. 2019. "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy." NBER Working Paper 25695. <https://doi.org/10.3386/w25695>.
- C2PA. 2026a. "C2PA and Content Credentials Explainer." Coalition for Content Provenance and Authenticity. <https://c2pa.org/specifications/specifications/2.2/explainer/Explainer.html>.
- C2PA. 2026b. "Guiding Principles for C2PA Designs and Specifications." Coalition for Content Provenance and Authenticity. <https://c2pa.org/principles/>.
- Chesney, Robert, and Danielle Keats Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107: 1753–1820.
- Ching, Didier, John Twomey, Matthew P. Aylett, Michael Quayle, Conor Linehan, and Gillian Murphy. 2025. "Can Deepfakes Manipulate Us? Assessing the Evidence via a Critical Scoping Review." *PLOS ONE* 20 (5): e0320124. <https://doi.org/10.1371/journal.pone.0320124>.
- Cisternas, Gonzalo, and Jorge Vásquez. 2026. "Misinformation in Social Media: The Role of Verification Incentives." Federal Reserve Bank of New York Staff Report 1028, revised July 2026. https://www.newyorkfed.org/research/staff_reports/sr1028.
- Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44. <https://doi.org/10.1086/466560>.

- Creane, Anthony, Thomas D. Jeitschko, and Kyoungbo Sim. 2022. “Welfare Effects of Product Certification under Latent Adverse Selection.” *International Journal of Industrial Organization* 81: 102826. <https://doi.org/10.1016/j.ijindorg.2022.102826>.
- Federal Communications Commission. 2024. “FCC Fines Man Behind Election Interference Scheme \$6 Million for Sending Illegal Robocalls That Used Deepfake Generative AI Technology.” September 26, 2024. <https://docs.fcc.gov/public/attachments/DOC-405811A1.pdf>.
- Federal Reserve Board. 2025. “Foreign Exchange Rates—G.5A Annual.” January 6, 2025. <https://www.federalreserve.gov/releases/g5a/20250106/>.
- Goldstein, Josh A., Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. 2024. “How Persuasive Is AI-Generated Propaganda?” *PNAS Nexus* 3 (2): pgae034. <https://doi.org/10.1093/pnasnexus/pgae034>.
- Google DeepMind. 2026. “SynthID: A Tool to Watermark and Identify Content Generated through AI.” <https://deepmind.google/models/synthid/>.
- Hong Kong Government. 2024. “LCQ9: Combating Frauds Involving Deepfake.” June 26, 2024. <https://www.info.gov.hk/gia/general/202406/26/P2024062600192.htm>.
- Kattwinkel, Deniz, and Jan Knoepfle. 2023. “Costless Information and Costly Verification: A Case for Transparency.” *Journal of Political Economy* 131 (2): 504–548. <https://doi.org/10.1086/721618>.
- Leica Camera AG. 2026. “Content Credentials.” <https://leica-camera.com/en-int/photography/content-credentials>.
- Merlino, Luca P., Paolo Pin, and Nicole Tabasso. 2023. “Debunking Rumors in Networks.” *American Economic Journal: Microeconomics* 15 (1): 467–496. <https://doi.org/10.1257/mic.20200403>.
- Nguyen-Le, Hong-Hanh, Van-Tuan Tran, Dinh-Thuc Nguyen, and Nhien-An Le-Khac. 2026. “Deepfake Detection across Image, Video, and Audio: A Comprehensive Survey with Empirical Evaluation of Generalization and Robustness.” *Artificial Intelligence Review* 59: 203. <https://doi.org/10.1007/s10462-026-11608-4>.
- OECD. 2026. *Digital Government Outlook 2026*. Paris: OECD Publishing.
- OpenAI. 2026. “Advancing Content Provenance for a Safer, More Transparent AI Ecosystem.” May 19; updated July 31, 2026. <https://openai.com/index/advancing-content-provenance/>.
- Palo Alto Networks Unit 42. 2024. “The Emerging Dynamics of Deepfake Scam Campaigns on the Web.” <https://unit42.paloaltonetworks.com/dynamics-of-deepfake-scams/>.
- Pigou, A. C. 1920. *The Economics of Welfare*. London: Macmillan.
- Rao, Justin M., and David H. Reiley. 2012. “The Economics of Spam.” *Journal of Economic Perspectives* 26 (3): 87–110. <https://doi.org/10.1257/jep.26.3.87>.
- Rathje, Steve, Jay J. Van Bavel, and Sander van der Linden. 2021. “Out-group Animosity Drives Engagement on Social Media.” *Proceedings of the National Academy of Sciences* 118 (26): e2024292118. <https://doi.org/10.1073/pnas.2024292118>.
- Robertson, Claire E., Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J. Van Bavel, and Stefan Feuerriegel. 2023. “Negativity Drives Online News Consumption.” *Nature Human Behaviour* 7: 812–822. <https://doi.org/10.1038/s41562-023-01538-4>.
- Rodríguez Alemán, Luis Ángel. 2016. “La informática en un mundo interdisciplinario.” In *Reflexiones de la Academia Mexicana de Informática a los 40 años de su fundación*. Universidad de Guadalajara. [Final page range to be verified before publication.]

- Rodríguez Alemán, Luis Ángel. 2026a. The Right to Information Authenticity. AliPsé White Paper No. 1. English edition.
- Rodríguez Alemán, Luis Ángel. 2026b. El derecho a la autenticidad de la información. AliPsé White Paper No. 1. Spanish edition.
- Rozin, Paul, and Edward B. Royzman. 2001. “Negativity Bias, Negativity Dominance, and Contagion.” *Personality and Social Psychology Review* 5 (4): 296–320. https://doi.org/10.1207/S15327957PSPR0504_2.
- Salvi, Francesco, Manoel Horta Ribeiro, Riccardo Gallotti, et al. 2025. “On the Conversational Persuasiveness of GPT-4.” *Nature Human Behaviour* 9: 1645–1653. <https://doi.org/10.1038/s41562-025-02194-6>.
- SEC (U.S. Securities and Exchange Commission). 2024. “Statement on Unauthorized Access to the SEC’s @SECGov X.com Account.” January 12, 2024. <https://www.sec.gov/newsroom/speeches-statements/gensler-x-account>.
- Stockwell, Sam. 2024. AI-Enabled Influence Operations: Threat Analysis of the 2024 UK and European Elections. Centre for Emerging Technology and Security, The Alan Turing Institute. September 19, 2024. <https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-threat-analysis-2024-uk-and-european-elections>.
- Tan, Yiheng, Xiyang Huang, and Wei Li. 2023. “Does Blockchain-Based Traceability System Guarantee Information Authenticity? An Evolutionary Game Approach.” *International Journal of Production Economics* 264: 108974. <https://doi.org/10.1016/j.ijpe.2023.108974>.
- Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The Spread of True and False News Online.” *Science* 359 (6380): 1146–1151. <https://doi.org/10.1126/science.aap9559>.
- World Bank. 2016. *World Development Report 2016: Digital Dividends*. Washington, DC: World Bank.