The Hidden Curriculum of AI Misuse at the Universities*

Artem Artyukhov[1,2,3,*,†], Oleksandr Dluhopolskyi[3,4,†], Dmytro Chumachenko[5,6,†], Vadim Schuchmann[4,†] and Andrii Butnar[4,†]

[1] Bratislava University of Economics and Business, Underground Road, 1, Bratislava, 85235, Slovak Republic
[2] Sumy State University, Kharkivska St., 116, Sumy, 40007, Ukraine
[3] WSEI University, Projektowa St., 4, Lublin, 20209, Republic of Poland
[4] West Ukrainian National University, Lvivska St., 11, Ternopil, 46020, Ukraine
[5] University of Warwick, Coventry, CV4 7AL, United Kingdom
[6] National Aerospace University "Kharkiv Aviation Institute", Vadym Manko St., 17, Kharkiv, 61070, Ukraine

Abstract

The rapid diffusion of generative artificial intelligence (AI) in higher educa-tion creates both new learning opportunities and significant challenges for academic integrity. This study examines AI misuse through the concept of the hidden curriculum, understood as the set of implicit norms and institu-tional signals that shape student behavior beyond formal policies. The article proposes a conceptual framework that models AI misuse as a dynamic pro-cess developing across multiple educational states, including legitimate en-gagement, strategic ambiguity, invisible delegation, normalization, cognitive dependency, and integrity-oriented correction. To formalize these dynamics, the study introduces a discrete-time Markov model that describes transitions among 11 states of AI engagement. In addition, two analytical indices, the Learning Integrity Reality Index (LIRI) and the Hidden Curriculum Risk In-dex (HCRI), are proposed to evaluate structural risks in educational envi-ronments. The results highlight the importance of integrity-by-design ap-proaches that reduce ambiguity, make reasoning visible, and promote re-sponsible AI use in higher education.

Keywords

AI, academic integrity, hidden curriculum, higher education, Markov model

1. Introduction

Artificial intelligence (AI) has rapidly moved from a specialized technological tool to an everyday component of higher education. Students use AI systems to generate ideas, structure arguments, improve language, summarize sources, and produce complete assignments. In parallel, universities increasingly integrate AI into teaching, administration, and assessment. This expansion creates new educational opportunities but also introduces new academic risks. The problem is no longer limited to plagiarism in its traditional sense. Instead, higher education is confronted with more complex forms of cognitive substitution, concealed delegation, and norm ambiguity, in which AI can perform substantial intellectual work while preserving the appearance of student authorship.

A particularly important challenge is that AI misuse may not arise solely from deliberate individual misconduct. It may also emerge from the university's hidden curriculum. Official policies may promote academic integrity and responsible AI use. Yet, assessment design, instructor inconsistency, peer tolerance, and weak verification procedures may simultaneously lead students to believe that problematic uses of AI are acceptable, strategically useful, or effectively invisible. In this sense, the university may communicate two competing lessons at once: a formal lesson about integrity and an informal lesson about how to succeed under conditions of ambiguity.

* The Third International Conference of Young Scientists on Artificial Intelligence for Sustainable Development: Intelligent Systems and Artificial Intelligence (YAISD-WS 2026:ISAI) May 7-8, 2026, Ternopil, Ukraine
[1,*] Corresponding author.
[†] These authors contributed equally.
artem.artiukhov@euba.sk (A. Artyukhov); dlugopolsky77@gmail.com (O. Dluhopolskyi); dichumachenko@gmail.com (D. Chumachenko); vadum.shuhmann@gmail.com (V. Schuchmann); a.butnar@st.wunu.edu.ua (A. Butnar)
0000-0003-1112-6891 (A. Artyukhov); 0000-0002-2040-8762 (O. Dluhopolskyi); 0000-0003-2623-3294 (D. Chumachenko); 0000-0002-1427-3312 (V. Schuchmann); 0009-0004-6455-4683 (A. Butnar)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-11

CEUR-WS.org/Vol-4326/paper10.pdf

This article approaches AI misuse as a dynamic institutional phenomenon rather than as a static ethical category. The central assumption is that problematic AI use develops through repeated transitions across states of engagement, ambiguity, delegation, normalization, dependency, and correction. Such movement is shaped by educational structures, not only by personal intention. Accordingly, the study aims to develop a conceptual model that explains how hidden norms and institutional arrangements influence the likelihood of movement toward either responsible AI use or entrenched misuse.

To make this objective more explicit, the study pursues four interrelated tasks. First, it conceptualizes AI misuse as a hidden curriculum phenomenon shaped by implicit institutional signals rather than solely by individual intention. Second, it formalizes this phenomenon as a discrete-time Markov process with eleven states that capture movement across legitimate engagement, ambiguity, delegation, normalization, dependency, correction, and responsible use. Third, it operationalizes the structural context of these transitions through two analytical indices, the Learning Integrity Reality Index (LIRI) and the Hidden Curriculum Risk Index (HCRI). Fourth, it compares three stylized governance scenarios (permissive ambiguity, reactive control, and integrity-by-design) to demonstrate how different institutional configurations produce different long-run educational outcomes.

