

Epistemological Friction in Artificial Intelligence Use among Humanities Faculty and Students in Western Mindanao, Philippines

Javier B. Sarmiento¹ , Norvie T. Jalani² 

Author Information:

¹ Western Mindanao State University,
Zamboanga City, Philippines

² Ateneo de Zamboanga University,
Zamboanga City, Philippines

Correspondence:

jav7sarmiento@gmail.com

Article History:

Date Received: August 8, 2026

Date Revised: September 7, 2026

Date Accepted: September 15, 2026

Date Published: September 30, 2026

DOI: <https://doi.org/10.67903/ijroms0405>

Document ID: 2026IJROMS0405

Recommended Citation:

Sarmiento, J. B., & Jalani, N. T. (2026). Epistemological friction in artificial intelligence use among humanities faculty and students in Western Mindanao, Philippines. *International Journal of Research on Multidisciplinary Studies*, 1(8), 426-442. <https://doi.org/10.67903/ijroms0405>

Abstract. Generative artificial intelligence (AI) has entered higher education faster than institutional policy can regulate it, a gap visible in Philippine humanities programs, where interpretive originality and authorial voice are central to disciplinary practice. Humanities programs in regional, resource-constrained Philippine settings remain largely absent from the AI-adoption literature, a gap this study addressed by examining AI adoption, perceived usefulness and ease of use, perceived risk, AI literacy, and behavioral intention among humanities faculty and undergraduate students at a state university in Zamboanga City, Philippines, where no institutional AI policy exists. A convergent mixed-methods design combined survey data from 200 students and 23 faculty with ten in-depth interviews; quantitative data were analyzed using descriptive statistics, Pearson correlation, multiple regression, and Welch's t-test, while qualitative data underwent reflexive thematic analysis, integrated through joint display analysis. Students used AI moderately and selectively, favoring comprehension and organization over original composition. Faculty reported higher AI familiarity than students but markedly lower classroom use ($M = 3.59$ versus $M = 2.67$), a pattern proposed here as epistemological friction: principled disciplinary resistance rather than technical inability. Both groups perceived high risk, yet risk predicted only the intention to limit AI use ($r = .531, p < .001$), not the intention to discontinue it ($r = .047, ns$), producing concealed adoption tentatively labeled a shadow AI ecosystem. Perceived usefulness ($\beta = .427$) and ease of use ($\beta = .365$) were the strongest predictors of continued use, while students significantly outperformed faculty in AI literacy ($d = -0.943$), reversing the assumed top-down transmission of competence. AI acceptance here was shaped as much by disciplinary values and unclear institutional guidance as by usefulness or ease of use, supporting tiered governance, faculty development addressing disciplinary and ethical concerns, and structured reverse mentoring to narrow the literacy gap and reduce concealed AI use.

Keywords: artificial intelligence adoption; artificial intelligence literacy; humanities education; perceived risk; technology acceptance model

Introduction

Generative AI tools, large language models, automated writing assistants, AI-powered research platforms, have moved into global higher education at a pace that has unsettled both enthusiasts and skeptics (Mohammadi et al., 2025; Schei et al., 2024). The bulk of empirical work so far has examined STEM disciplines and well-resourced institutions. Humanities programs, especially those situated in regional contexts where epistemological tradition, cultural norm, and infrastructural limitation interact, remain thinly documented (Bearman et al., 2023).

Philippine higher education makes this gap visible. AI uptake here is uneven and shaped by long-standing regional disparities in digital infrastructure (Clamor & Saloma, 2023). Zamboanga City sits at one such margin: an urban center in Western Mindanao that is home to one of the country's largest Muslim minority populations alongside a Christian majority, linguistically among the most diverse in the country, and almost entirely missing from Asia-Pacific educational technology literature. Most acceptance research has built on Davis's (1989) model and its later extensions (Venkatesh et al., 2012), but Southeast Asian humanities settings have received comparatively little of this attention, especially where AI-driven automation strikes at the work of the discipline itself (Chai et al., 2020). Interpretive judgment, original analysis, authorial voice, these are constitutive of humanities education, not optional features (Bridges, 2019). The conditions under which

humanities scholars and students take up or refuse AI, therefore, differ from those in STEM, and they merit separate study.

Despite growing research on artificial intelligence (AI) adoption in higher education, limited empirical evidence explains how humanities faculty and students in regional Philippine universities use and respond to AI, particularly in settings where institutional AI policies remain unclear or absent. Studies that compare faculty and student experiences within a single Western Mindanao institution are similarly scarce. This study addresses these gaps by examining both groups within a single humanities institution.

This study examines AI tool adoption, perceived usefulness, perceived ease of use, perceived risk, AI literacy, and behavioral intention among humanities faculty and undergraduate students at a state university in Zamboanga City, Western Mindanao, pairing quantitative analysis of adoption predictors with qualitative inquiry into how participants navigate AI use in the absence of institutional guidance.

Research Questions

Guided by these objectives, the study addressed the following research questions:

1. To what extent and ways do humanities faculty and students adopt AI tools?
2. How do both groups perceive AI's usefulness and ease of use, and how do these perceptions translate into practice?
3. What risks do they associate with AI use, and how are those risks ranked?
4. How do perceived risks shape decisions to adopt, limit, or avoid AI?
5. Where do faculty and student perspectives converge or diverge?

Scope and Delimitation of the Study

This study focused on undergraduate college students enrolled in humanities courses, namely English Language Studies, Filipino, History, Mass Communications, and Journalism, at a single state university in Zamboanga City. It also included humanities faculty members teaching within the same institution. The quantitative strand drew on survey responses from 200 students, and the qualitative strand drew on in-depth interviews with five faculty members and five students selected for variation in discipline, rank, and AI use patterns. The investigation was limited to the use and perceptions of generative and assistive AI tools employed for academic purposes, including writing support, translation, idea generation, summarization, and research assistance. Specialized AI systems used for advanced computational research, large-scale data analytics, or institutional administrative functions fell outside the scope of the study. The study was delimited to a single state university, and its findings are not intended to generalize to private institutions, elite or well-resourced universities, or humanities programs outside Western Mindanao. Faculty participation in the interview strand was limited to five purposively selected members, chosen for variation in discipline, rank, and years of service rather than for statistical representativeness. The study examined AI use and perceptions at a single point in time; it did not track how these patterns might change as AI tools, institutional policy, or regulatory guidance evolve. A fuller discussion of the study's methodological limitations appears in the Limitations subsection below.

Literature Review

Global Discourses on Artificial Intelligence in the Humanities

The integration of artificial intelligence into the humanities has sparked ongoing global discussion that differs notably from debates in STEM fields. In STEM, AI is commonly viewed as a tool that boosts efficiency and precision, whereas its role in the humanities raises deeper epistemic questions. Humanities education traditionally involves interpretive tasks such as critical reading, argument development, ethical deliberation, and creative expression, precisely the skills that large language models now imitate with growing complexity (Chan & Hu, 2025). A key concern across literature is not just whether AI can generate humanities-like outputs, but whether its use changes the way students process texts, ideas, and meanings. This issue has significantly influenced empirical research on AI and learning within humanities settings.

Critical Thinking, Cognitive Presence, and the "Resolution Gap"

A consistent pattern from international research shows AI's uneven impact at different stages of cognitive engagement. Using mixed-methods research based on the Community of Inquiry framework, studies indicate that AI tools such as ChatGPT support early learning stages, specifically "triggering events" and "exploration," by helping students brainstorm topics, find sources, and develop initial ideas (Nasr et al., 2025). Conversely, these studies note a decline in performance during the "resolution" phase, in which students analyze evidence, justify their claims, and develop original arguments. This pattern, often called a resolution gap, is linked to

automation bias, where students tend to trust AI-generated answers as final without critically examining their accuracy or coherence (Second Congressional Commission on Education [EDCOM 2], 2025). The literature suggests that, without targeted instructional scaffolding, AI use may encourage superficial fluency that masks gaps in deep understanding. Although students may appear more productive, their capacity for independent reasoning and critical reflection could be undermined. Across these studies, there is consistent agreement that AI assists early-stage idea generation but produces a resolution gap during evidence-based argumentation; what remains unresolved is whether this gap reflects a stable cognitive effect of AI use or is remediable through instructional scaffolding, a question this study's qualitative strand speaks to by examining how faculty and students describe using AI for organizational versus generative tasks (RQ1-RQ2).

