The Hijacking of Open Access and the Betrayal of Public Trust: OpenAI under Sam Altman and Artificial Intelligence Integrity Assessment in Higher Education

Brian Bantugan, PhD
St. Paul University Manila

Abstract: This study examined the strategic transformation of OpenAI from a nonprofit entity committed to open-access artificial intelligence (AI) into a closed, profit-driven platform deeply integrated with Microsoft's corporate ecosystem. Utilizing an interpretive case study approach and political economy theory, the paper analyzed the "mechanics of hijacking" employed by OpenAI leadership, including structural reconstitution into a capped-profit model, the use of safety rhetoric to justify model secrecy, and the creation of ecosystem lock-ins through exclusive cloud and IP agreements. The research highlighted how withholding critical technical data—such as model architecture and training corpora—undermines the core pillars of scientific trust: reproducibility, independent auditability, and peer review. These developments exacerbate global inequality and concentrate power among a few corporate actors, creating a "digital divide" and an "AI oligopoly". Furthermore, the study explored the specific implications for Higher Education Institutions (HEIs), noting that the lack of transparency in tools like ChatGPT poses significant risks to academic integrity. To mitigate these risks, the paper proposed a multi-dimensional integrity assessment framework for HEIs, focusing on transparency, ethical alignment, and auditability. The study concluded that stakeholders must shift away from inherent trust in proprietary AI, instead adopting rigorous regulatory measures, community-driven open alternatives, and enhanced digital literacy to ensure AI serves the public good.

Keywords: OpenAI, Open Access Ethos, Political Economy Theory, AI Integrity, Higher Education Institutions (HEIs)

I. Introduction

Open access has long been the normative foundation for trustworthy AI research: transparent methods, reproducible results, and widely shared artifacts (code, data, and model weights) that enable community scrutiny and collective progress. OpenAI's 2015 launch explicitly aligned with that ethos, pledging a nonprofit mission to "advance digital intelligence… unconstrained by a need to generate financial return," with a commitment to publish and share patents "with the world" (OpenAI, 2015). Over the past several years, however, OpenAI—under CEO Sam Altman—engineered a structural and strategic pivot: from nonprofit to a hybrid "capped-profit" LP (2019) and, more recently, toward a public-benefit corporate form that preserves investor equity and deep, exclusive commercial entanglements with Microsoft, while retaining mission-language oversight by a nonprofit foundation (Microsoft, 2025; OpenAI, 2025; TechCrunch, 2019).

This corporate metamorphosis coincided with a decisive turn away from openness at the model level. The GPT-4 technical report intentionally withholds core details—architecture, training compute, and dataset construction—explicitly citing competitive and safety concerns, a choice widely criticized by researchers as antithetical to scientific transparency (OpenAI, 2023; Xiang, 2023).

OpenAI's conversion under Sam Altman from an open-access, nonprofit lab to a closed, profit-driven platform represents a hijacking of the open ethos that the organization once championed. Through selective disclosures, ecosystem lock-ins, and narrative reframing under the banners of "safety" and "scale," OpenAI has centralized AI capabilities and market power while eroding reproducibility, public oversight, and user autonomy (Microsoft, 2025; OpenAI, 2025; TechCrunch, 2019; Xiang, 2023).

Open-Source Ethos: Definition and Principles

The open-source ethos is grounded in three core principles: **transparency, collaboration, and accessibility**. Transparency ensures that source code, datasets, and methodologies are openly available for scrutiny, enabling reproducibility and accountability in technological development (Stallman, 2002). Collaboration fosters a participatory environment where developers, researchers, and organizations contribute collectively to innovation, reducing duplication of effort and accelerating progress (Feller & Fitzgerald, 2002). Accessibility guarantees that technological resources are not restricted to elite institutions or corporations, democratizing knowledge and empowering diverse stakeholders to engage in development (Raymond, 1999). These principles collectively form the foundation of open-source culture, which has historically driven advancements in software and, more recently, AI.