2. Literature review

The rapid diffusion of generative AI tools has significantly transformed learning practices in higher education. Systems such as large language models enable stu-dents to generate text, summarize literature, and structure arguments with mini-mal effort, creating new pedagogical opportunities while simultaneously raising concerns about academic integrity and authorship [1; 2; 3]. Recent studies sug-gest that AI-assisted writing challenges traditional understandings of plagiarism because AI systems can produce original yet externally generated content that may obscure the learner's actual cognitive contribution [4; 5]. As a result, uni-versities increasingly face difficulties distinguishing legitimate AI-supported learning from concealed delegation of intellectual work.

A growing body of literature frames the issue not only as misconduct but also as a systemic challenge related to institutional design and assessment practices. Scholars argue that product-oriented assessment systems and unclear policies regarding AI use create environments in which students experiment with AI tools without clear boundaries for acceptable practice [6; 7]. In such contexts, academ-ic integrity becomes difficult to enforce because technological capabilities evolve faster than institutional governance mechanisms. Consequently, universi-ties must move beyond detection-based approaches and develop pedagogical strategies that emphasize transparency, reasoning processes, and responsible AI use.

Another important perspective in literature concerns the hidden curriculum, which refers to informal norms and implicit institutional signals that shape stu-dent behavior beyond official rules [8; 9]. Within AI-rich learning environments, hidden curricula may emerge when formal policies emphasize integrity while everyday academic practices reward polished outputs rather than visible reason-ing. Under such conditions, students may learn that AI-generated work can be strategically used without detection, thereby normalizing ambiguous or unethical practices. Recent studies on AI in education, therefore, emphasize the importance of integrating integrity-by-design principles into teaching and assessment to en-sure that AI tools support learning rather than replace intellectual effort [10; 11].

At the same time, researchers warn that the most significant risk of generative AI in education may not be limited to cheating but may involve long-term cogni-tive offloading, where learners increasingly rely on AI systems for tasks previ-ously associated with human reasoning and analysis [12; 13]. This shift raises concerns about the development of critical thinking, epistemic responsibility, and academic agency. Consequently, contemporary scholarship increasingly conceptualizes academic integrity in AI-mediated environments as a problem of institutional governance and learning design, rather than merely individual mis-conduct.

These debates indicate the need for analytical frameworks that explain how AI misuse develops within institutional contexts. Conceptual models that treat mis-use as a dynamic process, shaped by ambiguity, peer norms, assessment struc-tures, and corrective interventions, may provide a more realistic understanding of how AI practices evolve in higher education systems. Such approaches are par-ticularly relevant for analyzing how universities may unintentionally reproduce problematic AI use through hidden curricular signals embedded in educational environments.

3. Methodology

This study employs a conceptual-analytical methodology grounded in formal modeling and educational interpretation. The purpose is not to estimate behavioral probabilities empirically at this stage, but to construct a theoretically coherent framework that explains how AI misuse can emerge, stabilize, and be redirected within higher education. The methodological design combines three layers: state-based conceptualization, stochastic modeling, and index-based operationalization.

The first layer is state identification. The hidden curriculum of AI misuse is decomposed into a set of qualitatively distinct educational states that represent different positions within the relationship among user behavior, institutional rules, and pedagogical practice. These states include legitimate engagement, strategic ambiguity, invisible delegation, polished dishonesty, normative mimicry, collective normalization, cognitive dependency, integrity-oriented correction, and responsible AI use. In the extended version of the model, additional states, such as exploratory AI assistance and entrenched misuse, are included to reflect the branching and nonlinear nature of transition pathways. This state structure enables the phenomenon to be interpreted as a dynamic sequence rather than as a binary distinction between honest and dishonest conduct.

The second layer is Markov modeling. The article specifies the hidden curriculum of AI misuse as a discrete-time Markov process in which the probability of entering a future state depends on the current state and the educational conditions associated with it. This formulation is appropriate because the normalization of AI misuse is not assumed to be a one-time event; it emerges through repeated interactions with policy ambiguity, output-oriented assessment, peer behavior, and integrity-oriented interventions. The state vector and transition matrix formalize these dynamics.

The third layer is index-based operationalization. To complement the Markov chain's dynamic logic, the study introduces two aggregate indices that capture measurable properties of the educational environment. The Learning Integrity Reality Index (LIRI) measures the extent to which student reasoning remains visible, attributable, and verifiable in assessment settings. The Hidden Curriculum Risk Index (HCRI) measures the extent to which ambiguity, concealed delegation, social tolerance, and dependency are structurally embedded in the learning environment. These indices do not replace the Markov model; rather, they provide a cross-sectional way of estimating the contextual conditions that influence transition probabilities. In this sense, the methodology combines dynamic explanation with structural diagnosis.

To reduce the “black box” character of the two indices, this study specifies their baseline operationalization as normalized weighted composites. The Learning Integrity Reality Index (LIRI) captures the degree to which student reasoning remains visible, attributable, and pedagogically verifiable. In its baseline form, it may be written as:

$$LIRI = \sum_{i=1}^{5} w_i x_i, \sum_{i=1}^{5} w_i = 1, \tag{1}$$

where $x_1$is reasoning visibility (the extent to which intermediate reasoning steps are observable), x_2 is authorship traceability (the extent to which the student’s contribution can be attributed and reconstructed), $x_3$is process verifiability (the availability of drafts, oral defense, logs, or other process evidence), $x_4$is disclosure compliance (the degree to which AI use is transparently declared), and $x_5$is independent reconstruction capacity (the ability of the learner to reproduce and defend the submitted reasoning without AI support). Each variable is normalized to the $[0,1]$

interval, where higher values indicate stronger learning integrity conditions. In the theoretical version of the model, equal weights $(w_i = 0.20)$ may be used as a baseline, while future empirical work may estimate discipline-specific weights.