Human-AI Co-Participation and the Process-Oriented Turn

In response to concerns about cognitive erosion, several institutions advocate reframing AI not as a replacement for human intellectual work but as a partner in learning. Initiatives such as those supported by the University of California, Irvine's School of Humanities have reframed AI as a pedagogical partner rather than a substitute for human intellectual work, emphasizing thoughtful integration that serves core humanities learning goals and requiring students to critically evaluate AI's capabilities and limitations rather than defer to its outputs (University of California, Irvine, School of Humanities, 2025). While this view reflects a normative consensus in the literature, empirical findings reveal tension between theory and practice. Studies report an illusion of knowing, where students overestimate their mastery because they can produce polished outputs with AI help (Malczyk & Klichowski, 2025). Thus, although scholars widely agree on the need for co-participatory models, there is less consensus on how these ideals are implemented in classrooms, especially under assessment pressures and large class sizes. The literature therefore agrees on the normative goal of co-participatory AI use but has not established how such co-participation is actually achieved under real assessment conditions; this study's qualitative interviews were designed to surface these lived classroom conditions rather than idealized models (RQ1, RQ5).

The Philippine Context: Policy, Infrastructure, and Socio-Cultural Constraints

The adoption of AI in Philippine higher education takes place within a distinctly different socio-technical environment compared to that of the Global North. Literature consistently emphasizes three interconnected challenges: regulatory uncertainty, infrastructural inequality, and culturally influenced attitudes toward technology.

Policy Fragmentation and Institutional Uncertainty

As of 2024–2025, the Philippine higher education sector exhibits what many authors describe as a policy vacuum. Although proposed legislation such as the Artificial Intelligence Education Act of 2025 and anticipated CHED memorandum orders signals emerging national interest, a comprehensive and enforceable framework remains absent (EDCOM 2, 2025). In response, universities have developed institution-specific policies, resulting in considerable fragmentation. Elite institutions such as the University of the Philippines and De La Salle University have articulated principles for responsible AI use that emphasize transparency, data privacy, and human oversight (University of the Philippines Open University, 2024). In contrast, many regional and state universities adopt reactive approaches focused primarily on academic integrity and prohibition (Bearman & Luckin, 2025). The literature agrees that this inconsistency generates uncertainty among faculty and students, who must navigate ambiguous boundaries between acceptable assistance and academic misconduct. Studies at both the elite and regional level agree that policy fragmentation exists, but they disagree on its consequence: some frame it mainly as a compliance risk for institutions, while others frame it, as this study does, as a condition that reshapes how individual faculty and students behave day to day, a distinction that has received little direct empirical testing in Philippine settings (RQ3-RQ4).

Digital Divide and Unequal Access in the Regions

A key area of agreement in Philippine and Global South studies is the role of infrastructural inequality in shaping AI adoption. Research on public universities in the Zamboanga Peninsula shows a persistent gap between digital transformation goals and physical conditions, including unstable internet, limited hardware access, and insufficient digital literacy training (Sula-sula, 2024). These limitations create stratified access to AI technologies: students with reliable internet and paid subscriptions to advanced models gain significant academic advantages over peers who depend on free versions and intermittent connectivity. Albano and colleagues (2025) identify facilitating conditions as a crucial factor in AI adoption, emphasizing that pedagogical models developed in resource-rich environments may not be easily transferred to regions such as Western Mindanao. While there is consensus that infrastructural inequality shapes access, less is known about

how it interacts with the disciplinary and risk-related factors examined in this study, a gap RQ3 and RQ4 begin to address by treating access-related constraints as one facet of the broader perceived-risk construct.

Demographic and Cultural Variations in AI Attitudes

Unlike common assumptions in the global literature, Philippine research uncovers complex demographic trends. Data from a state university in Zamboanga Peninsula indicates that older and non-traditional learners tend to hold more positive views of AI, seeing it as a tool to boost productivity, while younger undergraduates often experience anxiety about skill loss and employability (Co, 2025). Related research also documents gender differences, with male respondents generally reporting more favorable perceptions of AI than female respondents. These results challenge simple narratives of "digital natives" and underscore the need for nuanced, contextual, and intersectional analysis. Nevertheless, most current studies remain largely descriptive, with limited theoretical framing connecting demographic variables to learning outcomes in the humanities. Demographic variables are not the focus of the present study's research questions, but this literature is reviewed briefly because it cautions against assuming uniform AI attitudes within the faculty and student groups examined here, and it motivates the disaggregated faculty-student comparison undertaken in RQ5.

Discipline-Specific Impacts within the Humanities

The literature converges on the observation that AI's impact is not uniform across humanities disciplines. Literature, history, and philosophy emerge as particularly vulnerable given their reliance on authorship, epistemic rigor, and ethical reasoning.

Literature and Creative Writing: Voice, Language, and Authorship

Studies on creative writing in Southeast Asia reveal a shared tension between AI-assisted idea generation and concerns about originality. While AI tools are commonly used to help overcome writer's block, higher-quality student work tends to contain less AI-produced content, indicating that skilled writers still favor their personal voice (Khristianto et al., 2025). A concern specific to the Philippine context is linguistic bias: large language models perform better in English than in Tagalog, Bisaya, or Chavacano, raising fears of linguistic marginalization and cultural uniformity (Cariaga et al., 2025). Despite general agreement on this issue, the research lacks detailed studies on how AI use affects language choice and cultural expression in student writing. This study does not analyze linguistic choice directly, but it contributes adjacent evidence on how humanities students in a linguistically diverse Western Mindanao setting describe balancing AI assistance with their own voice (RQ1-RQ2).

History Education and the Risk of Epistemic Distortion

In history education, AI's tendency toward hallucination presents heightened risks. Given ongoing efforts to counter historical revisionism in the Philippines, particularly regarding the Martial Law era, scholars warn that uncritical reliance on AI-generated content could unintentionally reinforce false narratives (Hipolito Jr., 2025). There is consensus that using AI requires a pedagogical shift toward historiography and source criticism, with an emphasis on verifying claims against primary sources, a concern reflected in initiatives such as a national online course on Wikipedia editing developed specifically to counter historical distortion (University of the Philippines Open University, 2025). However, current research remains mainly conceptual, with few empirical studies on how students actually perform source auditing when AI tools are used. Because empirical evidence on how students actually verify AI-generated historical claims remains scarce, the present study does not test source-verification behavior directly but documents faculty and student risk perceptions around AI-related misinformation as a related concern (RQ3).

Philosophy and the Ethics of Delegated Reasoning

Philosophy educators often frame AI as a moral and cognitive issue rather than merely a technical one. Literature agrees that relying on algorithms for ethical reasoning can produce what some scholars term an unthinking learner, one who reasons through arguments without moral engagement (Malbas & De Guzman, 2025). While philosophy departments are increasingly positioned as centers for AI ethics education, studies remain predominantly normative, and there is limited empirical evidence assessing whether exposure to AI ethics instruction meaningfully alters student reasoning or decision-making. Because empirical evidence on how AI ethics exposure affects student reasoning is still limited, this study's comparative focus on faculty and student risk perception (RQ3-RQ5) offers indirect, exploration evidence relevant to this normative debate rather than a direct test of it.

Faculty Responses and Institutional Readiness

Across settings, humanities faculty display mixed feelings about adopting AI. A common refrain, described as “not ready, not now,” reflects feelings of being unprepared and uncertain about their professional roles (Shata, 2025). To address this, faculty often redesign assessments to become more resistant to AI, such as by using oral defenses and handwritten exams. The literature reveals a divided perspective: while faculty are generally willing to use AI for administrative and preparatory activities, they remain cautious about its educational benefits for students (Alhur et al., 2025). This ongoing tension highlights unresolved questions regarding professional identity and authority in learning environments shaped by AI. This literature agrees that faculty caution is widespread but disagrees on its source, attributing it variably to unpreparedness, workload, or professional-identity concerns; the present study's qualitative strand was designed to examine which of these explanations best characterizes faculty accounts in this setting (RQ1, RQ5).

