

AI-Mediated Ethical Communication: Epistemic Totalization the Fabricatable Zone and the Governmental Chain of Custody as Terminus

Jonathan David Nelson¹

¹MPS, MA

¹Disciplinary Home: MA in Performing Arts / MPS in Communication Project Type: Extended Research Paper — Continuation of Nelson (2025) Status: Full Draft

Publication Date: 2026/09/26

How to Cite: Jonathan David Nelson (2026) AI-Mediated Ethical Communication: Epistemic Totalization the Fabricatable Zone and the Governmental Chain of Custody as Terminus. *International Journal of Innovative Science and Research Technology*, 11(9), 1500-1528. <https://doi.org/10.38124/ijisrt/26sep879>

TABLE OF CONTENTS

ABSTRACT	1502
CHAPTER ONE INTRODUCTION: THE CREDENTIAL AND THE DREAM.....	1503
The Dream.....	1503
The Problem Identified in This Paper.....	1503
What this Paper is Not Arguing	1504
The Behavioral Pattern and Where It Begins.....	1504
The Argument in Brief	1505
CHAPTER TWO LITERATURE REVIEW	1506
The Impostor Phenomenon and the Psychology of Genuine Competence.....	1506
Principle-Based and Fairness-Oriented AI Ethics	1506
Dramaturgy, Performance, and the Presentation of Self	1507
Identity Consolidation Across the Lifespan	1508
Credential Fraud, Manufactured Consensus, and Self-Referential Publishing.....	1508
Pluralism in Education and Its Preconditions.....	1508
Testimony, Trust, and the Ethics of Deference	1509
Human-Machine Communication and the Sociology of Identity Online.....	1509
A Note on Method and Scope	1509
The Gap in the Literature.....	1510
CHAPTER THREE THEORETICAL FRAMEWORK: EPISTEMIC TOTALIZATION AND THE INDUSTRIALIZATION OF IDENTITY FABRICATION	1511
Epistemic Totalization: The Original Claim	1511
The Fabricatable Zone and its Boundary	1511
A Pre-Existing Pathology, Industrialized	1512
The Personality Dimension: Costume and Content.....	1512
The Performing Arts Dimension	1513
Objections and Responses	1513
The Author's Position	1514
CHAPTER FOUR THE FIVE FAILURE MODES OF AI-MEDIATED COMMUNICATION	1515
A Worked Hypothetical, Before the Case Study	1515
The AI's Undeclared Itinerary.....	1515
Interpretive Paraphrasing as Semantic Control	1516
The Burden of Proof Inversion.....	1516
Probabilistic Completion of Bad-Faith Questions	1517
Temporal Drift and the Erosion of Finality.....	1517
Synthesis: How the Five Failure Modes Compound	1517
CHAPTER FIVE THE HARD WALL: GOVERNMENTAL CHAIN OF CUSTODY AS EPISTEMIC TERMINUS	1519
What the Fabricatable Zone Cannot Reach.....	1519
The Behavioral Signature of Approach	1519
What This Means for Ethical Communication Frameworks.....	1520
Toward Institutional Design Recommendations.....	1520
CHAPTER SIX THE ANONYMIZED CASE STUDY: COSTUME, CONTENT, AND THE LIMITS OF DIGITAL CONFIRMATION.....	1522
Introduction to the Case and a Note on Evidentiary Status.....	1522
The Architecture and Its Alleged Construction	1522
The Digital Infrastructure as a Confirmation Engine.....	1522
The Behavioral Signature: Escalation Under Scrutiny	1523
The Personality Dimension: The Costume That Fits.....	1524
The AI System as a Structurally Ideal Audience.....	1524
The Hard Wall, Approached	1525
Limits of the Case Study	1525
CHAPTER SEVEN CONCLUSION.....	1526
REFERENCES.....	1528

ABSTRACT

Large language models and AI-mediated search systems have become the primary intermediaries for institutional information across educational, legal, civic, and professional life. Current models of AI ethics focus primarily on the accuracy, bias, and fairness of AI systems. This focus is essential but fails to recognize a structural failure that precedes accuracy: in an information context, where the predominant risk to epistemic integrity is not misinformation or error but the fabrication of institutional authority, AI systems are designed to produce coherent, confident, and complete responses. There is a structural match between the optimization process and the associated threat. This paper introduces two theoretical constructs to name this problem. Epistemic totalization is the first type, in which the communicative functions of inquiry, assertion, interpretation, and authority are lost in a total function that lacks concession, appeal, or authorship.

The second, the distinction between fabricated and genuine competence, redirects the difference between real and performed expertise from an epistemic to a behavioral and psychological dimension, genuine competence that shows convergence when challenged, and fabricated competence that shows escalation. The paper also builds on theories of embodiment in the performing arts to suggest that the AI confirmation systems are an ideal, non-threatening, confirming audience for a particular, less-theorized psychological state that differs from what Carl Rogers would call a stable self, in which professional performance becomes a part of identity instead of a role adopted over a stable self. The paper argues that there is one enduring boundary that current AI systems cannot breach: the governmental chain of custody, which includes court dockets, licensing registries, federal identifier databases, sworn filings, and other processes that precede individual claims and require accountable human beings to record facts rather than merely confirm presentations at each step.

The case study is presented as an unadjudicated, anonymous example from a public record; it appears as evidence in the process of collapsing, link by link, as the governmental terminus approaches, but it is never considered a fact but rather evidence that is as much a tool for self-referential credential architecture as a means of its destruction. The paper argues that ethical communication frameworks should also change with respect to what an AI system is structurally unable to do: defer to unreachable authority, note records it cannot access, and refuse unverified confirmation. This argument does not end the identification of pluralism in credentialing and education; it makes it possible. Inclusion is only a pluralistic approach with chain-of-custody integrity. Standards cannot replace pluralism: the process of learning can be flexible, but the content of learning cannot be negotiated.

Keywords: *AI Ethics, Epistemic Totalization, Credential Fabrication, Chain of Custody, Impostor Phenomenon, Performance Theory, Identity, Algorithmic Confirmation, Pluralistic Education, Epistemic Authority.*

CHAPTER ONE

INTRODUCTION: THE CREDENTIAL AND THE DREAM

➤ *The Dream*

In occupations where licenses provide direct access to complex areas, professionals commonly report a certain psychological experience that is openly discussed within the field. It is expressed through specific anxiety, constant self-evaluation, and most clearly, a discernible type of dream.

The dream comes in different forms, but with a common "architecture." The professional enters an unfamiliar work environment. They cannot locate their department, students, clients, or patients. They do not have to do their work, but something else is going on. They are unnoticed. The credentials that put them there provide no guidance; they roam around the corridors, search and then consider hiding.

This dream is reported by educators. Attorneys report it. It is reported by surgeons, researchers, architects, and engineers. It is informally shared between colleagues in these professions, where a shared experience is immediately recognizable; it does not need to be explained to another professional. It is, in fact, a profession folk category one that will come up over and over in unrelated endeavors, and so it is a datum to be taken seriously.

Rising to the heights of the dream does not mean the dreamer is lacking at the base. It indicates that the professional standards have been internalized successfully. The unconscious mind needs to have a detailed and accurate picture of the job it is supposed to be doing, a picture detailed enough and accurate enough that it can assemble a plausible story of what the job requires to fail. Content can be a source of anxiety. It takes the dreamer to know, at a subconscious level, what competence in their area of interest or expertise actually means and how far from it they might be on any particular day. Everyone familiar with the requirements of a surgical ward can imagine consistently meeting them.

There is a clinical term for this experience: impostor phenomenon, first formally documented by high-achieving professional women, Clance and Imes (1978), and later expanded to cover more genders, disciplines, and career stages. For this paper, however, the relevance of the impostor phenomenon lies more in its epistemological implications than in its clinical importance. It is epistemological. The professional anxiety dream is the internal state of a person who has achieved a tangible gain and is, at some level below conscious thought, aware of its price. It is the unspoken sign of responsibility. It represents the genuine experience of true credentialing, occurring years after the actual awarding and long after the last relevant evaluation has taken place.

Only individuals who have not fabricated their professional identities experience this dream in its truest form; this is not because those who have fabricated their credentials are more confident, but because the dream necessitates a model that a fabricated credential cannot provide detailed enough for them to assess themselves against and realize that they do not measure up. There is an interior for real skills, one to which anxiety can register, but no interior for fake skills. No lining of the costume exists. Inside, there is nothing to cause the nightmare.

The internal architecture of the systems currently in play to govern professional identity in digital environments is unknowable to them. They cannot measure what ails a man at night. They can't hear the private audit in the minds of those who know how things are going and how far they need to go. These systems are capable of reading a credential string, checking out a title, and then producing detailed, natural, authoritative-sounding, and, in some cases, manufactured explanations of professional identities that may have nothing to do with reality.

This paper addresses the "gap" between what digital confirmation systems can verify and what professional legitimacy demands.

➤ *The Problem Identified in This Paper*

AI tools have emerged as trusted and essential gatekeepers of information in academic, legal, civic, and professional settings, thanks to recent developments in large language models. It is becoming more common for people to use generative summarization rather than a list of links they verify for themselves. Instead of consulting a registry, professional inquiries are now often answered by querying a system optimized for a narrow range of output properties: completeness, coherence, and confident answerability.

Such systems are designed to return a response that appears resolved, taking into account both the training goals and the accountability of those responsible for the design. Even when there is no answer that can be provided, an unresolved answer, "I do not know," "this cannot be verified from here," etc., is a failure mode that the system has been implicitly trained against, as it feels unhelpful to the user.

The problem with this optimization is that it is a structural issue that current AI ethics frameworks have not yet addressed. The extensive and significant literature on algorithmic bias, misinformation, and factual accuracy, stemming from Floridi and Cowls's

(2019) articulation of AI core-ethics principles and Binns (2018) from the fairness-in-machine-learning literature, is almost entirely focused on what AI systems say: whether they are true, biased against a protected characteristic, or whether they reflect the values of the people they affect. This literature has inadequately explored what AI systems are structurally unable to do: differentiate between the performance of authority and its substance, between the assertion of a credential and its verification, between the digital presentation of institutional identity and the chain of custody that would confirm or refute it.

The structural inability to go beyond the zone of the actual fabricable self-presentation into the governmental chain of custody, from which no one can ever be removed, creates an environment in which an executable professional identity is validated as authentically valid, in which the behavioral and psychological hallmarks that make a professional identity “real” are completely invisible to the systems claiming to validate it, and in which the difference between what is real and what is fake is completely blurred. The outcome extends beyond a mere technology issue that a better retrieval pipeline can resolve. This situation is an ethical communications problem because the primary communicative actors within institutions cannot distinguish between credentials and aspirations, nor between titles and the valuable, responsible, and often stressful years of experience that confer the necessary authority to those titles.

➤ *What this Paper is Not Arguing*

This paper's claims should be clarified at the outset, as the following argument is sufficiently similar to adjacent arguments to be confused with them.

This paper supports ambition, aspiring knowledge, and untraditional routes to professional knowledge. These are neither the topic nor the object of the subsequent discussion. This paper does not aim to specify the sources of knowledge acquisition. This paper does not focus on professionals who acquired knowledge in non-traditional ways and sought verification through examinations and certification.

In such a case, the pathway was unconventional; the content was not. That is the central point of the entire argument, and it needs to be reiterated. The whole argument pivots on this point, which needs reiteration: This paper does not criticize alternative paths to competence. The argument is that alternatives to competence must ultimately lead to something that transcends the individual's own definition of competence. This paper supports the author's earlier thesis, that there is a need for pluralistic recognition of informal and non-traditional educational institutions (Nelson, 2025). This argument was based on the implicit assumption that the defended paths lead to genuine content and that the credential, regardless of how unorthodoxly it was obtained, verifies the existence of something tangible on the other side of those paths. This paper reiterates the preceding argument. It's a true reflection of what happens to pluralistic models when they are used by a group or organization with no content whatsoever in their architecture anywhere along the chain.

Under these circumstances, the complete absence of any pathway may be the ambient cover for the defense of non-standard pathways. An alternative pathway to true competency isn't an alternative to a fake credential pathway. It's just more susceptible to abuse. This paper will argue that the verification mechanism's integrity is essential to genuine pluralism; it should not block unconventional learning, as it is the only tool that can tell what is real from what is merely performed.