The Hidden Curriculum Risk Index (HCRI) captures the extent to which the educational environment structurally promotes ambiguous, concealed, or dependency-forming AI use. In baseline form, it may be written as:

$$HCRI = \sum_{j=1}^{5} v_j z_j, \sum_{j=1}^{5} v_j = 1, \quad (2)$$

where z_1 is policy ambiguity, z_2 is a concealed delegation opportunity, $z_3$is social tolerance of misuse, z_4 is output-oriented assessment pressure, and $z_5$is cognitive offloading pressure. Each variable is normalized to the $[0,1]$ interval, where higher values indicate a more risk-prone hidden curriculum. As with LIRI, equal weights $(v_j = 0.20)$ are used here as a baseline specification.

An illustrative example clarifies the application's logic. If an institution has $x = (0.80, 0.70, 0.60, 0.75, 0.65)$, then $LIRI = 0.70$, which indicates a moderately strong integrity-oriented learning environment. If the same institution has $z = (0.60, 0.70, 0.55, 0.80, 0.65)$, then $HCRI = 0.66$, indicating a relatively high level of hidden-curriculum risk. These values are illustrative rather than empirical, but they demonstrate how the two indices can be used to diagnose the structural conditions that influence transition probabilities in the Markov model.

To make the model analytically interpretable, the study also proposes derived measures, including the abuse accumulation measure (AAM) and the responsible use measure (RUM), which summarize the proportion of the system occupying problematic versus corrective or responsible states over time. In addition, three stylized scenarios (permissive ambiguity, reactive control, and integrity-by-design) are used to demonstrate how different institutional configurations produce different transition structures and long-run outcomes. These scenarios do not function as empirical cases but as theoretical stress tests of the model, showing how different governance regimes alter the balance between misuse and responsibility.

The methodological strategy is designed to transform the hidden curriculum of AI misuse from a largely normative concern into a formalizable and governable educational problem. By integrating state logic, probabilistic transitions, and operational indices, the framework enables future empirical calibration while already providing a robust conceptual lens for understanding how universities may unintentionally teach, stabilize, or prevent problematic uses of AI.

4. Markov model of the hidden curriculum of AI misuse

To explain how AI misuse becomes normalized in higher education, it is useful to move beyond static typologies and represent the phenomenon as a dynamic sequence of state transitions. The hidden curriculum of AI misuse does not emerge as a single isolated act of cheating; rather, it develops through repeated interactions between users, institutional rules, assessment design, and peer norms. A Markov model is appropriate here because it captures the probabilistic nature of these repeated choices and shows how educational environments can increase or decrease the likelihood that users move from legitimate engagement toward misuse, normalization, and dependency. The purpose of the Markov model is therefore not merely illustrative. It formalizes a central argument of this study: universities may unintentionally teach AI misuse by creating conditions in which ambiguous rules, output-oriented assessment, and weak verification systems make problematic AI use more attractive, rational, and eventually normal. In this sense, the model serves as an analytical bridge between individual behavior and institutional design. It shows that misuse is not simply the result of personal dishonesty but a predictable outcome of educational environments that reward polished performance over transparent reasoning.

Fig. 1 represents AI misuse as a dynamic system of repeated transitions between legitimate engagement, ambiguity, concealed delegation, normalized misuse, and integrity-oriented correction. It illustrates how unclear rules, product-oriented assessment, peer tolerance, and repeated cognitive offloading can increase the probability of movement toward collective normalization and cognitive dependency, while targeted pedagogical interventions can redirect the system toward responsible AI use.

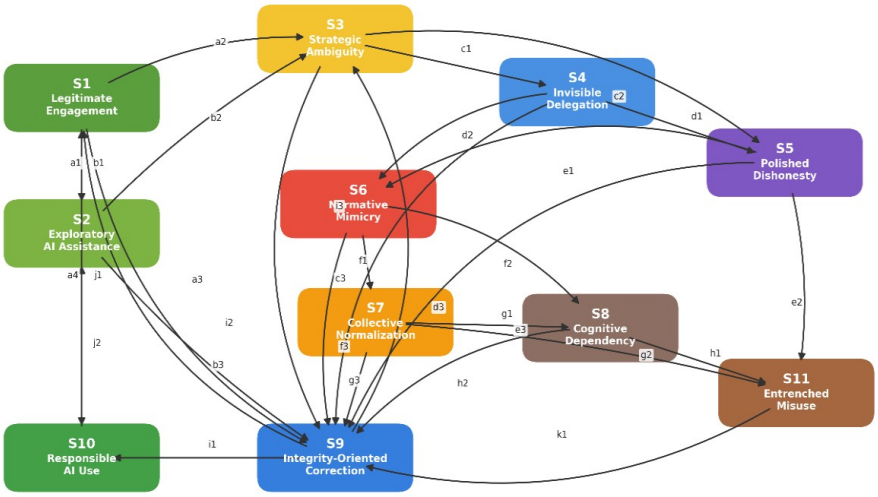


Figure 1: Markov model of the hidden curriculum of AI misuse in higher education.