Student Practices and the Emergence of a “Shadow Curriculum”

Students, in contrast, have incorporated AI into their academic routines at a faster pace than institutional policies can develop. Studies consistently describe a shadow curriculum, in which AI is informally used for summarization, grammar improvement, and idea generation (Bancoro, 2024). This widespread use comes with anxiety, especially fears of false positives from AI-detection tools and concerns about long-term skill loss (Dumandan & Ibojo, 2025). The literature agrees that students' firsthand experiences with AI remain largely unexplored, particularly in humanities fields within resource-limited settings. Taken together, research on faculty and student AI practices has more often examined the two groups separately than comparatively, and few studies situate both within the same institution; this gap directly motivates RQ5, which compares faculty and student perspectives within a single setting.

Synthesis and Identified Gaps

Across global and Philippine literature, there is broad agreement that AI fundamentally reshapes humanities education, particularly in relation to critical thinking, authorship, and ethical reasoning; the literature agrees markedly less, however, on how these effects unfold once specific disciplinary and institutional conditions are taken into account. Three gaps follow from this review. First, most studies are either normative or descriptive, with limited empirical work examining how AI reshapes learning processes in specific humanities disciplines; this gap motivates RQ1 and RQ2, which examine actual adoption behavior and usefulness/ease-of-use perceptions rather than general commentary. Second, the Philippine context, especially in regional and under-resourced universities, remains underrepresented in theory-driven research on perceived risk and AI literacy; this gap motivates RQ3 and RQ4, which examine how risk is experienced and how it shapes adoption decisions in exactly such a setting. Third, faculty and student AI practices have more often been studied separately than comparatively, and institutional conditions have typically been inferred from participants' accounts rather than examined as a distinct perspective; the present study is accordingly bounded to student and faculty perspectives situated within an institutional context characterized by limited AI guidance, rather than claiming to integrate a separate institutional-actor perspective. This third gap motivates RQ5, which directly compares faculty and student experiences within the same institution. These gaps underscore the need for context-sensitive, humanities-focused research that moves beyond questions of permissibility toward understanding how AI can be pedagogically integrated without undermining the epistemic foundations of humanities education. The present study addresses this need by examining AI adoption, perceived usefulness, perceived ease of use, perceived risk, and AI literacy among humanities faculty and students at a state university in Zamboanga City, Western Mindanao, using the convergent mixed-methods design and theoretical framework described in the sections that follow.

The literature reviewed above consistently supports perceived usefulness and perceived risk as central to AI adoption in this setting: usefulness-related studies (e.g., Wang et al., 2024; Kim et al., 2025) converge on selective, task-specific use, while risk-related studies (e.g., Gorichanaz, 2023; Figueroa et al., 2024) converge on academic-integrity concerns as the dominant risk category. Perceived ease of use and AI literacy, in contrast, receive comparatively less systematic attention in the humanities-specific literature reviewed here; ease of use is typically assumed once basic digital access is established rather than measured directly (Sula-sula, 2024), while AI literacy is discussed mainly in normative terms (Ng et al., 2021) rather than through comparative empirical measurement across faculty and student groups. This asymmetry in the existing literature is itself part of the rationale for including both constructs, alongside usefulness and risk, in the conceptual framework and regression model that follow.

Methodology

Research Design

A convergent mixed-methods design was employed (Creswell & Plano Clark, 2018), in which quantitative and qualitative strands were given equal methodological priority. Both strands were planned, and data were collected, in parallel during the same data-collection window; each strand was analyzed independently using its own analytic conventions, and integration occurred only at the interpretation stage, after independent quantitative and qualitative results were finalized. Integration followed a joint-display procedure: for each research question, the corresponding quantitative result (e.g., a mean, correlation, or regression coefficient) was placed alongside the qualitative theme addressing the same construct, and the two were compared and classified as convergent (pointing to the same conclusion), complementary (each adding a different facet of the same phenomenon), or divergent (pointing to different or contradictory conclusions), following the classification procedure described by Creswell and Plano Clark (2018).

Sampling Design

Two groups were recruited. For the student survey, stratified random sampling was used, with the five humanities programs and four-year levels serving as strata. The number of respondents drawn from each program-by-year stratum was allocated proportionally to that stratum's share of the total humanities' undergraduate population, so that larger programs and year levels contributed proportionally more respondents. Within each stratum, a random starting point was generated, and every third name on the class list was selected from that point onward; a student who declined to participate or could not be located within the survey period was replaced by the next name on the same class list to preserve the intended stratum size. Of 234 students approached, 200 returned usable responses, an 85.5% response rate, satisfying the 50 + 8k rule for multiple regression with four predictors (Tabachnick & Fidell, 2013). The faculty sample was drawn from 23 humanities faculty members across the five programs included in the study. Faculty were eligible if they held a regular or part-time teaching appointment in a humanities program during the semester of data collection; faculty on leave or teaching exclusively graduate courses were excluded. Purposive selection then prioritized variation in discipline, academic rank, and years of teaching experience, and eligible faculty were invited through department chairs until 23 faculty members had consented and completed the survey. Because this n was small relative to the eligible population, faculty data were treated descriptively and used only in Welch's t-test comparisons rather than in the regression model. For the qualitative strand, five faculty members and five students were purposively selected from among survey respondents who had indicated willingness to be contacted for a follow-up interview. Faculty interviewees were selected to vary in discipline, academic rank, and years of service; student interviewees were selected to vary in program, year level, and self-reported frequency of AI use, so that both frequent and infrequent users were represented. Survey responses were used to identify this variation before interviewees were approached. Sample size in reflexive thematic analysis was determined by data richness rather than by statistical power; data collection continued until the research team judged that additional interviews were unlikely to surface substantially new codes relevant to the study's research questions, a judgment reached after the ten interviews yielding roughly nine hours of transcript material had been reviewed (Braun & Clarke, 2022).

Research Locale

Data collection took place at a state university in Zamboanga City, Philippines, where the College of Liberal Arts offers undergraduate humanities programs in English Language Studies, Filipino, History, Mass Communications, and Journalism.

Research Participants

The study participants comprised 200 students and 23 faculty members, with their sociodemographic and academic characteristics. Among the students, 71.0% were female and 29.0% were male, while the faculty consisted of 43.5% female and 47.8% male participants. The students were distributed across year levels, with 41.0% in the first year, 16.5% in the second year, 21.0% in the third year, and 21.5% in the fourth year, whereas most faculty members had more than 10 years of teaching experience (60.9%), followed by those with 6–10 years (30.4%) and 5 years or less (8.7%). In terms of academic discipline, the students were primarily enrolled in BA English (44.5%) and BA Filipino (29.5%), while 26.0% were from other programs; the faculty represented varied humanities disciplines. Regarding internet access, 51.8% of students reported always having access, while 39.2% reported having access, compared with 21.7% of faculty members who reported often having access.

Research Instrument

The student survey covered six domains: sociodemographic and academic profile; AI familiarity ($\alpha = .814$) and frequency of AI use ($\alpha = .841$); perceived usefulness (13 items; $\alpha = .947$); perceived ease of use (9 items; $\alpha = .903$); perceived risk (5 items; $\alpha = .867$); and AI literacy (6 items; $\alpha = .927$). The perceived usefulness and perceived ease of use scales were adapted from Davis's (1989) original Technology Acceptance Model items and reworded to specifically address the use of AI tools in academic tasks, while the perceived risk scale was adapted from Featherman and Pavlou's (2003) risk-facets instrument and the AI literacy scale from Ng et al.'s (2021) AI literacy framework. Items measuring AI familiarity and usage patterns were researcher-developed to capture the specific AI tools and academic tasks relevant to humanities coursework at the study site. All scaled items employed a 5-point Likert scale. The faculty version paralleled these constructs and included additional items on institutional support, while two parallel semi-structured interview guides, also developed by the researchers based on the same construct definitions, explored participants' experiences with AI use, perceived benefits and risks, ethical concerns, and institutional policies. Content validity was assessed by two experts in humanities education and AI integration, who evaluated the items for clarity and construct relevance and recommended revisions, particularly to item wording and the AI familiarity set; these recommendations were incorporated before pilot testing. The revised instruments were pilot-tested with 10–20 students and two to three faculty members from comparable humanities programs who were not included in the main study, and pilot feedback resulted in shortening several items for readability and clarifying the response-scale anchors. The final instruments were then administered during the main data collection, with the student scales demonstrating acceptable to excellent internal consistency and the corresponding descriptive results: AI familiarity ($M = 3.25$, $SD = 0.84$), frequency of AI use ($M = 3.02$, $SD = 0.77$), perceived usefulness ($M = 3.44$, $SD = 0.69$), perceived ease of use ($M = 3.43$, $SD = 0.65$), perceived risk ($M = 3.90$, $SD = 0.80$), and AI literacy ($M = 4.09$, $SD = 0.80$).

