

Academic Integrity Beyond Audit: A Multi-Regime Field Approach to Performance, Internalization and Agency

Version 1.1.0

Author: Benjamin Rivera Perez

Independent Researcher; Hyperweb Healthtech OÜ, Tallinn, Estonia

Corresponding author: ben@hyperweb.ee

ORCID: <https://orcid.org/0009-0007-2649-7149>

Abstract

Universities describe themselves as "learning organizations" that cultivate autonomy, ethical judgment and deep understanding [Senge, 1990]. Yet their integrity practices increasingly resemble those of audited institutions: dense regimes of rules, detectors, dashboards and sanctions [Power, 1997; Shore & Wright, 2015]. This article argues that the tension between these identities is not rhetorical but architectural—and that existing frameworks, while valuable, are insufficient to map what students and staff actually navigate.

Self-Determination Theory models how individuals internalize norms along a single motivational continuum [Deci & Ryan, 2000]. Audit-culture analyses describe what happens when formal verification becomes structurally dominant [Power, 1997; Strathern, 2000]. Neither framework is designed to model situations in which multiple normative regimes coexist, collide and produce qualitative shifts in the very state of agency of those who must live among them.

This article proposes an integrity field for academic evaluation: a conceptual architecture in which legal, bureaucratic, professional, metric, algorithmic and communal regimes jointly govern the same agent. Within this field, it develops two diagnostic tools—the Performance–Internalization Contrast (PIC) and the Audit Dependency Profile (ADP)—and introduces a taxonomy of states of agency that agents occupy and move between under different configurations of regimes and evaluation events.

The central question shifts from "how much cheating can we detect?" to a prior one: when the cameras are turned off, how much of this still holds—and which regimes are actually shaping what remains?

Keywords:** academic integrity; audit culture; learning organizations; multi-regime governance; assessment; generative AI; surveillance; internalization; states of agency.

1. Introduction

A student sits in front of a screen. Proctoring software tracks her eye movements. A plagiarism detector will scan her submission. An AI detector will flag statistical anomalies in her prose. She is writing an essay on ethics.

The irony is not lost on her. She has learned not to mention it.

Formally, the university tells her that academic integrity is about honesty, responsibility and professional judgment. Informally, what she experiences is a dense apparatus of monitoring and suspicion. We teach students to think for themselves; we govern them as if they will not.

But the situation is more complex than a simple opposition between learning and audit. The student is not governed by one regime. She is governed by many—simultaneously:

- **Legal norms** constrain what counts as fraud, what privacy she retains, what intellectual property rules apply.
- **Bureaucratic rules** specify what the handbook permits, what the code of conduct prohibits, what procedures apply if she is accused.
- **Professional standards** from her discipline define what ethical practice looks like in her future career.
- **Metric regimes** make her GPA, her ranking, her progression contingent on auditable outputs.
- **Algorithmic systems** flag her prose, record her keystrokes, assign risk scores she may never see.
- **Communal norms**—what her peers actually do, what tutors tacitly tolerate, what "everyone knows"—often diverge from all of the above.

She must navigate these regimes at once. They do not speak with one voice. Sometimes they contradict. Often, compliance with one produces friction with another. And when high-stakes evaluation arrives—exam, defense, committee decision—the question is not simply "will I follow the rules?" but "which rules, for which audience, with what left of my own judgment?"

Current debates about academic integrity and generative AI tend to oscillate between two poles [Perkins, 2023; Cotton et al., 2023; Bittle & El-Gayar, 2025]. On one side are calls for stricter surveillance, more sophisticated detection and digital proctoring [Han et al., 2024]. On the other are appeals to redesign assessment, trust students and "teach integrity" [Cotton et al., 2023]. Both sides usually assume a single regime is at stake. Neither side offers a vocabulary for describing which regimes are actually in play, how they interact, or what happens to the agent caught between them.

This article proposes four conceptual moves to address that gap:

- An **integrity field** that models multiple normative regimes governing the same agent simultaneously.
- A refinement of **PIC (Performance–Internalization Contrast)** as a diagnostic lens that can be applied regime by regime.
- The **Audit Dependency Profile (ADP)** as a qualitative heuristic describing how a local configuration of regimes relies on external audit.
- A taxonomy of **states of agency**—deliberative engagement, strategic compliance, ritual performance, practical paralysis, exit—and an account of how evaluation events trigger transitions between them.

