A Multisectoral and Democratized AI Governance Policy for St. Paul University Manila Countering Global Techno-Authoritarianism and Abuse

Brian Bantugan, PhD

St. Paul University Manila

DOI: https://dx.doi.org/10.51244/IJRSI.2026.13010093

Received: 22 January 2026; Accepted: 27 January 2026; Published: 03 February 2026

ABSTRACT

This study investigates the influence of generative artificial intelligence (AI) on higher education governance, with a focus on creating a multisectoral and democratized policy framework for St. Paul University Manila (SPUM). It discusses the ethical and practical implications of AI adoption, highlighting the increasing control of AI by global corporations and the risks associated with techno-authoritarianism. The paper asserts that universities, particularly Catholic institutions, must assert leadership in the governance of AI by creating policies that align with their moral and civic responsibilities. By examining both open-source and for-profit AI systems, the study emphasizes the need for universities to manage AI's integration through participatory, transparent governance that incorporates faculty, students, administrators, IT professionals, and community stakeholders. Key mechanisms for user influence on AI, such as prompt engineering and system-level personalization, are also explored, alongside the economic pressures shaping AI's architecture and functionality. The proposed framework offers a balance between centralized control for risk management and decentralized user involvement in decision-making processes. The paper concludes by recommending a localized AI governance policy for SPUM, grounded in Catholic social teachings, and emphasizes the necessity of ongoing review and adaptation to ensure AI's responsible use in teaching, research, and administration.

Keywords: AI governance, Generative AI, Ethical frameworks, Higher education policy, Catholic institutions

INTRODUCTION

The rapid global diffusion of artificial intelligence (AI) systems has transformed how knowledge is produced, governed, and exercised across societies. From education and research to public administration, security, and media, AI increasingly mediates decision-making processes that were once firmly under human institutional control (Floridi et al., 2018; Zuboff, 2019). While these technologies promise efficiency, scale, and innovation, they also concentrate power in the hands of a limited number of states and corporations, giving rise to what scholars have described as *techno-authoritarianism*: the use of advanced digital technologies to surveil, discipline, and control populations while reducing transparency, accountability, and democratic participation (Feldstein, 2019; Morozov, 2011).

Global trends demonstrate that AI governance is frequently shaped by geopolitical competition, corporate interests, and security-driven logics rather than by ethical deliberation or public accountability (Crawford, 2021). Authoritarian regimes increasingly deploy AI for mass surveillance, predictive policing, and information control, while even democratic societies risk adopting opaque, automated systems that normalize exclusion, bias, and epistemic injustice (Benjamin, 2019; Eubanks, 2018). Within this context, universities face a critical crossroads: they may either become passive adopters of externally governed AI systems or assert their role as normative institutions capable of modeling ethical, participatory, and human-centered governance.

As a Catholic higher education institution grounded in the Paulinian tradition, St. Paul University Manila occupies a distinctive moral and civic position in this global debate. Its mission emphasizes human dignity, social responsibility, community engagement, and the formation of critical, ethical leaders. These values stand

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

in direct tension with governance models that treat AI as a purely technical instrument or as an uncontestable authority. The challenge, therefore, is not merely how to *use* AI within the university, but how to *govern* it in ways that resist domination, exclusion, and abuse while advancing the common good (Floridi, 2023).

This paper argues that AI governance in higher education must move beyond compliance-based or vendor-driven frameworks toward a **multisectoral and democratized governance policy**. Multisectoral governance recognizes that AI affects—and must therefore be shaped by—multiple stakeholders, including administrators, faculty, students, researchers, IT professionals, ethicists, and community partners (UNESCO, 2021). Democratized governance, in turn, insists that AI-related decisions should not be confined to technical elites or external platforms but should remain open to deliberation, contestation, and shared accountability within the academic community (Rahwan, 2018).

Positioned against the backdrop of global techno-authoritarianism, this paper proposes an institutional AI governance policy framework for St. Paul University Manila that foregrounds human oversight, transparency, proportionality, and ethical responsibility. It seeks to articulate how a university in the Global South can exercise moral leadership by embedding AI governance within its educational mission, safeguarding academic freedom, protecting vulnerable populations, and affirming the primacy of human judgment over automated systems. In doing so, the paper contributes to emerging scholarship on democratic AI governance while offering a context-sensitive policy model that resists both technological determinism and authoritarian misuse.

User Influence on AI Mechanisms

Generative artificial intelligence (AI) provides limited direct control to users. While core model architecture, training datasets, and interface logic are managed by developers and their organizations, users can shape AI outputs indirectly. Techniques such as **prompt engineering**, system-level personalization, custom instructions, and feedback loops allow users to influence AI responses within the constraints of the pre-trained model (Buhmann & Berendt, 2022; Wei et al., 2023). Such indirect influence is particularly relevant in academic contexts where precision, epistemic framing, and methodological consistency are required.

Organization-Controlled Aspects of AI

AI developers retain control over three primary dimensions: **core model and training**, **safety and policy mechanisms**, and **interface logic**. The core model operates as a predictive engine trained on large-scale text corpora without true understanding, generating outputs probabilistically rather than cognitively (Floridi et al., 2018; OpenAI, 2023). Safety and policy layers enforce ethical, legal, and reputational safeguards, including restrictions on unsafe content and compliance with institutional or governmental regulations. Interface logic governs user interactions, memory retention, response formatting, and adherence to user-specified preferences, ensuring standardized and reliable performance.

Economic Influences on AI Architecture and Mechanisms

Economic pressures heavily shape the architecture and functioning of AI applications. For-profit AI systems prioritize revenue generation, scalability, and institutional adoption, leading to centralized control, standardized safety protocols, and polished interfaces designed for mainstream usability (Crawford, 2021; Feldstein, 2019). Conversely, open-source AI models are shaped by community funding, contributor engagement, and survival imperatives, often prioritizing transparency, adaptability, and methodological freedom over polished user experiences or integrated safety layers (Hao, 2022; Rae et al., 2021).

Open-Source vs. For-Profit Generative AI Applications

Current open-source AI applications include **LLaMA, GPT-NeoX, BLOOM, Stable Diffusion, CLIP, Whisper**, and related tools accessible via Hugging Face, LM Studio, or Ollama (Hugging Face, 2023). Open-source systems allow local deployment, customization, and methodological experimentation but require technical expertise and researcher-led governance. For-profit AI applications such as **ChatGPT, Claude, DALL·E, MidJourney, GitHub Copilot**, and commercial text, image, and audio tools provide polished

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

interfaces, subscription-based access, and built-in safety features but offer limited transparency and user control (OpenAI, 2023; Anthropic, 2023). Each type supports distinct educational objectives: open-source AI fosters methodological transparency, reproducibility, and epistemic control, while for-profit AI emphasizes efficiency, language quality, and instructional or editorial support (Zawacki-Richter et al., 2019).

Risk Management for AI Applications

Managing risks is critical for both open-source and for-profit AI. Open-source AI requires researcher-led oversight, including **data anonymization, bias mitigation, documentation of prompts and model versions**, and validation protocols (Buhmann & Berendt, 2022; Floridi, 2023). For-profit AI requires **data de-identification, use limitation, disclosure, and human oversight** to prevent overreliance and ensure accountability. Institutional policies can provide frameworks for governance, ensuring AI remains a methodological tool rather than an autonomous decision-maker (Rahwan, 2018).

Human Governance Protocols and Policy

Effective AI governance emphasizes **human accountability, transparency, proportionality, oversight, and contestability** (Floridi et al., 2018; UNESCO, 2021). For open-source AI, governance focuses on procedural rigor, including authorization, access control, secure data handling, methodological validation, and competence training. For for-profit AI, governance emphasizes boundary management, vendor oversight, disclosure, and ethical usage limits. Both frameworks prioritize maintaining human responsibility for decisions and outputs, particularly in research, teaching, and publishing contexts.

Institutional Governance for Catholic Higher Education

In Catholic higher education, AI policies must integrate ethical, pedagogical, and mission-driven principles. Recommended institutional governance frameworks include **AI classification, tiered risk management, approved-use matrices, disclosure requirements, training, and multi-stakeholder oversight**, ensuring AI serves the institution's academic mission while upholding ethical standards (Floridi, 2023; UNESCO, 2021; Zuboff, 2019).

Policy Adaptation for Ubiquitous AI

As AI becomes routine in research, teaching, and administrative workflows, policy revisions should focus on **tiered AI-use declarations, continuous auditability, and risk-tiered governance**, distinguishing incidental, moderate, and substantive contributions. Human oversight remains essential, but institutional review can shift from micro-level approvals to standardized guidelines and audit logs, enabling responsible AI integration without stifling innovation (Wei et al., 2023).

Democratized and Localized AI Governance

To counter centralized techno-authoritarian risks, AI governance should adopt a **multisectoral and democratized approach**, involving administrators, faculty, students, IT personnel, ethicists, and community stakeholders (Rahwan, 2018; UNESCO, 2021). Localizing this framework for a Paulinian Catholic university entails aligning policies with institutional mission, Catholic social teaching, and pedagogical goals, operationalized through instruments such as AI-use disclosure forms, IRB checklists, syllabus clauses, and training programs (Floridi, 2023; Zuboff, 2019). Such an approach fosters ethical, accountable, and inclusive AI use in academic contexts.

Theoretical Framework

This study draws on **socio-technical systems theory, institutional governance, and ethical AI frameworks** to examine the development of a multisectoral AI governance policy at St. Paul University Manila. At its core, AI is a **methodological and pedagogical tool**, whose behavior emerges from the interaction between **user inputs**, **organization-controlled mechanisms**, and **economic and institutional contexts** (Trist, 1981; Orlikowski, 2007).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

User influence on generative AI occurs through prompt design, parameter adjustments, and feedback loops, shaping outputs without altering the underlying algorithm (Buhmann & Berendt, 2022; Wei et al., 2023). However, AI developers retain **organizational control** over the core model architecture, training, safety policies, and interface logic, which collectively govern performance, ethical constraints, and user experience (Floridi et al., 2018; OpenAI, 2023).

The **economic context** further shapes AI design and deployment. For-profit AI models prioritize revenue, scalability, institutional adoption, and regulatory compliance, resulting in centralized control, standardized safety, and polished interfaces (Crawford, 2021; Feldstein, 2019). Open-source models, in contrast, respond to funding availability, community norms, and contributor engagement, emphasizing transparency, flexibility, and local control (Hao, 2022; Rae et al., 2021). Both modalities embed values and operational priorities that influence educational and research applications.

In higher education, AI typologies have distinct **educational relevance**. Open-source AI supports **methodologically transparent, researcher-controlled analysis**, enhancing reproducibility, epistemic rigor, and local data sovereignty. For-profit AI facilitates **writing, synthesis, and pedagogical scaffolding**, providing efficiency, polish, and institutional risk management (Vygotsky, 1978; Jonassen, 1999; Hugging Face, 2023). Effective AI governance thus requires **risk management and human oversight**, including secure data handling, methodological validation, disclosure, and accountability, aligned with UNESCO's (2021) ethical AI guidelines.

Institutional policy frameworks for Catholic higher education integrate **mission-aligned principles**, tiered risk assessment, approved-use matrices, training, and disclosure protocols (Floridi, 2023; Scott, 2014; Zuboff, 2019). As AI becomes ubiquitous, governance must adapt to tiered levels of AI contribution—incidental, moderate, and substantive—ensuring **human accountability, transparency, and ethical compliance** even in routine AI use (Wei et al., 2023; Floridi, 2023).

Finally, the framework emphasizes **democratized, multisectoral governance**, combining stakeholder participation, local adaptation, and ethical oversight to counter centralized, techno-authoritarian risks. This approach operationalizes critical systems thinking by embedding **human responsibility, institutional oversight, and ethical alignment** at every stage of AI deployment in teaching, research, and administration (Midgley, 2000; Rahwan, 2018).