Data Gathering Procedure

Data collection proceeded only after securing ethics clearance from the Western Mindanao State University Research Ethics Board (Protocol No. 2026-EF-0172) and permission from the College of Liberal Arts and the relevant program chairs. Data was collected over a period of approximately two to four weeks. Student surveys were distributed on paper during regular class sessions, after the class instructor had left the room, to reduce any perception that participation was linked to course standing; a member of the research team explained that participation was voluntary, that students could decline or stop at any time without penalty, and that non-participation would not affect their grades, before distributing the survey and collecting written consent. Faculty surveys were distributed individually and returned directly to the research team. Incomplete surveys with more than 10% of items unanswered were excluded prior to analysis; isolated missing values in otherwise complete surveys were retained and handled as described under Data Analysis. Each interview ran between 45 and 60 minutes, was conducted in English and Filipino all conducted primarily face-to-face in private settings on campus and was audio-recorded with the participant's prior written consent. Recordings were transferred immediately to a password-protected drive accessible only to the research team and were deleted once verified transcripts had been produced. Transcription was performed by the researchers, and participant names and other identifying details were replaced with participant codes (e.g., Faculty 1-5, Student 1-5) at the point of transcription.

Results and Discussions

Research Question 1: To what extent and ways do humanities faculty and students adopt AI tools?

Table 1. Multiple Regression: Predictors of Intention to Continue Using AI

Predictor	β	SE	t	p
Perceived Usefulness	0.427	0.092	4.666	<0.001***
Perceived Ease of Use	0.365	0.098	3.712	0.0003***
Perceived Risk	-0.177	0.067	-2.619	0.0095**
AI Literacy	0.102	0.073	1.392	0.166 (ns)

Students reported moderate familiarity ($M = 3.25$) and moderate use frequency ($M = 3.02$). Summarizing readings was the most common task; generating original drafts was the least common. Faculty showed the inverse profile: higher familiarity ($M = 3.59$) coupled with lower teaching-related use ($M = 2.67$). Where faculty did use AI, they used it most for lecture preparation and least for assessment design, signaling an awareness-

practice gap. The qualitative strand converged on this picture. Students framed AI as routine academic support; faculty described their use as cautious and exploratory. This selective, comprehension-oriented pattern of student use, and the inverse profile among faculty, converges with findings that students favor AI for cognitive support over full automation (Wang et al., 2024) but departs from reports of near-universal, full-task AI uptake among students at well-resourced institutions (Contractor & Reyes, 2025; Kim et al., 2025), suggesting that the moderate, task-selective adoption observed here reflects disciplinary calibration specific to this resource-constrained humanities setting rather than narrower access alone; this interpretation is developed further below.

Research Question 2: How do both groups perceive AI's usefulness and ease of use, and how do these perceptions translate into practice?

Students rated AI's usefulness positively overall ($M = 3.44$), most strongly for understanding complex topics, and rated it lower for tasks demanding original thinking. Faculty usefulness scores were lower ($M = 3.17$), particularly for items linked to core humanities competencies. Ease of use was nearly identical across groups (faculty $M = 3.42$; students $M = 3.43$). The qualitative data sharpened this finding: faculty hesitation did not stem from technical accessibility but from concern about disciplinary fit. That faculty usefulness scores were lower than students' despite equivalent ease-of-use ratings indicate that faculty hesitation is not a matter of technical accessibility. This is consistent with Mah and Groß (2024), who found similar awareness-practice gaps among humanities faculty internationally and attributed them mainly to professional-development deficits; the qualitative data reported here point instead toward a more principled, discipline-grounded source of hesitation, elaborated below as epistemological friction.

Research Question 3: What risks do they associate with AI use, and how are those risks ranked?

Both groups expressed high perceived risk, with no significant difference between them (students $M = 3.90$; faculty $M = 3.70$; $p = .429$). Student concern centered on being falsely accused of academic dishonesty. Faculty concern centered on student overdependence and integrity breaches. Across interviews, both groups described the same uncertainty about acceptable AI use. The absence of a faculty-student difference in perceived risk, alongside each group's distinct source of concern, is consistent with Gorichanaz's (2023) account of academic-integrity precarity among students and with Gustilo and colleagues' (2024) findings on Philippine educators' anxiety about distinguishing authentic from AI-assisted work; the mechanisms behind each group's risk experience are discussed further below.

Research Question 4: How do perceived risks shape decisions to adopt, limit, or avoid AI?

Table 2. Pearson Correlation: Key Variable Relationships

Variables	r	p	Interpretation
PU → Continue AI Use	0.548	<.001***	Strong positive
Risk → Limit/Avoid AI	0.531	<.001***	Strong positive
PEOU → Continue AI Use	0.521	<.001***	Strong positive
Risk → Attitude (overall)	0.336	<.001***	Moderate positive
AI Literacy → Continue AI Use	0.293	<.001***	Weak positive
Risk → Continue AI Use	0.047	.512 (ns)	No significant relationship

In the regression, perceived usefulness ($\beta = 0.427$, $p < .001$) and perceived ease of use ($\beta = 0.365$, $p < .001$) were the strongest predictors of intention to continue AI use. Perceived risk emerged as a smaller but

significant negative predictor ($\beta = -0.177$, $p = .010$). AI literacy did not reach significance. The correlation matrix revealed something more telling. Perceived risk was strongly associated with intention to limit or avoid AI ($r = 0.531$, $p < .001$) yet was uncorrelated with intention to continue using AI ($r = 0.047$, $p = .512$). This split, the same risk perception driving one outcome but not the other, was the study's most consequential quantitative finding. Risk did not stop AI use; it changed how AI was used.

This apparent contrast between a significant regression coefficient for risk ($\beta = -0.177$, $p = .0095$) and a nonsignificant bivariate correlation between risk and continued use ($r = 0.047$, $p = .512$) reflects a difference in what each statistic estimates rather than a contradiction in the data. The regression coefficient captures the unique association between perceived risk and intention to continue AI use after statistically holding perceived usefulness, perceived ease of use, and AI literacy constant; once these more strongly related predictors are accounted for, risk's small negative unique contribution becomes detectable. The bivariate correlation, in contrast, captures the unadjusted, zero-order relationship between risk and continued use alone, without controlling for the other predictors, and this simple relationship is close to zero. The two statistics are therefore not measuring the same thing, and neither is more correct than the other; together they indicate that perceived risk has little relationship with continued AI use on its own, but a small, detectable one once the more strongly predictive usefulness and ease-of-use variables are already accounted for.

Qualitative material gave the concealment pattern its texture. Students used AI for background reading, brainstorming, and structuring ideas, and then carefully removed any detectable trace from the submitted version, aiming for zero-percent AI detection scores. Faculty issued blanket bans on AI in formal evaluations while permitting it for preparation, on condition that students disclosed the use. Both responses were sensible adaptations to a setting where acceptable AI use had never been formally defined. These concealment strategies mirror Lund and colleagues' (2025) finding that a majority of AI-using students engage in strategic concealment under detection-oriented policy regimes, and they support Figueroa and colleagues' (2024) conclusion that prohibition-only policies tend to produce concealment rather than reduced use. Read together with the regression and correlation results above, this qualitative evidence supports interpreting perceived risk as a behavioral modulator rather than a deterrent in this setting, a claim developed further below and offered here as an emerging, context-specific interpretation rather than an established finding.

Research Question 5: Where do faculty and student perspectives converge or diverge?