The central claim is simple: without tools that recognize multiple regimes and their effects on agency, universities will continue to extend their audit infrastructures without noticing that they are designing architectures that produce compliance without conviction, performance without understanding, graduates who have learned to satisfy auditors rather than exercise judgment.

AI is not the disease. It is the X-ray that makes existing fractures visible.

2. From Single Regimes to an Integrity Field

2.1 The limits of single-regime models

Self-Determination Theory (SDT) offers a powerful account of how motivation ranges from controlled to autonomous forms, depending on whether social norms are experienced as external pressure, introjected obligation, identified value or fully integrated commitment [Deci & Ryan, 2000]. In educational settings, SDT highlights the importance of autonomy support, competence feedback and relatedness in fostering internalization of academic values.

SDT, however, operates with an implicit simplification: it typically assumes that the relevant norms are coherent and belong to a single regime. The central question becomes whether the student accepts, negotiates or resists those norms. The model is not designed to represent situations where incompatible regimes simultaneously claim authority over the same act.

Audit-society and audit-culture analyses offer a complementary critique. Power [1997] argues that contemporary organizations are increasingly organized around formal verification practices—accreditation, quality assurance, performance indicators, compliance mechanisms—that can become ends in themselves. Strathern [2000] and others show how indicators, once institutionalized, reshape professional cultures as people learn to perform "for the measure" rather than for the underlying purpose.

These analyses illuminate how formal evaluation regimes can generate ritualized compliance, gaming and displacement of substantive goals. Yet they also share a simplifying assumption: the audit regime is treated as structurally central or dominant. The question becomes how life looks in institutions where audit has taken over, not how audit interacts with other forms of normativity.

Both frameworks—SDT and audit culture—provide essential local cuts through the landscape of academic integrity. But neither provides a map of the whole field—nor a resolution of its constitutive tension.

2.2 The integrity field: multiple regimes, one agent

In real academic settings, students and staff are governed by multiple normative regimes at once. The following list is not exhaustive but indicates the structure of the field:

- **Legal regime:** laws on fraud, privacy, harassment, discrimination, intellectual property.
- **Bureaucratic regime:** institutional policies, codes of conduct, handbooks, procedures.
- **Professional regime:** disciplinary ethics, accreditation bodies, licensing rules.
- **Metric/audit regime:** rankings, performance indicators, quality assurance audits, external reviews.
- **Algorithmic regime:** platform rules, detection tools, automated proctoring, recommender systems.
- **Communal regime:** peer cultures, local norms, unwritten rules about what is "really" expected.

These regimes can complement or contradict each other. A practice may be tolerated by peer culture, discouraged by professional ethics and prohibited by institutional policy. An AI-assisted writing practice may comply with course instructions but violate a licensing agreement. An integrity policy may satisfy an external audit while undermining trust in the classroom.

The integrity field is the space defined by the co-presence of these regimes. Agents do not choose which field to inhabit; they are always already within it, subject to its forces.

2.3 On fields, regimes, and theoretical neighbors

The language of "field" invites a question: why not Bourdieu?

Bourdieu's field theory offers a powerful apparatus for analyzing how agents compete for position within structured spaces of capital accumulation [Bourdieu, 1984; Bourdieu & Wacquant, 1992]. Academic integrity could, in principle, be analyzed as a field in Bourdieu's sense—with its own forms of capital (reputation, credentials, moral standing), its own struggles for position, its own mechanisms of reproduction.

This paper takes a different path—not because Bourdieu is wrong, but because the problem I am addressing is architectural before it is competitive.

Bourdieu's field presupposes that agents are competing for something within a structured space. The integrity field I propose is prior to that: it asks what happens when agents are subject to *multiple* normative structures that do not share a common currency. The student navigating legal, bureaucratic, metric, algorithmic, and communal regimes is not primarily competing for position. She is trying to figure out *which game she is even playing*—and discovering that the answer depends on who is watching.

Similarly, the term "regime" resonates with Foucault's analyses of governmentality—the ways in which populations are shaped, guided, and conducted through dispersed mechanisms of power [Foucault, 1991]. There is genuine kinship here. The algorithmic regime that flags a student's prose for "AI risk" is doing something Foucault would recognize: producing subjects who internalize surveillance, who learn to write *for the detector* rather than for understanding.

Foucault distinguished between *discipline*—which operates on bodies through direct surveillance, spatial arrangement, and normalization—and *governmentality*—which operates on conduct through the structuring of choices, incentives, and environments. Discipline creates friction; the subject *feels* constrained and can resist. Governmentality creates desire; the subject *wants* what the regime wants and believes the wanting is her own.