In summary, this theoretical framework positions AI as a socially embedded tool, whose governance and policy must balance **user influence, organizational control, economic pressures, educational purpose, and ethical accountability**. This integrative lens informs the development of a **Paulinian Catholic policy of AI governance** that is transparent, democratized, and mission-aligned.

Statement of the Problem

The rapid proliferation of generative artificial intelligence (AI) applications—both open-source and for-profit—has transformed teaching, research, and administrative practices in higher education. While these tools offer unprecedented opportunities for efficiency, creativity, and data-driven insights, they also pose significant challenges regarding algorithmic control, organizational governance, economic influence, ethical use, and institutional policy. Current understanding of how AI users can influence model behavior, the organization-controlled components of AI systems, and the ways in which developers' decisions are shaped by funders remains fragmented (Floridi et al., 2018; Crawford, 2021; Rae et al., 2021). Similarly, the distinctions between open-source and for-profit AI applications, their respective educational relevance, and the risks inherent to each type are not systematically mapped in the context of higher education, particularly within Catholic academic institutions (Zawacki-Richter et al., 2019; UNESCO, 2021).

Moreover, existing AI governance frameworks often do not account for the growing ubiquity of AI in academic workflows, the need for tiered human oversight, or the ethical implications of centralized, global AI control prone to authoritarian misuse (Rahwan, 2018; Zuboff, 2019). There is also a lack of guidance on how institutions can develop **mission-aligned, multisectoral, and democratized AI governance policies** that

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

balance innovation, human accountability, transparency, and protection of sensitive data while supporting teaching, research, and publication activities.

The research aimed to develop a localized, Paulinian framework for AI governance that addresses global techno-authoritarian threats while fostering innovation, accountability, and ethical practice in Catholic higher education.It sought to answer the following research questions: (1) What can generative AI users do to influence its mechanism/algorithm?; (2) What constitutes the AI organization-controlled aspects of the AI application (Core Model and Training, Safety and Policy, and Interface Logic)?;(3) How are the developers of AI applications' architecture and mechanisms influenced by the economic interests of their funding agencies?; (4) How are open source and for-profit generative AI models influenced by their funders?; (5) What are the open source and for-profit generative AI applications to date?; (6) What are the educational relevance of each AI type?; (7) What are the ways to manage risks of open source and for profit generative AI applications?; (8) What human governance AI-use protocols and policy can be recommended for open source and for-profit AI applications?; (9) What institutional governance framework for Catholic higher education schools?; (10) How should said policy be revised in the context where AI use becomes ubiquitous?;)? (11) How can the above policy revision be modified to enable the development of a multisectoral and democratized AI governance model that will counter centralized global governance subject to authoritarianism and abuse?; (12) How can the framework be localized for a Paulinian Catholic university?; (13) and What corresponding instruments for implementation can be proposed for the university?

METHODOLOGY

Paradigm, Approach, and Design

This study **adopted an interpretive qualitative case study** approach to examine generative artificial intelligence (AI) applications, their governance mechanisms, and the development of a multisectoral and democratized AI policy for St. Paul University Manila. A qualitative case study was particularly appropriate because it allowed for an in-depth exploration of complex phenomena within their real-world context, emphasizing meaning-making and interpretive understanding rather than quantification (Stake, 1995; Yin, 2018). By situating AI governance within the unique institutional, ethical, and educational context of a Catholic higher education institution, this research **prioritized holistic comprehension over generalization**.

The study **employed an interpretive paradigm**, recognizing that knowledge about AI governance was socially constructed, situated, and context-dependent. The interpretive lens enabled the researcher to consider multiple perspectives, including organizational policies, ethical frameworks, funding structures, and educational practices, acknowledging that AI use and governance were shaped by human decisions, institutional priorities, and broader socio-technical systems (Schwandt, 2014).

Data Construction

Primary data for this study **consisted of an online literature review** and document analysis. Sources included peer-reviewed academic journals, policy reports, white papers, and technical documentation relating to open-source and for-profit AI applications, AI governance frameworks, and higher education policy (Creswell & Creswell, 2018). The literature review focused on identifying mechanisms by which users influenced AI outputs, core organizational control aspects of AI systems, economic pressures on developers, educational relevance of AI types, and existing risk management and governance protocols.

Documents were retrieved using systematic keyword searches in academic databases such as Scopus, Web of Science, and Google Scholar, complemented by grey literature including institutional policy guidelines, UNESCO AI ethics recommendations, and relevant technical reports. Inclusion criteria emphasized relevance, currency (2018–2026), and applicability to higher education and governance contexts.

Data were analyzed using **thematic analysis**, which allowed for the identification, coding, and interpretation of patterns or themes within qualitative material (Braun & Clarke, 2006). Thematic analysis followed a six-phase procedure: familiarization with the data, initial coding, searching for themes, reviewing themes, defining

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

and naming themes, and producing the final narrative (Braun & Clarke, 2013). Themes were developed in alignment with the study's research questions, such as "user influence on AI behavior," "organizational control and economic influence," "risk management strategies," "educational applications," and "institutional governance protocols."

To ensure **rigor and trustworthiness**, the study applied strategies including prolonged engagement with the literature, triangulation of sources (e.g., academic, technical, and policy documents), audit trails for coding decisions, and reflexive memos documenting the researcher's interpretive choices (Lincoln & Guba, 1985). This approach aligned with the ethical and epistemological standards appropriate for research in educational and institutional governance contexts.

ChatGPT was employed as a tool to assist in drafting policy frameworks according to the specifications derived from the constructed dataset. The **data construction phase** involved systematically gathering, coding, and organizing information from a comprehensive online literature review, institutional guidelines, existing AI governance frameworks, and relevant theological and ethical documents pertinent to Catholic higher education. These sources were organized into thematic categories aligned with the research questions, including human governance protocols, risk management strategies, educational relevance, open-source and for-profit AI typologies, and institutional accountability mechanisms.

ChatGPT (v5-mini) was used to synthesize these data points into coherent policy text while adhering to the specifications established during data construction. These specifications included the inclusion of Catholic ethical principles, multisectoral governance, risk-tiered AI-use declarations, and distinctions between open-source and for-profit AI protocols. Prompts were carefully constructed to ensure that outputs reflected the synthesized dataset accurately, highlighting both procedural and participatory governance dimensions, while remaining consistent with the interpretive qualitative approach of the study. Outputs generated by ChatGPT were treated as **draft material**; all content was critically reviewed, edited, and contextualized by the research team to ensure alignment with institutional norms, ethical principles, and the research objectives.

The use of ChatGPT in this context exemplifies **controlled AI-assisted drafting**: it expedited the synthesis of diverse data sources into structured policy language, supported iterative refinement, and enhanced consistency across sections while maintaining human oversight as the ultimate decision-making authority. This approach reflects the methodological principle that AI serves as a tool for **language generation and structural coherence** rather than as an independent author or decision-maker, preserving the interpretive and ethical integrity of the research process (Creswell & Poth, 2018; Bender et al., 2021).

Thus, the integration of ChatGPT in the data construction phase allowed the research team to efficiently translate complex, multi-source data into draft policy text, while safeguarding human oversight, transparency, and accountability. AI-assisted drafting was fully documented, including prompt templates, model versions, and validation processes, ensuring replicability and auditability within the interpretive qualitative case study framework.

Ethical Considerations

Although the study relied on publicly available literature and documents, ethical considerations were maintained by ensuring accurate representation, proper attribution of sources, and critical engagement with sensitive issues, such as AI bias, privacy, and techno-authoritarian risks. The interpretive framework emphasized that the researcher remained accountable for analysis and interpretation, reflecting human governance principles mirrored in AI-use policy itself.

RESULTS

What can generative AI users do to influence its mechanism/algorithm?

Regular users of generative AI, such as ChatGPT, do not have the ability to directly modify the underlying architecture, training algorithms, or safety mechanisms of the system. The core model, whether a GPT-4 or

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

GPT-5 variant, is fully managed by OpenAI, which maintains control over the training data, retraining schedule, reasoning algorithms, safety filters, and global response optimization (OpenAI, 2023a; OpenAI, 2023b). Consequently, users cannot rewrite the algorithm, adjust neural network weights, or disable built-in safety protocols unless they possess specialized access through enterprise APIs or research partnerships (OpenAI, 2023b). In essence, the fundamental "engine" of ChatGPT is not user-configurable.

Despite these constraints, users can meaningfully influence the outputs and behavior of generative AI through several indirect mechanisms. The most powerful of these is **prompt engineering**, which involves carefully designing instructions to guide the model's responses. For example, users can assign roles ("Act as a qualitative research supervisor"), define constraints ("Use APA 7th edition only; no bullet points"), control tone and epistemology ("Write in an academic, post-positivist voice"), and request process transparency ("Explain your reasoning step by step") (Brown et al., 2020; OpenAI, 2023a). In academic work, prompt engineering effectively serves as the closest practical equivalent to algorithmic control.

Another mechanism is **system-level personalization** through user memory and preference settings. ChatGPT adapts to individual users based on saved preferences—such as writing style, citation format, and disciplinary lens—as well as interaction patterns, including which outputs are accepted, revised, or rejected (OpenAI, 2023b). For instance, a user consistently requesting APA-formatted, theory-grounded academic responses reinforces these patterns. Explicit instructions, such as "Remember that I always prefer APA style and academic tone," can strengthen this adaptive behavior, though the memory is selective and constrained.

Custom instructions provide an additional layer of persistent behavior shaping. Within ChatGPT's settings, users can specify how the AI should respond and what contextual information it should consider. This layer can influence the level of detail, disciplinary assumptions, and language register of outputs (OpenAI, 2023a). A researcher might, for example, enter: "I am an academic researcher in communication and education. Responses should be theoretically grounded, citation-aware, and suitable for publication drafts." These instructions act as a soft configuration layer that personalizes AI behavior without modifying the underlying model.

Feedback loops also play a role in shaping AI behavior at the interface level. By providing explicit feedback—using thumbs up/down, follow-up corrections, or requests to revise outputs using specific frameworks—users reinforce preferred patterns of output over time. While this does not retrain the model itself, it informs how future interactions are interpreted and displayed (OpenAI, 2023b).

Advanced users can leverage **tool-based controls** to extend influence further. Custom GPTs allow for pre-configured instructions, uploaded reference materials, or specialized behavior (e.g., an "APA Research Editor GPT"). Through API access, users can manipulate parameters such as temperature, maximum tokens, and system prompts, and even chain prompts programmatically. While these approaches do not alter the core algorithm, they effectively change how the AI is deployed and how it performs specific tasks (OpenAI, 2023a; OpenAI, 2023b).

Conceptually, AI control can be viewed in layers. OpenAI governs the core model, training, safety, and interface logic, while users exercise control over prompt design, context and memory shaping, and task framing (Brown et al., 2020; OpenAI, 2023b). Though the underlying engine cannot be reprogrammed, users can "drive the vehicle" with precision, steering outputs through these indirect but powerful levers.

For academic users, practical strategies include maintaining a standardized opening prompt for research tasks, explicitly framing epistemological approaches (e.g., constructivist or post-positivist), reiterating formatting and citation rules upfront, and correcting outputs decisively rather than silently editing them. Consistently applying these practices trains the interaction system to align with user expectations and produces outputs that better support scholarly work (OpenAI, 2023a; OpenAI, 2023b).

Table 1. Conceptual Model of AI Algorithm Control

Layer	**Who controls it**

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Core model & training	AI organization
Safety & policy	AI organizationI
Interface logic	AI organization
Prompt & instruction layer	User
Context & memory shaping	User
Task framing & constraints	User

What constitutes the AI organization-controlled aspects of the AI application (Core Model and Training, Safety and Policy, and Interface Logic)?