Table 3. Welch's t-test: Faculty vs. Student Comparisons

Construct	Faculty M (SD)	Student M (SD)	t (df)	p	d	Effect
AI Familiarity	3.59 (1.01)	3.25 (0.84)	1.554 (25.5)	.133	–	ns
Frequency of AI Use	2.67 (1.00)	3.02 (0.77)	-1.624 (27.4)	.117	–	ns
Perceived Usefulness	3.17 (1.04)	3.44 (0.69)	-1.215 (24.5)	.236	–	ns
Perceived Ease of Use	3.42 (0.93)	3.43 (0.65)	-0.050 (27.5)	.960	-0.015	Negligible
Perceived Risk	3.70 (1.16)	3.90 (0.80)	-0.805 (25.8)	.429	–	ns
AI Literacy	3.30 (1.12)	4.09 (0.80)	-3.288 (24.6)	.003**	-0.943	Large
Continue AI Use	3.26 (1.01)	3.12 (0.82)	0.641 (26.2)	.527	–	ns
Limit/Avoid AI	3.13 (1.18)	3.76 (0.90)	-2.479 (25.0)	.020*	-0.676	Medium

Welch's t-test surfaced two significant differences. Students reported higher AI literacy ($M = 4.09$) than faculty ($M = 3.30$), with a large effect ($d = -0.943$, $p = .003$). Students also reported a stronger intention to limit AI use ($M = 3.76$) than faculty ($M = 3.13$; $d = -0.676$, $p = .020$). The literacy reversal documented here runs counter to the top-down sequence commonly assumed in educational-technology models, in which institutional policy and faculty modeling precede student competence (Mishra & Koehler, 2006); it is more consistent with accounts of informal, peer-mediated learning outperforming structured curricular approaches (Laupichler et al., 2022). This pattern, and its implications for reverse mentoring, is discussed in more theoretical depth below.

Joint display analysis showed convergence, complementarity, and divergence across the strands. Both data sources confirmed moderate adoption used mainly as a support tool, shared high-risk perception centered

on academic integrity, and a marked gap between faculty and student engagement with AI. The most consequential pattern was divergent: quantitatively, risk reduced intention to continue AI use, while qualitatively, students did not stop but concealed their use, supporting an interpretation of behavioral modulation rather than cessation.

Table 4. Qualitative Themes and Representative Quotes

Theme	Group	Description	Representative Quote
Assistive student adoption	Students	Students draw on AI as an on-call tutor for clarifying lessons and organizing ideas, but not for producing original work.	"I use it to clarify lessons and organize my thoughts... it's like a tutor available at any time." (Student 1)
Epistemological faculty resistance	Faculty	Faculty hesitation traces to a principled view that AI is structurally at odds with humanistic inquiry, not to unfamiliarity.	"For me, it is not compatible... AI now uses its capacity to think deeper. And less, so they depend more on AI." (Faculty 4)
Academic integrity anxiety	Both	Faculty fear undetectable AI submissions; students fear being wrongly accused.	"The number one is plagiarism... when you use AI, it can perfect your work and without your knowledge that the student uses AI." (Faculty 3)
AI concealment behaviors	Students	Students wash AI traces from submitted work, a sensible response to zero-tolerance regimes lacking clear boundaries.	"AI can really hallucinate... they can create fake citations." (Student 4)
Inverted literacy hierarchy	Both	Students show stronger practical AI-critical skills than faculty.	"I think it's the privacy because AI now can, you can already upload images, and then the AI will have the repository of that." (Student 5)
Institutional regulatory vacuum	Both	Neither group has been given clear institutional guidance on acceptable AI use.	"There is no way for us to find out whether the essays are authentic. So the risk is students will just rely on AI." (Faculty 3)

Table 5. Joint Display: Integration of Quantitative and Qualitative Findings

RQ	Quantitative Finding	Qualitative Theme	Integration	Inference
RQ1: Adoption	Faculty use M = 2.67; student familiarity M = 3.25	Faculty articulate do's-and-don'ts on acceptable AI use	Complementary	Quant shows moderate uptake; qual reveals faculty role-modeling
RQ2: Usefulness/PEOU	PU: students 3.44, faculty 3.17; PEOU d = -0.015	Epistemological friction: principled, not technical resistance	Convergent / Complementary	Real adoption barrier is epistemological
RQ3: Risks	Risk M = 3.90 yet use continues (M = 3.02)	Academic integrity anxiety; cognitive erosion concerns	Convergent / Divergent	Risk is felt but does not halt adoption
RQ4: Risk and adoption	Risk → limit r = .531***; Risk → continue r = .047 (ns)	Strategic concealment; calls for formal rules	Complementary / Divergent	Students conceal rather than stop
RQ5: Faculty-student	Faculty knowledge 3.59 vs. use 2.67; AI literacy d = -0.943	Governing authority (faculty); regulated autonomy (students)	Complementary / Convergent	Knowledge-practice gap; literacy inversion confirmed

The discussion that follows distinguishes two kinds of labels. Qualitative themes such as "Epistemological faculty resistance" and "AI concealment behaviors" were derived directly from coding the interview data using Braun and Clarke's (2022) procedure. Broader interpretive constructs such as epistemological friction, the shadow AI ecosystem, adversarial literacy, and erosion of the assessment contract are, in contrast, higher-level concepts proposed by the authors during interpretation, building on but extending beyond the coded themes; they are presented throughout as proposed or emerging interpretations requiring further validation, not as findings directly established by the data alone.

The Pragmatic Adopter and the Theoretically Aware Non-User

The student adoption pattern observed in this study, moderate familiarity at $M = 3.25$ paired with selective use focused on comprehension and organization rather than content generation, departed from the near-universal AI uptake reported in resource-rich settings (Contractor & Reyes, 2025; Kim et al., 2025). Studies of elite institutions describe students using generative AI across the full task spectrum, drafting and original composition included. The present sample showed something narrower. The picture aligns with Wang and colleagues (2024), who reported that 61% of undergraduates use AI for cognitive support rather than automation, but it sharpened that picture by anchoring the same behavior in an under-resourced provincial setting. Scaffolding-oriented use, on this evidence, appeared to be a stable behavioral pattern across resource contexts rather than an artifact of access constraints.

What set the present case apart was its disciplinary location. AI adoption research in Southeast Asia has clustered around STEM and business education (Chai et al., 2020), where AI-assisted automation aligns more naturally with curricular goals. Humanities programs operate under a different logic. Interpretive originality, authorial voice, and authentic authorship are not optional features here; they constitute the discipline (Bridges, 2019; Haslam, 2024). The students in this study appeared to register this boundary intuitively. They confined AI to preparatory and organizational work and kept generative tasks for themselves. This was not technological avoidance. It was disciplinary calibration.

The faculty pattern carried more theoretical weight. The awareness-practice gap documented above cannot be explained through the digital literacy or technology anxiety frameworks that have long dominated this literature (Brosnan, 1998; Venkatesh et al., 2003). Faculty in this sample knew AI tools, found them operationally accessible (PEOU $M = 3.42$, statistically equivalent to students), and still declined to integrate them. Mah and Groß (2024) document similar awareness-practice gaps among humanities faculty internationally and attribute them to professional development deficits. The data pointed to a deeper dynamic, proposed here, tentatively, as epistemological friction: the discipline-grounded resistance that emerges when AI's probabilistic outputs run up against humanities practitioners' commitments to interpretive originality and authentic authorship. The construct sits apart from technophobia (Brosnan, 1998), which assumes anxiety, and from generic resistance-to-change formulations (Oreg, 2003). It connects instead to what Petriglieri (2011) calls identity threats in professional contexts, where new tools are read as undermining the foundational commitments that constitute professional identity itself. Faculty in this study did not decline to integrate AI because they could not use it; the qualitative data instead suggest a disciplinary concern that using it, in its present forms, would compromise what they took humanities education to be, a proposed interpretation offered here as epistemological friction rather than an established fact about faculty motivation.

The qualitative material supported this reading. Faculty described AI outputs not as inadequate due to technical limitations but as structurally incompatible with the kind of thinking the humanities require. Barassi (2024) calls these flattening errors: outputs that come out grammatically polished but lack interpretive depth. Revell and colleagues (2024), comparing ChatGPT and human essayists in humanities contexts, found that even technically competent AI outputs failed to reproduce the interpretive moves that mark humanities scholarship. Faculty resistance, on this view, was not a failure of adoption. It was an empirically defensible position about the limits of current generative AI in disciplines where interpretation is the work itself.

Risk as Emotional and Institutional Experience

Featherman and Pavlou (2003), in their foundational integration of risk into technology acceptance work, identified six risk facets and assumed risk operates as a rational deterrent: as risk rises, adoption falls. The data gathered here complicated that assumption. Both groups reported high perceived risk. AI use continued anyway. The deterrent mechanism did not behave as the model predicts.