Role in AI Development

Open-source practices play a critical role in AI development by **preventing monopolization** and **accelerating scientific progress**. When AI models and tools are openly shared, barriers to entry are lowered, allowing smaller organizations and academic institutions to participate in cutting-edge research (Feller & Fitzgerald, 2002). This openness mitigates the risk of concentrated power among a few corporations, which could otherwise dominate AI capabilities and dictate ethical norms (Whittlestone et al., 2019). Furthermore, open-source AI fosters rapid innovation by enabling peer review, collaborative debugging, and iterative improvements, which are essential for addressing complex challenges such as bias and safety (Bengio, 2020). Initiatives like Hugging Face and EleutherAI exemplify how open collaboration accelerates progress by providing accessible large language models and datasets to the global research community (Bommasani et al., 2021).

OpenAI's Original Commitment

OpenAI was founded in 2015 with a mission explicitly aligned with the open-source ethos: to ensure that artificial general intelligence (AGI) benefits all humanity. In its inaugural statement, OpenAI pledged to "freely collaborate" and share research, patents, and code with the world (OpenAI, 2015). Early actions reflected this commitment, such as the release of GPT-2 in 2019, which—despite initial concerns about misuse—was eventually made publicly available along with its training methodology (Radford et al., 2019). These steps positioned OpenAI as a champion of openness and transparency in AI development, reinforcing public trust and signaling a commitment to democratizing access to advanced AI technologies. However, subsequent shifts toward proprietary models and closed releases have sparked debate about whether OpenAI has deviated from its founding principles, raising questions about the sustainability of openness in a competitive AI landscape (Xiang, 2023).

Political Economy Theory

Political economy theory examines the interplay between politics and economics, focusing on how institutions, power relations, and resource distribution shape economic outcomes and societal structures. Rooted in classical thought, political economy emerged as a framework to understand how economic systems are embedded within political and social contexts (Gilpin, 2001).

Foundations of Political Economy. Classical theorists such as Adam Smith, David Ricardo, and Karl Marx laid the groundwork for political economy by analyzing the relationship between markets, state intervention, and class dynamics. Smith emphasized the role of markets and the "invisible hand" in promoting efficiency, while Marx critiqued capitalism for perpetuating inequality and exploitation (Smith, 1776/1994; Marx, 1867/1990).

Core Principles. Political economy theory posits that economic decisions are not purely technical but deeply political, influenced by power structures and institutional arrangements (Mosco, 2009). It challenges the notion of market neutrality, arguing that policies reflect competing interests among social classes, corporations, and governments (Gilpin, 2001).

Modern Applications. The contemporary political economy extends beyond national boundaries to address globalization, trade regimes, and the role of multinational corporations. Scholars highlight how neoliberal policies—emphasizing deregulation and privatization—have reshaped global economic governance, often privileging capital over labor and deepening inequality (Harvey, 2005; Stiglitz, 2002).

Political economy also informs debates on technology and AI governance, illustrating how innovation is shaped by corporate interests and regulatory frameworks rather than purely technical considerations (Zuboff, 2019). This perspective underscores the need for democratic oversight to prevent monopolistic control and ensure equitable access to transformative technologies.

II. Statement of the Problem

The rapid transformation of OpenAI from an open-source, nonprofit research lab to a closed-source, profit-driven platform under CEO Sam Altman represents a significant "hijacking" of the open-access ethos. While OpenAI's 2015 charter pledged to share research and patents for the benefit of humanity, the organization has since shifted toward a "capped-profit" model and a public-benefit corporation with deep, exclusive commercial ties to Microsoft. This pivot has replaced transparency with model secrecy, exemplified by the withholding of core technical details for GPT-4 and GPT-5 under the guise of "safety". Consequently, this shift erodes scientific reproducibility, consolidates market power within a Big Tech ecosystem, and creates ethical and integrity risks for Higher Education Institutions (HEIs) that rely on these increasingly opaque technologies. This paper sought to describe the hijacking in the context of Altman's OpenAI and its implications in higher education. Specifically, it answered the following questions:

1. How has OpenAI's organizational structure and mission evolved from its founding in 2015 to its current corporate state?
2. What rhetorical and strategic mechanisms are used to justify the shift from open access to proprietary, closed models?
3. What are the consequences of this "betrayal" of open-source principles for developers, users, and broader society?
4. What are the ethical implications of the hijacking of the open access ethos by OpenAI?
5. What safeguards and solutions can help mitigate said hijacking?
6. How can political economy theory explain the interplay between OpenAI's economic incentives and its control over AI innovation?
7. What structured evaluation mechanisms should HEIs implement to assess the integrity of proprietary AI applications?