Generative AI applications such as ChatGPT consist of multiple layers, each controlled and maintained by the organization that develops the system. Understanding these layers—core model and training, safety and policy, and interface logic—clarifies what is organizationally managed versus what users can influence (OpenAI, 2023a; OpenAI, 2023b).

The **core model and training** can be thought of as the "brain" of the system. At its foundation, ChatGPT is a predictive language model that has processed vast amounts of text to learn statistical patterns in word sequences. It does not retrieve information in real time, comprehend meaning like a human, or possess thoughts, emotions, or opinions (Brown et al., 2020). Instead, it analyzes user input, predicts likely subsequent words, and generates responses sequentially. Training occurs prior to deployment and includes three major phases. First, during the reading phase, the model is exposed to large corpora of text to learn language structures. Second, in the teaching phase, humans provide examples of high-quality responses. Finally, in the correction phase, human evaluators rank outputs to reinforce preferred behaviors (OpenAI, 2023a). It is important to note that user interactions do not modify the underlying model permanently; the model cannot be "taught" by regular conversations. Limitations of the core model include its potential to produce confident-sounding but inaccurate information and its dependence on the specificity and clarity of user prompts.

The **safety and policy layer** functions as the system's "rulebook." This component dictates what the AI is permitted to generate, independent of the core model's capabilities (OpenAI, 2023b). The safety mechanisms can block, redirect, or modify responses according to predefined guidelines. Topics typically restricted include content related to violence, self-harm, hate speech, sexual content, illegal activities, and professional advice in fields such as medicine, law, or finance. Because these rules are enforced at the organizational level, users cannot disable or bypass them. Their presence explains why AI outputs may sometimes refuse a query or provide inconsistent responses based on context, effectively acting as guardrails to prevent harm or misuse.

The **interface logic**, often described as the "conversation manager," governs how users experience interactions with the AI. This layer determines which prior messages are retained during a session, applies user-specified preferences such as tone, style, or formatting, and manages response structure and length (OpenAI, 2023a). The interface can adapt to recent interactions, giving the impression of a responsive conversation, yet it operates under strict organizational rules. Instruction priority is hierarchical: platform rules take precedence, followed by built-in system instructions, and finally user inputs. Consequently, while users can guide outputs through careful phrasing and structured prompts, they cannot override the organizational constraints embedded in the system.

Thus, the organizationally controlled aspects of generative AI include the core predictive model, the safety and policy framework, and the interface logic that manages conversation flow. These components work in concert to ensure the AI operates safely, reliably, and within ethical and legal boundaries. Users influence behavior indirectly through prompt design, system preferences, and interaction patterns, but the fundamental architecture, operational rules, and conversational management remain under the control of the AI-developing organization (Brown et al., 2020; OpenAI, 2023a; OpenAI, 2023b).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

How are the developers of AI applications' architecture and mechanisms influenced by the economic interests of their funding agencies?

The development and deployment of generative AI applications such as ChatGPT are heavily influenced by the economic realities faced by their creators. Building, training, and maintaining large-scale AI systems involves substantial financial investment, which in turn shapes architectural, safety, and interface design decisions (OpenAI, 2023a; OpenAI, 2023b). Developers must ensure that the system operates reliably at scale, complies with legal requirements, appeals to diverse markets, and generates predictable revenue streams. These economic pressures influence both the structural design of AI systems and their observable behavior.

Funding for AI applications typically originates from multiple sources, including venture capital investors, strategic corporate partners, paying users via subscriptions or API access, and institutional adopters in education, business, or government (OpenAI, 2023b). While funders do not usually dictate specific outputs, their economic priorities shape what kind of system is feasible and acceptable to deploy. In particular, funding imperatives favor systems that minimize legal and reputational risk, maximize scalability, and appeal to broad user bases.

Economic considerations directly affect architecture decisions. For instance, the need to support millions of users encourages the development of a single, general-purpose model rather than numerous niche variants. This design prioritizes consistency and centralized governance, ensuring that rules and behavior apply uniformly across the user base. Allowing full user-level reprogramming would be prohibitively expensive and risky, as fragmented or inconsistent behavior could damage credibility and usability (OpenAI, 2023a; OpenAI, 2023b). Similarly, safety layers are integrated as a form of risk management. Unsafe outputs could lead to lawsuits, regulatory penalties, or loss of institutional clients, making conservative refusals and overblocking in ambiguous cases economically prudent. Beyond ethics, these safeguards serve as financial insurance for the organization.

Economic pressures also influence the training process and resulting behavior. Human feedback plays a crucial role in model fine-tuning, but evaluators are hired at scale and operate under standardized guidelines. As a result, the system tends to reflect mainstream professional norms, emphasizing neutrality, politeness, and risk-averse language. High-liability domains, such as medical, legal, and financial advice, are heavily restricted through disclaimers or outright refusals, protecting both users and the organization's bottom line (OpenAI, 2023b). Consequently, outputs are generally cautious and designed to be institutionally acceptable, rather than promoting radical or niche perspectives.

The interface logic of AI applications is similarly shaped by economic imperatives. Interfaces are designed to maximize user engagement, support subscriptions, and provide reliable performance for professional tasks. To balance utility with legal and financial constraints, memory and personalization features are limited, retaining only short-term context or superficial user preferences. Deep, long-term modeling of individual users would be valuable but would also increase costs, legal risk, and privacy compliance burdens (OpenAI, 2023a).

It is important to clarify that economic influence does not imply that funders control specific outputs or secretly enforce ideological agendas. Rather, these pressures structure the system to operate safely, legally, and profitably in diverse markets. A useful analogy is that of a commercial airplane: engineers optimize performance, airlines prioritize safety and profitability, regulators enforce risk reduction, and passengers expect reliability. The final aircraft is not the fastest or most customizable, but it is safe, rule-bound, and capable of serving millions reliably. ChatGPT is similarly designed to survive in complex economic, regulatory, and institutional environments while remaining broadly useful and commercially viable (OpenAI, 2023b).

Hence, the economic interests of funding agencies influence AI architecture and mechanisms by shaping scalability, safety, training practices, and interface design. These pressures result in systems that are cautious, polite, general-purpose, rule-governed, and difficult to deeply customize, reflecting the priorities of large

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

markets, institutional clients, and legal compliance rather than the preferences of individual funders (OpenAI, 2023a; OpenAI, 2023b).

How are open source and for-profit generative AI models influenced by their funders?

For-profit AI systems must prioritize sustainability, scale, and profitability, necessitating centralized control, limited user-level customization, and strong safety and policy layers (OpenAI, 2023b). Human feedback during training is standardized, reflecting mainstream professional norms to produce neutral, polite, and institutionally acceptable outputs. Interface design emphasizes reliability, productivity, and moderate personalization, balancing user engagement with privacy and legal compliance. Economic influence in these systems is structural rather than prescriptive; funders shape what is safe, deployable, and marketable, not the specific content of outputs (OpenAI, 2023b).

In contrast, **open-source generative AI models** are influenced by funding survival, community values, and the need to attract contributors. These models often provide access to weights for fine-tuning or repurposing, employ lighter or optional safety layers, and rely on social rather than technical enforcement of norms (Stiennon et al., 2020). Training data may be smaller, less curated, and more transparent, resulting in greater variance in tone, potential bias, and experimental use cases. Deployment is decentralized, with users managing local or server-based interfaces, creating a trade-off between transparency, flexibility, and user control versus polish, safety, and standardization (Bender et al., 2021). Crucially, open-source AI does not exist free from economic influence; its design and priorities are shaped by the need to sustain funding, maintain community legitimacy, and ensure ongoing access and openness, rather than by corporate revenue and brand protection.

These dynamics have significant implications for knowledge production and education. For academic users, for-profit AI favors outputs aligned with institutional language, conservative framing, and citation-friendly formats, making it well-suited for formal research, teaching, and professional contexts. Open-source AI, by contrast, supports methodological experimentation, epistemic diversity, and greater researcher control, facilitating innovative or unconventional approaches. Neither model is inherently superior; each serves distinct intellectual ecosystems and reflects the economic, ethical, and social priorities of its funders and maintainers.

Table 2. Key Comparison Between Open Source and For-Profit AI Funding Influence

Aspect	**For-profit models**	**Open-source models**
Main economic driver	Revenue & risk control	Funding & community
Control	Centralized	Distributed
Safety	Mandatory & strong	Optional & uneven
Customization	Limited	High
Transparency	Low	High
Legal risk	Minimized internally	Shifted to users
Innovation style	Incremental	Experimental

What are the open source and for-profit generative AI applications to date?

Generative AI applications today fall broadly into two categories: open-source systems, which make code and model weights publicly available, and for-profit commercial systems, which restrict access to models and provide controlled interfaces. Open-source generative AI allows researchers, developers, and institutions to

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

modify, self-host, and experiment with models, though safety and moderation depend on the user (Bender et al., 2021; Touvron et al., 2023). Prominent open-source text and language models include Meta's LLaMA and its derivatives (e.g., LLaMA 2 and LLaMA 3), Mistral 7B and Mixtral, Falcon (Technology Innovation Institute), BLOOM (BigScience consortium), GPT-Neo, GPT-J, and GPT-NeoX (EleutherAI), as well as open variants like Qwen (Alibaba) and Yi models (01.AI). These models are commonly accessed through platforms such as LM Studio, Ollama, or Hugging Face Transformers. Open-source image generation applications include Stable Diffusion and its SDXL variant, Kandinsky, OpenJourney, and DeepFloyd IF, typically used via user interfaces like AUTOMATIC1111, ComfyUI, or InvokeAI. In audio and speech, notable open-source tools include Whisper, Coqui TTS, ESPnet, and the legacy Mozilla DeepSpeech. Multimodal research tools such as CLIP, BLIP/BLIP-2, and OpenFlamingo further demonstrate the diversity and experimental potential of open-source AI.

By contrast, for-profit generative AI applications prioritize user accessibility, safety, and scalability, often targeting institutions and mainstream audiences through polished, subscription-based platforms (OpenAI, 2023a; OpenAI, 2023b). For text and multimodal outputs, commercial examples include OpenAI's ChatGPT, Anthropic's Claude, Google's Gemini, Microsoft's Copilot, Perplexity AI, Jasper AI, and Notion AI. Image and video generation applications encompass OpenAI's DALL·E, Midjourney, Adobe Firefly, Runway, Pika, and Synthesia for AI video avatars. In audio and voice, ElevenLabs, Descript, Murf AI, and Soundraw offer production-ready solutions. For coding and productivity, platforms such as GitHub Copilot, Replit AI, Cursor, and Tabnine provide integrated AI-assisted programming support. These for-profit systems emphasize strong safety mechanisms, subscription or usage fees, and limited transparency, reflecting their commercial and institutional design priorities.

Some generative AI applications occupy a hybrid or "open-core" space, combining elements of openness with commercial offerings. Examples include Meta's LLaMA, which provides open model weights under a restricted license, Mistral's open models coupled with paid services, Stability AI's open models integrated with commercial platforms, and Hugging Face's ecosystem, which combines open-source frameworks with enterprise solutions. These cases blur the line between free experimentation and controlled commercialization, reflecting a spectrum of accessibility, safety, and monetization strategies.

A comparative perspective illustrates key distinctions between open-source and for-profit AI. Open-source models prioritize freedom, transparency, and user control, but often require technical expertise to implement and rely on user-managed safety protocols. For-profit models emphasize ease of use, polished interfaces, managed safety, and predictable institutional deployment, though at the expense of transparency and deep user control. In practice, most users—particularly in academic, research, and creative contexts—employ a combination of both types depending on the task, balancing experimental flexibility with safety, reliability, and accessibility (Bender et al., 2021; OpenAI, 2023a; Touvron et al., 2023).

What is the educational relevance of each AI type?