Three conditions explain the divergence. First, students experienced risk as immediate and punitive, not abstract. They navigated zero-tolerance plagiarism-detection regimes without any guidance on acceptable AI use, resulting in what Gorichanaz (2023) describes as ongoing academic precarity. The highest-rated risk item, fear of false accusation ($M = 4.07$), carried concrete material consequences. Second, the affective quality of student risk perception aligns with what Paustian-Underdahl and Slinger (2024) document across multiple national contexts: students using AI under detection regimes feel anxious because they cannot predict how their use will be evaluated. Lund and colleagues (2025) found that 68% of AI-using students engaged in strategic concealment, mirroring the patterns observed in this study.

The third condition concerned faculty. Their risk perception ran through a different mechanism, provisionally described here as erosion of the assessment contract: a collapse of epistemic confidence in distinguishing authentic student work from sophisticated AI output. This was not anxiety about their own AI use. It was anxiety about losing the interpretive authority on which humanities pedagogy rests. Gustilo and colleagues (2024) document parallel concerns among Philippine educators.

Risk in this setting worked as a behavioral modulator rather than a deterrent. The Featherman-Pavlou model treats risk and adoption intention as inversely related. As reported above, risk was strongly tied to the intention to limit AI use but not to the intention to continue it. The model needs bifurcation: when policy environments fail to provide acceptable-use pathways, risk predicts modulation, not cessation. Governance built only on prohibition and detection will, on this logic, push AI use into concealed forms that institutional pedagogy cannot reach (Figueroa et al., 2024).

The Inverted Literacy Hierarchy and the Shadow AI Ecosystem

The literacy inversion documented above was among the study's most theoretically consequential findings. Standard accounts of educational technology integration assume a top-down sequence: institutional policy shapes faculty practice, faculty model competent use, and students acquire literacy through guided exposure (Mishra & Koehler, 2006). The data gathered here reversed that sequence. Students had built practical AI evaluation skills through direct engagement and peer exchange. Faculty, who used AI less and approached it with epistemological caution, had not accumulated the same adaptive knowledge.

Ng and colleagues (2021) identified four AI literacy domains: knowing, applying, evaluating, and ethical use. Their framework presumes these competencies develop through structured curricular intervention. The data gathered here suggested a different pathway. Students appeared to have acquired evaluative and ethical competencies through clandestine engagement, with no curricular scaffolding at all. Laupichler and colleagues (2022) make a similar observation: informal learning networks frequently outperform structured curricular approaches in producing functional AI literacy.

The finding speaks to debates about the digital native's hypothesis. Bennett and colleagues (2008) argued that generational claims about technology fluency obscure within-cohort variation and overstate young users' competencies. The data gathered here offered a partial corrective. Students did not show uniform AI fluency. Their literacy emerged specifically in evaluative domains, developed under conditions of risk and constraint, a pattern provisionally termed here adversarial literacy: competencies apparently acquired by navigating prohibition rather than through guided exposure, offered as a descriptive label requiring further validation rather than an established construct. Faculty literacy scores ($M = 3.30$) reflected deliberate non-engagement rooted in epistemological friction rather than skill deficits. Murphy (2012) argued that reverse mentoring succeeds where senior members hold structural authority and contextual wisdom, while junior members hold technical fluency. The configuration observed in this study met these conditions, providing the structural foundation for meaningful reverse mentoring.

When this inverted hierarchy, combined with punitive policy and the absence of acceptable-use guidance, emerges, it is tentatively termed here, as an emerging and context-specific concept requiring validation beyond this institution, the shadow AI ecosystem: an underground, peer-mediated AI use operating outside institutional oversight. The concept extends the shadow of IT literature (Silic & Back, 2014) into higher education. Silic and Back identified three structural conditions under which shadow IT emerges: organizational technology offerings lag user needs, formal policy is unclear or punitive, and informal peer networks provide functional alternatives. All three conditions are held in this setting. The shadow AI ecosystem extends the framework in two ways. While shadow IT typically operates with tacit organizational tolerance, shadow AI operates in the face of active prohibition. And shadow AI generates a distinctive form of intra-institutional digital inequality, invisible to detection-based monitoring. Chan (2023) argues that the absence of clear acceptable-use guidance functionally guarantees underground ecosystems. Figueroa and colleagues (2024) found that prohibition-only policies consistently produced concealment without reducing actual use rates.

The implication is structural, not remedial. Institutions cannot teach what they prohibit. Critical AI literacy, ethical use, and equitable access all require an institutional posture that acknowledges actual use patterns. The faculty literacy gap is not the primary problem. The policy gap is.

Theoretical Contributions and Institutional Implications

This study makes two principal theoretical contributions and offers two further descriptive constructs that warrant later validation.

As the first principal theoretical contribution, epistemological friction offers a more conceptually precise alternative to technophobia and resistance-to-change frameworks for explaining faculty AI non-engagement, and it carries Petriglieri's (2011) identity threat framework into educational technology contexts. Where prior TAM extensions have largely treated faculty resistance as a deficit to be remediated through training (Mah & Groß, 2024), epistemological friction recognizes resistance as potentially principled and empirically defensible.

The second principal contribution is the shadow AI ecosystem: an underground, peer-mediated pattern of AI use operating outside institutional oversight. The concept extends the shadow IT literature (Silic & Back, 2014; Haag & Eckhardt, 2017) into higher education and specifies two distinguishing features: operating

against active prohibition rather than tacit tolerance and generating intra-institutional digital inequality that is invisible to detection-based monitoring.

Two empirical patterns warrant descriptive labels for further work. The first, erosion of the assessment contract, names the structural dimension of faculty risk perception: a collapse of epistemic confidence in distinguishing authentic student work from sophisticated AI output. The second, adversarial literacy, names the developmental pathway by which student AI competencies emerge under conditions of constraint rather than through guided exposure. Both warrant validation in comparable institutional contexts.

Beyond these constructs, the study makes two additional contributions. It applies and partially validates TAM in a humanities-specific, resource-constrained context ($R^2 = 0.369$), demonstrating that the model's core constructs retain predictive validity while requiring extension to account for principled epistemological resistance and bifurcated risk responses. It also documents a bottom-up AI literacy inversion ($d = -0.943$) that empirically challenges top-down models of educational technology adoption (Mishra & Koehler, 2006), with implications for reverse mentoring as a structural intervention rather than a supplementary practice (Murphy, 2012).

Three institutional implications follow. First, prohibition-and-detection approaches to AI governance are likely counterproductive in contexts resembling this one. High perceived risk paired with high usefulness and accessible tools produces concealment, not cessation. Institutions in similar conditions may benefit from acceptable-use frameworks that distinguish AI use across assessment types: AI permitted for preparatory and organizational work, restricted for generative tasks where it remains pedagogically central. The approach aligns with Chan's (2023) tiered policy framework and is empirically supported by Lund and colleagues' (2025) finding that students in clear-policy environments report substantially lower concealment rates. Second, faculty professional development built only on technical training will not address epistemological friction. Effective development for humanities faculty requires engagement with the disciplinary and ethical dimensions of AI use, not tool familiarization alone. Ng and colleagues' (2021) framework, with its emphasis on evaluative and ethical competencies alongside applied skills, offers a more appropriate model. Third, the literacy inversion has direct structural implications. The assumption that faculty serve as primary AI literacy models is not supported by the data gathered here. Reverse mentoring, in which AI-literate students and ethically grounded faculty exchange knowledge under structured institutional support, offers a configuration that respects both faculty disciplinary expertise and student practical competencies. At the national level, the absence of CHED guidance leaves Philippine institutions navigating adoption without a regulatory framework. A national framework might productively address tiered acceptable-use policies, faculty development integrating disciplinary considerations, data governance standards, and equity provisions for the inequalities documented in the shadow AI ecosystem.