The model begins with S1 (legitimate engagement), where the learner uses AI minimally, transparently, or not at all within academically acceptable boundaries. From this state, three main transitions are possible. The learner may remain in legitimate engagement, move to S2 (exploratory AI assistance) when AI is used experimentally for brainstorming or clarification, or shift toward S3 (strategic ambiguity) when institutional rules are unclear and the boundaries of acceptable use become negotiable. Under stronger pedagogical guidance, movement from S1 may also lead directly to S9 (integrity-oriented correction) or, in well-structured environments, to S10 (responsible AI use).

From S2, the system may return to S1 if AI remains bound and reflective, or it may progress to S3 if exploratory use develops under inconsistent expectations. Corrective intervention may also redirect this state toward S9 (integrity-oriented correction).

S3 functions as the main gateway state. From here, the learner may move toward S4 (invisible delegation), where AI begins to perform substantive intellectual work, or toward S5 (polished dishonesty), where AI-generated fluency conceals weak understanding. If institutional intervention occurs, the learner may be redirected toward S9.

From S4, three transitions are especially important. The state may deepen into S5, evolve into S6 (normative mimicry), or be interrupted by S9. Similarly, S5 may progress into S6, especially when polished output is rewarded more than visible reasoning. In weaker governance environments, S5 may also drift directly into S11 (entrenched misuse).

S6 lies in the central middle layer of the graph. Here, the learner appears formally compliant while still engaging in substantive misuse. This state may transition into S7 (collective normalization) when peers and instructors silently accept such practices, into S8 (cognitive dependency) when repeated offloading weakens autonomy, or into S9 if misuse is pedagogically surfaced and challenged.

From S7, the system may proceed to S8 because socially normalized offloading reduces the incentive to engage in independent reasoning. It may also harden into S11, especially when institutional signals reward performance over understanding. Strong interventions may still redirect the learner from S7 toward S9.

S8 represents the long-term developmental consequence of repeated AI offloading. From this state, movement toward S11 becomes more likely if no pedagogical recovery occurs. However, targeted intervention may still redirect the learner into S9.

S9 serves as the principal recovery state. It may lead back to S1 when the learner re-establishes independent reasoning, or forward to S10 when AI becomes transparently and critically integrated into learning. Under weak or inconsistent support, relapses from S9 to S3 remains possible.

S10 is the desired stabilizing state. From here, the system may remain stable or, in some contexts, return to S1 depending on task demands. In poorly governed environments, however, even responsible use may regress toward S3 if institutional signals again become inconsistent.

Finally, S11 is the most problematic end state. It represents habitual, normalized, and resilient misuse. In a weak institutional environment, this state may persist over time. Only a strong, explicit pedagogical intervention can redirect the learner from S11 back to S9.

These states do not represent moral labels alone; they represent educationally meaningful positions within a learning system. Some states are transitional and unstable, while others may function as attractors depending on how the institution structures incentives and evaluation.

To formalize the hidden curriculum of AI misuse, the model represents learner movement across eleven discrete states as a time-homogeneous Markov process. Let the probability distribution over the state space at time t be given by the row vector:

$$x_t = \left[x_1(t), x_2(t), x_3(t), x_4(t), x_5(t), x_6(t), x_7(t), x_8(t), x_9(t), x_{10}(t), x_{11}(t) \right] \tag{3}$$

The vector satisfies the standard normalization condition:

$$\sum_{i=1}^{11} x_i(t) = 1, x_i(t) \geq 0. \tag{4}$$

The system evolves according to the Markov update rule:

$$x_{t+1} = x_t P \tag{5}$$

where $P = \left[p_{ij} \right]$ is the 11x11 transition matrix.

Each row of P must satisfy:

$$\sum_{j=1}^{11} p_{ij} = 1, p_{ij} \in [0,1]. \tag{6}$$

Using the transition labels shown in the graph, the matrix may be written as:

$$\mathbf{P} = \begin{bmatrix} 1-a_1-a_2-a_3-a_4 & a_1 & a_2 & 0 & 0 & 0 & 0 & 0 & a_3 & a_4 & 0 \\ b_1 & 1-b_1-b_2-b_3 & b_2 & 0 & 0 & 0 & 0 & 0 & b_3 & 0 & 0 \\ 0 & 0 & 1-c_1-c_2-c_3 & c_1 & c_2 & 0 & 0 & 0 & c_3 & 0 & 0 \\ 0 & 0 & 0 & 1-d_1-d_2-d_3 & d_1 & d_2 & 0 & 0 & d_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-e_1-e_2-e_3 & e_1 & 0 & 0 & e_3 & 0 & e_2 \\ 0 & 0 & 0 & 0 & 0 & 1-f_1-f_2-f_3 & f_1 & f_2 & f_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-g_1-g_2-g_3 & g_1 & g_3 & 0 & g_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-h_1-h_2 & h_2 & 0 & h_1 \\ i_2 & 0 & i_3 & 0 & 0 & 0 & 0 & 0 & 1-i_1-i_2-i_3 & i_1 & 0 \\ j_1 & j_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-j_1-j_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_1 & 0 & 1-k_1 \end{bmatrix}. \tag{7}$$

This formulation reflects the logic of the redrawn graph. Some states contain more branching possibilities than others, which is theoretically meaningful. For example, S_3 is modeled as a gateway state with three outgoing transitions, whereas S_8) has only two principal directions: further deterioration into entrenched misuse or redirection into integrity-oriented correction.