III. Methodology

Interpretive Paradigm

The interpretive paradigm emphasizes understanding social phenomena through subjective meanings and contextual interpretation rather than objective measurement (Schwandt, 1994). It seeks to uncover underlying values, narratives, and ethical tensions reflected in scholarly and industry texts rather than testing hypotheses (Willis, 2007).

Case Study Approach

A case study approach provides an in-depth examination of a bounded system—in this case, OpenAI's organizational transformation and its implications for public trust (Yin, 2018). By focusing on a single case, the literature review synthesizes diverse perspectives from academic articles, policy reports, and media analyses to construct a holistic understanding of the phenomenon. This method allows for rich contextualization of events, decisions, and discourses shaping OpenAI's trajectory.

Qualitative Design

The literature review adopts a qualitative design, prioritizing interpretive analysis over numerical data. Qualitative inquiry is appropriate for examining complex ethical and governance issues because it captures nuanced meanings embedded in texts (Creswell & Poth, 2018). Rather than quantifying trends, the review interprets language, framing, and thematic patterns to reveal how stakeholders conceptualize openness, ethics, and commercialization in AI.

Purposive Data Source Selection

Purposive sampling guides the selection of sources based on relevance to the research questions (Palinkas et al., 2015). The review draws from peer-reviewed journals, organizational charters, technical reports, and credible media outlets discussing OpenAI's mission, structural changes, and ethical debates. This strategy ensures that the data set reflects authoritative and diverse viewpoints, enabling a comprehensive synthesis of the discourse.

Textual Analysis

Textual analysis is employed to examine the content, structure, and rhetorical strategies within selected documents (Fairclough, 2013). This involves identifying key concepts such as "openness," "safety," and "public trust," and analyzing how these terms are framed across sources. Textual analysis reveals the tension between OpenAI's stated mission and its operational practices, highlighting contradictions and legitimizing narratives.

Thematic Analysis and Synthesis

The final stage involves thematic analysis, which organizes findings into coherent themes such as **loss of transparency**, **ethical justification through safety rhetoric**, and **concentration of power** (Braun & Clarke, 2006). These themes are synthesized to construct an interpretive narrative that explains the ethical significance of OpenAI's pivot and its broader implications for AI governance. Thematic synthesis enables integration of diverse perspectives into a structured argument, supporting critical reflection and theory-building.

Ethical Considerations

Conducting an online literature review involved several ethical responsibilities to ensure credibility, fairness, and respect for intellectual property. These considerations aligned with broader principles of research ethics and academic integrity.

Intellectual Property and Proper Attribution. One of the primary ethical obligations was for the researcher to acknowledge original authors and sources. Plagiarism—using ideas or text without proper citation—

violated academic standards and undermined trust in scholarly work (American Psychological Association [APA], 2020). The researcher provided accurate in-text citations and a complete reference list for all sources consulted, including journal articles, organizational reports, and online content.

Accuracy and Honesty in Representation. Ethical literature reviews required the researcher to faithfully represent sources. Misquoting, selective reporting, or distorting findings to fit a preconceived argument constituted unethical practice (Resnik, 2020). The researcher summarized and synthesized information objectively, ensuring that interpretations reflected the original context and meaning.

Transparency in Methodology. The researcher clearly described the search strategy, inclusion criteria, and analytical approach to maintain transparency and reproducibility (Booth et al., 2016). For an interpretive, qualitative design, this included explaining purposive sampling and thematic synthesis methods. Transparency help readers assess the rigor and credibility of the review.

Avoiding Bias and Ensuring Fairness. Bias could occur through selective inclusion of sources that supported a particular viewpoint while ignoring contradictory evidence. The researcher provided balanced coverage of diverse perspectives, especially in contested areas such as AI ethics and governance (Whittlestone et al., 2019). This ensured that conclusions were grounded in comprehensive evidence rather than ideological preference.