Ethical Considerations

The study followed the 2022 National Ethical Guidelines for Research Involving Human Participants. The Western Mindanao State University Research Ethics Board issued clearance under Protocol No. 2026-EF-0172 before data collection began. Before completing the survey or interview, each participant was informed, verbally and in writing, that participation was voluntary, that they could decline to answer any question or withdraw at any point without penalty or effect on their academic standing, and that their responses would not be shared with course instructors, department chairs, or any other institutional authority. Student surveys were administered only after the class instructor had left the room, so that no faculty member was present while consent was sought or surveys were completed, reducing the likelihood that students would feel pressured to participate because of the classroom setting. Participants then gave written informed consent. All participants were 18 years of age or older; no participant required parental or guardian consent. Confidentiality was protected in several ways. Survey responses and interview transcripts were anonymized using participant codes (e.g., Faculty 1-5, Student 1-5) rather than names, and the key linking codes to identities was stored separately from the data in a password-protected file accessible only to the two researchers. Interview audio recordings were transferred immediately after each session to a password-protected drive, were accessed only by the research team, and were permanently deleted once each recording had been transcribed and the transcript verified for accuracy; names and other identifying details (e.g., specific course sections or instructors) mentioned during interviews were removed or generalized at the point of transcription. Because several participants described sensitive behavior, including concealment of AI use and concern about being accused of academic dishonesty, quotations reported in this manuscript were reviewed to ensure that no participant could be identified by course instructors, department chairs, or other institutional personnel from the content of the quotation itself. No monetary incentive or compensation was provided to participants for either the survey or the interview. Data were stored in password-protected files accessible only to the research team, in line with the approved protocol.

Conclusion

This study identified patterns that complicate prevailing assumptions in the educational technology adoption literature. Students in this sample adopted AI selectively, treating it as a scaffold while preserving generative tasks as their own work, a pattern best read as disciplinary calibration rather than enthusiasm or avoidance. Faculty in this sample did not decline to integrate AI from a technical deficit but from what this study proposes to call epistemological friction: a principled, discipline-grounded caution that is best understood as an emerging, context-specific interpretive concept requiring validation beyond this single institution, rather than an established theoretical category. Both responses make sense given the conditions in which participants operated. The study's central contribution is the finding that AI acceptance in this humanities setting was shaped not only by usefulness and ease of use, as the Technology Acceptance Model predicts, but also by disciplinary beliefs about authorship and interpretation, by uncertainty over acceptable use in the absence of institutional policy, and by a literacy gap running counter to the direction usually assumed in the educational technology literature. This combination, rather than any one finding in isolation, is what this study adds to the literature. Because the design was cross-sectional, the statistical associations reported here indicate prediction and association, not causation: perceived usefulness and ease of use were the strongest statistical predictors of intention to continue using AI, and perceived risk was associated with a stronger intention to limit AI use than with any relationship to the intention to discontinue it ($r = .531$ versus $r = .047$), a pattern this study interprets as risk functioning more as a behavioral modulator than as a deterrent within this sample. Under the prohibition-oriented policy conditions observed at this institution, this modulation was associated with a pattern of concealed use that this study tentatively terms a shadow AI ecosystem; this label should be read as bounded to the participants and institutional context examined here rather than as a claim that such concealment characterizes Philippine humanities education generally. Similarly, the AI literacy inversion documented here ($d = -0.943$) reverses the direction of competency transmission typically assumed in the literature, with implications discussed in the Recommendations, for reverse mentoring as one possible structural response. Future research should test whether epistemological friction, the shadow AI ecosystem, and the literacy inversion documented here recur in other Philippine humanities programs, particularly those with different institutional policy environments, before these constructs are treated as generalizable features of humanities AI adoption rather than as findings specific to this setting. Risk in this setting works as a behavioral modulator rather than a deterrent. The strong association between risk and intention to limit AI use ($r = 0.531$), set against the absence of any association between risk and intention to discontinue ($r = 0.047$), shows that risk shapes how AI is used, not whether. Under prohibition-only policy regimes, this modulation produces the concealed use pattern of the shadow AI ecosystem, one that detection-based governance cannot reach. The AI literacy inversion ($d = -0.943$) reverses the assumed direction of competency transmission, with direct implications for reverse mentoring as a structural intervention.

Recommendations

The findings support five sets of recommendations, presented here with the responsible party and a concrete first action for each, consistent with the bounded, single-institution scope of this study. First, the university's academic affairs office, in consultation with program chairs, should pilot a tiered, assessment-specific AI-use policy rather than a blanket ban: permitting AI for preparatory and organizational tasks (e.g., brainstorming, summarizing background readings) across all programs, while requiring case-by-case disclosure for tasks closer to original composition, with each humanities program specifying which of its own major assessments fall into which tier within one academic year of adoption. Second, faculty development, coordinated by the College of Liberal Arts, should move beyond general AI-tool orientation, which the data suggest faculty do not primarily lack, toward discipline-specific training in assessment redesign, verification of AI-generated content, and articulating disclosure expectations to students; a practical first step is a one-semester workshop series co-designed with faculty from each of the five humanities programs represented in this study. Third, given that students significantly outperformed faculty in measured AI literacy in this sample, the College should pilot a structured reverse-mentoring program pairing AI-literate student volunteers with interested faculty for one semester, with participation voluntary and outcomes evaluated before any decision to scale the program; because this study did not test reverse mentoring directly, it is recommended here as a pilot to be evaluated, not as an established solution. Fourth, university administration should address the digital-access disparities documented in this study through concrete, low-cost measures within its control, such as extending free campus internet access to computer laboratories during evening hours, maintaining a list of no-cost or low-cost approved AI tools, and offering non-AI-dependent alternative assessment options for students who lack reliable access, rather than through broad equity statements alone. Finally, the university's research and extension office, in coordination with counterpart offices at other Philippine higher education institutions, should extend this line of inquiry to other Philippine humanities programs, particularly

those with different institutional policy environments, to test whether epistemological friction, the shadow AI ecosystem, and the observed literacy inversion are transferable findings rather than features specific to this institution. Any resulting recommendations for national policy should await this cross-institutional evidence rather than treating the tiered-governance approach proposed here as immediately applicable to all Philippine higher education institutions.