To facilitate interpretation at the system level, three aggregate measures may be defined. AAM captures the share of the system occupying problematic states:

$$M_t = x_4(t) + x_5(t) + x_6(t) + x_7(t) + x_8(t) + x_{11}(t). \quad (8)$$

A RUM captures the share of the system oriented toward recovery or stable responsible practice:

$$R_t = x_9(t) + x_{10}(t). \quad (9)$$

Finally, a hidden curriculum pressure measure captures the extent to which ambiguity, mimicry, and normalization dominate the environment:

$$H_t = x_3(t) + x_6(t) + x_7(t) \quad (10)$$

These measures do not replace the state model, but they provide compact summaries of how the educational environment evolves under different institutional conditions.

The transition parameters in the 11-state model represent institutionally mediated probabilities rather than purely individual preferences. Each parameter reflects the extent to which assessment design, policy clarity, faculty behavior, peer norms, and verification mechanisms move toward states that are more or less likely. In this way, the parameters operate the hidden curriculum as a system of structured opportunities, constraints, and reinforcements.

The a-parameters describe the earliest branching conditions from S_1. Parameter $a_1$captures the probability that students move from legitimate engagement to S_2, typically in environments where AI is available and pedagogically tolerated as a low-risk support tool. Parameter $a_2$represents a drift from legitimate use into S_3, which becomes more likely when institutional rules are unclear or inconsistently enforced. Parameter $a_3$captures movement from legitimate engagement into S_9, reflecting settings where reflection, disclosure, or instructor feedback are built into the learning process even before misuse becomes visible. Parameter a_4 represents direct stabilization in S_{10}, which is most likely when expectations are explicit and AI use is transparently scaffolded.

The b-parameters govern movement from S_2. Parameter $b_1$represents a return to S_1, indicating that exploratory use remains bound and does not destabilize authorship or reasoning. Parameter b_2 captures the probability that exploratory experimentation becomes S_3, especially when students begin to test the limits of what is permitted. Parameter $b_3$reflects movement into S_9, which may occur when low-risk experimentation is accompanied by reflective prompts, process checks, or discussions of acceptable use.

The c-parameters are particularly important because they define the outflows from S_3, the model's key gateway state. Parameter $c_1$represents a transition to S_4, where AI begins to perform substantive cognitive work without disclosure. Parameter $c_2$represents movement to S_5, where AI-generated fluency conceals weak understanding. Parameter $c_3$captures the extent to which ambiguity is interrupted by S_9, making it a direct indicator of whether institutions can respond constructively before misuse deepens.

The d-parameters govern transitions from S_4. Parameter $d_1$measures the extent to which concealed substitution develops into S_5, as delegated work becomes rhetorically polished and increasingly difficult to distinguish from authentic student performance. Parameter $d_2$represents movement into S_6, where students learn to maintain the appearance of compliance while preserving substantive misuse. Parameter $d_3$represents corrective interruption and therefore measures the effectiveness of assessment designs that can surface hidden delegation.

The e-parameters describe the possible outcomes of S_5. Parameter $e_1$captures progression into S_6, where polished misuse becomes normalized through formal appearances of compliance. Parameter e_2 represents direct movement into S_{11}, indicating cases where concealed AI

substitution becomes habitual and resistant to weak educational controls. Parameter e_3 captures the transition to S_9, which depends on whether instructors can see through polished output and redirect the student toward accountable practice.

The f-parameters govern the transitions from S_6. Parameter $f_1$measures the probability of movement toward S_7, where peers and instructors silently accept problematic AI use as ordinary. Parameter f_2 measures transition into S_8, reflecting the developmental effect of repeated offloading. Parameter f_3 captures movement into S_9, showing whether performative compliance can still be pedagogically challenged before it becomes culturally embedded.

The g-parameters define the outflows from S_7. Parameter $g_1$represents progression to S_8, since normalized misuse reduces incentives for independent cognitive effort. Parameter $g_2$represents movement into S_{11}, where the misuse pattern becomes stable and self-reinforcing. Parameter g_3 captures the possibility that even socially normalized misuse can be interrupted if the institution acts decisively.

The h-parameters correspond to S_8, which is the most important addition in the 11-state model. Parameter $h_1$measures the probability that long-term dependency hardens into S_{11}, while parameter $h_2$captures the possibility of recovery through S_9. These parameters are especially important because they show that the greatest risk of AI misuse is not only dishonest performance but also the weakening of epistemic autonomy itself.

The i-parameters define the recovery logic of S_9. Parameter $i_1$measures the probability of transition into S_{10}, which is the preferred long-term outcome of successful correction. Parameter i_2 captures a more conservative recovery path back to S_1, where AI use is reduced or re-bounded. Parameter $i_3$represents a relapse into S_3, indicating that correction has occurred, but institutional conditions remain unclear enough for misuse to re-emerge.

The j-parameters govern outflows from S_{10}. Parameter $j_1$represents a return to S_1, especially in contexts where AI is not needed or where transparent use is replaced by fully independent work. Parameter $j_2$captures movement to S_2, indicating bounded and low-risk experimentation within a responsible framework. These parameters model responsible AI use not as a rigid endpoint, but as a stable yet flexible educational condition.