Respect for Open Access and Copyright. When using online sources, the researcher respected copyright laws and licensing terms. Open-access materials could be freely used with attribution, but paywalled or proprietary content might have required permission for reproduction (APA, 2020). Ethical scholarship involved verifying usage rights and adhering to fair-use guidelines.

Data Privacy and Confidentiality. Although literature reviews typically analyzed published materials, ethical considerations extended to protecting sensitive information. If sources included personal data or case-specific details, the researcher avoided disclosing identifiable information unless it was already public and ethically permissible (Resnik, 2020).

Academic Integrity in Synthesis. Finally, thematic synthesis had to reflect the researcher's authentic engagement with sources, not superficial aggregation. Ethical scholarship valued critical analysis, originality, and contribution to knowledge rather than mere compilation (Booth et al., 2016).

IV. Results

The Strategic Pivot: From Open Access to Closed Profit

Timeline of Transformation. OpenAI was founded in 2015 as a nonprofit research organization with a mission to ensure artificial general intelligence (AGI) benefits all humanity. Its founding charter emphasized openness, pledging to share research and collaborate broadly (OpenAI, 2015). This commitment initially manifested in the release of models like GPT-2, which, despite initial hesitation over misuse risks, was eventually made publicly available along with its training methodology (Radford et al., 2019).

However, in 2019, OpenAI introduced a **capped-profit model** through the creation of OpenAI LP. This structural shift was justified as necessary to attract the billions of dollars required for computing resources and talent acquisition (TechCrunch, 2019). The capped-profit structure allowed investors to earn returns up to 100 times their investment, signaling a significant departure from the original nonprofit ethos.

The pivot toward commercialization became more pronounced with the **closed release of GPT-4 in 2023**. Unlike earlier models, GPT-4's technical report withheld critical details such as architecture, training data, and compute resources, citing competitive and safety concerns (OpenAI, 2023). GPT-5, expected to follow the same pattern, further entrenches this closed approach, marking a full transition from open access to proprietary control (Xiang, 2023).

Narrative Framing. OpenAI has consistently framed its move away from openness under the banner of **safety and misuse prevention**. Sam Altman and other executives argue that releasing full model details could enable malicious actors to weaponize AI systems, creating risks such as automated cyberattacks or disinformation campaigns (Tech Policy Press, 2023). This narrative positions secrecy as a responsible choice, aligning with broader calls for AI governance and risk mitigation.

While safety concerns are legitimate, critics contend that this framing also serves as a **strategic justification for competitive advantage**. By limiting transparency, OpenAI maintains control over frontier models, securing its position in the AI arms race and reinforcing its dominance in the market (West, 2023).

Reality Check. Despite its safety rhetoric, OpenAI's pivot has coincided with aggressive monetization strategies. The launch of **ChatGPT Plus**, a subscription service offering enhanced access to GPT models, marked the first major revenue stream (OpenAI, 2023). This was followed by enterprise-grade APIs, enabling businesses to integrate GPT capabilities into their workflows for a fee.

The most significant commercial move came through OpenAI's deep partnership with Microsoft. Beginning with a $1 billion investment in 2019 and expanding to a multi-billion-dollar deal in 2023, this alliance granted Microsoft exclusive rights to integrate OpenAI models into its Azure cloud platform and products like Microsoft 365 (Microsoft, 2023). In 2025, the partnership evolved further, with Microsoft acquiring a substantial equity stake and extending exclusive IP rights through 2032 (Microsoft, 2025).

These developments underscore a stark reality: OpenAI's transformation is not merely about safety—it is about **consolidating market power and monetizing AI at scale**. The organization that once championed openness now operates as a commercial entity deeply embedded in the competitive dynamics of Big Tech.

V. Mechanisms of Hijacking

1. Structural Reconstitution: From Nonprofit to Investor-Aligned Hybrid. OpenAI's 2019 creation of **OpenAI LP**—a "capped-profit" subsidiary governed by a nonprofit board—was framed as necessary to fund large-scale compute and attract talent (TechCrunch, 2019). In 2025, Microsoft and OpenAI announced a renewed partnership tied to OpenAI's recapitalization into a **public benefit corporation (PBC)**, confirming Microsoft's stake of roughly 27% and extending exclusive IP rights and Azure API exclusivity through 2032—formalizing long-term commercial capture of frontier models and APIs (Microsoft, 2025; OpenAI, 2025).