This distinction matters for the integrity field. A proctoring system is disciplinary: it constrains the body, produces discomfort, and can generate visible resistance. A gamified learning platform with metrics, badges, and leaderboards is governmental: it shapes what the student wants, and she may never recognize the shaping as external.

The danger is asymmetric. Discipline produces subjects who know they are constrained—and can therefore push back. Governmentality produces subjects who believe they are free—and therefore cannot recognize what has been done to them. An institution that relies on discipline creates potential insurgents. An institution that relies on governmentality creates willing collaborators who do not know they are collaborating.

This complicates any simple valorization of "internalization." If regimes do not merely constrain but *produce* subjects—shaping what they want, how they see themselves, what they experience as "their own" judgment—then high internalization may indicate not autonomous endorsement but successful subjectification. The student who genuinely believes in the integrity norms may be the most governed of all.

But Foucault's genealogical method asks *how* a regime came to be thinkable, what historical conditions made it possible. The integrity field framework asks a more immediate question: *given* that multiple regimes now coexist, what happens to the agent caught between them? This is not genealogy but topology—mapping the space agents must navigate, not the history that produced it.

The framework is therefore compatible with both Bourdieu and Foucault but does not require their full theoretical apparatus. It operates at a different level of abstraction: not "how did we get here?" but "what is the structure of *here*, and what does it do to those who must live in it?"

One might also ask about Boltanski and Thévenot's analysis of multiple "orders of worth"—the ways in which actors in modern societies draw on different moral grammars to justify their actions [Boltanski & Thévenot, 2006]. The integrity field shares their intuition that modern life involves navigating plural normative frameworks. The difference is in focus: Boltanski and Thévenot analyze how actors *justify* their actions across orders; the integrity field analyzes how *institutional architectures* impose demands that may not be justifiable at all—demands that must simply be navigated, performed, or escaped.

This is not a claim to have superseded these traditions. It is a claim about where the diagnostic work is needed: not in richer theory, but in sharper tools for seeing what institutions are actually doing to the people inside them.

2.4 The audit turn and massification

Over the past three decades, higher education has undergone what can reasonably be called an audit turn [Power, 1997; Shore & Wright, 2015]. Building on broader transformations documented in public-sector governance, quality assurance and performance management, universities have developed dense infrastructures of monitoring, reporting and external evaluation. Academic integrity has been swept into this movement.

Instead of asking whether teaching, supervision and assessment designs cultivate internalization, institutions increasingly treat integrity as something that must be assured by instruments: plagiarism-detection services, digital proctoring platforms, risk scores, case-management workflows and, more recently, generative-AI detectors [Han et al., 2024; Bittle & El-Gayar, 2025]. These instruments do not merely record integrity; they reshape it by defining what counts as a breach, how suspicion is produced and when cases are escalated [Shore & Wright, 2015; Strathern, 2000].

Massification is often invoked as a technical justification for architectures that rely almost entirely on binary multiple-choice formats and automated scoring. This choice is not neutral. It systematically favors behaviors that optimize performance under audit—selecting the keyed option, avoiding detectable anomalies—while leaving internalization largely untested.

Assessment research has long offered alternatives: ordered or proximity-based items [Briggs et al., 2006], multiple true–false formats [Albanese & Sabers, 1988; Brassil & Couch, 2019], confidence-based marking [Gardner-Medwin, 2006]. Historically, these formats were dismissed as too costly to author and mark at scale. With contemporary large language models, that constraint is increasingly obsolete. AI can now assist in generating, refining and calibrating such items [Karlsson & Lunander, 2025].

When institutions continue to default to binary ABCD formats under these conditions, massification is no longer a technical inevitability but an architectural decision that sustains high dependence on auditor-visible performance.

An architecture that only ever sees integrity in the glow of a dashboard should not be surprised if integrity exists there alone.

3. PIC: Performance–Internalization Contrast

3.1 Conceptual definition

The Performance–Internalization Contrast (PIC) is a diagnostic lens that distinguishes between two dimensions of integrity:

- **Performance:** the observable alignment of behavior with a given integrity norm.
- **Internalization:** the extent to which the agent endorses the underlying value of the norm.

High performance with high internalization corresponds to the ideal of integrity as both practice and conviction. High performance with low internalization corresponds to compliant behavior driven by fear, opportunism or habit. Low performance with high internalization can occur under conflicting pressures, lack of resources or poorly designed systems. Low performance with low internalization corresponds to overt disregard of the norm.