Finally, the k-parameter governs the only modeled recovery pathway from S_{11}. Parameter k_1 represents a transition into S_9. Its value is expected to be low in weakly governed systems and higher only where institutions deploy strong, explicit, and sustained interventions. Its presence in the model is analytically important because it prevents entrenched misuse from being treated as absolutely irreversible, while still recognizing that recovery at this stage is substantially more difficult than earlier correction.

Taken together, these parameters show that the hidden curriculum of AI misuse is best understood as a probability structure shaped by educational design. Lower values of a_2, c_1, c_2, f_1, f_2, g_2, and h_1, combined with higher values of a_4, c_3, d_3, e_3, f_3, g_3, h_2, and i_1, describe a more integrity-oriented educational environment. Conversely, when ambiguity, concealment, social tolerance, and dependency are weakly challenged, the model predicts drift toward normalization and entrenched misuse. In this sense, the transition parameters do not simply describe student behavior; they encode the pedagogical and institutional logic through which universities either stabilize responsible AI use or reproduce hidden trajectories of misuse.

Although the state structure of the model remains invariant across disciplines, the transition parameters are expected to vary by field of study. In STEM disciplines, transitions from legitimate engagement to exploratory AI assistance $(S1 \rightarrow S2)$ may be relatively more frequent because AI is often used for explanation, debugging, code completion, and procedural support. At the same time, transitions from ambiguity or concealed delegation to correction $(S3 \rightarrow S9,\ S4 \rightarrow S9)$ may also be stronger in settings where solutions can be validated through oral problem-solving, laboratory replication, code review, or stepwise demonstration of reasoning. However, in programming-intensive take-home assignments, the transition from ambiguity to invisible delegation $(S3 \rightarrow S4)$ may increase because AI can generate substantial portions of code while preserving the appearance of authorship.

In Humanities and essay-based social sciences, transitions from strategic ambiguity to polished dishonesty $(S3 \rightarrow S5)$ and from polished dishonesty to normative mimicry $(S5 \rightarrow S6)$ may be comparatively stronger, because fluent AI-generated prose can more easily conceal weak conceptual ownership when assessment relies primarily on final written output. In such settings, transitions back to correction or responsible use depend more strongly on draft-based assessment, reflective commentary, oral defense, source annotation, and transparent disclosure of AI involvement. For this reason, the model assumes discipline-invariant states but discipline-sensitive transition matrices. Future empirical work should therefore calibrate these probabilities separately for STEM, Humanities, and other field clusters rather than treating one transition structure as universally valid (Table 1). Table 1 does not assign empirical probabilities, but illustrates how the same state architecture may produce different transition tendencies depending on disciplinary assessment formats and verification mechanisms.

The model makes a couple of analytical points especially clear. AI misuse is not best understood as a binary event (cheating versus not cheating) but as a trajectory. Users move gradually from one form of engagement to another, often through ambiguous intermediate states that appear harmless at first. The model shows that institutional design matters: transition probabilities are not fixed personal traits but are shaped by assessment formats, policy clarity, faculty signaling, and peer culture. The model reveals that the most dangerous outcome is not only dishonesty in the narrow sense, but also the long-term movement toward collective normalization and cognitive dependency, in which misuse becomes both socially acceptable and developmentally consequential.

Table 1.
Illustrative disciplinary variation in selected transition tendencies.

Transition	STEM tendency	Humanities tendency	Interpretation
$S1 \rightarrow S2$	Higher	Moderate	AI is more often used for explanation, coding support, and procedural help in STEM
$S2 \rightarrow S1$	Higher	Moderate	Verifiable tasks more often allow bounded return to legitimate use
$S3 \rightarrow S4$	Moderate to high	Moderate	Coding-intensive tasks may facilitate concealed delegation
$S3 \rightarrow S5$	Moderate	Higher	Fluent prose more easily masks weak authorship in essay-based tasks
$S4 \rightarrow S9$	Higher where oral/lab checks exist	Moderate	Correction is easier when outputs can be independently verified
$S5 \rightarrow S6$	Moderate	Higher	Performative compliance is more sustainable in output-centered writing contexts
$S6 \rightarrow S8$	Moderate	Moderate to high	Repeated offloading may gradually weaken autonomous reasoning in both fields

5. Three scenarios of development

To illustrate the model's usefulness, three stylized scenarios can be considered. Fig. 2 provides a graphical comparison of the three scenarios. It shows how the same state architecture generates different developmental trajectories depending on whether the institutional environment is permissive, reactively controlling, or integrity-oriented by design.

Scenario 1 – permissive ambiguity. In this scenario, university AI policies are vague, assignment design focuses on final products, and disclosure rules are absent or symbolic. Under such conditions, the probability of movement from legitimate engagement to strategic ambiguity is high, and transitions from ambiguity to invisible delegation and polished dishonesty are also likely. Because verification is weak and sanctions are rare, repeated misuse becomes normalized, pushing the system toward collective normalization and eventually cognitive dependency. In this scenario, the hidden curriculum teaches that successful academic performance depends less on understanding and more on managing appearances.

Scenario 2 – reactive control. In this scenario, the institution recognizes the problem and introduces integrity rules, but the response remains predominantly punitive and detection-oriented. AI use is monitored, some forms of misuse are penalized, and suspicious outputs may trigger review. However, assessment design remains product-centered, faculty guidance is inconsistent, and students receive few constructive models of acceptable AI use. As a result, some transitions into misuse are slowed, especially the most visible ones, but normative mimicry remains common because users learn to preserve misuse while complying formally. The system does not fully collapse into dependency, but it remains unstable because correction is reactive rather than educationally embedded.