Governance complexity and equity allocation institutionalize profit incentives, subordinating openness to capital access and exclusive commercial arrangements, transforming the lab into a platform whose primary accountability is to capital and partners rather than the open research community (Microsoft, 2025; TechCrunch, 2019).

2. Model Secrecy and Selective Disclosure. With **GPT-4**, OpenAI declined to disclose architecture, training corpus, and compute—citing "competitive landscape and safety implications"—a marked departure from earlier openness (OpenAI, 2023). Journalistic and research commentary characterized the release as "closed" and "shrouded in secrecy," noting that opacity precludes independent evaluation of bias, environmental cost, and replication—core pillars of scientific trust (Xiang, 2023).

Mechanism: By withholding model details while publishing performance claims and system cards, OpenAI preserved the legitimacy benefits of academic-style publication without conceding the practical openness required for reproducibility—maintaining trade secrets that concentrate value capture (OpenAI, 2023; Xiang, 2023).

3. Narrative Reframing: Safety and "Responsible" Scale as Justification. Altman's testimony frames centralized licensing and closed releases as necessary for **"safe and beneficial AI,"** urging model licensing regimes for powerful systems (OpenAI, 2023; Tech Policy Press, 2023). By 2025, public remarks emphasized avoiding regulations that could "slow down" U.S. efforts versus China—aligning openness decisions with geopolitical competitiveness rather than community reproducibility (AOL/Goldman, 2025). Independent analysis from Brookings underscored lawmakers' concerns about opacity and disinformation risks, reinforcing that transparency is central to trust—precisely where OpenAI has reduced disclosure (West, 2023).

Mechanism: Safety rhetoric functions as a legitimizing shield for proprietary control, deflecting calls for openness by arguing closed access prevents misuse, while the competition narrative reframes secrecy as patriotic prudence—both narratives anchoring a business model predicated on exclusive commercialization (OpenAI, 2023; AOL/Goldman, 2025; West, 2023).

4. Ecosystem Lock-In: Exclusive Cloud, IP Rights, and API Stickiness. OpenAI's long-term partnership with Microsoft secured exclusive Azure provisioning and IP rights—first in 2019 and then via a multi-year, multi-billion expansion in 2023—embedding OpenAI workloads within Microsoft's infrastructure and product stack (TechCrunch, 2023). The 2025 agreement extended IP rights and **Azure API exclusivity** until AGI (with verification by an expert panel), further entrenching single-cloud dependence for frontier models and third-party API products (Microsoft, 2025; OpenAI, 2025). Industry reporting repeatedly warned of **vendor lock-in** for startups and enterprises tightly coupled to OpenAI APIs—exposed during leadership crises—prompting calls for multi-provider strategies (Miller, 2023).

Mechanism: By standardizing proprietary APIs and exclusive cloud/IP terms, OpenAI turns developer adoption into switching costs; toolkits and integrations optimize for OpenAI's stack, making migration cumbersome and reinforcing commercial dependency on OpenAI's pricing and policy decisions (Microsoft, 2025; TechCrunch, 2023; Miller, 2023).

5. Token Openness and Timing: Strategic, Symbolic Releases. Recent statements—Altman acknowledging being "on the wrong side of history" in the open-source debate—coexist with continued proprietary control of frontier models, suggesting openness is **selective, symbolic, and strategically communicative**, not foundational (Nuñez, 2025). Reports describe internal discussion about open-sourcing older models, often in response to pressure from open-weights ecosystems (e.g., DeepSeek), rather than as a principled return to open access (Nuñez, 2025).

Mechanism: Selective openness manages public perception and community goodwill while protecting the revenue core; it borrows the moral halo of open access without surrendering competitive assets (weights, data, training recipes) that constitute the firm's defensible moat (Nuñez, 2025).

6. Mission Branding vs. Material Practices. OpenAI's public pages maintain a mission narrative—"benefits all of humanity" under nonprofit oversight—even as material practices (equity stakes, exclusive IP, closed models) prioritize commercial capture and partner advantage (OpenAI, 2025; Microsoft, 2025). Legal and policy analysis warns that evolving corporate forms and governance intricacies risk diluting fiduciary duty to a charitable mission, making **shareholder-aligned decision rights** primary in practice (Frazier, 2025).