References

- Albano, R. A. R., Basconcillo, A. P. H., Dacayan, A. M. M., Orme, J. L. B., & Kurata, Y. B. (2025). Analysis of Generation Z Filipinos' perception of AI using multiple regression. In *Proceedings of the 6th African Industrial Engineering and Operations Management Conference, Rabat, Morocco*. IEOM Society International. <https://ieomsociety.org/proceedings/morocco2025/109.pdf>
- Alhur, A. A., Khlaif, Z. N., Hamamra, B., & Hussein, E. (2025). Paradox of AI in higher education: Qualitative inquiry into AI dependency among educators in Palestine. *JMIR Medical Education*, 11, e74947. <https://doi.org/10.2196/74947>
- Bancoro, J. C. M. (2024). The relationship between artificial intelligence (AI) usage and academic performance of business administration students. *International Journal of Asian Business and Management*, 3(1), 27–48. <https://doi.org/10.55927/ijabm.v3i1.7876>
- Barassi, V. (2024). Toward a theory of AI errors: Making sense of hallucinations, catastrophic failures, and the fallacy of generative AI. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.ad8ebbd4>
- Bearman, M., & Luckin, R. (2025). Integrating generative AI in university teaching and learning: A model for balanced guidelines. *Online Learning*, 29(1), 1–19. <https://doi.org/10.24059/olj.v29i1.4508>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86(2), 369–385. <https://doi.org/10.1007/s10734-022-00937-2>
- Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. <https://doi.org/10.1111/j.1467-8535.2007.00793.x>
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.
- Bridges, W. (2019). A brief history of the inhumanities. *History of Humanities*, 4(1), 1–26. <https://doi.org/10.1086/701981>
- Brosnan, M. J. (1998). *Technophobia: The psychological impact of information technology*. Routledge.
- Cariaga, R., El Halaissi, M., Refugio, C. N., Dagunan, M. A. S., Sabidalas, M. A. A., Cariaga, V. B., Cubero, G. D., Cofino, C. L., Limutan, J. J. N., Pospos, R. S., & Gerodias, E. G. (2025). Local voices, global technologies: AI integration barriers in K–12 classrooms. *International Journal of Interdisciplinary Viewpoints*, 1(5), 672–680. <https://doi.org/10.64612/ijiv.v1i5.45>
- Chai, C. S., Wang, X., & Xu, C. (2020). An extended theory of planned behavior for the modelling of Chinese secondary school students' intention to learn artificial intelligence. *Mathematics*, 8(11), Article 2089. <https://doi.org/10.3390/math8112089>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chan, C. K. Y., & Hu, W. (2025). University students' perceptions of generative AI for critical thinking and creativity: The influence of self-efficacy and disciplinary differences. *Innovations in Education and Teaching International*. Advance online publication. <https://doi.org/10.1080/14703297.2025.2600476>
- Clamor, W. L., & Saloma, C. (2023). Digital capital and belonging in universities: Quantifying social inequalities in the Philippines. *Frontiers in Education*, 8, Article 1137121. <https://doi.org/10.3389/feduc.2023.1137121>
- Co, S. J. (2025). Artificial intelligence in Philippine education: A narrative review of applications, perceptions, and challenges. *International Research Journal of Science, Technology, Education, and Management*, 5(2), 25–38. <https://doi.org/10.5281/zenodo.15902950>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Contractor, Z., & Reyes, G. M. (2025). Generative AI in higher education: Evidence from an elite college. arXiv. <https://doi.org/10.48550/arxiv.2508.00717>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Dumandan, D. D., & Ibojo, D. T. (2025). The evolving role of English teachers: Experiences on determining AI-generated language output. *International Journal for Multidisciplinary Research*, 7(4). <https://www.ijfmr.com/papers/2025/4/53017.pdf>
- Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies*, 59(4), 451–474. [https://doi.org/10.1016/S1071-5819\(03\)00111-3](https://doi.org/10.1016/S1071-5819(03)00111-3)
- Figueroa, B. M., Eaton, S. E., Pethrick, H., Hayden, A., Brennan, R. W., Wiens, J., & McDermott, B. (2024). Academic integrity and artificial intelligence in higher education contexts: A rapid scoping review. *Canadian Perspectives on Academic Integrity*, 7(3). <https://doi.org/10.55016/ojs/cpai.v7i3.78123>
- Gorichanaz, T. (2023). Accused: How students respond to allegations of using ChatGPT on assessments. *Learning: Research and Practice*, 9(2), 183–192. <https://doi.org/10.1080/23735082.2023.2254787>
- Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in AI era. *International Journal for Educational Integrity*, 20(1), Article 3. <https://doi.org/10.1007/s40979-024-00153-8>
- Haag, S., & Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering*, 59(6), 469–473. <https://doi.org/10.1007/s12599-017-0497-x>
- Haslam, R. J. (2024). "This wonderful machine": How should we teach humanities texts like Gulliver's Travels in the time of ChatGPT? *Critical Humanities*, 2(2). <https://doi.org/10.33470/2836-3140.1040>
- Hipolito, L. S., Jr. (2025). Transforming Philippine history education: The game-changing effects of Department of Education Order No. 20, series of 2014. ResearchGate. <https://www.researchgate.net/publication/393273407>
- Khristianto, K., Monteza, A. N., Dhyaningrum, A., & Rabbani, I. (2025). Maintaining authenticity of AI-assisted works for Indonesian and Philippines creative writing learners: Students' experience and reflection. *Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia*. <https://ejournal.uinmadura.ac.id/index.php/ghancaran/article/view/21628>
- Kim, J., Klopfer, M., Grohs, J., Eldardiry, H., Weichert, J., Cox, L., & Pike, D. R. (2025). Examining faculty and student perceptions of generative AI in university courses. *Innovative Higher Education*, 50(4), 1281–1313. <https://doi.org/10.1007/s10755-024-09774-w>
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, Article 100101. <https://doi.org/10.1016/j.caeai.2022.100101>
- Lund, B., Mannuru, N. R., Teel, Z. A., Lee, T. H., Ortega, N., Simmons, S., & Ward, E. M. G. (2025). Student perceptions of AI-assisted writing and academic integrity. *AI in Education*, 1(1), Article 2. <https://doi.org/10.3390/aieduc1010002>
- Mah, D.-K., & Groß, N. (2024). Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, 21(1), Article 58. <https://doi.org/10.1186/s41239-024-00490-1>
- Malbas, M. H., & De Guzman, M. J. (2025). Between innovation and tradition: A narrative inquiry of students' and teachers' experiences with ChatGPT in Philippine higher education. *Social Sciences*, 14(6), Article 359. <https://doi.org/10.3390/socsci14060359>
- Malczyk, R., & Klichowski, M. (2025). The impact of generative AI on critical thinking skills: A systematic review, conceptual framework and future research directions. *Information Discovery and Delivery*. Advance online publication. <https://doi.org/10.1108/IDD-05-2025-0125>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mohammadi, E., Thelwall, M., Cai, Y., Collier, T., Tahamtan, I., & Eftekhari, A. (2025). Is generative AI reshaping academic practices worldwide? *Information Processing & Management*, 63(1), Article 104350. <https://doi.org/10.1016/j.ipm.2025.104350>
- Murphy, W. M. (2012). Reverse mentoring at work: Fostering cross-generational learning and developing millennial leaders. *Human Resource Management*, 51(4), 549–573. <https://doi.org/10.1002/hrm.21489>
- Nasr, N. R., Tu, C.-H., Werner, J., Bauer, T., Yen, C.-J., & Sujo-Montes, L. (2025). Exploring the impact of generative AI ChatGPT on critical thinking in higher education: Passive AI-directed use or human–AI supported collaboration? *Education Sciences*, 15(9), Article 1198. <https://doi.org/10.3390/educsci15091198>

- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Oreg, S. (2003). Resistance to change: Developing an individual differences measure. *Journal of Applied Psychology*, 88(4), 680–693. <https://doi.org/10.1037/0021-9010.88.4.680>
- Paustian-Underdahl, S., & Slinger, B. (2024). Students are using large language models and AI detectors can often detect their use. *Frontiers in Education*, 9, Article 1374889. <https://doi.org/10.3389/feduc.2024.1374889>
- Petriglieri, J. L. (2011). Under threat: Responses to and the consequences of threats to individuals' identities. *Academy of Management Review*, 36(4), 641–662. <https://doi.org/10.5465/amr.2009.0087>
- Revell, T., Yeadon, W., Cahilly-Bretzin, G., Clarke, I., Manning, G., Jones, J., Mulley, C., Pascual, R. J., Bradley, N., Thomas, D., & Leneghan, F. (2024). ChatGPT versus human essayists: An exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities. *International Journal for Educational Integrity*, 20(1), Article 18. <https://doi.org/10.1007/s40979-024-00161-8>
- Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, 14(8), Article 922. <https://doi.org/10.3390/educsci14080922>
- Second Congressional Commission on Education. (2025). *Navigating the generative artificial intelligence era: Charting the course for curricular reform in higher education in the Philippines*. <https://edcom2.gov.ph/publications/navigating-the-generative-artificial-intelligence-era-charting-the-course-for-curricular-reform-in-higher-education-in-the-philippines/>
- Shata, A. (2025). "Opting out of AI": Exploring perceptions, reasons, and concerns behind faculty resistance to generative AI. *Frontiers in Communication*. Advance online publication. <https://doi.org/10.3389/fcomm.2025.1614804>
- Silic, M., & Back, A. (2014). Shadow IT: A view from behind the curtain. *Computers & Security*, 45, 274–283. <https://doi.org/10.1016/j.cose.2014.06.007>
- Sula-sula, J. (2024). *Smart public universities in the Philippines' Zamboanga Peninsula region* (MPRA Paper No. 118044). Munich Personal RePEc Archive. <https://mpra.ub.uni-muenchen.de/118044/>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- University of California, Irvine, School of Humanities. (2025, February 19). *Humanities and AI*. <https://www.humanities.uci.edu/news/humanities-and-ai>
- University of the Philippines Open University. (2024). *Guidelines on the use of AI in teaching and learning*. <https://www.upou.edu.ph/wp-content/uploads/2024/01/Guidelines-on-Use-of-AI-in-Teaching-and-Learning.pdf>
- University of the Philippines Open University. (2025, January 17). *UPOU launches free online course on Wikipedia editing and fighting historical distortions*. <https://www.upou.edu.ph/news/upou-launches-free-online-course-on-wikipedia-editing-and-fighting-historical-distortions>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending UTAUT. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Wang, K. D., Wu, Z., Tufts, L., Wieman, C., Salehi, S., & Haber, N. (2024). Scaffold or crutch? Examining college students' use and views of generative AI tools for STEM education. arXiv. <https://doi.org/10.48550/arxiv.2412.02653>