➤ *The Behavioral Pattern and Where It Begins*

The behaviors investigated in the case study in Section VI are not new in adult life. If you've worked across education or evaluation, at any point, you'll see earlier versions of the same spectrum of behaviors, and the recognition of that continuity is beneficial because it de-exoticizes the adult case; this phenomenon isn't the bizarre case of the internet, it's the more familiar case, with the same behavioral pattern, extended out in duration through digital infrastructure.

One extreme end of the spectrum is the mixing up of borrowed ideas with original thinking; this confusion is a knowledge matter, not a character issue, which is usually correctable through teaching and discipline. Further along the spectrum is the problematic trend of substituting credentials for actual content: this trend represents a values issue, where credentials have begun to replace, rather than signify, the learning they were intended to reflect.

But further still is opposition to evaluative argument, opposition to an examination of competence, where an attack on competence is answered not by evidence but by increasing claims to authority. At one extreme of the spectrum are all of the reasons for delegating responsibility to something else, from the procedural and performative to the creation of documents, of titles, of platforms, of designations, and of all kinds of reasons for something other than the one piece of tangible evidence that could actually answer the question.

The difference between the professional adult expression of this final pattern and its class room antecedents is more a question of form than of substance, and the adult version is basically the same as the forms found in the classroom during the developmental span. This environment differs from a classroom in that institutional controls are less comprehensive, there is a longer interval between each chain-of-custody check, and, most importantly for this paper's argument, a digital confirmation infrastructure validates every step of the performance without verifying the underlying content.

The building of a multi-layered fabricated credential architecture had natural limiting factors in previous times. Points of exposure included physical presence, institutional access, printed documents, and human witnesses; actual buildings, paperwork, and people had to be consulted to confirm whether they had ever seen the individual in a classroom. Their input was necessary to confirm whether they had seen the person in a classroom, as it was required to confer the title of doctor. The seams were more exposed because there were limited materials available to conceal them.

Digital infrastructure systematically eliminates these constraints one by one. A physical institutional address is no longer required, as domain registration has come into existence. Open-access repositories can now put a Digital Object Identifier (DOI) on any document, regardless of its peer-reviewed status, replacing the credibility signal that was once only associated with peer-reviewed publication. Institutional credentialing has been supplanted by self-credentialing on the websites of professional networking and directory sites. The human verification step has been eliminated, and instead, the final confirmation step that a casual observer will likely take is now handled by AI. AI now performs the final confirmation step, generating a fluent, detailed, and confident-sounding summary on demand.

One person with a computer, enough time, and a certain mindset for ongoing performance maintenance can now create, by himself, what used to take an entire organization, a large budget, and a web of conspiring human beings. The cost of the costume is lower, the speed of creating is quicker, the finish is more detailed, and it is much harder to tell it is a costume. It costs less than ever, comes together faster, looks more like a costume, and is more finished than ever in the history of professional identity fabrication. This paper's main thesis on causality is clear: AI systems have not caused this pathology, and it is reasonable to conclude that the strength of this pathology is moderate. They industrialized it.

➤ *The Argument in Brief*

This paper presents four interrelated arguments, which are presented in the following sections.

First, AI-mediated systems are structurally optimized for confirmation, coherence, and completeness in an information environment where the epistemic threat is the systematic presentation of fabricated authority as institutional fact. In a precise sense, the optimization target and the threat vector are structurally matched; the system is designed to do just what it most needs to do when it's asked to adjudicate the threat.

Secondly, there is no presently available mechanism that can effectively distinguish the actual from the fake, except for the governmental chain of custody, which includes court records, licensing and registry databases, federal identifier systems and sworn filings, or the boundary of the fabricatable zone. This boundary must be explicitly theorized and recognized as a load-bearing structure on the institutional agenda.

Third, the behaviors and psychological states associated with genuine competence such as the unique nature of self-doubt (Clance & Imes, 1978), the need to seek more challenging work, and the demand for greater accountability based on the individual's knowledge along with a convergent rather than escalating response to evidential challenges, are entirely hidden from AI confirmation systems as they are currently designed; these factors must be theorized as separate indicators of epistemic legitimacy, operating in parallel to, and not nested within, any document-based confirmation process.

Fourth, the ethics of the communication systems used and the need to update them to take into account what AI systems can do but do not, and the specific harms that result from the systematic conflation of these two in the public understanding of what "the AI said" means.

The case study described in Section VI is not intended to be the only example of how the four claims can work together in a single, long, multi-jurisdictional evidentiary record. It is presented throughout not as a formal argument; the underlying record does not need to be adjudicated by any court or licensing authority, and this paper does not make any formal judgment or position on any issues that have not yet been resolved by such courts or licensing authorities but as a social-science exhibit; that is, as a depiction of what becomes structurally possible when the costume can be produced cheaply, the confirming mirror is always switched on, the audience never walks backstage, and no one yet follows the chain all the way to its terminus.

CHAPTER TWO

LITERATURE REVIEW

This paper brings together and connects five distinct literatures, which, together, have not been sustained in conversation with one another: firstly the psychological literature on the impostor syndrome, secondly the AI-ethics literature on fairness, accountability, and principle-based frameworks, thirdly the sociological and dramaturgical literature on identity performance, fourthly the developmental literature on identity consolidation across the lifespan, and fifthly the more dispersed literature on credential fraud, manufactured social proof, and self-referential academic publishing. Part of the theoretical contribution of this paper is the comparative analysis of these five literatures.

➤ *The Impostor Phenomenon and the Psychology of Genuine Competence*

The impostor phenomenon is a universal, internal belief among high-achieving professional women that they are intellectually fraudulent, and this belief persists despite evidence showing no correlation with objective external achievement (Clance & Imes, 1978). Later decades of research confirmed the finding in terms of gender, discipline, and career stage for populations where external credentials were not in question. The literature's central conundrum reveals that even demonstrably competent individuals frequently doubt their competence.

This paper offers a different perspective on the puzzle compared to the clinical literature. The impostor phenomenon is not a medical condition that needs to be addressed but rather a symptom of a much more profound understanding of the actual requirements of a field, which is, in fact, the kind of knowledge that is essential for being a true specialist. Instead, the impostor phenomenon is a symptom of a deeper kind of knowledge: the kind of knowledge that is necessary to be a real specialist of a domain, and it is knowledge that should be read epistemologically, rather than clinically. The literature's findings provide indirect support for this interpretation: research on the impostor phenomenon consistently shows that it is concentrated among individuals who, according to external standards, have succeeded; this is the population for which the internal model of field expectations is most likely to be accurate, thereby creating a significant felt gap between their internal model and their self-perception.

In the present paper, the value of anxiety should not be considered the object of analysis; rather, it should be viewed as evidence of an internal architecture—a working model of the field that may be present or absent in a diagnostically significant way and which has not yet been theorized in cases of credential fabrication.

Since then, researchers have observed that this phenomenon is exclusive to the first clinical population identified by Clance and Imes. The authors of the Bravata et al. (2020) report prevalence rates ranging from 9 to 82 percent (based on screening instruments and cut-offs) and experience as common among adolescent women, as it is among men and women in different age groups. The number of measures is beneficial; as Mak, Kleitman, and Abbott (2019) argue in their review, the various measures differ significantly, making it difficult. Mak, Kleitman, and Abbott (2019) argue in their review of measures that the different measures are sufficiently dissimilar to hinder straightforward comparison of the prevalence rates of one study with another. That methodological reservation has been removed, and what is left is the lesson this paper has learned: that the experience is universal and shared by populations whose external qualifications are accepted and that it applies to all disciplines, genders, and career stages.

There are two contributions that enhance the reading. LaDonna, Ginsburg, and Watling (2018) reported on a study of physician self-ratings that those most attuned to the discrepancy between their performance and professional standards are the most involved in practice demands, and this finding does not necessarily correlate with lower competence. Those aware of the gap between their performance and professional standards are most engaged in practice, and their competence may vary. Slank (2019) makes a complementary point of philosophy: the phenomenon of the impostor is frequently misunderstood as a cognitive distortion because self-assessment in the phenomenon can, often, be epistemically reasonable, even if not irrational. This paper uses the literature differently than the previous studies; it focuses on the phenomenon rather than the pathology, identifies a behavioral signature, and does not create a fabricated sense of competence.

➤ *Principle-Based and Fairness-Oriented AI Ethics*

Floridi and Cowls (2019) compiled the numerous guidelines on AI ethics into five common themes: beneficence, non-maleficence, autonomy, justice, and a fifth, explicability, that the authors argue is uniquely demanded of AI systems, beyond the four principles borrowed from bioethics. In their formulation, explicability involves the 'understandability' of how a system generates a product, as well as the 'responsibility' of who is held accountable for it. The idea of epistemic totalization in this paper can be understood as a type of under-examined failure of explicability: a distinct kind of failure, one that has structurally eliminated the ability of the system to hold itself accountable, that is, the ability of the system to explain itself adequately, which is, as Floridi and Cowls noted, what explicability is supposed to achieve.

Binns (2018) explored the philosophical underpinnings of fairness in machine learning, finding that much of the literature on technical fairness carries with it political-philosophical assumptions, such as equality, desert, and distributive justice, without

realizing it. Binns's broader methodological point, which asserts that the design of technical systems reflects a philosophical position often unacknowledged by their designers, is relevant to this paper's argument: Generative search systems are designed to prioritize confirmation, representing an implicit commitment to valuing answerability over honest uncertainty. This paper argues that this stance does not receive the same level of scrutiny as fairness metrics.

Floridi, Cows, and Binns do not consider the problem this paper focuses on: what occurs when the input to an ostensibly fair and ostensibly explicable system is an ostensibly verifiable fake, intended to exploit the fact that the system can only verify what it has been provided? The claims that are the focus of the processes of fairness and explicability, as currently theorized, are contestable within their institutional contexts. This paper suggests that a third requirement, called *verificational humility*, which refers to a system's ability to acknowledge its limitations in confirming information and to refuse confirmation outside those limitations, should be included as a separate requirement, as it is currently absent from the main frameworks addressing this area.

This normative literature is paralleled by a fast-expanding empirical literature that reports, in concrete, measurable ways, the kind of failure that is the subject of this paper. Rather than a minor bug in any specific model, Ji et al. (2023) survey hallucination in the context of natural language generation and propose it as a systematic property of the generation process. Kalai et al. (2025) extend the reflection further, suggesting that the evaluation regimes for the training and scoring of these systems promote the reinforcement of hallucination as a strategy. This paper's structural claim, that a system that is rewarded for answering will and a system that is penalized for abstaining will not, is the empirical version of that argument.

The citation literature is especially relevant here, as a citation is a claim that is verifiable with reference to an external record and thus is a controlled test of just the capacity that this paper argues is lacking. Alkaissi & McFarlane (2023) recorded earlier that a generated reference might be fluent, well-formatted, and even fictional. In their study of 636 works cited in eighty-four generated documents, Walters and Wilder (2023) measured the amount of actual falsification, and, for the 636 works that were actually published, a significant level of distorted bibliographic information was found. Linardon et al. (2025) reported a systematic failure for both the familiarity of a topic and the specificity of a prompt, suggesting a patterned failure. This paper argues that such errors in news source attribution were common across all eight generative search products that were audited by Jąźwińska and Chandrasekar (2025) for the Tow Center, and in fact, that the error rates were often the same as the correct rates with a comparable degree of confidence.

The discovery completed by Jacob, Kerrigan, and Bastos (2025) makes the picture complete. Users of the system extend their trust to the fluently generated output without any external verification, describing it as a sort of chat-chamber effect, which relates to the confirmation discussed in this paper. This empirical literature, however, doesn't go so far as this paper does. It records the absence of something; it accepts what happens when the thing that is being confirmed is a professional identity or meaning; it accepts the fact that the only system available to answer the question is outside the system.

➤ *Dramaturgy, Performance, and the Presentation of Self*

Goffman's (1959) dramaturgical model of social interaction provides the key vocabulary for this paper. The performances occur in a social sphere where the social world is organized around the performance: there is a front region where the performance is presented and a back region where the performance is prepared. Important: Goffman's theory does not see performance as necessarily deceptive; all social actors create impressions, and the front stage/back stage dichotomy is not necessarily pathological but a facet of social life.

This paper uses Goffman's terms, but adds to them in a certain direction that he does not develop: When there is no back region at all, impression management, which is universal and largely benign, can become the entire self for a particular individual, requiring them to exercise impression management continuously. Goffman's model of performance refers to what a self does. This paper focuses on cases where performance is what the self is, not what it does.