Mechanism: Mission language and foundation oversight mask a redistribution of control toward investors and strategic partners; this discursive veneer sustains legitimacy while normalizing closed, profit-maximizing behavior that conflicts with open-access expectations (OpenAI, 2025; Frazier, 2025).

7. Scientific Capital Without Scientific Duties. By publishing benchmark results and system cards while withholding artifacts necessary for replication, OpenAI harvests **scientific prestige** (citations, perceived leadership) without accepting the duties of **open science** (auditability, reproducibility, community verification) (OpenAI, 2023). Researchers warn that this practice impedes independent safety assessment and bias auditing—core to trustworthy deployment in high-stakes domains (Xiang, 2023).

Mechanism: A publish-without-open-artifacts strategy converts research communication into marketing collateral, enabling rapid ecosystem adoption while foreclosing external scrutiny that would otherwise discipline model claims and risks (OpenAI, 2023; Xiang, 2023).

VI. Consequences of Betrayal

The shift from open access to proprietary control in AI development—exemplified by OpenAI's strategic pivot—has profound consequences for developers, users, and society at large. This section synthesizes scholarly and industry perspectives on these impacts.

For Developers. One of the most immediate consequences for developers is the **loss of transparency**. Open-source principles traditionally enable reproducibility and peer review, which are essential for scientific progress (Stallman, 2002; Raymond, 1999). However, OpenAI's decision to withhold critical details about GPT-4 and subsequent models—such as architecture and training data—has created significant barriers for independent researchers (OpenAI, 2023; Xiang, 2023). This opacity undermines trust and limits the ability to audit models for bias and safety.

The move toward closed systems also introduces **innovation bottlenecks**. When access to cutting-edge models is restricted, smaller organizations and academic institutions struggle to compete, slowing the pace of collaborative breakthroughs (Whittlestone et al., 2019). Furthermore, the concentration of talent within a few dominant firms exacerbates this problem. Reports indicate that OpenAI's aggressive hiring and compensation strategies have drawn top researchers away from academia and smaller startups, consolidating expertise within a narrow corporate ecosystem (Bengio, 2020).

For Users. For end-users, the betrayal of open-source ideals results in an **erosion of trust**. Users initially embraced platforms like ChatGPT under the assumption that they were built on principles of openness and accountability. The shift to proprietary models, coupled with limited transparency about data usage and safety protocols, has fueled skepticism about whether these systems serve public interest or corporate profit (West, 2023).

This opacity also introduces **ethical risks**. Without clear disclosure of training data and bias mitigation strategies, users cannot reliably assess whether AI outputs are fair or accurate (Bommasani et al., 2021). Additionally, the lack of accountability mechanisms—such as independent audits—means that harmful behaviors or misinformation generated by AI systems may go unchecked (Whittlestone et al., 2019).

For Society. At a societal level, the consequences are even more far-reaching. The centralization of AI capabilities within a handful of corporations creates a **concentration of power** that rivals historical monopolies in technology (Zuboff, 2019). This dynamic not only influences economic structures but also geopolitical strategies, as nations compete for access to proprietary AI systems (West, 2023).

The pivot away from openness also **widens global inequality**. Developing countries and smaller enterprises lack the resources to license proprietary models, leaving them dependent on dominant players and deepening the digital divide (Bengio, 2020). Finally, the rapid commercialization of AI amid limited transparency has led to **regulatory uncertainty**. Policymakers struggle to craft effective governance frameworks when critical technical details are withheld, raising concerns about safety, accountability, and equitable access (Whittlestone et al., 2019; West, 2023).

VII. Ethical Implications

Why This Breach Matters. The shift from OpenAI's original mission of openness to a closed, profit-driven model represents a significant ethical breach because it **contradicts the organization's founding principles**. OpenAI's 2015 charter explicitly committed to ensuring that artificial general intelligence (AGI) benefits all of humanity and pledged to share research openly (OpenAI, 2015). This promise created public trust and positioned OpenAI as a steward of democratic access to transformative technology. However, the subsequent decision to restrict access to GPT-4 and GPT-5, withholding critical details such as architecture and training data, undermines this commitment (OpenAI, 2023; Xiang, 2023).