What AI type fits academic research best?. Neither open-source nor for-profit generative AI models alone can be considered universally ideal for academic research. Instead, the strongest practice is a hybrid approach that leverages the respective strengths of each type of AI for specific research, teaching, or publication tasks (Bender et al., 2021; OpenAI, 2023a; Touvron et al., 2023). Open-source AI is particularly well-suited for research contexts that require methodological transparency, replicability, and sensitive data handling. Researchers can document the model name, version, training data, and fine-tuning procedures, allowing others to replicate analyses and ensuring epistemic control through adjustable prompts, parameters, and datasets. Local deployment protects sensitive participant data, making open-source AI optimal for qualitative analysis, thematic coding, methodological experimentation, grant-funded projects, ethics-reviewed studies, and critical AI research (Bender et al., 2021).

Conversely, for-profit AI applications are best employed when research or academic workflows prioritize efficiency, language quality, synthesis, and pedagogical support. Commercial platforms such as ChatGPT, Claude, and Gemini offer polished interfaces, editorial assistance, and high-level summarization capabilities. These tools are particularly effective for drafting and revising manuscripts, generating teaching materials,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

conducting literature overviews (with verification), and providing academic writing support (OpenAI, 2023a; OpenAI, 2023b). Therefore, an optimal setup for researchers engaged in education studies, qualitative research, and curriculum development is to use open-source AI for data handling and analysis, while employing for-profit AI to support writing, teaching, and synthesis tasks.

What are the ethical implications of each type of AI?. The ethical implications of AI use vary according to model type. Open-source AI offers transparency, supports research integrity through reproducibility, preserves participant data sovereignty via local deployment, and enables non-Western or decolonized adaptations (Bender et al., 2021). However, it carries risks such as weaker safety controls, higher misuse potential, and amplified bias without guardrails. The ethical responsibility largely falls on the researcher, necessitating strong competence and institutional oversight. For-profit AI, in contrast, embeds safety protections, lowers the barrier to ethical use, and provides consistency across users through clear terms of service. Its ethical limitations include opacity in decision-making, uncertain data governance, potential reinforcement of institutional or cultural biases, and dependence on vendor platforms. Consequently, ethical use in for-profit AI emphasizes compliance and risk avoidance rather than methodological critique (OpenAI, 2023b). Table 1 summarizes the comparative ethical dimensions.

Table 3. Ethical Implications of AI Use in Academic Research

Dimension	**Open-source AI**	**For-profit AI**
Transparency	High	Low
User responsibility	High	Lower
Participant protection	Strong (if local)	Mixed
Bias control	Researcher-managed	Platform-managed
Academic freedom	High	Moderate

What are the teaching, publishing, or qualitative analysis workflows for each AI type?. Mapping AI applications to academic workflows clarifies their appropriate roles. In teaching, for-profit AI such as ChatGPT, Claude, or Gemini supports lesson planning, learning activity design, rubric creation, assessment generation, and scaffolds for differentiated instruction, provided that AI is used to augment teacher work rather than replace it. In publishing, for-profit AI facilitates structuring articles, enhancing academic tone, editing for clarity and coherence, drafting abstracts, and preparing reviewer responses, while open-source AI can provide secondary support for reproducible editing workflows. In qualitative research, open-source AI is primary for first-cycle coding, thematic clustering, memo writing, pattern identification, and translation (with verification), while for-profit AI assists in articulating emergent themes in academic prose. A typical workflow might involve using a local open-source model to code transcripts, with the researcher reviewing and refining codes, and employing a for-profit system to produce polished write-ups; final interpretation remains fully human-led.

A practical decision guide reinforces this hybrid approach. Open-source AI should be used when handling sensitive or ethics-reviewed data, when transparency and replicability are essential, or when methodological control is prioritized. For-profit AI is preferable for tasks that demand rapid generation of polished academic language, teaching materials, or clear communication outputs. Ultimately, open-source AI strengthens research integrity, while for-profit AI enhances academic communication. Ethical academic use lies not in the choice of tool alone but in the rigor with which researchers document, govern, and disclose AI-assisted processes (Bender et al., 2021; OpenAI, 2023a; OpenAI, 2023b).

What are the ways to manage risks of open source and for profit generative AI applications?

Managing risks in generative AI, whether open-source or for-profit, requires governance, documentation, and human oversight rather than reliance on the choice of a particular model. AI should be treated as a

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

methodological aid rather than an autonomous actor, with researchers or instructors retaining ultimate responsibility for ethical and accurate outputs (Bender et al., 2021; OpenAI, 2023a).

Open-source AI shifts substantial control to the researcher, granting both flexibility and increased responsibility. A major area of concern is data privacy and confidentiality. Local deployment of open-source models may involve processing sensitive participant information, such as interview transcripts or student data, without built-in safeguards. Improper storage or handling can expose participants to re-identification risks. Best practices include anonymizing data prior to AI processing, storing datasets on encrypted, access-controlled devices, limiting AI computation to secure offline environments, and explicitly prohibiting cloud syncing in research protocols (Touvron et al., 2023). Open-source AI should be treated like a human research assistant bound by confidentiality.

Bias and interpretive distortion are additional challenges. Open-source models inherently reflect biases in their training data, and fine-tuning may inadvertently amplify researcher assumptions. Outputs may appear objective even when they are not. To manage these risks, AI should be used primarily for exploratory or first-cycle coding. Researchers should cross-check AI-generated codes with manual coding, maintain reflexive memos documenting human decisions, and involve multiple reviewers where possible (Bender et al., 2021). AI assists in pattern detection rather than meaning-making.

Methodological opacity is another concern with open-source AI. Failure to document model versions, prompts, or parameters can compromise replicability. Researchers should maintain logs of model name and version, prompt templates, and parameter settings, explicitly include AI use in methodology sections, and archive configurations alongside research data (Touvron et al., 2023). Transparency serves as the ethical counterweight to flexibility.

Open-source AI also poses risks of misuse and scope creep. Models may be repurposed beyond approved research protocols, or students and assistants may use AI without supervision. Management strategies include defining explicit AI-use boundaries in protocols, requiring training for team members, and auditing AI-assisted outputs regularly.

For-profit generative AI applications reduce some risks through built-in safety and moderation, but they introduce new dependencies and opacity. Data governance and platform reliance pose notable risks: unclear retention policies, vendor lock-in, and cross-jurisdictional storage issues. These risks are mitigated by avoiding the upload of raw sensitive data, using de-identified or synthetic datasets, reviewing terms of service regularly, and preferring enterprise or education plans with stronger data controls (OpenAI, 2023a). Researchers should assume that anything uploaded may persist beyond their control.

The black-box nature of proprietary AI systems also presents interpretability challenges. Training data and safety interventions are generally uninspectable, making it difficult to explain AI behavior to reviewers. Risk management strategies include limiting AI use to editing, summarization, and structural support, avoiding proprietary AI for core analytic decisions, and clearly stating AI limitations in disclosures (OpenAI, 2023b). Closed systems are appropriate when interpretability is not central to the methodology.

Overreliance on for-profit AI can erode scholarly voice, homogenize academic writing, and encourage researchers or students to defer judgment to AI. To address this, human-authored claims should remain mandatory, AI outputs should serve only as drafts, and educational exercises should integrate AI critique. Clear authorship accountability rules reinforce responsible AI use.

Regulatory and compliance uncertainty is another area of concern. Policies may change, features may be removed, or regional legal conflicts may arise. Strategies include maintaining non-AI backups of workflows, avoiding exclusive reliance on a single vendor, and monitoring institutional guidance and regulations.

The comparative risk management matrix (Table 1) highlights distinctions between open-source and for-profit AI applications. Open-source AI requires higher user responsibility for privacy, bias, transparency, legal exposure, and governance, whereas for-profit AI offers managed safety layers and lower skill requirements but less transparency.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Table 4. Comparative Risk Management Matrix for Generative AI

Risk Area	**Open-source AI**	**For-profit AI**
Privacy	User-managed (high responsibility)	Platform-managed (limited control)
Bias	Visible but unmanaged	Managed but opaque
Transparency	High	Low
Legal exposure	On researcher	Shared with provider
Skill requirement	High	Moderate
Governance need	Explicit protocols	Usage policies

At the institutional level, effective risk management for both AI types requires formal policies, including AI ethics guidelines, mandatory disclosure statements, AI literacy training, and updates to IRB or ethics review frameworks. Institutions should distinguish among AI as an analytic tool, as an editorial aid, or as instructional support, ensuring that human oversight governs all stages of AI-assisted work.

In summary, open-source AI risks are mitigated through discipline, documentation, and methodological rigor, whereas for-profit AI risks are managed through restraint, disclosure, and careful selection of appropriate tasks. Ethical use of AI in academia depends less on the specific model chosen and more on human governance, training, and transparency practices (Bender et al., 2021; OpenAI, 2023a; OpenAI, 2023b; Touvron et al., 2023).

What human governance AI-use protocols and policy can be recommended for open source and for-profit AI applications?

Human governance in AI use refers to the set of rules, roles, and decision-making processes designed to ensure that artificial intelligence remains a supporting tool rather than an autonomous authority. In practice, this means that AI outputs are accountable to human judgment and aligned with ethical, legal, and academic standards (Floridi et al., 2020; OpenAI, 2023a). Human governance ensures that AI may assist decision-making, research, or teaching, but ultimate responsibility always resides with a named human actor.

Core principles of AI-use governance apply across open-source and for-profit applications. These include human accountability, transparency, proportionality, oversight, and contestability. Human accountability requires a responsible individual to be clearly identified for AI-assisted outputs. Transparency involves documenting AI use in research, teaching, or publishing. Proportionality ensures that the level of AI assistance corresponds appropriately to the sensitivity of the task. Oversight requires that all AI outputs be reviewed and validated by humans, and contestability guarantees that AI-generated content can be questioned or rejected when necessary (Jobin et al., 2019).

Governance protocols for open-source AI applications emphasize control and procedural rigor, reflecting the maximum flexibility and responsibility afforded to researchers. Authorization and access control policies require prior institutional approval for any AI deployment, limiting usage to trained personnel. Researchers must maintain a registry of approved models, including documentation of version, deployment environment, and intended purpose. Data handling and protection policies mandate that sensitive data never leave secure environments. Protocols include anonymization before AI processing, encryption of stored datasets, and strictly local deployment when human subjects are involved (Touvron et al., 2023).

Methodological governance for open-source AI reinforces that AI is an analytic support tool rather than an interpretive authority. AI outputs should be treated as preliminary, with human validation required for coding schemes, thematic development, and interpretive claims. Reflexive memos must document any influence AI

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

has on analysis decisions. Documentation and auditability policies further require that all AI-assisted processes be logged, including prompts, parameters, and model versions, with interaction logs archived alongside research data. Training and competence standards ensure that users of open-source AI have completed mandatory instruction in bias recognition, prompt governance, and ethical data handling (Bender et al., 2021).

Governance protocols for for-profit AI applications emphasize standardization and risk containment, with oversight focused on permitted uses rather than full procedural control. Approved-use policies clearly define the tasks commercial AI may perform, such as editing, summarization, instructional support, and non-sensitive drafting, while restricting high-stakes decisions, primary data analysis, and interpretation of sensitive human data. Data exposure safeguards require de-identification of human data, use of synthetic examples, and preference for enterprise or education accounts with strong privacy protections (OpenAI, 2023a).

Transparency and disclosure policies mandate that AI assistance be acknowledged in academic outputs, including methodology sections, course syllabi, and institutional reports. Dependency and authorship controls ensure that AI is not listed as an author, that human authors retain responsibility, and that automated outputs—including citations—are verified. Vendor governance requires institutions to review AI provider terms of service, data retention policies, and jurisdictional compliance at least annually (Floridi et al., 2020).

The **comparative governance focus** highlights differences between open-source and for-profit AI (Table 1). Open-source AI demands high control and intensive oversight, with the researcher bearing the primary risk and responsibility for internal data governance. For-profit AI emphasizes use limitation, with oversight moderated by platform policies and contractual arrangements, offering less transparency but lowering the individual skill requirement.