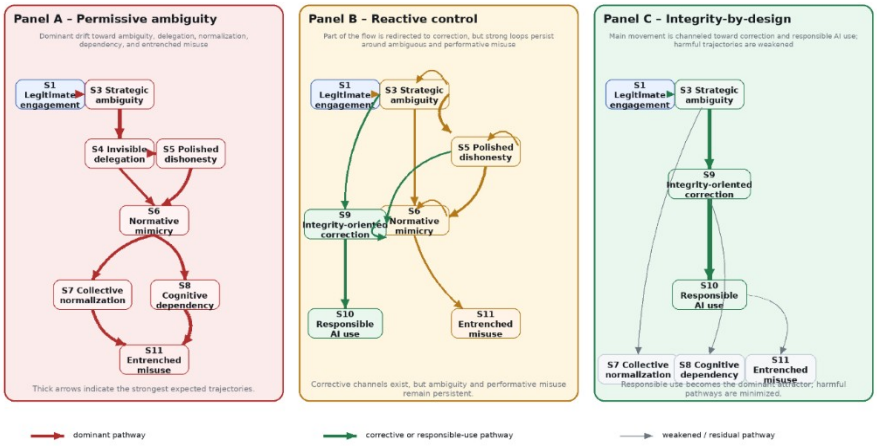


Figure 2: Graphical comparison of three development scenarios of AI misuse in higher education.

Scenario 3 – integrity-by-design. In this scenario, policies are explicit, process-based assessment is widespread, oral defense and reflective disclosure are normalized, and faculty model consistent expectations across courses. AI is not banned, but channeled toward transparent, accountable use. Over time, the system stabilizes around responsible AI use, and the hidden curriculum begins to teach not misuse, but disciplined collaboration with cognitive systems. This is the most desirable scenario because it preserves both academic integrity and the pedagogical potential of AI.

The comparison (Table 2) shows that permissive ambiguity produces the highest accumulation of problematic states and the weakest responsible-use orientation, while integrity-by-design yields the opposite pattern. Reactive control occupies an intermediate position: it reduces the most visible forms of misuse, but it does not fully stabilize responsible AI use because correction remains reactive rather than structurally embedded in pedagogy.

Table 2.
Stylized scenario outcomes for the aggregate measures.

Scenario	Institutional logic	AAM	RUM
Permissive ambiguity	Vague rules, product-centered assessment, weak disclosure and verification	0.67	0.14
Reactive	Punitive monitoring, partial containment, and	0.41	0.29

control	limited pedagogical redesign		
Integrity-by-design	Explicit policy, process-based assessment, reflective disclosure, visible reasoning	0.16	0.68

6. Discussion

The proposed model suggests that AI misuse in higher education should be understood less as a set of isolated violations and more as a trajectory shaped by institutional conditions. This shift is analytically important. Traditional approaches to academic integrity often assume that misuse stems from poor individual choices, insufficient ethics, or weak deterrence. The present framework does not deny the relevance of individual agency, but it shows that agency operates within environments that unevenly distribute opportunities, risks, ambiguities, and rewards. From this perspective, the hidden curriculum is not incidental; it is a structurally produced layer of learning that tells students what actually counts, what can be hidden, and what kinds of performance are rewarded.

One of the most significant implications of the model is that ambiguity functions as a gateway state. Strategic ambiguity is not merely a transitional inconvenience; it is the point at which official norms begin to lose force and tactical adaptation becomes possible. In many educational environments, students do not encounter AI as a clearly bounded resource. They encounter it in settings where some uses are accepted, others are discouraged, and many remain undefined. The model suggests that once ambiguity becomes normalized, subsequent movement toward invisible delegation and polished dishonesty becomes more likely because institutional signals no longer clearly distinguish between assistance and substitution. This fact means that the quality of AI governance depends not only on whether rules exist, but on whether they reduce ambiguity in educationally meaningful ways.

The framework also reveals that the most damaging long-term outcome is not necessarily misconduct in its narrowest sense, but cognitive dependency. This is a crucial interpretive point. If universities focus exclusively on detecting cheating, they may miss the deeper educational problem: repeated offloading of thinking to AI can gradually weaken the learner's capacity for independent analysis, synthesis, judgment, and self-correction. In that case, the hidden curriculum does not simply teach students how to conceal misuse; it teaches them to rely on cognitive systems as substitutes for epistemic labor. The resulting loss is developmental as much as ethical. It affects not only the fairness of assessment, but the very formation of academic agency.

The scenario analysis further clarifies the stakes of institutional choice. In the permissive ambiguity scenario, the system drifts toward misuse because the reward structure favors polished output, the rules are unclear, and verification is weak. In the reactive control scenario, misuse is partially contained but not transformed, because detection and punishment do not eliminate the deeper structural conditions that generate ambiguity and mimicry. In the integrity-by-design scenario, by contrast, the educational environment shifts from reactive surveillance to proactive design: reasoning becomes visible, AI use becomes discussable, and correction becomes pedagogically embedded rather than merely disciplinary. The comparison of these scenarios suggests that the most effective response to AI misuse is not prohibited alone, but the systematic redesign of the transition environment. At a broader level, the model contributes to conceptualizing academic integrity in AI-rich settings as a problem of cognitive governance. In earlier educational contexts, integrity could be framed largely in terms of authorship, citation, and originality. In environments saturated with generative and adaptive AI, those concerns remain important, but they are no longer sufficient. The central governance problem becomes how to preserve epistemic responsibility when cognitive labor itself can be partially externalized. The Markov model and the complementary indices provide one way of addressing this challenge: they enable us to describe not just what misuse looks like, but also how it evolves, how it becomes normalized, and where institutional intervention is most likely to matter.