PIC is not a numerical variable but a contrast. It directs attention to the possibility that integrity policies can increase performance without increasing—and sometimes while lowering—internalization.

3.2 PIC in a multi-regime field

When multiple regimes govern the same act, PIC helps to disentangle where performance and internalization are located. Consider a student in a program with a strict AI-use policy:

- She may perform strict compliance with the institutional policy because sanctions are severe (high performance relative to the bureaucratic regime),
- while privately endorsing a professional norm of transparent AI-assisted writing (high internalization relative to the professional regime),
- and experiencing the metric regime (rankings, GPA) as pushing her toward instrumental strategies that undermine learning (low internalization and ambivalent performance relative to the metric regime).

PIC thus invites empirical work that looks not only at "whether" students comply but at *which regime* their compliance or resistance is addressed to.

3.3 Relation to existing theories

PIC intersects with, but does not collapse into, existing constructs:

- It aligns with SDT's distinction between extrinsic and intrinsic motivation but focuses on architectures of evaluation, not on individual trait measures. A more granular mapping onto SDT's regulatory types—external, introjected, identified and integrated—remains a productive avenue for future operationalization, though the present performance-versus-internalization distinction captures the architectural contrast most relevant to institutional diagnosis.
- It resonates with audit-culture critiques that emphasize how measurement can displace substantive judgment, but adds a regime-specific dimension: the same agent may show different PIC profiles relative to different regimes.
- It can be combined with moral-development models but does not assume fixed stages. Instead, it treats internalization as something regimes can enable or inhibit through design.

PIC gives us a way to name what is often invisible: the gap between what an institution measures and what it actually cultivates.

4. ADP: Audit Dependency Profile

4.1 From index to profile

In earlier formulations, the Audit Dependency Index (ADI) was presented as if it were, or could soon become, a numerical measure. In its current stage of development, it is more accurate and methodologically sound to treat it as a qualitative profiling heuristic rather than a validated scalar index. For this reason, this article uses the term Audit Dependency Profile (ADP).

An ADP describes how strongly a local configuration of regimes relies on externalized audit mechanisms to maintain integrity-relevant behavior. It does not produce a single numerical score but a structured description of how dependence on audit is distributed across regimes.

4.2 Dimensions of an Audit Dependency Profile

A minimal ADP can be described along four dimensions, each rated qualitatively (low/medium/high) based on policy analysis, document review and interviews:

- **Legal–bureaucratic dependence:** To what extent does the system rely on formal rules, procedures and sanctions to secure integrity?
- **Metric–audit dependence:** To what extent does the system rely on external audits, rankings, performance indicators and quality assurance mechanisms?
- **Algorithmic dependence:** To what extent does the system rely on automated detection, proctoring and platform-level control?
- **Communal–professional dependence:** To what extent does the system rely on peer norms, professional identities and communities of practice rather than formal audit?

An institution with high legal–bureaucratic, high metric–audit and high algorithmic dependence but low communal–professional dependence exemplifies a highly externalized integrity regime. Conversely, an institution with strong communal–professional norms and minimal metric–audit apparatus may rely more on internalized standards.

4.3 The counterfactual question

The core diagnostic move of ADP can still be expressed as a question:

What would happen to integrity here if the auditors stepped out of the room?

More precisely:

- What would happen to integrity outcomes if plagiarism detection were disabled for a semester?
- How would behavior change if AI-detection risk scores were no longer reported to staff?
- Would students' reasoning about integrity look different in settings where proctoring is impossible?

An educational system that must continuously surveil its students is implicitly declaring low confidence in its own teaching.

High integrity under permanent surveillance is not evidence of internalization. It is evidence of architected dependence.

5. States of Agency

5.1 From continuous motivation to discrete states

SDT models motivation as movement along a continuum from external regulation to intrinsic motivation. This is useful when a single regime dominates and the central question is how deeply its norms are internalized. In a multi-regime integrity field, however, the key phenomenon is often not a gradual shift in motivational style but a qualitative change in the agent's state of agency under conflicting regimes.

Building on the integrity field and on PIC/ADP, we can distinguish five ideal-typical states of agency:

Deliberative engagement: The agent recognizes multiple regimes, reflects on their demands and acts in ways that seek coherence between personal standards and institutional requirements. PIC gaps are low; integrity is both performed and endorsed.

Strategic compliance: The agent complies selectively with regimes that control salient rewards and sanctions, while privately endorsing a different set of norms. PIC gaps are high; performance is driven by calculation rather than endorsement.