Table 5. Comparative Governance Focus for Generative AI

Governance Area	**Open-source AI**	**For-profit AI**
Control	High	Low
Risk locus	Researcher	Platform
Oversight need	Intensive	Moderate
Data governance	Internal	Contractual
Transparency	High	Limited
Policy emphasis	Process control	Use limitation

In practical academic settings, human governance is applied across teaching, research, and publishing. In teaching, students should submit AI-use declarations, permissible and prohibited uses should be defined, and AI literacy should be integrated into coursework. In research, AI-use protocols must be included in ethics review, audit trails maintained, and analytic support separated from interpretation. In publishing, journals' AI disclosure requirements should be followed, AI may assist with language support but not intellectual contribution, and author accountability must be retained (OpenAI, 2023b).

An effective **institutional governance framework** for AI should include classification of AI tools (open-source versus for-profit), risk tiers (low, medium, high), matrices of approved uses, mandatory disclosure statements, training and certification requirements, oversight committees or review boards, and sanctions for misuse. Open-source AI demands strong procedural governance, while for-profit AI requires clear boundary governance. Across all applications, human governance ensures AI remains a tool rather than a decision-maker, reinforcing that ethical AI use in academia is ultimately a human responsibility (Floridi et al., 2020; OpenAI, 2023a; OpenAI, 2023b; Touvron et al., 2023).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

What institutional governance framework can be recommended for Catholic higher education schools?

I. Preamble

Catholic higher education institutions are committed to the pursuit of truth (veritas), the formation of the whole person, and the service of the common good. As generative artificial intelligence (AI) technologies increasingly influence teaching, research, administration, and knowledge production, Catholic institutions are called to govern their use in a manner consistent with Gospel values, the Social Teachings of the Church, and the mission of Catholic education.

This framework provides a principled, human-centered governance structure for the ethical, responsible, and mission-aligned use of open-source and for-profit generative AI applications within Catholic higher education.

II. Theological and Ethical Foundations

AI governance in Catholic higher education is grounded in the following principles:

1. ***Human Dignity** – AI must serve the human person and never replace human moral agency.*
2. ***Common Good** – AI use should contribute to social responsibility, inclusion, and justice.*
3. ***Subsidiarity** – Decisions about AI use should be made at the most local competent level, with appropriate oversight.*
4. ***Solidarity** – AI should not exacerbate inequality, exclusion, or epistemic injustice.*
5. ***Stewardship** – Institutions are responsible for prudent, ethical management of emerging technologies.*

III. Scope and Applicability

This framework applies to all institutional uses of generative AI, including but not limited to:

- *Teaching and learning*
- *Research and knowledge production*
- *Administrative and operational functions*
- *Student support services*
- *External communications and publications*

The framework covers both:

- ***Open-source AI applications** (locally deployed, modifiable models)*
- ***For-profit AI applications** (commercial, platform-based systems)*

IV. Governance Structure

A. Institutional AI Governance Council

*Each institution shall establish an **AI Governance Council** composed of:*

- *Academic administrators*

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- *Faculty representatives*
- *IT and data protection officers*
- *Ethics or mission officers*
- *Legal or compliance officers*
- *Student representation (where appropriate)*

Responsibilities:

- *Approve institutional AI policies*
- *Classify AI tools by risk level*
- *Review emerging AI technologies*
- *Oversee compliance and ethical alignment*

B. AI Risk Classification Framework

AI applications shall be classified into three risk tiers:

1. ***Low Risk** – Editing, summarization, instructional support*
2. ***Moderate Risk** – Data analysis support, curriculum design, advising tools*
3. ***High Risk** – Human subjects research, automated decision-making, assessment of persons*

Higher-risk uses require heightened review and authorization.

V. Governance Protocols for Open-Source AI

A. Authorization and Accountability

- *Open-source AI deployment requires prior institutional approval*
- *A named faculty or staff member shall be accountable for AI use*

B. Data Protection and Confidentiality

- *Sensitive data must be anonymized prior to AI processing*
- *Local or secured environments are mandatory for human-subject data*
- *Compliance with data privacy laws and Church ethical standards is required*

C. Research and Academic Integrity

- *AI may support data organization and pattern identification*
- *Interpretive judgment remains a human responsibility*
- *Reflexive documentation of AI use is mandatory*

D. Documentation and Auditability

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- *Model versions, prompts, and parameters must be logged*
- *AI-assisted processes must be reproducible and reviewable*

VI. Governance Protocols for For-Profit AI

A. Approved Use Boundaries

Permissible uses include:

- *Language editing and proofreading*
- *Lesson planning and instructional design*
- *Non-sensitive drafting and summarization*

Prohibited uses include:

- *Uploading identifiable personal or student data*
- *Sole reliance on AI for evaluative or disciplinary decisions*

B. Vendor and Platform Oversight

- *Platforms must be reviewed for data retention and privacy practices*
- *Preference given to education or enterprise agreements*
- *Annual vendor compliance reviews are required*

C. Transparency and Disclosure

- *AI assistance must be disclosed in research, instruction, and publications*
- *Standardized AI disclosure statements shall be adopted institution-wide*

VII. Teaching and Learning Governance

- *Faculty shall define permissible AI use in course syllabi*
- *Students must disclose AI assistance where required*
- *AI literacy and ethical use shall be integrated into curricula*
- *AI may not replace formative human mentorship and evaluation*

VIII. Research Governance

- *AI use must be declared in research proposals and ethics reviews*
- *Human subjects research requires explicit AI-use approval*
- *AI shall not be treated as an author or intellectual agent*

IX. Institutional Accountability and Formation

- *Regular training in AI ethics and governance is mandatory*

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- *Institutions shall promote critical reflection on AI and human values*
- *Violations of AI policy are subject to existing academic and disciplinary processes*

X. Continuous Review and Discernment

In keeping with the Catholic tradition of discernment, this framework shall be:

- *Reviewed regularly in light of technological developments*
- *Assessed for alignment with institutional mission*
- *Updated through consultative and participatory processes*

XI. Concluding Statement

Generative AI, when governed wisely, can support the mission of Catholic higher education by enhancing learning, advancing knowledge, and serving society. This framework affirms that technology must remain at the service of the human person, guided by faith, reason, and moral responsibility.

How this framework is designed to help Catholic schools

- ***Mission-aligned****: Explicitly grounded in Catholic Social Teaching (human dignity, common good, subsidiarity, solidarity, stewardship)*
- ***Administratively usable****: Structured like an actual institutional policy (preamble, scope, governance bodies, protocols)*
- ***Technologically realistic****: Distinguishes clearly between* ***open-source*** *and* ***for-profit*** *AI*
- ***Ethics-board ready****: Suitable for IRB / research ethics review integration*
- ***Curriculum-compatible****: Includes teaching, research, and formation components*

How should the said policy be revised in the context where AI use becomes ubiquitous?

As artificial intelligence becomes increasingly ubiquitous, its integration into academic and administrative workflows is anticipated to transform routine practices. AI tools are likely to be embedded in a wide range of tasks, including writing, data analysis, coding, teaching, and administrative operations. Researchers and students may interact with AI continuously, often in ways that are largely invisible to the user, such as automatic summarization, auto-referencing, or language refinement (OpenAI, 2023a). This pervasive integration has significant implications for institutional policy: it will be necessary to distinguish between incidental AI assistance, such as spellchecking or minor grammatical corrections, and substantive AI contributions, including coding, thematic analysis, or interpretation of qualitative data.

The increasing ubiquity of AI also redefines the boundaries between human and AI contributions. While current frameworks emphasize human primacy with AI as a support tool, widespread AI adoption will frequently augment human creativity and analytical reasoning (Floridi et al., 2020). Governance protocols must therefore delineate levels of AI influence. Minor AI contributions may include spelling, formatting, or summarization; moderate contributions may encompass structuring content or synthesizing information; and major contributions may involve interpretation, pattern recognition, or drafting analytic conclusions. Although human oversight remains non-negotiable, disclosure thresholds may require adjustment to reflect the nature and depth of AI involvement.

At scale, the cumulative use of AI introduces additional risks, including increased exposure of sensitive data, propagation of biases, and standardization of outputs that may homogenize scholarly work. Monitoring every

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

AI interaction individually is impractical, necessitating a shift from case-by-case approval to standardized institutional guidelines and tiered risk management. Encouraging self-reporting and maintaining audit logs are recommended practices to manage these risks effectively (Touvron et al., 2023).

The anticipated changes in AI use also impact how AI contributions are disclosed in research. Currently, researchers are required to explicitly document AI use in methodology sections, acknowledgments, or writing assistance notes, with open-source AI often requiring more detailed disclosure due to its potential influence on analytic decisions. In a future with ubiquitous AI, declarations should be more nuanced, differentiating incidental from substantive use. Incidental AI use, such as grammar correction or reference formatting, might only require a generic note in the methodology section. Substantive AI use—such as coding, theme identification, data synthesis, or interpretive analysis—should be accompanied by detailed documentation including the model name and version, prompt and parameter specifications, and evidence of human validation. Standardized AI-use reporting forms, akin to conflict-of-interest disclosures, may become necessary to maintain transparency and accountability (Bender et al., 2021).

Ethical considerations remain paramount. Widespread AI adoption does not diminish human responsibility; researchers remain accountable for the accuracy, interpretation, and conclusions of their work. Documenting how AI shaped analytic or interpretive outputs is essential to maintain epistemic integrity and mitigate bias. Furthermore, equity and access considerations are important, as differential access to AI resources could exacerbate existing disparities. Transparent disclosure practices can help highlight such systemic inequities and reinforce ethical research standards.

In response to these anticipated changes, several revisions to existing AI governance frameworks are recommended. Policies should introduce a tiered AI-use declaration system distinguishing incidental, moderate, and major contributions. Human oversight must remain explicit, even when AI is seamlessly integrated into workflows. Institutions should maintain standardized AI-use logs or registries to track researcher engagement with AI tools. Governance protocols should transition from pre-approval of every AI interaction to continuous compliance audits for low-risk applications, with review committees focusing on moderate-to-high-risk AI tasks. Training programs must be updated to include best practices for ongoing AI augmentation in research and teaching contexts.

Research declarations should clearly identify the level of AI involvement, specify the model and version used, and describe human validation procedures. Statements should also include reference to ethical oversight, affirming that the human researcher retains ultimate responsibility. For example, a declaration might read: *"In this study, AI tools were used at multiple stages. Grammarly and ChatGPT (v5-mini) assisted in language refinement (incidental). An open-source AI model (LLaMA 3) was used for preliminary coding of qualitative data (substantive). All AI outputs were reviewed, validated, and interpreted by the research team. AI was not involved in final analytic conclusions, which remain the responsibility of the authors."*

As such, the ubiquity of AI necessitates nuanced disclosure practices rather than binary statements of AI use. Human governance remains central, but review processes should transition from micro-level approvals to tiered oversight, structured documentation, and continuous auditability. Institutional frameworks should emphasize transparency, human accountability, and alignment with ethical and mission-driven standards, ensuring that AI remains a tool to support, rather than replace, human judgment (Floridi et al., 2020; OpenAI, 2023a; Touvron et al., 2023).