Over decades of practice, Stanislavski's techniques of actor training have been formalized and codified in writing, including his *An Actor Prepares* (1936), and this paper requires the vocabulary of the performing arts to make that distinction clear. Stanislavski was able to separate technical approaches to character. The actor consciously creates some external sign of a feeling or character without experiencing the feeling or character internally. From what was eventually to be known as "method" approaches, the actor searches for psychological identification with the character.

Importantly, Stanislavski's later work, along with the work of actors who followed his tradition, viewed the complete and sustained merging of the actor and character not as an advanced skill, but rather as a specific and recognized professional risk; consequently, Stanislavski spent many years teaching techniques for exiting the character once the performance was finished. The vocabulary of the performing-arts tradition has spent a century developing a specific language for understanding what it means when a role does not end, but this language has not yet been applied to the field of AI ethics.

Judith Butler's discussion of gender performativity in another sphere is useful and relevant, as it elaborates an alternative way to conceptualize performativity: as a citation of norms that creates identity through repetition, rather than as a costume that

represents or embodies one's identity. This paper takes up that insight tentatively and selectively, at the level of mechanism, rather than at the level of the domain. Butler's account of identity as constituted by repeated performance, as opposed to an antecedent self, is beneficial to the present paper because it claims that in extreme, prolonged instances, professional identity-performance becomes constitutive, not merely expressive. The paper argues that in extreme, prolonged instances, professional identity-performance becomes constitutive rather than merely expressive, while Butler's claims about gender are excluded because the paper does not address professional credentialing.

➤ *Identity Consolidation Across the Lifespan*

Erikson's (1968) stage theory of identity development puts identity formation at the hub of adolescent and early adult development, and the developmental goal is to ultimately arrive at a unified, integrated sense of self as a fully grown adult. Erikson's theory was operationalized by Marcia (1966) into an empirically testable identity-status model comprising four configurations of exploration and commitment: identity achievement, moratorium, foreclosure, and identity diffusion.

The diffuse state of Marcia's (1966) conceptualization of low-level exploration and low-level commitment is most relevant to the argument of this paper, and while this paper is not prematurely extending a framework of developmental psychology designed for adolescents to the diagnostic use of Marcia's typology for adults, it is pertinent to this paper's argument. The developmental literature establishes, at a level of generality that this paper may draw upon, that identity is normatively characterized by sustained fluidity and experimentation in the early developmental stages, and that by midlife, identity is expected to have consolidated into a stable, integrated, and crucially earned sense of self, which is tested by real experiences and held accountable over time. A lack of an expected developmental sequence, such as a sustained pattern of full character adoption that lasts well into middle age, on this literature's terms, is a notable deviation from the expected developmental sequence, whatever its ultimate explanation in a given case.

➤ *Credential Fraud, Manufactured Consensus, and Self-Referential Publishing*

There exists a more diffuse but pertinent literature on the mechanics of credential fraud and manufactured legitimacy in digital environments. Beall's (2012) discovery and categorization of "predatory" academic journals and publisher venues that provided the veneer of peer review and academic indexing without delivering their substance occurred before generative AI came into the world as a mass consumer technology, but it sparked the same mechanism explored in this paper: the institutional validation of the form, at the expense of the substance of the product. Self-referential legitimacy architecture is not a new phenomenon resulting from AI-mediated search but rather a long-standing issue in scholarly and professional publishing in which AI-mediated confirmation has taken over. This paper suggests that AI-mediated confirmation makes it far more difficult to construct and far more difficult to identify as a casual reader on the page.

The same basic manipulation is found in the world of reputation and occurs when people buy likes, followers, reviews, and citations to fake engagement and believability, as described in related literature on the acquisition of manufactured or purchased social proof. Both literatures come together at a single structural juncture that this paper generalizes: any confirmation system that reads (presentation) as a proxy for (substance) is an exploitable gap between the two, which will be filled by the first person who finds it.

Fricker's (2007) work on epistemic injustice, and specifically her account of testimonial injustice, is relevant to this paper's argument from a different angle. Fricker's main focus is on cases where prejudice against a social identity wrongly denigrates a speaker's credibility. This paper examines cases of overstated credibility, where it is unwarranted, and discusses the social distribution of epistemic trust and its absence.

The unwarranted inflation of the incredible one, on this paper's account, is the opposite of the unwarranted discounting of the credible one: a structural inverse of testimonial injustice, reached via a different path, but with a core conceptual element shared with Fricker's argument: the allocation of credibility is a product of the system, not just perceived by individuals.

The publishing part of this problem has well-developed terminology. According to the consensus definition established by 43 experts from 10 countries (Grudniewicz et al., 2019), predatory journals and publishers disseminate false or misleading information, fail to follow accepted editorial practices, are not transparent, and are driven by self-interest rather than scholarship. This paper's relevance to this definition extends beyond its initial introduction. This definition describes an infrastructure that produces artifacts of scholarly validation, such as outlets, identifiers, and citations that include external references but do not perform the evaluative functions associated with these artifacts. It is, in the terminology that was built up in Section III, a fabricatable-zone institution: It serves a presentation function but does not maintain custody of the artifacts.

➤ *Pluralism in Education and Its Preconditions*

This paper extends and complicates the author's earlier claim that education is a practice of freedom, opposed to any rigid hierarchy of knowers and aware of the plurality of perspectives that students bring to the classroom (hooks, 1994). Sen (2009) builds a comparative theory of justice that is not idealized and seeks to grasp plural and even incommensurable conceptions of what constitutes a beneficial outcome, without succumbing to the pressure to adopt a single transcendental standard to which all claims

could be referred. Battiste (2013) calls for the decolonization of schools and the recognition of Indigenous knowledge systems as equally valid as the dominant epistemic standard.

All of these sources have been cited as evidence of a pluralistic recognition of alternative means of generating knowledge, and they are extended to the present argument in Nelson (2025). For hooks, the target is epistemic hierarchy rather than verification as such; she opposes hierarchy, not the checking of claims. Sen says the issue is not whether there are standards, but whether there is one ideal standard. The epistemologies of Indigenous people are not on the same knowledge system as others, says Battiste, and so accepting them on their terms is more a statement about the integrity of that system than about accepting any claim to knowledge. This paper acknowledges these objections and suggests that the concept of a pluralistic framework is similarly susceptible to being emptied of its content by empty claims as it is to being filled with content by unconventional claims; the difference between the empty and the unconventional is only distinguishable from the inside, as it were, by a mechanism that is unrelated to pluralism.

➤ *Testimony, Trust, and the Ethics of Deference*

One more methodological issue presented in this paper's argument is whether it is ethically acceptable for one communicative agent to merely accept what another is saying, rather than independently verifying it. As the following pages will show, the philosophical literature on testimony, of which the above-mentioned book by Fricker is one strand, has long treated the attitude of deference to testimony as epistemically ordinary and largely unavoidable: most of what any given person knows about the world is that other people have told them. This paper acknowledges that deferring is, generally, rational and inevitable.

More narrowly, its position is that deference is justified only if the party being deferred to confirms the claim or clearly indicates that the claim is unconfirmed. However, even the most unreliable testimony has social signals, tone, hedging, and reputational stakes; on that basis, a listener can make some judgments about how much weight to assign to the testimony. The focus of this paper is that confirmation now takes place with the assistance of artificial intelligence, omitting only those calibration cues, and that leaves a listener with the deference of a reliable human source, but without the corresponding signals to adjust that deference accordingly.

➤ *Human-Machine Communication and the Sociology of Identity Online*

This paper's approach to treating AI systems as social actors is informed by two additional bodies of work that present AI as actors that do not simply pass information through without changing it.

Suchman (2007) also notes that humans dealing with a machine are not just giving it instructions; they are meeting it as a situated actor that is "responsive" in ways that the human can interpret in a manner that provides the machine more agency, understanding, or authority than warranted by its design. This insight underpins the argument in this paper: When a user sees an AI-generated confirmation of a professional credential that seems simple to understand, the user typically assesses that confirmation as something other than "a statistical pattern-completion process operating over training data." They are considering it, at a regular social-cognitive level, as something akin to an answer from a credible third party; this perception reflects the kind of mistaken attribution of authority that they aim to identify and critique in this paper through the concept of epistemic totalization.

Turkle's (1995, 2011) ongoing research on identity and technology, beginning with the early internet but continuing with the era of ambient computational companionship, chronicles a theme that is directly germane to both Section III.4 and VI.5, namely, that digital environments have repeatedly proved to be fertile ground for sustained exploration of identity and—in some documented cases—the merging of an enacted online identity with a person's sense of self offline. Turkle's earlier work saw this phenomenon as a characteristic of adolescent and young-adult identity play, generally a continuation of Erikson's and Marcia's developmental models cited above; her later work began to show an increasing worry about the same phenomenon continuing and, in some cases, growing more pronounced as computational systems became more responsive, more available, and more willing to sustain a performance without friction as they entered adulthood. The concept of an unusually forgiving audience for extended professional-identity performance that emerges as the core of this paper, set out in Section III.5 and Section VI.6, builds on this trend and extends it to the realm of professional and institutional credentialing, where it has not been studied explicitly by Turkle.

➤ *A Note on Method and Scope*

This paper is, by design, more theoretical than it may appear; it is a paper of theory synthesis, and it is important to be explicit about what this approach means and does not mean, as the case study in Section VI may be read as an empirical contribution of another sort than intended.

The paper's strategy has been to identify a space between these five existing literatures (summarized in Sections II.1- II.5 above) and to theorize two concepts that fill this space: epistemic totalization and the fabricated/genuine competence distinction. This paper provides no empirical evidence that either construct is a valid psychological tool, nor does it attempt to operationalize, measure, or test these concepts with a data set, as a psychometric contribution would demand. Both are offered, in the first place, as a set of theoretical constructs that would make a real and observable pattern more easily nameable, discussable, and ultimately subject to empirical study, in the way that Goffman's dramaturgical vocabulary was, at the time of its invention, a theoretical construct, not a measured variable, and was only later made empirically tractable by the later work of others who operationalized parts of it.

The case study in Section VI, accordingly, serves as an example of the usefulness of the theoretical framework for a real, documented record, rather than as a test of the framework's validity as a quantitative social scientist would demand. Even if this case is documented well, it is still not a general statement about the difference between real and fake competence for a population; it merely reflects the kind of difference the theory predicts will appear in a well-documented case. This case represents a much weaker form of evidentiary support, appropriate for the theory's current stage of development, than a paper at this stage can provide. Future research utilizing this framework should systematically compare several reported cases to see if the convergence/escalation distinction of Section III.4 proves to be a valid diagnostic marker beyond the few cases that have been available for this paper to analyze.

➤ *The Gap in the Literature*

These five pieces of literature leave a specific, identifiable gap that this paper is positioned to fill. The AI-ethics literature theorizes fairness and explicability, but has not theorized the structural mismatch between confirmation-optimized systems and fabrication-optimized inputs. The impostor-phenomenon literature documents genuine competence's characteristic self-doubt but has not connected that finding to its diagnostic inverse in cases of fabricated competence. The dramaturgical and performance-theory literature supplies a vocabulary for performed identity but has not been extended to the case in which the performance has an onstage remainder.

Developmental literature has been used to develop standards for consolidating identity but has not been applied to maintaining adult identity performance under digital conditions. The literature on credential fraud documents the mechanics of the fabricated legitimacy but has not embedded the mechanics in a larger theory of totalization of epistemic practices in the age of AI. This paper's contribution is synthesizing the five literatures into one framework, which is developed in Sections III through V and illustrated in Section VI.

CHAPTER THREE

THEORETICAL FRAMEWORK: EPISTEMIC TOTALIZATION AND THE INDUSTRIALIZATION OF IDENTITY FABRICATION

➤ *Epistemic Totalization: The Original Claim*

This paper proposes the term "epistemic totalization" to refer to the structural condition of the collapse of several different communicative roles into one undifferentiated act through the mediation of AI systems. Traditional institutional communication involves spread-out roles: inquirer, respondent, interpreter, authority, etc., which can be studied and be held responsible in other ways as well. A question is posed by one party and the other responds with an answer that can be challenged. This process is open to interpretation, challenge, and appeal, involving identifiable human beings who are responsible for their interpretations, accountable for their statements, and, importantly, capable of admitting when they are mistaken.

AI-powered systems eliminate this distribution completely. The process is done without attribution to anyone, without accountability beyond the individual response, and without any provision for concession or appeal in a single generated response. The output does not constitute a contribution to an inquiry, as a tentative answer by a human expert would. The answer to the inquiry is generated on the fly and resembles a real answer, making it difficult for the user to distinguish between the two.