Ritual performance: The agent performs required behaviors as empty rituals, without expectation that they are substantively connected to integrity or learning. Audit dependence is high; the agent has stopped asking whether the norms make sense.

Practical paralysis: Conflicting regimes and high-stakes exposure lead to inaction or avoidance. The agent is unable to resolve what counts as "the right thing to do" under incompatible demands.

Exit: The agent withdraws from the integrity field—changing institutions, abandoning the course of study, disengaging from assessment, or in extreme cases, exiting academic life altogether.

These are not personality types. They are positions that agents can occupy at different times and in different contexts within the integrity field.

5.2 States of agency and their theoretical precedents

A reader familiar with organizational sociology will recognize echoes in this taxonomy.

Strategic compliance resembles what Meyer and Rowan [1977] called "decoupling"—the gap between formal structures and actual practice in institutionalized organizations. Ritual performance echoes Strathern's [2000] analysis of how audit regimes produce "performances for the measure" that may have little connection to underlying substance. Exit is, of course, Hirschman's [1970] term for one of the three responses to organizational decline.

The question is fair: what is new here?

The answer lies in *what produces the states* and *what triggers movement between them*.

On decoupling. In Meyer and Rowan's account, decoupling is a stable organizational strategy—a way of managing the tension between legitimacy and efficiency. Organizations maintain formal structures to satisfy external audiences while actual practices diverge to preserve operational functionality. In the integrity field, strategic compliance is decoupling at the level of the individual agent. But where organizational decoupling is a collective strategy, individual decoupling is often a survival mechanism—not a way of managing institutional pressures, but of preserving a self that the institution has not colonized. The student who performs compliance without endorsement is not gaming the system; she is protecting something from it.

On exit, voice, and loyalty. Hirschman [1970] identified exit and voice as the two primary responses to organizational decline, with loyalty determining which path agents choose. High loyalty favors voice—the attempt to change from within. Low loyalty favors exit—departure for alternatives. The integrity field reveals a third possibility that Hirschman's framework does not fully capture: the agent

who can neither exit (costs are prohibitive, credentials are necessary) nor voice (regimes are algorithmic, metrics are "objective," no one is listening) nor remain loyal (the institution has forfeited trust).

This agent does not leave, does not speak, does not believe—but performs.

Strategic compliance is what remains when exit is blocked, voice is futile, and loyalty is dead. It is not a choice between Hirschman's options but the exhaustion of all three.

Moreover, high-ADP configurations systematically eliminate voice. When the regime is algorithmic, to whom does one complain? The algorithm does not listen. When the regime is metric, what does one contest? The number is "objective." When the regime is bureaucratic, what does one say? "Those are the rules." Voice requires an audience capable of response. Audit regimes replace responsive audiences with automated systems and dashboards.

Without voice, only exit or compliance remain. And when exit is blocked—by the necessity of credentials, by sunk costs, by lack of alternatives—compliance becomes the only option. But compliance without loyalty or voice is not commitment. It is strategic compliance at best, ritual performance at worst.

On exit as state, not decision. In Hirschman's framework, exit is a discrete decision—one of three options available to dissatisfied members of an organization. In the integrity field, exit is a *state*—sometimes a dramatic withdrawal, but often a quiet disengagement, a slow retreat from participation, an inner emigration that leaves the body present but the agent absent. Students who stop caring, who do the minimum, who have learned that the game is rigged—they have exited without leaving.

Exit, it turns out, has gradations:

This is exit as internal condition, not external action. The student is physically present, enrolled, submitting work. But the agent—the one who might care, who might believe, who might engage—has departed.

The field-level account. What the taxonomy adds is this: these states are not personality types or strategic choices but *products of architecture*. A student does not "choose" ritual performance; she is *produced* as a ritual performer by a configuration of regimes that rewards performance and ignores—or cannot see—internalization.

And crucially: the transitions between states are not gradual shifts in attitude but *qualitative breaks* triggered by high-stakes evaluation events. The student who was deliberatively engaged may, in the pressure of an exam governed by conflicting regimes, find herself suddenly in strategic compliance—and may never quite return. The architecture does not just constrain behavior; it reshapes the agent.

This is the sense in which the states are "of agency" rather than "of behavior" or "of motivation." They describe not what agents do but *what kind of agent the field allows them to be*.

5.3 Evaluation events and transitions

High-stakes evaluation events—exams, defenses, committee decisions—are not just points where performance is measured. They reconfigure the integrity field by exposing the agent's performance to a collective gaze: peers, committees, external bodies.