Below is the tabular presentation of revisions that should be done to the previous draft:

Table 6. Revisions Considering AI Ubiquity

Section	Original Version	Revised Version (Ubiquitous AI)	Key Changes / Notes
Preamble	AI governed in context of Catholic values; general	Explicitly notes AI ubiquity; acknowledges pervasive AI	Added context for routine AI use; emphasizes need for

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

	reference to teaching, research, administration	integration in workflows	nuanced oversight
Theological & Ethical Foundations	Principles: human dignity, common good, subsidiarity, solidarity, stewardship	Same principles, with explicit note that governance must address AI ubiquity	Reframed to account for pervasive AI and continuous human responsibility
Scope & Applicability	Applies to teaching, research, admin, student support; covers open-source and for-profit AI	Same scope, plus explicit acknowledgment of **AI integration at scale and daily workflows**	Recognizes incidental vs. substantive AI use in daily tasks
Governance Structure	AI Governance Council to approve policies, classify tools, review tech, oversee compliance	Same council, plus explicit responsibility for **continuous AI use practices** and **tiered AI-use disclosure**	Added tiered disclosure responsibility and ongoing oversight for ubiquitous AI
AI Risk Classification	Low, Moderate, High risk, defined by task	Expanded definitions: Low = routine assistance, Moderate = workflow influence, High = human-subject research / evaluative tasks	Added specificity to distinguish levels of AI influence and oversight requirements
Tiered AI-Use Declarations	Not present in original	Introduced tiered declarations: Incidental, Moderate, Major/Substantive, with disclosure requirements	Major addition; distinguishes types of AI contributions and documentation expectations
Open-Source AI Protocols	Authorization, accountability, data protection, research integrity, documentation, training	Maintains original sections; adds emphasis on **logging AI interactions**, **auditability**, and tiered declaration	Updates reflect need for continuous documentation and oversight with pervasive AI use
For-Profit AI Protocols	Approved use boundaries, vendor oversight, transparency, human accountability	Adds tiered AI-use disclosure, continuous workflow audit, periodic vendor review; emphasizes human accountability despite ubiquitous use	Emphasizes documentation and oversight for widespread AI usage; explicit tiered disclosures
Teaching & Learning Governance	Faculty define AI use in syllabi; students disclose; AI literacy integrated; mentorship maintained	Adds tiered risk and declaration categories for students and faculty; reflects routine AI use	Policy now accounts for frequent, low-risk AI assistance without requiring full approval
Research	AI use declared in	Adds tiered disclosure system;	Introduces nuanced

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Governance	proposals, ethics review; human responsibility emphasized; AI not authors	specifies incidental/moderate/substantive AI use; logs and documentation required; human interpretation remains final	declaration based on AI contribution levels
Institutional Accountability	Mandatory training, reflection on AI, policy violations subject to academic procedures	Adds emphasis on continuous reflection and tracking cumulative AI influence in workflows	Updates reflect ongoing governance in ubiquitous AI context
Continuous Review & Discernment	Reviewed regularly in light of technology; updated via consultative processes	Same, with added focus on AI ubiquity, tiered disclosure refinement, oversight protocols	Explicitly addresses challenges of pervasive AI use
Concluding Statement	AI governed prudently; supports mission, learning, knowledge; human responsibility emphasized	Reframed for ubiquitous AI; AI as tool, human responsibility central; highlights tiered governance and transparency	Strengthened to acknowledge pervasiveness of AI and continued human accountability

Summary of overall changes:

- Added **tiered AI-use declarations** (incidental / moderate / substantive) across research, teaching, and publication.
- Emphasized **continuous oversight, logging, and auditability** due to pervasive AI use.
- Clarified **nuanced disclosure requirements** in methodology and publications.
- Adjusted governance language to reflect AI as **ubiquitous, routine, and integrated**, rather than exceptional.
- Minor refinements to training, accountability, and risk classification to ensure human responsibility remains central.

How can the above policy revision be modified to enable the development of a multisectoral and democratized AI governance model that will counter centralized global governance subject to authoritarianism and abuse?

The evolution of AI governance policy for St. Paul University Manila reflects a progression from an institution-centered, Catholic values–driven framework toward a model that accounts for ubiquitous AI integration and emphasizes democratized, multisectoral oversight. The original preamble focused on the alignment of AI governance with Catholic principles and its application in teaching, research, and administration. In the first revision, the framework was updated to address the reality of AI becoming a routine component of academic and administrative workflows, highlighting the need for nuanced oversight that differentiates between incidental and substantive AI contributions. The democratized version further expanded this focus, stressing the importance of participatory, multisectoral governance designed to counter centralized or potentially authoritarian control, thereby ensuring that AI governance remains inclusive, transparent, and aligned with both ethical Catholic and democratic principles.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Theological and ethical foundations were similarly reinterpreted. Initially, principles such as human dignity, the common good, subsidiarity, solidarity, and stewardship guided the policy. With the recognition of AI ubiquity, these principles were reframed to emphasize continuous human responsibility. The democratized version incorporated an explicit principle of democratization, ensuring ethical oversight by diverse stakeholders and preventing centralization of decision-making, thereby reinforcing the ethical and participatory dimensions of governance.

The scope and applicability of the policy likewise expanded. Originally focused on teaching, research, administration, and student support, covering both open-source and for-profit AI, the revised frameworks acknowledged the routine and pervasive use of AI in academic workflows. The democratized approach extended this further by encouraging multi-institutional collaboration, transparency, and shared responsibility, promoting cross-sector participation to mitigate risks associated with centralized control or authoritarian tendencies.

Governance structures were similarly enhanced. The original AI Governance Council approved policies, classified tools, and oversaw compliance. In the first revision, the council's responsibilities were extended to include continuous AI use practices and tiered AI-use disclosures. The democratized framework added civil society participation, external collaborators, and regional networks for peer oversight, ensuring that decision-making remained distributed and accountable, reducing the concentration of authority.

Risk classification was refined in parallel. Whereas the original system defined low, moderate, and high-risk AI based on task type, the revisions expanded these definitions to account for the influence of AI on workflows, including human-subject research and decision-making. The democratized framework emphasized multisectoral review for moderate- and high-risk applications, ensuring that higher-risk AI deployments were overseen by diverse stakeholders rather than a single institutional authority.

Tiered AI-use declarations, absent in the original policy, were introduced to categorize AI use as incidental, moderate, or major/substantive, with corresponding disclosure requirements. In the democratized model, stakeholder involvement in documentation and validation was added, thereby embedding participatory oversight and transparency, particularly for substantive AI use.

Protocols for open-source and for-profit AI were likewise updated. Original open-source protocols addressed authorization, accountability, data protection, research integrity, documentation, and training. These were expanded in the first revision to include logging interactions, auditability, and tiered disclosure. The democratized framework further incorporated multisectoral approval, external review options, and shared audit access, ensuring participatory governance and reducing the risk of misuse by a single actor. Similarly, for-profit AI protocols, initially focused on approved use boundaries, vendor oversight, transparency, and human accountability, were expanded to include tiered disclosure, continuous workflow auditing, periodic vendor review, and democratic oversight of vendors, thereby reducing the potential for opaque, centralized influence and aligning use with ethical and mission-aligned standards.

In teaching, learning, and research, the revised frameworks accounted for tiered risk and disclosure, routine AI use, and participatory governance. In the democratized approach, stakeholder engagement, community input, and democratic literacy were integrated, promoting participatory learning and oversight while preparing students and faculty to engage responsibly with AI technologies. Research governance protocols similarly emphasized multisectoral stakeholder review and external collaborator input, ensuring that outputs were ethically validated, transparent, and accountable beyond the institution. Institutional accountability mechanisms were strengthened through shared responsibility, continuous reflection, and cumulative AI influence tracking, mitigating risks of centralization and enhancing transparency.

Finally, continuous review and discernment processes were adapted to account for ubiquitous AI use. The democratized policy emphasized multisectoral review, cross-institutional coordination, and democratic evaluation, ensuring that policy refinement remained participatory, transparent, and responsive rather than concentrated in a single authority. The concluding statement, which originally stressed prudent AI governance aligned with mission and human responsibility, was revised to highlight democratic oversight, multisectoral

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

participation, and safeguards against centralization, framing participatory governance as essential for ethical, equitable, and mission-aligned AI use.

Table 7. Summary of Policy Revisions towards Greater Democratization

Section	Original Version	First Revision (Ubiquitous AI)	Democratized Version	Justification for Democratized Version Changes
Preamble	Focus on Catholic values, AI governance in teaching, research, administration	Added context for AI ubiquity and routine integration; need for nuanced oversight	Emphasizes multisectoral, participatory governance to counter centralized or authoritarian AI control	To ensure AI governance is inclusive, transparent, and resistant to concentration of power, aligning with ethical Catholic and democratic principles
Theological & Ethical Foundations	Human dignity, common good, subsidiarity, solidarity, stewardship	Reframed for AI ubiquity and continuous human responsibility	Adds **democratization** principle; multisectoral oversight	Democratization principle ensures ethical oversight by diverse stakeholders and prevents centralization of decision-making
Scope & Applicability	Applies to teaching, research, admin, student support; open-source & for-profit AI	Notes AI integration at scale; distinguishes incidental vs substantive use	Adds encouragement for **multi-institutional collaboration and transparency**, beyond routine use	Promotes cross-sector participation and shared responsibility to counter authoritarian tendencies
Governance Structure	AI Governance Council approves policies, classifies tools, reviews technology, oversees compliance	Adds responsibility for continuous AI use practices and tiered AI-use disclosure	Includes civil society, external collaborators; emphasizes multisectoral review, regional networks for peer oversight	Broader stakeholder involvement prevents centralization; regional networks enhance accountability and shared governance
AI Risk Classification	Low, moderate, high risk based on task	Expanded definitions: low = routine, moderate = workflow influence, high = human-subject research/decision-	Same tiers; adds **requirement for multisectoral review for moderate/high risk**	Ensures higher-risk AI applications are overseen by diverse stakeholders, reducing concentration of authority

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

		making		
Tiered AI-Use Declarations	Not present	Introduced: incidental, moderate, major/substantive with disclosure	Maintains tiers; adds **stakeholder involvement in documentation and validation**	Democratizes oversight and ensures transparency, particularly for substantive AI use
Open-Source AI Protocols	Authorization, accountability, data protection, research integrity, documentation, training	Adds logging AI interactions, auditability, tiered disclosure	Adds **multisectoral approval, external review options**, and shared audit access	Ensures open-source AI use is subject to participatory governance and reduces risk of misuse by a single actor
For-Profit AI Protocols	Approved use boundaries, vendor oversight, transparency, human accountability	Adds tiered AI-use disclosure, continuous workflow audit, periodic vendor review	Adds **democratic oversight of vendors**, transparency to external stakeholders, and cumulative impact tracking	Reduces risk of centralized vendor influence or opaque decisions; aligns with ethical and mission-aligned accountability
Teaching & Learning Governance	Faculty define AI use in syllabi; students disclose; AI literacy; mentorship	Adds tiered risk/disclosure categories; routine AI use permitted	Adds **stakeholder engagement, community input, and democratic literacy**	Promotes participatory learning and oversight; prepares students to understand democratic AI governance
Research Governance	AI use declared in proposals, ethics review; human responsibility; AI not authors	Adds tiered disclosure; documentation for moderate/substantive use	Adds **multisectoral stakeholder review**, community or external collaborator input	Ensures research outputs are ethically validated, transparent, and accountable beyond the institution, countering authoritarian misuse
Institutional Accountability	Training, reflection, policy violations	Adds continuous reflection, cumulative AI influence tracking	Adds **stakeholder participation in accountability mechanisms**	Shared responsibility increases transparency and reduces risk of centralized abuse

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Continuous Review & Discernment	Regular review; consultative updates	Adds AI ubiquity, tiered disclosure refinement, oversight protocols	Adds **multisectoral review, cross-institutional coordination, democratic evaluation**	Ensures continuous improvement is participatory and accountable, not concentrated in a single authority
Concluding Statement	AI governed prudently; supports mission; human responsibility	Reframed for ubiquitous AI; tiered governance and transparency emphasized	Adds **democratic oversight, multisectoral participation, safeguards against centralization**	Highlights participatory governance as essential for ethical, equitable, and mission-aligned AI use

Summary of Democratized Changes:

- Added **democratization principle** to ethical foundations.
- Expanded governance councils to include **students, civil society, and external collaborators**.
- Introduced **regional/institutional networks** for peer oversight.
- Required **multisectoral review for moderate and high-risk AI use**.
- Added **shared audit access, stakeholder input, and transparency mechanisms** across teaching, research, and vendor interactions.
- Purpose: To **counter centralized power, reduce risk of abuse, and ensure participatory, mission-aligned AI governance** in Catholic higher education.