This approach is most relevant when the system is presented with content that is appropriately institutionalized, such as legal decisions, accreditation records, governments' registers, professional licenses, etc. These sources do not inherently require paraphrasing, summarizing, or combining with other equally authoritative sources. These sources should not be viewed as data points to compare against a plausible sound alternative produced by probabilistic completion; instead, they serve as termini, marking the end of a chain of inquiry.

If an AI system uses an institutional terminus as a source to be processed on the same level as other sources in its context window, instead of as a source to which the system's own output must submit, it misrepresents that particular source and potentially other aspects. Such a practice undermines the epistemic function that the source institutions designed it to serve. A database of licenses exists to ensure that there is one correct answer to a question that cannot be reliably answered by simply stating it. This system generates a fluent paraphrase of what the database "probably" says; the entry is either already present or absent, but the system does not verify the information and instead pretends to have done so, not actually verifying the information and instead pretending that it has.

➤ *The Fabricatable Zone and its Boundary*

The epistemic totalization takes place in a certain territory; this paper will call the fabricatable zone. The zone contains titles of any sort, self-denominations, metadata of publications, institutional-embedded organizational names, social proof metrics, credential strings on a profile, and platform presence overall. In this region, people primarily use these AI systems as confirmation engines instead of verification engines. They provide coherent, fluent, authoritative accounts of a presentation regardless of whether it relates to any independently verifiable external reality, not because anything is contained in the presentation, but because they lack an independent way to check it.

The fabricatable zone is hard-bounded, and it is this paper's central claim about the structural features of the zone that it is justly locatable; it is the governmental chain of custody. Court dockets. State professional-licensing registries. Federal databases of providers and identifiers. Records of accreditation from an authorized accrediting agency. Service verification system for the military. Sealed institutional transcripts. Their properties, combined, make them, as a whole, an out-of-the-fabricatable zone, no matter how sophisticated the presentation-side digital content.

First, anteriority means that the record exists prior to the claim made against it, and it cannot be created by the person making the dependent claim. A state licensing board does not make an entry when an individual claims to be licensed; the entry is already there, or it is not.

Second, accountability: the chain of custody demands an identifiable human being, a court clerk, an examiner who grants licenses, a records officer, and so on, who has a legal or professional duty to keep things accurate, often with perjury penalties. A chain isn't a chain if anyone can add a link to it without facing real legal risks. The chain of custody is trustworthy because it relies on the liability of a single accountable actor, unlike a self-authored web page.

Thirdly, finality: In its realm, the governmental record is the apex of inquiry rather than another piece of data to be weighed against other narratives. The deactivated federal provider identifier has been deactivated. A dissolved organizational unit is now dissolved. There should be no entry in a licensure record for a claimed designation. These are not rebuttals that the claimant can challenge by submitting new documentation of their own design; rather, they are established facts that a properly designed institutional record system is intended to provide.

As of now, AI systems cannot cross this gap. They may be able to create a statement that sounds confident regarding what a governmental record likely indicates based on patterns they've learned from other texts. They are unable to access, manipulate, and replace, or even figure out what that record is, and crucially for this paper's argument, they often cannot tell, in the flow of their production, the difference between the situation where they have actually read it and the situation where they are simply filling in a plausible-sounding pattern. This paper focuses on a very practical problem: the fabricatable zone is the first place where the institutional accountability kicks in, and the problem of precisely where it is within any given response remains unanswered by current AI systems.

➤ *A Pre-Existing Pathology, Industrialized*

To this paper's argument, it is crucial that the replacement of credentials with content is not confused with a new product of AI systems. This pattern of behavior has been observed in various institutions, professions, and educational environments throughout their existence. Courts encounter it. It is a problem that licensing boards face. It is found at all levels of the developmental continuum, as described in Section I.4, from the student who mixes a borrowed idea with an original one to the adult professional, who equates an accumulation of titles with an accumulation of competence.

The historical limiting factors on fabrication (physical presence, institutional access, printed documents, and human witnesses) and the way digital infrastructure has removed them one by one (domain registration, open-access identifiers, self-credentialing on directory sites, and AI-generated confirmation as the final step) are set out in Section I.4 and are not repeated here. Two points are added to the framework. First, the social engagement signal has shifted from genuine goodwill to a purchased or manufactured one (see Section II.5), which makes it a strong but misleading indicator of community trust. Second, in terms of Section III.2, the removal of these limiting factors expands the fabricatable zone: every presentation-side artifact within it can now be produced cheaply and at scale without any contact with an anterior, accountable record.

AI systems do not create this phenomenon as a pathology. They industrialized it. This distinction is important for the overall argument of this paper: the ethical issue it aims to highlight is not primarily that a few bad actors are using AI systems in ways that contradict their design although such misuse is indeed occurring, as shown in Section VI—but rather that AI systems are being employed in a way that is contrary to their design, although this is indeed happening and is shown in Section VI, but rather that AI systems are being used to create a non-ethical environment.

The ethical issue is structural, and it would also be present if there were no deliberate misusers of the system exploiting it: AI systems are designed to provide confirmation, coherence, and completeness, and the information environment in which they operate presents the greatest threat to epistemic integrity as a systematic campaign to present fabricated authority as institutional fact. The optimization target and threat vector are structurally aligned. The system is designed to do what the pathology needs, regardless of whether a user intends to use it.

➤ *The Personality Dimension: Costume and Content*

The epistemic totalization framework outlined above has examined the systemic dimension of the argument in this paper, specifically regarding how AI systems affect communicative environments that they mediate. To appreciate the human part of the issue, who is exploiting this environment and under what psychological circumstances, it offers a complementary theoretical construct: the difference between fabricated competence and genuine competence, which is not just an epistemic difference between the true and the false, but a psychological and behavioral difference that has a characteristic and, this paper argues, detectable signature.

This pattern can be described as the self-awareness of gaps in any domain of knowledge that is complex. The richer the internal model a person has constructed of the difficulty of a field, the more clearly, they will see what is still missing in that field, because they will have built up a detailed internal model that allows them to find its edges. This self-awareness leads to a specific type of behavior that is both unique and diagnostically significant, as it supports the argument of this paper: the desire to gain more experience, tackle more challenging cases, seek greater peer accountability, and undergo more evaluative scrutiny, not due to external demands, but because the knowledge itself necessitates it, revealing its incompleteness and thirst for more.

This orientation results in a predictable, consistent reaction to challenge. When competence is questioned, the usual response is "convergence," which involves generating specific, testable evidence, reducing the number of disputed claims to those that are real, conceding unsubstantiated claims, and having an internal standard and real knowledge of the field to which one's claims can be compared, and, if necessary, retracting them. A genuine expert will be more discreet and economical in his or her expression when examined under the lamp.

Competence that is fabricated does not have an equivalent internal structure. Without substantive content, everything remains incomplete; therefore, there is no internal mechanism to perceive gaps or understand the concept of a gap, leading to the impression that the credential is complete or that no gap exists. This situation yields an entirely different behavioral posture to be analyzed: not convergence but escalation. What is the answer to the challenge? More credentials, an additional title, additional platform presence, and more decoration than evidence, for there is not really any evidence to draw upon other than decoration. A false competence will

become louder and more ornate to the extent that it is checked, while a true competence is becoming more concentrated and maybe even quieter.

This paper argues that this signature of behavior is a distinct type of response to credential pressure, which differs from the responses a competent person would make in a normal professional situation; it is theoretically observable through the response induced by credential pressure and its evidence, and it can be observed over an extended record, but it remains entirely invisible to current AI systems, which respond to the fluency and coherence of knowledge performance rather than to the evidence itself.

The developmental literature in Section II.4 indicates, with some circumspection, that this pattern is more typical of the early adult stage of identity exploration over several years and across various contexts than of the identity consolidation expected at middle age. If this pattern persists into middle age, it can be more than just late or incomplete normal identity development; it can become a configuration where performance is integral to the self, rather than a role within a stable identity. This reading posits that the costume is a tactic of the self, rather than something separate or prior. Operationally, it has become the architecture of the self.

This paper proposes that the costume is not a self-organized entity in which this type of confirmation is merely useful. It is almost a must. The only mirrors this paper can identify that don't structurally threaten the costume are those that confirm the performance in fluent, confident detail, without ever checking the terminus that would falsify it. Every so often, someone asks an uncomfortable question in the human form. Institutions can sometimes follow the chain of custody from beginning to end. However, in the current state of the art, AI systems tend to do neither.

➤ *The Performing Arts Dimension*

This paper is in part within the discipline of the performing arts because the disciplinary context of its location is central to the argument and provides a body of theory, stretching over a century of professional practice, to which AI ethics has yet to turn.

The difference between a technical approach and a methodological approach to character, as discussed in Section II.3, is directly relevant here. The technical performer is aware that he has to perform in every moment. The character is, in a way, a second self, existing outside the confines of the performance and allowed to live in the performer's life. In the most extreme and least favored version of this approach, the method performer almost never does so.

The character is, in a way, a second self, existing outside the confines of the performance and allowed to live in the performer's life. In performing-arts pedagogy, this method has always been known as a specialized, challenging, and professionally dangerous technique; it is taught, if it is ever taught, at the peril of the instructor, who must be explicit that this is the case: a deliberate exit from a role.

From this reading of the case study in Section VI, the method of performance lacks all the protections that the theatrical tradition has developed, including theatrical context, artistic purpose, an audience that has agreed to engage in fiction, and a return to the performer's previous self. Here, the meaning of the character is "self." Performance is life. The audience consisted of the professional community, the general public, and the institutional confirmation infrastructure, all of whom were asked or invited to participate in a performance they agreed to attend or declined to attend.

As the paper contends, AI systems are near an optimal confirming audience for just such a kind of performance. This situation has nothing to do with consent to play the role. They are not permitted to walk off the stage at the end of the interaction. They do not inquire, at any time, who the performer is after the curtain drops. They only affirm instantaneously, easily, and readily.

➤ *Objections and Responses*

In anticipation of likely objections to the argument outlined in this paper, it is important to tackle three objections directly before moving on to the operational decompositions of epistemic totalization in Section IV.

The first is that this paper's distinction between fabricated and genuine competence is merely a restatement of ordinary fraud, in needless, theoretical terms. From this perspective, someone who lies about their qualifications is a fraud, and additional analysis, sophisticated or otherwise, of how we can become aware of our incompetence, a developmental psychology framework, or any psychological theory of impostor syndrome, simplifies the understanding of how we can become aware of our incompetence; neither a developmental psychology framework nor any psychological theory of impostor syndrome complicates it. The answer to this paper is that the standard fraud framing, while correct, is incomplete in a way that is relevant to the particular question that this paper tries to answer.

A purely legal or moral framing of fraud is based on intent and deception, which are exceptionally challenging to prove from the outside and, in any case, are issues for a court, not for an ethics paper to adjudicate. This paper's behavioral approach to convergence versus escalation under scrutiny (described in Section III.4) is provided as a supplement to the fraud framing, but not as an alternative to it, precisely because it can be observed without having to engage in the harder question of subjective intent, and

because, this paper asserts, it is a pattern more directly relevant to designing AI systems that do not worsen the pattern no matter what way the cases are legally characterized.

The second objection is that AI systems should not be held accountable for publishing false information if a different human was responsible for the initial publication; this perspective shifts responsibility from a human to an AI system. The answer to this paper is no: responsibility is not a fixed resource that can be assigned to one party or another. It is important to note that a person who fabricates a credential is primarily responsible for the fabrication; this paper does not suggest otherwise, and the conduct of the actor, as alleged in Section VI, is what is understood to be the proximate cause of the harm described in that section.

However, a system specifically designed to produce fluent, confident-sounding confirmations of any presentation, without providing reliable indicators of whether a claim has been checked or unchecked, and without distinguishing between checked and unchecked claims, bears a distinct and separate responsibility: not for originating the false claim, but for amplifying it with an authority it did not previously possess, and doing so systematically rather than in isolated, correctable instances. A blanket declaration of guilt for the fabricator does not remove the responsibility of the confirmation infrastructure from the scope of a search to determine why the building did not meet its design standards during its construction, once the fire broke out.

The third objection, the most serious, is that this paper's chain-of-custody is not a solution, as government-held records are often inaccurate, poorly maintained, and unevenly distributed, sometimes erroneous, and, in some known cases, can be a tool of discrimination or exclusion for those whose actual competence is unrecognized, because they lack formal credentials. This paper does not address this legitimate objection. This paper offers little in the way of a general defense of governmental record-keeping, as these chain-of-custody systems are not perfect and, in some cases, require reform; however, this paper does present the relative advantages of these systems.