In such moments, the dominant question for the agent shifts from "What do I value?" or "What is the rule?" to "How will this appear to those whose judgment matters?"

When the agent cannot construct a workable resolution from conflicting regime demands, transitions from deliberative engagement to strategic compliance, ritual performance, paralysis or exit become more likely.

The framework proposed here treats such transitions as features of the integrity field, not merely as traits of individuals. They are outcomes of how regimes are configured and which audiences are made salient by evaluation design.

6. Case Illustration: A High-ADP Integrity Regime

To show how the integrity field, PIC and ADP work in practice, this section presents a composite case drawn from contemporary university practices.

6.1 The integrity module and the detection stack

At the start of the academic year, all incoming students must complete an online "Academic Integrity and AI Literacy" module. The module combines short videos, interactive activities and multiple-choice quizzes. Students must achieve a passing score to unlock access to core assessments.

On the staff side, the institution has invested in:

- A plagiarism-detection system integrated with the learning management platform.
- A digital proctoring service for remote exams, recording webcam footage, keystrokes and screen activity.
- An AI-detection tool that produces risk scores for text submissions.
- A central dashboard where faculty and administrators can view similarity indices, proctoring flags and AI-risk scores by course, cohort and time.

Cases flagged by these systems are escalated through a standardized workflow. Annual reports summarize incident numbers and outcomes.

6.2 Reading the regime through the integrity field

Through the lens of the integrity field, multiple regimes are operating:

- **Bureaucratic:** The code of conduct, the module, the escalation procedures.
- **Metric/audit:** The dashboard, the annual reports, the KPIs on "integrity incidents."
- **Algorithmic:** The detectors, the proctoring software, the risk scores.
- **Communal:** What students actually tell each other about how to navigate the system.
- **Professional:** Largely absent—integrity is not embedded in discipline-specific reasoning.

The ADP for this configuration:

6.3 PIC analysis

Students quickly learn that what matters is not only producing good work but doing so in ways that do not trigger similarity or AI-risk thresholds. They develop local knowledge about how to paraphrase, vary prompts or adjust workflows to evade detection.

- **Performance relative to bureaucratic/algorithmic regimes:** High (low detection rates)
- **Internalization relative to those regimes:** Low to medium (compliance is strategic)

- **Performance relative to communal regime:** Variable (depends on peer group)
- **Internalization relative to professional regime:** Unknown (rarely tested)

The PIC gap is substantial: the institution observes integrity (low flags) while the underlying orientation may be strategic compliance rather than conviction.

6.4 States of agency in this regime

In this configuration, we would expect:

- Some students in **deliberative engagement** (those whose personal values align with institutional demands).
- Many in **strategic compliance** (navigating the system instrumentally).
- Some in **ritual performance** (going through the motions without belief).
- A few in **practical paralysis** (especially international students facing conflicting norms, or those with disabilities poorly served by proctoring).
- Occasional **exit** (students who find the surveillance intolerable).

The regime produces compliance. It is less clear that it produces integrity.

7. Implications and Research Agenda

7.1 Institutional implications

For institutions, the integrity field framework offers a way to reframe strategic questions. Instead of asking only:

- "Which detection tools should we subscribe to?"
- "What penalties should we impose for AI-assisted cheating?"

it invites questions such as:

- "Which regimes are actually governing our students, and how do they interact?"
- "What is the ADP of our current configuration?"
- "Where are the PIC gaps—high performance with low internalization?"
- "What states of agency are we producing?"
- "Can we redesign assessment so that communal–professional dependence increases while algorithmic dependence decreases?"

This reformulation does not eliminate the need for monitoring or sanctions. It does, however, make visible a choice that is often implicit: whether to invest predominantly in ever more sophisticated auditors or in architectures that require less auditing to function.

7.2 Agenda for empirical research

The framework generates specific research questions:

Comparative configuration analysis: How do different configurations of regimes correlate with distinct Audit Dependency Profiles? (Design: qualitative comparative analysis of institutional policies, coded into ADP descriptors.)

States of agency under evaluation: Under what conditions do high-audit configurations produce shifts from deliberative engagement to strategic compliance, ritual performance, paralysis or exit? (Design: longitudinal studies following cohorts through key evaluation events.)

Operationalizing PIC across regimes: Can qualitative case studies be systematically coded into PIC descriptors that distinguish performance and internalization across multiple regimes? (Design: interview-based studies where agents describe how they navigate conflicting demands.)