How can the framework be localized for a Paulinian Catholic university?

St. Paul University Manila AI Governance Policy and Phased Implementation Framework

I. Executive Policy Summary

Purpose

To ensure the ethical, responsible, and mission-aligned use of generative artificial intelligence (AI) across teaching, research, administration, and communications at St. Paul University Manila (SPU Manila), grounded in Catholic Social Teaching, the Paulinian charism, and the principles of *Ex Corde Ecclesiae*.

Scope

This policy applies to all AI applications—open-source, institutionally developed, and for-profit systems—used in any university function.

Key Principles

- **Human Dignity** – AI must serve authentic human development and never replace moral agency or conscience.
- **Common Good & Solidarity** – AI use should promote inclusion, justice, and equitable access.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- **Subsidiarity** – AI-related decisions are made at the most competent and local level, with appropriate oversight.
- **Democratization** – Multisectoral governance prevents centralized, technocratic, or authoritarian misuse.
- **Stewardship** – Responsible management, transparency, and accountability in all AI deployment.

Governance Structure

- **AI Governance Council**
  - President (Chair or Delegate)
  - Vice President for Academics and Research
  - Faculty representatives
  - IT and Data Protection Officers
  - Ethics/Mission Officers
  - Student representatives
  - Community/External representatives
- **Regional and Network Collaboration**
  - Coordination with other Paulinian institutions for shared standards, peer review, and best-practice exchange.

AI Risk Classification

- **Low Risk** – Routine support tasks (grammar checking, formatting)
- **Moderate Risk** – Workflow or cognitive influence (data organization, lesson planning)
- **High Risk** – Human-subject research, evaluative decisions, policy formulation

Tiered AI-Use Declaration

- **Incidental Use** – Minor assistance; generic disclosure
- **Moderate Use** – Significant assistance; model/version disclosure and human verification
- **Major/Substantive Use** – Core analysis or interpretation; full documentation and stakeholder review

Implementation & Compliance

- Mandatory AI literacy and ethics training
- AI-use logs maintained and subject to audit
- Violations subject to disciplinary, academic, and ethical review

Alignment

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Fully aligned with *Ex Corde Ecclesiae*, emphasizing formation of the whole person, academic integrity, truth-seeking (*veritas*), and fidelity to the Catholic mission.

II. St. Paul University Manila AI Governance Framework (Localized)

A. Preamble

St. Paul University Manila affirms its commitment to ethical AI use in service of human dignity, communal welfare, and holistic formation, consistent with *Ex Corde Ecclesiae* and the Paulinian mission of education as service.

B. Governance Structure

- Multisectoral **AI Governance Council**
- Cross-campus and regional Paulinian collaboration
- Oversight of AI use in teaching, research, administration, and communications

C. Protocols

Open-Source AI

- Prior authorization and multisectoral approval
- Mandatory audit logs
- Human validation of outputs

For-Profit AI

- Vendor transparency and ethical review
- Tiered AI-use declarations
- Clear human accountability

Teaching and Learning

- Course-level AI-use policies
- Student disclosure requirements
- AI literacy integrated into curricula
- Human mentoring and evaluation remain central

Research

- AI-use declaration in proposals, ethics reviews, and publications
- Documentation of models, prompts, outputs, and validation
- Mandatory human interpretation and accountability

Institutional Accountability

- Ongoing training and formation

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- Continuous ethical reflection
- Proportionate disciplinary measures for non-compliance

D. Continuous Review

- Periodic review with multisectoral participation
- Updates responding to AI developments, democratization concerns, and mission alignment

III. Phased Implementation Roadmap

Phase I: Foundational Implementation (Short-Term: Year 1)

Governance Actions

- Constitute and formally mandate the AI Governance Council
- Approve initial AI Risk Classification and Tiered AI-Use Declaration system
- Issue interim guidelines for faculty, researchers, and students

Capacity Building

- Launch mandatory AI literacy and ethics orientation for faculty, staff, and students
- Develop standard AI-use disclosure templates and logging mechanisms

Policy Integration

- Embed AI governance clauses into syllabi templates, research protocols, and administrative workflows
- Begin pilot audits of AI use in selected units

Outcome

- Baseline ethical awareness
- Operational governance structures
- Institution-wide readiness without overregulation

Phase II: Institutionalization and Scaling (Medium-Term: Years 2–3)

Governance Actions

- Refine risk classifications and tiered declarations based on implementation data
- Establish standing subcommittees (Teaching, Research, Administration, Ethics)
- Formalize partnerships with Paulinian and Catholic higher education networks

Operational Expansion

- Integrate AI governance into:
  - Faculty evaluation and development

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- Research ethics review processes
- Quality assurance systems

Monitoring and Accountability

- Implement regular AI-use audits
- Publish annual AI Governance Reports to the University Board
- Introduce case-based ethical review mechanisms

Outcome

- Embedded governance culture
- Evidence-based refinement
- Strong accountability and transparency mechanisms

Phase III: Maturation and Mission Leadership (Long-Term: Years 4–5 and Beyond)

Governance Actions

- Position SPUM as a regional leader in Catholic AI ethics and governance
- Develop institutional AI ethics research and policy centers
- Contribute to national and ecclesial AI discourse

Sustainability and Innovation

- Continuous updating of governance frameworks
- Encourage ethically aligned AI innovation projects
- Institutionalize formation-focused AI ethics education

Mission Integration

- Evaluate AI governance impact on formation, equity, and academic integrity
- Strengthen AI governance as part of SPUM's Catholic identity and public witness

Outcome

- Mature, mission-centered AI governance ecosystem
- Institutional credibility and leadership
- Long-term ethical resilience

IV. Implementing Guidelines / Checklists

Faculty

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

- Define AI use per course using tiered categories
- Teach AI ethics and responsible use
- Maintain human-led mentoring and evaluation
- Record AI use for transparency and audit

Researchers

- Declare AI use in proposals, ethics reviews, and publications
- Maintain detailed AI-use logs
- Engage multisectoral review for moderate/high-risk applications
- Retain full human responsibility for interpretation and conclusions

Students

- Disclose AI use according to tiered requirements
- Adhere to ethical AI-use guidelines
- Participate in AI literacy and ethics formation
- Assume personal accountability for submitted work

V. Alignment with *Ex Corde Ecclesiae*

AI governance at SPUM:

- Supports formation of the whole person
- Upholds truth-seeking (*veritas*) and academic integrity
- Ensures technology serves the human and Christian vocation of education

VI. Board-Approved Policy Resolution Format

Resolution No.: [Insert Number]

Title: *St. Paul University Manila AI Governance Policy*

Whereas, the University recognizes the ethical, educational, and societal implications of generative AI;

Whereas, the University seeks to align AI use with Catholic Social Teaching, the Paulinian charism, and *Ex Corde Ecclesiae*;

Resolved, that the St. Paul University Manila AI Governance Policy, including its phased implementation roadmap, tiered AI-use declarations, and multisectoral governance structures, is hereby adopted;

Further Resolved, that the AI Governance Council is authorized to oversee implementation, monitor compliance, and submit regular reports to the University Board.

Effective Date: [Insert Date]

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

Signed:

Chair, Board of Trustees University President Secretary of the Board

This phased localized framework integrates **ubiquitous AI readiness**, **tiered AI-use declarations**, **democratized governance**, **multi-stakeholder oversight**, and **compliance with Ex Corde Ecclesiae**, making it actionable for administrators, faculty, researchers, and students at St. Paul University Manila.

Mapping Catholic Social Teaching Principles to AI Governance Mechanisms

This section clarifies how foundational Catholic Social Teaching (CST) principles are operationalized through specific AI governance structures, protocols, and decision-making processes at SPU Manila. It ensures that moral commitments are translated into concrete, auditable institutional practices.

1. Human Dignity

CST Principle: Every human person possesses inherent dignity and must never be reduced to a tool, data point, or algorithmic output.

AI Governance Translation to Policy Commitments

- AI shall support, not replace, human judgment, conscience, and moral agency.
- No AI system may make final decisions affecting academic evaluation, discipline, employment, or research interpretation.

AI Governance Translation to Governance Mechanisms

- Human-in-the-Loop Requirement for all moderate- and high-risk AI uses
- Prohibition of Fully Automated Decision-Making in:
    - Student assessment
    - Faculty evaluation
    - Disciplinary or admissions decisions
- Tiered AI-Use Declaration requiring explicit identification of human oversight for substantive AI use

AI Governance Translation to Operational Tools

- Mandatory documentation of human validation steps in AI-use logs
- Ethics review escalation for any AI system influencing persons or personal data

AI Governance Translation to Accountability Indicator

- Evidence that final judgments remain attributable to identifiable human agents

2. Common Good

CST Principle: Social conditions must allow individuals and communities to reach their fulfillment more fully and easily.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

AI Governance Translation in Policy Commitments

- AI adoption must demonstrably benefit the academic community as a whole.
- AI tools that increase inequality, exclusion, or harm institutional trust are restricted or prohibited.

AI Governance Translation to Governance Mechanisms

- AI Risk Classification System evaluating potential social and institutional impact
- Institutional Impact Assessment for moderate- and high-risk AI deployments
- Shared AI Resources Policy to avoid exclusive access for privileged units or groups

AI Governance Translation to Operational Tools

- Periodic review of AI tools for bias, accessibility, and educational value
- Prioritization of AI applications that enhance learning support, research capacity, and administrative efficiency for all

AI Governance Translation to Accountability Indicator

- Documented alignment between AI deployments and institutional mission outcomes

3. Solidarity

CST Principle: All members of the community are responsible for one another, especially the vulnerable.

AI Governance Translation to Policy Commitments

- AI systems must not reinforce discrimination, marginalization, or epistemic injustice.
- Vulnerable groups (students, early-career researchers, non-technical staff) receive special consideration in AI governance.

AI Governance Translation to Governance Mechanisms

- Multisectoral AI Governance Council including student and community representatives
- Bias and Harm Review Protocols for AI systems affecting teaching, research, or communication
- Mandatory AI Literacy Formation to prevent informational asymmetry

AI Governance Translation to Operational Tools

- Accessible training materials tailored to different user groups
- Ethical review of datasets, prompts, and outputs for representational harm

AI Governance Translation to Accountability Indicator

- Inclusive participation in AI governance processes and documented mitigation of harms

4. Subsidiarity

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

CST Principle: Decisions should be made at the most local and competent level, with higher authorities providing support—not domination.

AI Governance Translation to Policy Commitments

- Faculty, departments, and research units retain primary responsibility for AI use within their contexts.
- Central governance provides ethical guidance, oversight, and coordination rather than micromanagement.

AI Governance Translation to Governance Mechanisms

- Course-Level AI Policies developed by faculty within institutional parameters
- Unit-Based AI Declarations reviewed proportionately based on risk
- Escalation Pathways only for moderate- and high-risk applications

AI Governance Translation to Operational Tools

- Templates and guidance rather than prescriptive mandates
- Decentralized logging systems subject to centralized audit

AI Governance Translation to Accountability Indicator

- Evidence that local units exercise informed autonomy within ethical boundaries

5. Participation (Implicit in CST and Paulinian Charism)

CST Principle: All members of the community have the right and responsibility to contribute to social and institutional life.

AI Governance Translation to Policy Commitments

- AI governance is participatory, transparent, and dialogical.
- Stakeholders affected by AI decisions are included in deliberation.