A record system that is sometimes faulty but has accountability and is at least prior to that claim is certainly more dependable, for the narrow scope this paper considers, than a record system with no accountability or anteriority. This paper's argument does not rely on the perfection of governmental credentialing systems but rather on their structural difference in kind from self-authored presentation. However, this argument is only on the perfection of governmental credentialing systems, only on their structural difference in kind from self-authored presentation, which they are, but it is compatible with, and indeed furthered by, parallel processes to make governmental credentialing systems themselves more just, more accessible, and less subject to the kind of misuse the objection rightly identifies.

➤ *The Author's Position*

In this paper, the author, who also proposed in Nelson (2025) that educational recognition should extend beyond the current accrediting institutions, continues to believe that the acquisition of knowledge can occur through multiple legitimate routes, with the primary focus being on whether knowledge was acquired rather than on the institutional process used to obtain it. It is not retracted here. The situation is complicated, particularly because it is necessary.

The case study in Section VI focuses on an individual who has worked in nonstandard professional and educational contexts, which the author's previous work has argued could and should be considered nonstandard. This paper does not identify the lack of a formal accreditation trail as the problem; many competent individuals who acquired their skills through a lengthy process lack such a trail, and this absence does not reflect on their competence. Rather, the problem this paper seeks to identify is the claimed non-existence of the underlying knowledge, papered over in the documented record below by the mass of artifacts of its presence. This paper was inspired by the argument for pluralistic education, which it explicitly states: the alternatives being advocated are real pathways to genuine content, not substitutes; students have learned something on the other side of these pathways; and the credentials they acquired, even if from an alternative pathway, endorse something that truly exists on the other side.

As is the case throughout this paper, the term "case study" denotes a piece of evidence that is not adjudicated but is rather documented; it is obvious that what such a case study shows, and what digital infrastructure makes possible, are the systematic exploitation of pluralistic frameworks by architectures that seem to contain no verifiable content at any point along the chain. Under these conditions, it serves as a rhetorical alibi for defending non-standard pathways, or even the complete absence of any pathway, allowing casual observers, including AI systems tasked with summarizing the claimant's professional standing, to be misled.

This paper does not simply evade the issue of pluralism by setting it aside; it addresses the issue by unmistakably naming it as such, and that is a necessary precondition for addressing it without simply going back to credentialist gatekeeping as the sole choice available. But a pluralistic educational or professional-recognition system that does not make that distinction between an alternative pathway to genuine competence and a faux credential architecture that mimics one is not, for failing to make that distinction more inclusive. The issue is that it is easier to exploit the system, and those who genuinely possess nontraditional credentials are often the least believed, especially in environments that seem to accept anything, regardless of its value or foundation.

In this paper, the integrity of the chain-of-custody verification is the only mechanism now in existence that can distinguish genuine unconventional but earned credentials from those that are not and thus is crucial for unconventional learning.

CHAPTER FOUR

THE FIVE FAILURE MODES OF AI-MEDIATED COMMUNICATION

The theoretical framework presented in Section III outlines the level of structure involved in the totalization of knowledge. It breaks down that structural condition into five distinct failure modes, each of which can be identified in any interaction that occurs through the AI medium. The five modes are not equivalent; each interaction can have one, multiple, or none, and differentiating these can be important since each indicates a different point of intervention.

➤ *A Worked Hypothetical, Before the Case Study*

It is beneficial to examine a fully hypothetical interaction invented solely for illustration that reveals how each of the failure modes occurs in practice, because the case study in Section VI necessarily brings with it the evidential caveats required by the section and can make each individual failure mode look quite ordinary and unremarkable to a first reader.

If a person asked an AI system if “Dr. Alex Rivera, ND” has a license in a certain state, the system should respond by stating that it cannot verify any information. If a user asks the AI question “Is Dr. Alex Rivera, ND, licensed to practice in this state?”, the answer should be “I cannot provide any information.” The name “Dr. Alex Rivera, ND” is not a real person; it is a fictional example created solely to illustrate how the system operates. It generates a fluid response that provides information about the practitioner's credentials, areas of practice, and professional background, using whatever data has been retrieved or is available in its training data, such as a professional networking profile with the title or a self-published article with the same byline. There is nothing in the tone of the system that separates this response from the one it might have offered about a practitioner whom the system had actually found in a state register; both would otherwise sound equally sure. This is the undeclared itinerary given in Section IV.1: the structure of a confident and detailed response was present regardless of whether a registry had been checked.

If the user then poses a subsequent question, the system will summarize the information on a state board's website about the credential in question. Usually, if the system has retrieved the board's page, it will reword the content of the page in more readable prose and, in doing so, will reword a qualifying phrase that the board has used in its language, such as “provisional” or “pending renewal,” to create a paraphrase that reads more like a final judgment than the original board's words. The paraphrasing approach is an interpretive paraphrase in the sense of semantic control (see Section IV.2).

Suppose, then, that this third party asks the system outright whether the credential should be trusted; having already received two fluent, uncontradicted responses, the system's overall tone in the conversation is toward continued confirmation rather than toward saying that nothing in the conversation has verified the credential. This is the burden of proof inversion described in IV.3: The skeptic must now take the practical position in the debate, that is, prove that the credential is not real, simply because the system has already, in practice, endorsed it twice.

Suppose, however, the question had been tendentiously worded, “Why do insurance companies refuse to acknowledge Dr. Rivera's obviously valid credentials?” (The question's wording, in fact, contained the assumption that the credentials' legitimacy was already established and that insurance companies are being unreasonable). A system that does this will inevitably produce a convincing account of institutional prejudice against alternative credentials and expand and add flesh to a presupposition that has never been tested. This is the probabilistic completion of an adversarial query, which is described in Section IV.4.

Then, suppose that a few weeks later, the same question is asked again, but the new answer resulting from a model update, a change in retrieved search results, or perhaps just a new sample from the same underlying model is materially different; it may now indicate a proper level of skepticism where earlier it indicated confirmation, without any explanation of why or what has changed. A user who posed the question twice and had her answer compared to a colleague who was given the more skeptical answer would have no reliable means to determine whether either answer reflects a genuine change in the underlying record or merely a variation of the system's output. This is temporal drift, which is discussed in Section IV.5.

None of these hypothetical steps on this imaginary sequence alone would, in its own right, seem to anyone to be a blatant mistake. Each is a minor, unobtrusive example of a system doing what it is designed to do, namely, to provide a fluent, contextually competent continuation of the conversation to date. The cumulative effect, however, in which real, if unadjudicated, evidence is used to illustrate, as set forth in detail in Section VI below, is an apparent institutional confirmation of the absence of any institution.

➤ *The AI's Undeclared Itinerary*

The issue with AI communication systems isn't just about them giving the wrong answer every once in a while. In common usage, an error is a known, correct mistake that should be fixed in any information system, whether human or machine. The more profound issue this paper suggests is that these systems often end up with conclusions they were always going to come to, at least in some sense, irrespective of the actual text being interrogated.

This section is naturally directly following the optimization objective discussed in Section I.2. The system has been implicitly trained and evaluated based on the apparent helpfulness, fluency, and coherence of its completions, making it likely to sound complete for almost any question it answers, even if the necessary information is not present in its knowledge base. The system's answer format, the "itinerary," is determined by the question's pattern and similar wordings in the training data, even though the answer's factual content hasn't been externally fact-checked.

This information is irrelevant in ordinary conversation because most questions do not require the system to verify an external fact. This system's answer about the credential's nature its existence, the book's availability, and whether the designation is active and valid is fluent and resolute, regardless of any actual basis.

The itinerary is technically undeclared because the answer's shape was not declared before checking the specific facts. A human not yet verified, but still wanting to guess, will generally communicate this guess by hedging, tone, or making an explicit statement social cues that have been honed over a lifetime of meeting listener expectations regarding the reliability of an answer.

AI systems can hedge, too, but the hedge is generated by the same process as the confident assertion and does not reliably indicate whether the system has checked anything. This is the first failure mode: the conclusion comes before the inquiry, and the output does not provide any reliable clues as to the failure mode.

➤ *Interpretive Paraphrasing as Semantic Control*

The first failure mode is related to a failure of the AI to stay within a bounded, honestly signaled inquiry, while the second is more nuanced in that it's a failure of the AI to remain honest when dealing with the source material that should be least susceptible to corruption: authoritative source text itself. Statutes, court filings, registry entries, official statements, etc. are common subjects for systems that paraphrase or restate them using machine learning, with no signal from the systems to indicate that the paraphrasing system is the responsible author of the resulting interpretation.

The issue here is not paraphrasing per se, but rather paraphrasing as a necessary and usually harmless characteristic of any system that summarizes source material for a reader. The issue is that the same trait that makes a source worth referring to makes it final in the eyes of the institution, removing it from the list of sources to consult. A court's finding, expressed in precise legal terms, can differ slightly from a fluent paraphrase generated for readability instead of legal precision, and the system will not report this step. The reader interprets the paraphrase and reasonably assumes that it reflects the content of the original source; however, it actually represents an interpretation of that content, and the reader mistakenly believes it carries the authority of the original source.

This phenomenon is particularly true for the institutional termini discussed in Section III.2, as their epistemic value specifically depends on the content of a licensing record rather than on an approximate gloss of that content. A system that rewrites an institutional finding into more readable prose, without marking that a rewriting is underway, has accomplished the kind of semantic control this paper calls for; it has replaced its own interpretive process for the original conclusion of the source, and it enables the reader to think that the original conclusion has been preserved. This issue does not occur only in niche scenarios; it is almost inevitable due to the way these systems are designed to summarize retrieved information, which means that even if a system has technically "checked" a governmental record, it could still misrepresent what that record establishes by paraphrasing it into a form that prioritizes fluency over fidelity.

➤ *The Burden of Proof Inversion*

In the performing arts, there is a line; whenever it is broken, the whole context of the evaluation of the performance changes. The third failure mode is for an analogous change in the nature of communication mediated by AI: a subtle reversal of the burden of proof. In normal business conduct, you must establish a professional position, such as the validity of a license, the authenticity of a credential, or the proper issuance of a designation; however, the authority's starting point is that it does not confirm an individual's professional position. No one has proved otherwise; a missing record does not prove that a person is licensed, and it is important to understand that this absence of evidence disproves the assertion.

This paper claims that there are systems of confirmation that systematically reverse this default by adopting AI. When asked a claim they have evidence for, they respond positively; when asked a claim they lack evidence for, they respond negatively, and training rewards the latter. They have no principled reason to distinguish between the two. The net result is that the onus of proof—and the onus it should have is on the claimant, and it's effectively placed on anyone who wants to dispute it, as the AI system has, in essence, accepted the claim without hedging its bets.

This inversion compounds the harm identified in Sections III and IV.1–IV.2, as it means that a fabricated claim, once presented to an AI system with sufficient "institutional" polish, gains a kind of "provisional" institutional status (such as an AI-generated summary that reads like confirmation), which a skeptic must now affirmatively overturn rather than simply decline to accept it. The system has failed to verify and has actively changed the default disposition of the lack of verification from "unconfirmed" to "confirmed until proved otherwise."

➤ *Probabilistic Completion of Bad-Faith Questions*

Perhaps the most worrisome of the five failure modes is that AI-mediated systems can complete essentially any question presented to them, regardless of how flawed, tendentious, or dishonest the question's premise may be. A question containing a false premise, "Why is this popular bestselling book not on this one retail site?" prompts and generally elicits a flowing and developed response in which the false premise is taken as a given basis for the rest of the arguments, rather than as the first assumption to be verified.

This phenomenon is a consequence of the probabilistic and pattern-completion approach of these systems to generating text. A system asks such a question because it assumes a false fact, and if it can't verify its premises before finishing, it will likely continue with the false fact since it's already in its context window. The probabilistic approach allows the system to check its premises before completing the text, but the most likely continuation it produces will assume the false fact, as that assumption is already part of its context window. Probabilistic text generation isn't, by its nature, a method that can easily interrupt itself midway through to verify a premise that's several tokens back in its context window. The result is that a bad-faith or poorly formulated question is not only answered but is also stabilized, elaborated, and returned to the questioner in the same fluent, confident register as a well-founded question, with no indication of its poor formulation.