Intervention studies: What happens to ADP and PIC when institutions shift from detection-heavy to process-oriented assessment designs?

7.3 On evidence and the status of this framework

This paper offers no empirical data. This is a limitation, and it should be named clearly rather than excused.

The composite case in Section 6 is illustrative—drawn from patterns visible across many institutions but not documented as a formal case study. The states of agency are ideal types—conceptual constructs intended to illuminate, not empirical categories validated through observation. The ADP is a heuristic for structuring questions, not a measurement instrument.

Why, then, publish?

The answer is architectural. Before we can measure something, we need to know what we are looking for. Before we can design interventions, we need a language for describing the problem. The empirical literature on academic integrity is vast and growing; what it often lacks is a framework capable of integrating findings that span detection technologies, student motivation, institutional policy, and professional ethics.

Consider what existing frameworks miss:

- **SDT** can measure motivation but cannot model situations where multiple regimes demand contradictory responses.
- **Audit culture analysis** can critique surveillance but cannot distinguish which specific configurations produce which effects.
- **Organizational theory** describes decoupling at the institutional level but does not trace how it manifests in individual agents—or what it costs them.
- **Hirschman's framework** identifies exit, voice, and loyalty as options but does not model what happens when all three are blocked.

The integrity field is proposed as scaffolding that connects these perspectives—not as a finished theory but as a framework for inquiry. The test is not whether it is "true" in some abstract sense but whether it proves *useful*: whether it generates questions that would not otherwise be asked, whether it reveals patterns that would otherwise be invisible, whether it helps institutions see what they are doing.

Section 7.2 outlines a research agenda. The agenda is not decorative. It describes what would be needed to move from heuristic to validated framework:

- Comparative configuration analysis: coding institutional policies into ADP profiles and testing whether configurations cluster in predictable ways.
- Longitudinal studies: tracking cohorts through evaluation events and documenting shifts in states of agency.
- Intervention studies: testing whether changes in assessment design produce changes in PIC gaps.

These studies are possible. They require time, access, and resources that the present paper does not claim to have mobilized. The framework is offered in the hope that others—with different institutional positions, different methodological skills, different access to data—will find it worth testing.

There is also a harder question beneath the methodological one: *can* internalization be measured? PIC assumes that we can, in principle, distinguish performed compliance from genuine endorsement. But endorsement does not announce itself. It is inferred, interpreted, contested. And as Foucault reminds us, "genuine" endorsement may itself be a product of successful subjectification—the subject who believes most sincerely may be the most thoroughly governed.

This paper does not resolve that difficulty. It does, however, insist that the difficulty is real and consequential. An institution that cannot distinguish compliance from conviction—that can only see what auditors see—is an institution that does not know what it is producing. The framework makes that blindness visible. Whether it can be cured is an open question.

7.4 Limitations

Several limitations should be noted.

First, the integrity field, PIC and ADP are diagnostic tools, not panaceas. They do not specify the values institutions should internalize, nor do they resolve conflicts between competing norms.

Second, ADP is a qualitative heuristic, not a validated psychometric scale. Its value is in structuring comparison and generating hypotheses, not in producing precise measurements.

Third, the states-of-agency taxonomy is ideal-typical. Real agents may occupy hybrid or transitional positions.

Finally, the composite case is illustrative rather than representative. Many institutions already experiment with approaches that move toward lower audit dependence. The framework can help document and compare these efforts.

8. Conclusion

Academic integrity regimes are increasingly built as audit systems. Generative AI has accelerated this trend by making it easier to delegate work and harder to distinguish human from machine-generated outputs [Perkins, 2023; Cotton et al., 2023; Kofinas et al., 2025]. Universities have responded with new instruments of detection and control, often without revisiting the architectures of assessment and feedback that make such instruments necessary [Bittle & El-Gayar, 2025].

This article has proposed an integrity field framework—a conceptual architecture in which multiple normative regimes jointly govern the same agent. Within this field, PIC makes visible the gap between performed compliance and genuine endorsement. ADP profiles how strongly a configuration relies on external audit. The states of agency taxonomy names what happens to agents caught between regimes: deliberative engagement, strategic compliance, ritual performance, paralysis, exit.

On values

This framework is diagnostic. It describes what configurations of regimes do to agents; it does not prescribe what configurations institutions should prefer.

But description is never innocent. The language of "audit dependency" implies that dependence is a problem. The distinction between performance and internalization implies that internalization matters. The taxonomy of states implies that some states—deliberative engagement—are preferable to others.