AI Governance Translation to Governance Mechanisms

- Multistakeholder Representation in the AI Governance Council
- Consultative Review Processes for policy updates and high-impact AI use
- Public-Facing AI Guidelines for transparency

AI Governance Translation to Operational Tools

- Feedback mechanisms for reporting AI-related concerns
- Periodic consultations and open forums on AI governance

Accountability Indicator

- Documented stakeholder input and responsiveness in policy revisions

6. Stewardship of Creation and Resources

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

CST Principle: Human beings are entrusted with the responsible care of resources, knowledge, and technological power.

AI Governance Translation Policy Commitments

- AI resources are managed responsibly, sustainably, and ethically.
- Data, intellectual property, and institutional reputation are safeguarded.

AI Governance Translation to Governance Mechanisms

- Vendor Vetting and Data Protection Reviews
- Audit Logs and Lifecycle Management of AI systems
- Regular Compliance and Performance Audits

AI Governance Translation to Operational Tools

- Data minimization and security protocols
- Periodic cost–benefit and risk–benefit analyses

`AI Governance Translation to Accountability Indicator

- Transparent records of AI procurement, use, and retirement decisions

7. Truth (*Veritas*) and Academic Integrity

CST and Ex Corde Ecclesiae Principle: Universities exist to pursue truth through disciplined inquiry and moral responsibility.

AI Governance Translation to Policy Commitments

- AI must never obscure authorship, accountability, or intellectual honesty.
- Human interpretation remains essential in teaching and research.

AI Governance Translation to Governance Mechanisms

- Mandatory AI-Use Disclosure in coursework, research, and institutional outputs
- Prohibition of AI-Generated Misrepresentation of authorship or originality
- Ethics Review for AI-Assisted Research Interpretation

AI Governance Translation to Operational Tools

- Standardized disclosure statements
- Academic integrity enforcement mechanisms

Accountability Indicator

- Clear attribution and traceability of intellectual work

Table 8. Summary Matrix (Policy-Ready)

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

CST Principle	Governance Mechanism	Policy Effect
Human Dignity	Human-in-the-loop, ban on automated decisions	Protects moral agency
Common Good	Risk classification, impact assessment	Prevents harm, promotes equity
Solidarity	Multisectoral governance, AI literacy	Protects vulnerable groups
Subsidiarity	Local AI policies with proportional oversight	Preserves academic freedom
Participation	Consultative review, transparency	Democratic governance
Stewardship	Vendor vetting, audits	Responsible AI management
Truth (*Veritas*)	Disclosure, integrity protocols	Protects academic mission

What corresponding instruments for policy implementation can be proposed for St. Paul University Manila?

These three instruments ensure that AI use at SPUM is **ethical, transparent, accountable, and mission-aligned**, providing actionable guidance for administrators, faculty, researchers, and students.

1. Title: St. Paul University Manila AI-Use Policy

Purpose: To establish ethical, responsible, and mission-aligned standards for the use of generative AI across all academic, research, administrative, and communication activities at SPUM, in alignment with Catholic Social Teaching, Ex Corde Ecclesiae, and the Paulinian charism.

Scope: All AI applications (open-source and for-profit) used in SPUM operations, including teaching, research, administrative functions, student support, and external communications.

Policy Statements:

1. **Human Dignity & Academic Integrity:** AI use must respect human moral agency and cannot replace faculty, researcher, or student judgment.

2. **Tiered AI-Use Declarations:** AI use must be disclosed according to the following tiers:

   - **Incidental:** Minor assistance (e.g., grammar, formatting) – brief disclosure
   - **Moderate:** Significant support (e.g., data organization, summarization) – model/version and human verification required
   - **Major/Substantive:** Core analytical or interpretive contributions – detailed documentation and multisectoral review required

3. **Governance & Oversight:** The SPUM AI Governance Council shall oversee all AI applications, approve policies, classify AI tools, and monitor compliance.

4. **Training & Competence:** All faculty, researchers, and students must complete AI ethics, literacy, and responsible-use training prior to moderate or major AI use.

5. **Data Privacy & Security:** All AI use must comply with applicable data protection laws, institutional policies, and ethical standards, especially for human-subject data.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue I January 2026

6. **Accountability:** Faculty, researchers, and students remain responsible for the outcomes of AI-assisted work. Violations will be subject to disciplinary and ethical review.

7. **Continuous Review:** The AI Governance Council shall review the policy periodically, updating it for technological advances and ensuring alignment with Catholic mission and democratized governance principles.

Effective Date: [Insert Date]

Approved by: Board of Trustees, SPUM

2. Title: Institutional Research Ethics Board (IREB) AI Checklist

Purpose: To guide IREB review of AI use in human-subject research.

Checklist Items:

1. **AI Tool Description:** Model name, version, developer (open-source or commercial), and intended use.

2. **Risk Classification:** Identify tier (low, moderate, high) based on impact on participants or study outcomes.

3. **Data Handling:** Anonymization, encryption, storage location, compliance with privacy regulations.

4. **Human Oversight:** Documentation of human review, interpretation, and validation of AI outputs.

5. **Transparency:** How AI use will be disclosed to participants and in research outputs.

6. **Bias and Fairness:** Assessment of potential algorithmic bias, discriminatory effects, or ethical concerns.

7. **Informed Consent:** Whether participants are informed of AI involvement and its implications.

8. **Auditability:** Maintain logs of AI prompts, outputs, and human corrections for IRB review.

9. **Multisectoral Oversight:** Evidence of review or consultation with other faculty, ethics officers, or external collaborators for moderate/high-risk AI use.

10. **Compliance Verification:** Alignment with SPUM AI Governance Policy and Catholic ethical principles.

IREB Decision Options:

- Approved
- Approved with Conditions
- Request Revision
- Denied

3. Title: Course Syllabus Clause

AI Use in This Course:

1. **Permissible Use:** Students may use AI tools for incidental or moderate assistance, such as grammar checking, summarization, or research organization.

2. **Disclosure Requirement:** All AI-assisted work must be disclosed according to SPUM's tiered AI-use system:

   - Incidental AI use: brief acknowledgment in assignments
   - Moderate AI use: specify AI tool, version, and how outputs were verified
   - Major/Substantive AI use: prior faculty approval required, full documentation of AI contributions

3. **Ethical Responsibility:** Students are responsible for ensuring that AI use does not violate academic integrity, misrepresent work, or replace personal judgment.

4. **Verification:** Faculty may require students to provide evidence of AI prompts, outputs, and human validation.

5. **Training:** Students are expected to complete the SPUM AI literacy and ethics module prior to using AI tools in assignments.

DISCUSSION

In the formulation of AI-use policies, particularly within Catholic higher education, it is critical to consider the economic interests and funding considerations of AI developers. Generative AI systems, whether open-source or for-profit, are not value-neutral technologies; their design, architecture, safety mechanisms, and operational behavior are shaped by the financial priorities of their funders, investors, and institutional clients (Floridi, 2023; Bender et al., 2021). For-profit AI models, such as ChatGPT or Claude, prioritize scalability, market adoption, and risk mitigation, which results in centralized control, conservative safety layers, and standardized interfaces. These features, while ensuring reliability and commercial viability, may limit transparency, user customization, and epistemic flexibility—factors that are essential in academic research and the cultivation of critical thinking (Hao, 2023). Open-source models, conversely, are influenced by the availability of grants, community support, and contributor labor, often favoring adaptability and methodological experimentation but relying heavily on user competence for safety and ethical application (Stiennon et al., 2023).

For Catholic higher education institutions, these economic influences bear significant ethical and pedagogical implications. AI systems funded for market profitability may inadvertently perpetuate dominant cultural norms or reinforce epistemic biases, thereby conflicting with Catholic commitments to human dignity, social justice, and the preferential option for the marginalized (Pontifical Council for Culture, 2020). Policies that fail to account for the underlying economic drivers of AI development risk over-reliance on opaque systems, potentially ceding intellectual and moral oversight to commercial actors. By explicitly considering developers' funding motivations and market pressures, universities can design governance structures that ensure AI remains a tool in service of ethical pedagogy, research integrity, and institutional mission rather than an instrument of market-driven priorities.

In practice, this entails distinguishing between AI outputs influenced by economic imperatives and those that are academically robust, establishing risk-tiered protocols, mandating transparency and disclosure, and integrating critical reflection into curriculum and research workflows. By doing so, Catholic higher education can maintain human-centered, morally accountable engagement with AI, aligning technological use with principles of subsidiarity, solidarity, and the common good (Vanderburg, 2022). Understanding the economic context of AI development therefore becomes a foundational component of responsible policy formulation, safeguarding both ethical standards and the academic mission.

REFERENCES

1. Anthropic. (2023). Claude: The next generation AI assistant. https://www.anthropic.com
2. Beer, S. (1981). The brain of the firm: The managerial cybernetics of organization. Allen & Unwin.
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity Press.

4. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922
5. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
6. Braun, V., & Clarke, V. (2013). Successful qualitative research: A practical guide for beginners. Sage.
7. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. https://arxiv.org/abs/2005.14165
8. Buhmann, A., & Berendt, B. (2022). Human-centered AI: Managing risk and governance. AI & Society, 37(3), 763–777. https://doi.org/10.1007/s00146-021-01315-4
9. Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage.
10. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
11. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
12. Feldstein, S. (2019). The global expansion of AI surveillance. Carnegie Endowment for International Peace.
13. Floridi, L. (2023). Ethics, governance, and policies for artificial intelligence. Springer.
14. Floridi, L. (2023). The logic of AI governance: Principles for responsible artificial intelligence. Oxford University Press.
15. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
16. Hao, K. (2022). Open-source AI: Benefits, risks, and the future. MIT Technology Review. https://www.technologyreview.com
17. Hao, K. (2023). AI is shaping up to be a tool of institutional power. MIT Technology Review. https://www.technologyreview.com
18. Hugging Face. (2023). Transformers library. https://huggingface.co/transformers
19. Jonassen, D. H. (1999). Computers as mindtools for schools: Engaging critical thinking. Prentice Hall.
20. Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage.
21. Midgley, G. (2000). Systemic intervention: Philosophy, methodology, and practice. Springer.
22. Morozov, E. (2011). The net delusion: The dark side of Internet freedom. PublicAffairs.
23. OpenAI. (2023a). ChatGPT: Optimizing language models for dialogue. https://openai.com/research/chatgpt
24. OpenAI. (2023b). GPT-4 technical report. https://openai.com/research/gpt-4
25. OpenAI. (2023c). ChatGPT and GPT-4 technical overview. https://openai.com/research
26. Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. Organization Studies, 28(9), 1435–1448. https://doi.org/10.1177/0170840607081138
27. Pontifical Council for Culture. (2020). Ethics in the digital age: Human dignity and artificial intelligence. Vatican Press.
28. Rae, J., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., … Irving, G. (2021). Scaling language models: Methods and insights. arXiv preprint arXiv:2106.04554. https://arxiv.org/abs/2106.04554
29. Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8
30. Schwandt, T. A. (2014). The Sage dictionary of qualitative inquiry (4th ed.). Sage.
31. Scott, W. R. (2014). Institutions and organizations: Ideas, interests, and identities (4th ed.). Sage Publications.
32. Stake, R. E. (1995). The art of case study research. Sage.
33. Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., … Christiano, P. (2023). Learning to summarize with human feedback. Journal of Machine Learning Research, 24(1), 1–42. https://www.jmlr.org/papers/v24/22-0474.html

34. Touvron, H., Martin, L., Stone, E., Albert, P., Almahairi, A., Babaei, Y., … Douze, M. (2023). LLaMA 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288. https://arxiv.org/abs/2307.09288
35. Trist, E. (1981). The evolution of socio-technical systems. In A. Van de Ven & W. Joyce (Eds.), Perspectives on organization design and behavior (pp. 283–292). Wiley.
36. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000373434
37. Vanderburg, W. H. (2022). Technology and human responsibility: Catholic perspectives on digital ethics. Georgetown University Press.
38. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
39. Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., … Le, Q. (2023). Emergent capabilities of large language models. Transactions on Machine Learning Research, 10, 1–42.
40. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0
41. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.