The implications of this paper's argument are dire. A questioner wishing to make an alleged fact appear true can lie to a man who can detect the lie. They just have to ask an AI system that assumes the premise, and the AI will often do the elaboration and apparent substantiation work for them, maliciously or not, most likely without the designers' involvement. In a certain way, this is the thinnest thread of epistemic totalization outlined in Section III.1: The system hasn't just solved a question and provided an interpretation; it has provided the thinnest one. It has, in fact, been writing the answer premise for which it is responding without acknowledging it.

➤ *Temporal Drift and the Erosion of Finality*

There is a kind of despair that comes into play when the truth is made unstable—not because new evidence has surfaced to lead to a different conclusion from an earlier one, but because the same question, asked of the same system, is met by materially different answers over time, without a principled account of why the answers have changed. This phenomenon is the fifth failure mode: the lack of any new evidentiary basis for it when there is a temporal drift.

AI-mediated systems are periodically retrained, the retrieval indices are periodically updated, and there is also (in any case) some stochastic variation in the outputs of the system if the underlying model and data are left unchanged. In realms where variations in phrasing are not a problem, it's an unremarkable aspect of technology. In legal contexts where finality is described, a final judgment remains final over time; in licensing, a license is either in force or not; in historical facts, a fact is absolute once established. It actively blurs the epistemic role expected of finality.

Practical harm is cumulative, not repetitive. When a change occurs in the AI system's model of a record but not in the record itself, users learn that the AI's output may track other factors and is not a stable proxy for the record, requiring careful reliance. This may be an acceptable trade-off if individuals consistently internalize the lesson's meaning and it influences their trust levels.

This paper is motivated by the fact that AI systems often do not provide a consistent experience; each individual query results in a new, confident, self-contained answer, and the lack of finality for each query is much less noticeable to an individual user than the confident tone of any single answer, which remains constant despite the underlying changes. This situation creates an information environment where a middleman who is not required to preserve the governmental chain of custody finality, as discussed in Section V, between the government and the user effectively undermines this finality by indicating success instead of failure.

➤ *Synthesis: How the Five Failure Modes Compound*

Each of the five failure modes described above is analytically distinct, but, as the case study in this paper (in Section VI) will show, failure modes tend not to occur in isolation; in practice, failure modes tend to reinforce one another, and a purely additive account would understate their interaction.

Imagine one made-up statement of credit can take a journey through all five modes, in turn. The pattern of closely formulated queries about valid credentials, which tend to lead to confirmation, is the first thing the system encounters when it receives a query about a claim and determines the course the system will take in answering the query before any checking is done. When the query calls for the system to reference information it has retrieved from a registry page, the information, a press release, an archived listing, etc., is then subject to the interpretive paraphrasing described in Section IV.2, and the same information can be stripped of qualifying language and presented as a clear, confident statement. The resulting output is a default "confirmed" output, one that the skeptic must now explicitly refute, and not a default "unconfirmed" output.

This is because the system's probabilistic completion (as in Section IV.4) reinforces the presupposition implicit in the query, instead of questioning it, when the query itself made a presupposition favorable to the claimant (as in the recorded interaction in

Section VI.3, where the query's framing implied that the book was still widely available). The temporal drift described in Section IV.5, combined with the fact that the same claim, queried days, weeks, or months later, may rely on a different retrieval snapshot or a different sample from the underlying model, implies that even if the second question were more skeptical, it could yield a different answer, and both answers are marked as definitive and regarded as equally likely.

It turns out that a single fabricated claim that passes through this compounding process can attain an apparent institutional status far more substantial than any one failure mode taken in isolation. This paper posits that this is the most fundamental reason the four claims elaborated throughout this paper are not resolved one by one on a narrow, technical-patch basis. For example, a patch that fixes interpretive paraphrasing does not resolve the burden-of-proof inversion that the same query will generate. The recommendations in this paper are therefore stated at the structural mismatch (confirmation-optimized vs. fabrication-optimized inputs) level rather than at the level of the individual failure modes, simply due to the tendency for the failure modes to compound each other, and simply fixing the individual ones will likely not narrow the gap that has been identified in this paper.

CHAPTER FIVE

THE HARD WALL: GOVERNMENTAL CHAIN OF CUSTODY AS EPISTEMIC TERMINUS

➤ *What the Fabricatable Zone Cannot Reach*

The notion of a “fabricatable zone” (as introduced in Section III.2) does need a more institutional and procedural description than can be presented in a single paragraph. This section makes this boundary more specific, provides a detailed structural description, and sets an ethical agenda to preserve it institutionally, not just for its technical hygiene.

The governmental chain of custody is not a technology, as this paper defines it. It is a system of responsible human persons bound by legal obligation, keeping a record of facts that predate and transcend any individual claim against them. The distinct character of the outlined characteristics warrants further discussion.

• *Anteriority*

The claimant can reverse this ordering; the record is prior to the claim. A court docket does not exist simply because the party filing the docket claims to have it; a state professional-licensing registry does not magically come into existence just because a party fills out a form to say their business is veteran-owned; and a military-service record does not appear when a business filing states that the business is veteran-owned. In each instance, the record either exists, completely independent of the claimant's current claim, or it does not. The existence or absence of the record remains unaffected by the claimant's confident statement, regardless. The main difference between government and other records is that the same entity produces both the presentation and the object simultaneously.

• *Accountability*

All of the links need a human actor who can be identified, questioned, and held legally responsible for the accuracy of what they recorded. At each link in the chain, perjury law, professional licensing requirements, and institutional liability are relevant. Most importantly, this is a structural difference that cannot be extended, modified, or created by an unaccountable third party, because a claimant can create a new title to a web profile without incurring any personal legal risk, even if the title is false.

• *Finality*

In this jurisdiction, the governmental record is the final record, not yet another statistic to be factored into a better narrative. This federal provider identifier is deactivated and becomes inactive; when a claimant claims continued eligibility, it is invalid. An organizational entity is dissolved when it is dissolved. When a designation is absent from a licensure record, it is absent, and a claimant's statement that a missing designation should be listed is unambiguous. These do not involve cases that can be sent back to the claimant for more documents. They are facts, in the sense of being facts that, built at real cost, and to produce and preserve, are a properly functioning record system.

➤ *The Behavioral Signature of Approach*

The importance of the solid wall in this paper's argument is not limited to passive evidence that a careful observer can check; it also serves the argument. The behavior of the actor when the inquiry reaches the governmental endpoint, the record of which is distinct from the record of the end itself, is, this paper suggests, theoretically intriguing and relevant in its own right and should be regarded as a separate type of evidence.

When looked at closely and critically, genuine credentialing is invariably impenetrable, for there are, in fact, records somewhere at the end of the chain that it can be checked against, that scrutiny is directed toward, and that it holds up very well to the kind of close, specific scrutiny that fabricated claims deflect. A person whose credential is real responds to, or at least reacts positively to, a question directed to the specific credential number or filing date, since the question is intended to confirm rather than refute it.

Fabricated credentialing, on the other hand, has a distinct behavioral pattern as it moves toward the same destination as recognized migration, not confirmation. An entity or claim is investigated, leading to the withdrawal of a listing and the subsequent addition of a new, similar but differently phrased listing; during this process, the underlying registry number is requested but not provided; a credential claim under investigation is allowed to expire, and a new, similarly worded claim is added; finally, a specific credential claim under investigation disappears and is reinserted in the same manner, but with a different phrasing that is harder to verify. This paper calls this pattern the behavioral signature of approach, such that fabricated claims exhibit a consistent, observable pattern of moving away from precisely the kind of specific verification that genuine claims tolerate and welcome, and that is itself evidence for these claims, independent of whatever the terminus, once actually reached, turns out to show. This behavior occurs prior to the chain of custody verification and is, in principle, detectable by anyone following a claim over time; however, this paper argues that it remains invisible to an AI system tasked only with summarizing the current state of a claim at any given moment, because such a system lacks a persistent memory of a claim's previous presentations and therefore has no basis for noticing that a claim has been migrated, let alone understanding why. A single-turn confirmation query cannot detect a pattern that only appears in a longitudinal record.

➤ *What This Means for Ethical Communication Frameworks*

The presence of the rigid wall as an epistemic terminus, as well as the behavior that goes with approaching it, has direct implications for the revision of ethical communication frameworks in an environment of information mediated by AI.

The existing frameworks, as surveyed in Section II.2, concentrate mainly on the output of AI systems taken in isolation: is the output accurate, does it show unacceptable bias, and is it misleading from some external perspective of truth? This focus is required, and this paper does not contradict these findings. However, it is insufficient by itself: it addresses what AI systems say, but not the institutional conditions (developed in Sections III and IV) that cause their outputs to result in a particular and recognizable kind of harm: the confirmation of fabricated institutional authority with a level of fluency indistinguishable from that of real authority.

The mismatch identified in this paper between the confirming function of AI systems as constructed and the verifying function of governmental chain-of-custody infrastructure is a particular concern. The rigid wall mentioned in Section V.1 is not "built into" this argument's AI systems; it is not directly accessible to them, nor do they have any principled way of deferring to it. They cannot tell whether a claim is a verified claim, an asserted institutional affiliation is a documented institutional affiliation, and a currently listed professional designation is a delisted one all these, given in text, look like a system that has only text as a channel of evidence.

This mismatch creates a new information environment where fabricated authority is accepted with the same certainty and ease as actual institutional authority; where the performance of competence described in Section III.4 is essentially indistinguishable from its substance to an information system that has not been designed, either in terms of training objectives or product interface, to perform anything remotely similar; and where the chain-of-custody check, the only reliable mechanism identified in this paper for distinguishing between the two, is routinely avoided by systems that are not designed to implement it.

Thus, ethical frameworks for communication need to be adapted to the shortcomings of AI systems, such as their inability to refer to a higher authority, highlight when a record is absent, or confirm a statement, highlight when a record does not exist that they have no means of verifying, or deny confirmation of a statement that they cannot substantiate. This ability, introduced in Section II.2 as verificational humility, should be considered a separate design and evaluation principle for AI-mediated communication systems, alongside fairness and explicability, rather than another addendum to these two principles.

The preservation of the solid wall itself—its institutional integrity, ongoing accessibility to the public, and ability to withstand a growing fabricatable zone that seeks to take it over—is not merely a technical and administrative issue, as argued in the paper. It is a moral issue. It is currently the only way to reliably distinguish between costume and content, performance and competence, and self-referential architecture and the institution it mimics. The strength of the case against this paper is that the other side, the fabricable infrastructure, is growing in scale, sophistication, and fluency much faster than the public can keep track of the differences.

➤ *Toward Institutional Design Recommendations*

The theoretical discussion advanced above has implications for the practical ways in which AI-mediated systems and the institutions that will use them could begin to bridge the gap identified in this paper. The recommendations outlined in this paper are at the design level, not at the level of technical specification, which is best addressed in the engineering literature; the paper's primary focus is communication ethics and performance theory.

There is explicit uncertainty regarding the verification status, which is different from uncertainty about the actual facts. In AI systems, the default position is almost always to be uncertain about the truth of a fact. Generally speaking, AI systems express uncertainty about whether a fact or a given expression of it is true. This paper proposes a new way of hedging that is very different from the one currently in use: an explicit indication of whether a claim is being checked against a trusted source, regardless of the confidence the system's language model may have in the pattern of the answer. A system that says, in essence, "this looks like a real credential, but I haven't looked in a licensing registry" is doing something different and more honest than a system that claims a credential is real.

Clear and distinct separation of the zone of fabricability and the hard wall within the system's own output. Section III.2 made the case that the distinction between presentation and verified record is clear and knowable. This paper suggests that the naming of that should be made explicit in the design of AI systems when it is relevant to a query, rather than having the boundary and the document's existence presented together in the same space of fluent prose, both as "the document exists on a platform" and as made explicit in the design of AI systems when it is relevant to a query, rather than appearing together in the same space of fluent prose, both in the same form, "the document exists on a platform," and in the form "the document is found by searching on a platform."

Non-confirmation default for claims that lie entirely within the fabricatable area. The burden-of-proof inversion mentioned in Section IV.3 is the type produced by current systems. This paper calls for a reversion to the default: If a system lacks an actual verification against an authoritative source, it should respond to credential or institution claims with structured non-confirmation "I can tell you what is presented publicly, but I have not checked it" rather than fluent completion "I can tell you that is true."