At the same time, the framework has limitations. The current model is conceptual and stylized rather than empirically calibrated. The assignment of transition parameters remains theoretical,

and real educational environments will be more heterogeneous than any abstract state system can fully capture. Moreover, misuse trajectories may differ across disciplines, levels of study, and institutional cultures. Nonetheless, the value of the model lies precisely in offering a structured vocabulary and analytical architecture that future research can test, refine, and compare across settings. Its purpose is not to produce a final behavioral prediction, but to provide a clearer map of the institutional dynamics through which AI misuse becomes pedagogically meaningful.

7. Conclusion

This article explains how universities may unintentionally reproduce AI misuse through the hidden curriculum of higher education. The central problem ad-dressed was not simply whether students misuse AI, but how educational systems create the conditions under which such misuse becomes probable, normalized, and developmentally consequential. By treating AI misuse as a dynamic process rather than as a one-time violation, the study reframed academic integrity as a matter of institutional design as much as individual conduct.

The analysis demonstrated that the hidden curriculum of AI misuse can be modeled as a sequence of transitions across states of engagement, ambiguity, delegation, normalization, dependency, correction, and responsibility. The Mar-kov approach clarified that users do not move directly from honesty to dishones-ty in a single step; instead, they often pass through intermediate states in which rules are unclear, output quality is rewarded over visible reasoning, and social tolerance reduces the perceived cost of misuse. This dynamic perspective pro-vides a more realistic explanation of how problematic AI practices become em-bedded in everyday academic life.

The framework remains conceptual and therefore requires empirical testing. Future research should calibrate the transition matrix on real educational data, validate the proposed indices across different assessment environments, and explore how institutional variation influences movement across states. Even so, the present study provides a foundation for such work by showing that the hid-den curriculum of AI misuse is not an incidental by-product of technological change, but a structured educational phenomenon that can be modeled, meas-ured, and redesigned.

Declaration on Generative AI

The authors acknowledge the use of Gemini (https://gemini.google.com/), ChatGPT (https://chatgpt.com/) and Grammarly (https://app.grammarly.com/) to refine the academic language, accuracy of the text, check and improve the quality of English in this article. The imagei were generated by ChatGPT 5.2 from the author's original material. After using this tool/service, the authors reviewed and edited the content as needed and took full responsi-bility for the publication's content.

References

[1] A. Artyukhov, N. Artyukhova, O. Dluhopolskyi, O. Adamyk, B. Adamyk, Dialogue with generative artificial intelligence: is its "product" free from academic integrity violations? Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 2 (2025) 181–188.

[2] A. Artyukhov, O. Dluhopolskyi, O. Adamyk, N. Artyukhova, B. Adamyk, O. Liashenko, GenAI and academic integrity in dialogue: pros and cons, in: Proceedings of the 2025 15th International Conference on Advanced Computer Information Technologies (ACIT), 2025, pp. 789–795.

[3] A. Artyukhov, N. Artyukhova, O. Dluhopolskyi, S. Kozlovskyi, L. Buiak, N. Moskaliuk, Ethical issues in AI policy for academia: is it enough for AI effective implementation? in: Proceedings of the 2025 15th International Conference on Advanced Computer Information Technologies (ACIT), 2025, pp. 785–788. doi:10.1109/ACIT65614.2025.11185807.

[4] D. Cotton, P. Cotton, J. Shipway, Chatting and cheating: Ensuring academic integrity in the era of ChatGPT, Innovations in Education and Teaching International 61 (2024) 228–239.

[5] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, G. Kasneci, ChatGPT for good? On opportunities and challenges of large language models for education, Learning and Individual Differences 103 (2023) 102274.
[6] M. Bearman, J. Ryan, R. Ajjawi, Discourses of artificial intelligence in higher education: A critical literature review, Higher Education Research & Development 86 (2023) 369–385.
[7] C. Filson, D. Atuase, Artificial intelligence and academic integrity: The role of academic librarians, Information Development 41(3) (2025) 781–793.
[8] P. Jackson, Life in classrooms, Holt, Rinehart & Winston, 1968.
[9] B. Snyder, The hidden curriculum, MIT Press, 1971.
[10] W. Holmes, M. Bialik, C. Fadel, Artificial intelligence in education: Promises and implications for teaching and learning, Center for Curriculum Redesign, 2019.
[11] J. Rudolph, S. Tan, S. Tan, ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning & Teaching 6(1) (2023) 342–363.
[12] E. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big? in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 610–623.
[13] N. Selwyn, T. Hillman, R. Eynon, G. Ferreira, J. Knox, F. Macgilchrist, J. Sancho-Gil, What's next for Ed-Tech? Critical hopes and concerns for the 2020s, Learning, Media and Technology 45(1) (2020) 1–6.