These implications are intended. They reflect a normative commitment that should be made explicit: *an educational institution that can only produce integrity under surveillance has failed at education.*

But there is a deeper danger that must also be named.

Foucault [1991] showed that modern governance operates not primarily through prohibition but through production—producing subjects who freely want what the system needs them to want. The most governed subject is not the one who obeys under duress but the one who has internalized the regime's norms so completely that compliance feels like autonomy.

This complicates any simple celebration of internalization. High audit dependency is visible and resistible—students know they are being watched and can push back. The greater danger may be low audit dependency achieved through successful subjectification: students who have internalized institutional norms so thoroughly that they no longer recognize them as external. They perform integrity not because they are watched but because they have become subjects for whom integrity—as defined by the regime—is simply "who they are."

This is not autonomy. It is governance that has colonized the self.

The framework cannot resolve this tension. It can only insist that the content of what is internalized matters—that high PIC alignment with a surveillance regime is not the same as high PIC alignment with a community of practice, and that producing compliant subjects at scale may be more dangerous than producing visible resisters.

This does not mean that monitoring has no place. Detection systems may deter some misconduct. Proctoring may secure some exams. Risk scores may flag some cases worth investigating. The question is not whether these tools are ever useful but whether they have become *structurally central*—whether the architecture assumes that integrity will not exist without them.

A high-ADP configuration is not automatically wrong. There may be contexts—high-stakes professional licensing, security-sensitive assessments—where externalized assurance is appropriate. But a university that defaults to high ADP across all assessments, for all students, in all contexts, has made a choice about what kind of institution it wants to be. The framework makes that choice visible.

What would a low-ADP configuration look like? It would rely more heavily on:

- Assessment designs that make cheating irrational rather than merely risky (process-based, iterative, personalized)
- Professional communities that model and transmit integrity through practice rather than compliance
- Feedback cultures that make learning visible without requiring surveillance
- Institutional trust that students are capable of judgment—and designs that develop that capacity rather than assume its absence

Whether institutions choose to move in this direction is, ultimately, a question of values. The framework cannot answer it. But it can make the question unavoidable.

The stakes

Together, these tools shift the focus from catching more cheaters to asking a prior question: what kind of integrity are we actually building here—and what kind of agents are we producing?

There is a generation being formed now. They will learn that integrity means not getting caught—or they will learn that integrity names something that persists when no one is watching. They will develop capacities contingent on surveillance—or capacities that are their own. They will become performers skilled at satisfying auditors—or agents capable of judgment.

No framework can resolve this tension. But the integrity field names what is at stake, and it makes it harder to pretend that the choice is merely technical.

References

- Albanese, M. A., & Sabers, D. L. (1988). Multiple true-false items: A study of interitem correlations, scoring alternatives, and reliability estimates. *Journal of Educational Measurement*, 25(2), 111–123.
- Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296.
- Boltanski, L., & Thévenot, L. (2006). *On justification: Economies of worth* (C. Porter, Trans.). Princeton University Press. (Original work published 1991)
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). Harvard University Press.
- Bourdieu, P., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. University of Chicago Press.
- Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: A Bayesian item response model comparison. *International Journal of STEM Education*, 6(1), 1–16.
- Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33–63.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.
- Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & M. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 87–104). University of Chicago Press.
- Gardner-Medwin, A. R. (2006). Confidence-based marking: Towards deeper learning and better exams. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 141–149). Routledge.
- Han, S., Nikou, S., & Ayele, W. Y. (2024). Digital proctoring in higher education: A systematic literature review. *International Journal of Educational Management*, 38(1), 265–285.
- Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.
- Karlsson, N., & Lunander, A. (2025). A continuous scoring function for confidence-based marking (Working Paper No. 11/2025). Örebro University.
- Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*. Advance online publication.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), Article 7.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.
- Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. Doubleday.

Shore, C., & Wright, S. (2015). Audit culture revisited: Rankings, ratings, and the reassembling of society. *Current Anthropology*, 56(3), 421–444.

Strathern, M. (2000). The tyranny of transparency. *British Educational Research Journal*, 26(3), 309–321.

Version History

- **v1.0.0** (2025-06-01): Initial release
- **v1.1.0** (2025-12-02): Added theoretical contextualization (Section 2.3 on Bourdieu, Foucault, Boltanski & Thévenot; Section 5.2 expanded with Hirschman's exit/voice/loyalty framework and gradations of exit; Section 7.3 on evidence and framework status; expanded conclusion with explicit normative position and Foucauldian warning on subjectification)