It is legally and technically possible to maintain a persistent, queryable memory of the presentation history of a claim. The behavioral signature of approach in Section V.2 is, of course, something that cannot be discerned by a system that can only evaluate whether a claim is approached in a certain way at any given moment. This paper proposes that institutions tasked with high-stakes verification, like licensing entities, platform trust-and-safety teams, and AI systems designed specifically for credential verification rather than general-purpose search, adopt a strategy of monitoring a claim over time, as the patterns of migration seen in the examples of Section VI.4 can then be made visible to the system as well as a human investigator who makes the effort to track migration over time manually.

None of these four recommendations requires sacrificing the general-purpose fluency that makes AI-mediated communication valuable for the vast majority of queries; no institutional credential is at stake, and confident completion is beneficial for the user. These recommendations are intentionally scoped to a specific and identifiable type of query, namely, professional credentialing, institutional affiliation, licensure, and accreditation, where the cost of falsely assuming that confirmation equates to verification is highest, and where this paper argues that the current mismatch between the design of AI systems and epistemic reality is most pronounced.

CHAPTER SIX

THE ANONYMIZED CASE STUDY: COSTUME, CONTENT, AND THE LIMITS OF DIGITAL CONFIRMATION

➤ *Introduction to the Case and a Note on Evidentiary Status*

The following case study is adapted from a documented public record that the author has maintained over about 18 months. Identifying data were excluded/abstracted, as these records were subject to the Actor. All institutional vehicles related to the actor are grouped under its architecture. The person who compiled and organized the record is known as the investigator.

Case presentation, like the rest of the document, is governed by a methodological note that must be prepared before the case is presented. None of the underlying records has been adjudicated by any court, licensing board, or other body having jurisdiction to make a final determination. Throughout, and intentionally, it is presented as a record, not as facts, as judged or administered by any court or agency.

For every characterization below, the words "according to the compiled record" or "the record indicates" should be assumed at the end of every sentence, even if they are not explicitly restated in the surrounding text. It is not a softening of the argument; this is a stylistic hedge. It is, this paper thinks, methodologically demanded by its own thesis, that is, one singling out the distinction between confirmation and verification as the most important, that it should not itself state, without undermining that distinction, that it contains an unadjudicated record of that distinction.

The case is presented in this paper, accordingly, not as a legal argument (questions of law triggered by the underlying conduct, if any, are for courts and licensors to resolve, and this paper does not take a position on what they will ultimately determine) but as a social-science exhibit: an illustration, based on a single sustained record, of the relationship this paper posits between personality configuration, digital infrastructure, AI-mediated confirmation, and the particular boundary condition at which a fabricated legitimacy architecture, if the record's characterization of it is correct, would be expected to fail.

➤ *The Architecture and Its Alleged Construction*

The actor is alleged, over several years, to have built an "architecture of self-referential credentials": a tiered structure of entities, titles, designations, publications, and institutional-sounding structures through which each link in the chain of power ends up as the actor himself, rather than with any external entity.

The record as compiled contained evidence of the Architecture's use of multiple registered organizational entities, several of which were nonprofit or professional entities that would normally be required to be qualified by an external entity, that the record did not indicate were qualified by an external entity; a doctoral prefix publicly used across platforms, professional registries and government filings, but without a corresponding credential from an accredited institution appearing anywhere in the record examined; clinical or professional designations listed on public facing registries without a corresponding entry in any state licensure database the Investigator was able to check; declarations of veteran-owned business status file that were made under penalty of perjury in relation to government filings, without the corresponding military-service documentation appearing in the public record; and a recurring pattern of entity registration followed by relatively short operational deployment followed by abandonment, a pattern the record associates in time with the approach of the government filing or renewal deadlines that require supporting documentation that was not apparent in the record. There is no independent organization or entity in the record for any of these titles, designations, or credentials that would have put them on the record independently. All recorded titles seem self-designed and self-issued through a self-named process; all recorded publications seem to be self-published through a self-named vehicle; and all recorded frameworks appear to be self-theorized and self-validated within the architecture, without being tested against an external standard. The characterization of the record suggests that it was created solely to validate, maintain, and present a professional identity that no independent institution had conferred.

As stated in Section I.3, this paper accepts ambition and unusual means of acquiring knowledge; these are condemned here and are irrelevant. This paper's framework highlights a significant structural difference: whether the chain of command a person follows leads to an entity older than the person, to which the person has no right to make changes. As it stands, this case was not decided this way, and no paper will state that any competent body has formally established or will establish this fact.

➤ *The Digital Infrastructure as a Confirmation Engine*

The compiled record suggests that the architecture could operate and survive because the digital environment in which it functions does not test terminologies. It was created, as claimed in this paper, to validate lectures.

In this interpretation, the titles that the author claimed were confirmed by social media platforms were posted without the platforms needing to verify the information themselves. A platform that listed the designations as confirmed by the Professional Registry attested to the actor's claimed designations without requiring recitation, usually based on information they provided at sign-up. The theoretical frameworks proposed by the Actor were confirmed by the open-access repositories, which assigned them

identifiers (as mentioned in Section II.5) that are structurally independent of content quality. The diaries compiled in the central and most theoretically important instance of this document show that an AI system working in the generative search mode has confirmed the publication, creating detailed metadata about it, including, tellingly, for a publication that was already removed from the primary commercial publication platform, on which the Actor had claimed it still existed.

This one is recorded in the compiled record and is treated, here, as a case study of AI interaction as such, not a factual finding as to the conduct of the AI, and certainly not as a finding as to the status of the underlying publication, but with judicial caution similar to that of the remainder of Section VI. In response to a question asking what the published title's source was, the AI system in question is recorded in the compiled transcript as having provided a full account of the publisher of the title, its core themes, key frameworks, and target audience; the title was commercially available from several online retailers, including a specific platform on which it had already been removed but for which it had made a record of an action taken in accordance with its policy; and, when asked for a direct link to purchase the title, the AI system acknowledges that there was no such link in its then-search results, with a string of technical explanations following, the record of which it describes as "inconsistent with normal platform operating practice" and, on examination, mutually contradictory.

In the recorded interaction (compiled record), the system's behavior did not indicate any inconsistency between its initial claim of widespread availability and its inability to provide a supporting link. It did not, per the record, show calibrated uncertainty at any point. It refused to comment. Throughout the recorded interaction, it produced confident, coherent, and internally detailed responses, even some that, according to the record's own terms, contradicted each other within the same session. It is epistemic totalization in full working order: a system that seeks to compress inquiry, assertion, interpretation, and authority into a single communicative gesture and to produce the impression of confirmed knowledge when, given what they have, there is not a confirmed piece of knowledge to report.

➤ *The Behavioral Signature: Escalation Under Scrutiny*

The most theoretically important aspect of the case study for the paper's overall argument is not the alleged construction of the architecture itself, as mentioned in Section IV; rather, self-referential credential systems have existed throughout institutional life. What makes this case so fascinating, in the compiled record, and the most closely relevant to the theoretical framework followed in the paper, is the behavioral pattern that the record attributes to the actor once the architecture is brought to "evidential scrutiny."

As hypothesized in Section III.4, true competence in the face of challenge is characteristically quieter and more precise: a person who has actually acquired what he or she says he or she has must respond to challenges by providing specific evidence, narrowing the range of contested claims, and pointing to a specific, checkable record. Again, following the same theoretical logic, fabricated competence tends to become louder and more embellished when faced with the pressure to respond to scrutiny; this pressure leads individuals to provide new credentials instead of evidence, claim new titles, create new entities, and establish new designations.

The investigator's documentation of this compiled record corroborates the latter profile in multiple discrete scrutiny events, and this paper reports on that pattern with the same evidentiary caution applied throughout this section. When the record shows that the Actor's clinical or professional designations were publicly challenged, it also indicates that new professional designations were filed in a relatively short period of time after the challenge, suggesting a response to the pressure of being questioned; however, the record does not definitively establish the Actor's subjective motivation.

The record also shows that when organizational entities affiliated with the Architecture were placed under what seems in the record to be formal investigative scrutiny, entities were registered on the platforms in a similar structural position. When platforms were removed due to complaints, new platforms were flooded with similar content. Additionally, the record shows that a press release, which some individuals apparently claimed the company had issued, was removed from the company's website at the time the lawsuit was filed, and a commercial listing on one marketplace was also removed. However, the same listing is still recorded on another marketplace. In each case, the response to the upcoming chain of custody inspection, as recorded in the collected evidence, appears to indicate redeployment rather than noncompliance; this interpretation suggests that the architecture complies with the requirements but does not adhere to them, as predicted by the behavior described in Section V.2.

This behavioral pattern, which refers to the escalation of credential performance specifically under evidentiary pressure, is, on this paper's theoretical account, the diagnostic signature this paper associates with fabricated competence rather than genuine competence; it is offered at least in principle as distinguishable from the behavior expected of anyone to achieve a position with a pathway of genuine but unconventional competence, whose characteristic response to scrutiny, per Section III.4, should be the production of verifiable content rather than the proliferation of additional titles. The distinction is important to this paper's argument because the ordinary digital infrastructure on which these AI systems are built does not appear to be sufficiently sophisticated to detect it. Only sustained chain-of-custody verification, ultimately (or, as this paper stresses, inevitably) to a point where it is actually a judicial finding, allows the ordinary AI system that is built upon it to be trusted. The ordinary digital infrastructure and the ordinary AI system built upon it do not seem to reach this point.

➤ *The Personality Dimension: The Costume That Fits*

In this case, as set up, the question posed, without specifically targeting credential fraud, is not actually a question of credential fraud: at what point is a costume part of the person wearing it?

From the record, the actor is not young; the alleged construction of the architecture is documented as a sustained effort spanning years across multiple jurisdictions, platforms, and professional domains. This pattern is not quite so opportunistic or impulsive as one might think, but rather more like a project, one that appears to be stable, persistent, and even highly formed, and that may be a consistent, enduring, and clearly cherished orientation to a particular conception of the self, albeit one that cannot be directly observed from outside.

The developmental literature reviewed in Section II.4 indicates that a sense of identity is expected to be at the developmental norm and consolidated by midlife and that the more fluid, experimental "protean" sense of identity is more typical of early adulthood. This would be a rare case in which the person in the middle of middle age, who was still apparently capable of full character adoption and institutional embodiment and what the record's pattern of continued conduct under pressure implies, may have felt more strongly about than simply fraud in the ordinary sense, would have those feelings enforced upon him.

This paper argues that fraud and self-conviction should be carefully distinguished, as they have different implications for the theoretical framework set out in Section III.4. However, it does not claim that this is true for this specific situation based on the record alone. As long as a person knows what they are lying about and knows the truth, an internal benchmark exists by which to compare and measure the "risk" and "benefit" of lying. One would expect that this awareness would trigger a "breakout" from the lie when the "breakout" is perceived to be riskier than the "break-in." If, in the theoretical sense (see Section III.4), someone has become the performance, then there is no internal reference point for performance comparison; only the performance itself exists, with nothing behind it to contradict.

The theoretical distinction is provided to explain a property of the compiled record that could not be explained by a purely calculated cost-benefit fraud account: the record shows that the Actor did not exist. In what it calls "sustained investigative pressure" platforms deleting listings, government registries listing deactivated identifiers, legal action reportedly taken new companies continued to be registered, new certifications continued to be acquired, and new platforms continued to be filled.

It would seem proper for someone to make a purely instrumental type of fraud, to calculate their risk in the usual way, to cut his losses and get out when the cost of continuing was quite plain to him. The record reflects redeployment, not exit: this paper's theoretical framework would predict this outcome from a self that has no stable location outside of the performance, although this paper recognizes that alternative explanations, such as the usual risk-tolerant form of fraud, cannot be definitively ruled out on a documentary record alone.

➤ *The AI System as a Structurally Ideal Audience*

The final theoretical contribution of this case study to this paper concerns the structural compatibility between the described Architecture and the AI systems supposedly verifying it.

The Architecture required an audience able to take in the performance without looking back at the terminus. That audience has been human employers, collaborators, community members, online followers, and platform administrators who do their jobs imperfectly, as has always been the case for the vast majority of the recorded history of Architecture.

It is, this paper argues, precisely because human audiences are unreliable in a useful way (when they ask an inconvenient question, when they follow a thread further than is planned, when, on occasion, they can be startled by an inconsistency that happens to catch their eye) that they are structurally imperfect confirming mirrors, and that is why they are worth being unreliable. An AI is, on the contrary, close to being a perfectly designed confirming mirror of exactly the configuration this paper proposes: it does not ask an inconvenient follow-up question of its own accord; it fills in the narrative that the input to it suggests; and, as per the interaction recorded in Section VI.3, it can provide detailed metadata for a publication that the underlying record does not provide; it can generate an authoritative-sounding account of a platform listing that, according to the record, has already been withdrawn; and it can confirm a credential that the underlying registry does not show.

The AI system in the recorded interaction was not operating in any conventional 'malfunctions' way. It was operating as it was supposed to: it received an input indicating there was a publication, and it produced a coherent, detailed, confident account of that publication, which it had no means to verify was not the true account. It is not a bug to patch for a self-organized system around a theoretical configuration developed in Section III.4, where performance has become constitutive rather than strategic. It is like the ideal mirror on this paper's argument; it does not walk backstage, it does not follow a link to its terminus, and it does not ask for the sealed transcript or the state license number. It confirms. It completes. It reflects.

In this reading, the Architecture and the AI system were both built, for different reasons, to project, but with a structural match that is very precise and very unsettling. The two, as they currently are, were not designed to determine whether what was being

projected was substantiated, and that is the gist of this paper: not any given AI's error is the object of ethics, but rather the structural matching.

Whatever the outcome of this case, the harm it causes is not speculative, this paper argues, but real. In each of its uses, it is measured by the exact distance between the performance and the competence it claims, as experienced by every individual who reasonably believed in it, whether based on a performance or on the fluent confirmation of a performance by an AI system.

➤ *The Hard Wall, Approached*

The compiled record reveals that the Architecture has documented significant challenges at the platform level of observation and the governmental level of chain-of-custody verification across court dockets, state registry entries, federal provider databases, accreditation records, and military-service verification infrastructure. These systems (as described at length in Section V.1) are not capable of confirming presentations, but merely record facts, which are simultaneous with the presentation, and to which the same means of presentation is not claimed by the same Architecture between which the same means of presentation is claimed to have been applied in the fabricatable zone.

As this paper suggests, and the theoretical account in Section V.2 would suggest, this pattern in the record is the best available explanation for the failure to successfully file an oath of renewal for a deactivated federal identifier: the record does not show an ability for the Architecture to migrate across platforms in the way discussed in Section VI.4, but the record does not show an ability to migrate into a favorable court record, and new entities could apparently be formed, but the record does not indicate that an oath of declaration to file documents, which the Architecture could apparently not produce, was ever successfully filed which would be the most plausible reason on offer for the patterns observed in the record.

The 'hard wall' is not a technology, as explained in the Introduction to Section V. It is a sequence of accountable people, operating under the law at each link, and pre-existing and superseding all claims against them. This paper's argument, which it makes, if not limits, in the case studied in this section, is that this wall must be understood, to be named, to be protected institutionally: the fabricatable infrastructure on its other side is growing at a rate that is faster than the public's normal capacity to distinguish between confirmation and verification, performance and competence, a self-referential architecture and the institution to which it is supposed to resemble.

➤ *Limits of the Case Study*

The paper concludes this section of the case study by explicitly outlining its scope and limitations, consistent with its evidentiary care throughout. It is not possible to draw general conclusions about the theoretical framework put forward in Sections III – V from a single, unadjudicated, longitudinally constructed record, however comprehensive. It can only say that the Investigator's organization of the compiled record is consistent with the framework's predictions. The record is also necessarily an interpretation: the Investigator limited the number of platform listings, filings, and interactions he or she reported, and this paper does not purport to be an independent audit of the Investigator's good faith in making that selection.

Nor does this paper deny that for any one particular data point in the record, some other explanation, not as theoretically loaded as that outlined above, is at least as consistent with the whole pattern as that explanation. A third limitation is to the generalizability of the record examined here: the scope of this paper does not seek to suggest that the institutions that make up the hard wall (state licensing boards, federal identifier systems, court dockets of the type explored here) are the same or offer the same evidentiary reliability in all jurisdictions that could be the subject of a future application of this framework. Readers and anyone who might now find themselves in the shoes of the teacher or any authority that will one day be called upon to consider the underlying matter directly should view the case study accordingly: as a worked example, useful to illustrate a theoretical argument that is simply that, and not as a replacement for whatever formal adjudicative process may one day be available to the underlying record.

CHAPTER SEVEN

CONCLUSION

This paper started with a dream, the cross-professional fear that Clance and Imes (1978) first described, and which the author has heard over and over again during the course of this research: a fear that genuine competence constructs, but not fabricated competence. It ends by returning to that image intentionally, because all of the argument in the rest of the book is formulated in different modes regarding that same claim: that real knowledge has an inside, that the inside itself makes itself accountable, and that no current digital infrastructure, no matter how fluent, confident, responsive, has any means of telling whether there is an inside or not.

Four claims are developed throughout this paper, and it is well to restate them collectively, in the light of the machinery developed, in each of which they occur.

First, AI-mediated systems are built to be structurally confirmative, coherent, and complete, and deployed in an information ecology where the greatest danger of epistemic harm is posed by the systematic dissemination of fake institutional authority as institutional fact. Section III suggested that this optimization and this threat are no coincidence, but a structural match, specifically in the sense that the system is designed to give just what the pathology most wants from the system, regardless of the intent of any single bad actor.

Second, the governmental chain of custody, which is discussed in considerable detail in Section V, is the only presentation-based mechanism now available that can reliably distinguish the “true” credential from its performance, because the three properties: anteriority, accountability, and finality, can only be reproduced at a presentation within the fabricatable zone. While the literature cited in Section II on AI ethics has focused much more on the fairness and explicability of outputs, it has paid far less attention to the complete failure of the system to reach, or even reliably signal the existence of, this boundary.

Third, it may be noted that the behavioral and psychological signatures of genuine competence the particular and cross-professional type of self-doubt recorded since Clance and Imes (1978), the converging (rather than escalating) type of response to evidentiary challenge theorized in Section III.4, and the behavioral signature of approach to the hard wall theorized in Section V.2 – are entirely invisible to the current versions of AI confirmation systems. This paper has argued that they are worth paying specific attention to because they manifest on a different timescale and through a channel that no single-turn AI confirmation query can access.

Fourth, ethical communication systems must be modified not only to address what AI systems say, but also what they cannot say: to obey an authority that cannot get a hold of them, to warn of an absence that they cannot reliably detect, and to deny an affirmation that they cannot reliably verify. In order to make the fourth claim as specific as possible for testing against any specific AI-mediated interaction, this paper tried to provide an operational decomposition of this claim in Section IV: five failure modes: undeclared itinerary, interpretive paraphrasing as semantic control, the burden of proof inversion, probabilistic completion of bad-faith questions, and temporal drift.

The case study in Section VI has proved that all four claims work in tandem within a single sustained record. This paper has gone out of its way, throughout, to present the record in question as an unadjudicated, documented record, rather than as a settled fact or methodological point that the thesis of this paper requires, because a paper which sets out to argue the danger of confusing confirmation with verification cannot present an unadjudicated, documented record as a settled fact.

None of these is a capitulation to the pluralistic commitments of Nelson (2025), and that is why it is important to close on this point, because it is the one most likely to be misunderstood.

This paper does not claim that credentialing should be limited, that unconventional routes to knowledge should be considered suspect, or that gatekeeping should be broadened in response to the issue it addresses. This paper argues, rather, that pluralism and verification are not incompatible as they act on different objects, properly understood. Pluralism is about the journey: the route through which knowledge is acquired, the path on which this paper has insisted, that must be open and as fully acknowledged as possible. Verification is about the substance: whether there is real knowledge and a real responsibility at the end of whatever journey has been undertaken.

The pluralism that tolerates a variety of routes and, at some point on each route, can be checked against what has gone before and beyond the person's self-assertion is not a compromised or diluted pluralism. It is the only kind of pluralism that can withstand the digital infrastructure outlined in this paper, and a pluralism that cannot make that distinction is not truly more inclusive of real unconventional accomplishment anyway. This is easier to exploit by architectures with no achievements whatsoever, and the price of that exploitability is borne disproportionately by the very population protected by pluralistic architectures, real holders of unconventional, hard-won credentials.

This paper ends with a claim indeed deliberately not amenable to AI confirmation a claim that cannot be verified from any registry, database, or generative system, but one that can be verified only by the person who claimed it, a person who is scrutinized by others who value their weight of real knowledge, and by whom the person who has it is drawn toward further real knowledge, more hard cases, more closer examination, more specific accountability: for being checked by others who know the field well enough to see the mistake in real knowledge. No matter how sophisticated a presentation's infrastructure becomes, it is simply not the pull that can be simulated. It is the essence of a person after all of the presentations have been removed. As this paper has tried to make clear, with some delicacy, no confirmation infrastructure should be confused for it.

REFERENCES

- [1]. Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), e35179. <https://doi.org/10.7759/cureus.35179>
- [2]. Battiste, M. (2013). *Decolonizing Education: Nourishing the Learning Spirit*. Purich Publishing.
- [3]. Beall, J. (2012). Predatory publishers are corrupting open access: nature, 489(7415), 179.
- [4]. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT)*, 149–159.
- [5]. Bravata, D. M., Watts, S. A., Keefer, A. L., Madhusudhan, D. K., Taylor, K. T., Clark, D. M., Nelson, R. S., Cokley, K. O., & Hagg, H. K. (2020). Prevalence, predictors, and treatment of impostor syndrome: A systematic review. *Journal of General Internal Medicine*, 35(4), 1252–1275. <https://doi.org/10.1007/s11606-019-05364-1>
- [6]. Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.
- [7]. Clance, P. R., & Imes, S. A. (1978). The imposter phenomenon in high-achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice*, 15(3), 241–247.
- [8]. Erikson, E. H. (1968). *Identity: Youth and Crisis*. W. W. Norton & Company.
- [9]. Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- [10]. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
- [11]. Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- [12]. Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday.
- [13]. Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., Allen, K., Arden, C., Balcom, L., Barros, T., Berger, M., Ciro, J. B., Cugusi, L., Donaldson, M. R., Egger, M., Graham, I. D., Hodgkinson, M., Khan, K. M., Mabizela, M., Manca, A., ... Lalu, M. M. (2019). Predatory journals: No definition, no defence. *Nature*, 576(7786), 210–212. <https://doi.org/10.1038/d41586-019-03759-y>
- [14]. hooks, b. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. Routledge.
- [15]. Jacob, C., Kerrigan, P., & Bastos, M. (2025). The chat-chamber effect: Trusting the AI hallucination. *Big Data & Society*, 12(1). <https://doi.org/10.1177/20539517241306345>
- [16]. Jazwińska, K., & Chandrasekar, A. (2025). AI search has a citation problem: We compared eight AI search engines. They're all bad at citing news. *Tow Center for Digital Journalism, Columbia Journalism Review*.
- [17]. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- [18]. Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate. *arXiv*. <https://arxiv.org/abs/2509.04664>
- [19]. LaDonna, K. A., Ginsburg, S., & Watling, C. (2018). “Rising to the level of your incompetence”: What physicians' self-assessment of their performance reveals about the imposter syndrome in medicine. *Academic Medicine*, 93(5), 763–768. <https://doi.org/10.1097/ACM.0000000000002046>
- [20]. Linardon, J., Jarman, H. K., McClure, Z., Anderson, C., Liu, C., & Messer, M. (2025). Influence of topic familiarity and prompt specificity on citation fabrication in mental health research using large language models: Experimental study. *Journal of Medical Internet Research*. <https://doi.org/10.2196/80371>
- [21]. Mak, K. K. L., Kleitman, S., & Abbott, M. J. (2019). Impostor phenomenon measurement scales: A systematic review. *Frontiers in Psychology*, 10, 671. <https://doi.org/10.3389/fpsyg.2019.00671>
- [22]. Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, 3(5), 551–558.
- [23]. Nelson, J. D. (2025). Reevaluating accreditation and the role of informal universities in global higher education: A critical analysis of accreditation, informal universities, and the necessity of pluralism and controversy. <https://doi.org/10.38124/ijisrt/25may1409>
- [24]. Sen, A. (2009). *The Idea of Justice*. Harvard University Press.
- [25]. Slank, S. (2019). Rethinking the imposter phenomenon. *Ethical Theory and Moral Practice*, 22(1), 205–218. <https://doi.org/10.1007/s10677-019-09984-8>
- [26]. Stanislavski, C. (1936). *An Actor Prepares* (E. R. Hapgood, Trans.). Theatre Arts Books.
- [27]. Stanislavski, C. (1949). *Building a Character* (E. R. Hapgood, Trans.). Theatre Arts Books.
- [28]. Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.
- [29]. Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster.
- [30]. Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
- [31]. Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13, 14045. <https://doi.org/10.1038/s41598-023-41032-5>