This breach also **undermines democratic access to transformative technology**. Openness in AI research is not merely a technical preference—it is a mechanism for equitable participation in innovation (Whittlestone et al., 2019). By centralizing control within a few corporations, the pivot toward proprietary systems exacerbates global inequality and limits opportunities for smaller organizations and academic institutions to contribute to AI development (Bengio, 2020). Scholars argue that such concentration of power risks creating an "AI oligopoly," where decisions about safety, ethics, and deployment are made by a handful of actors rather than through inclusive, multi-stakeholder processes (Zuboff, 2019).

Moral Responsibility of Technology Leaders. The ethical implications of OpenAI's transformation extend to the **moral responsibility of technology leaders**. Founders and executives in the AI industry hold a duty to uphold commitments that safeguard public interest, particularly when those commitments form the basis of trust and collaboration (Floridi & Cowls, 2019). Sam Altman's leadership illustrates the tension between this duty and the pursuit of profit. While Altman has publicly framed secrecy as necessary for safety and misuse prevention (Tech Policy Press, 2023), critics argue that these narratives often mask commercial motives, enabling monopolistic practices under the guise of responsible governance (West, 2023).

The literature on tech ethics emphasizes that moral responsibility in AI development involves more than compliance with regulation; it requires proactive transparency, fairness, and accountability (Jobin et al., 2019). When leaders prioritize investor returns over openness, they risk eroding public trust and setting precedents that normalize secrecy in high-stakes technologies. This dynamic underscores the need for ethical frameworks that bind corporate actors to their stated missions, ensuring that transformative technologies serve humanity rather than narrow commercial interests (Whittlestone et al., 2019).

VIII. Safeguards and Solutions

The ethical and societal risks associated with the commercialization of AI underscore the need for comprehensive safeguards and solutions. Scholars and policy experts emphasize three critical domains: regulatory measures, community-driven initiatives, and education and advocacy.

Regulatory Measures. Regulatory frameworks are essential to ensure accountability and transparency in AI development. **Mandatory transparency**—including disclosure of model architecture, training data sources, and safety protocols—has been widely recommended to enable independent audits and prevent deceptive practices (Whittlestone et al., 2019; West, 2023). Public audits conducted by third-party organizations can verify compliance with ethical standards and mitigate risks such as bias and misinformation (Jobin et al., 2019). Additionally, **ethical licensing** models, similar to Creative Commons but tailored for AI, can enforce responsible use while discouraging monopolistic control (Floridi & Cowls, 2019). These measures collectively aim to balance innovation with public interest, ensuring that AI systems remain subject to oversight rather than unchecked corporate discretion.

Community Action. Beyond regulation, **community-driven initiatives** play a vital role in preserving openness and democratizing access to AI technologies. Open-source platforms such as **Hugging Face** and **DeepSeek** exemplify efforts to provide transparent, accessible models that foster collaboration and accelerate innovation (Bommasani et al., 2021). These initiatives counterbalance proprietary ecosystems by enabling researchers and developers worldwide to experiment, audit, and improve AI systems without prohibitive costs. Scholars argue that supporting such projects through funding and institutional partnerships is critical to maintaining a pluralistic AI landscape and preventing concentration of power (Bengio, 2020).

Education and Advocacy. Finally, **education and advocacy** are indispensable for cultivating ethical awareness among future AI users and developers. Integrating AI ethics into curricula ensures that students understand principles such as fairness, accountability, and transparency (Jobin et al., 2019). Teaching vigilance—such as how to critically evaluate AI claims and identify deceptive practices—empowers individuals to make informed decisions and resist manipulation (Floridi & Cowls, 2019). Advocacy efforts, including public campaigns and multi-stakeholder dialogues, further promote responsible AI governance by engaging civil society in shaping norms and policies (Whittlestone et al., 2019).

Collectively, these safeguards and solutions represent a holistic approach to mitigating the risks of AI commercialization. By combining regulatory oversight, community action, and ethical education, stakeholders can ensure that AI development aligns with societal values and serves the public good.

IX. Discussion

In this context, OpenAI's "mechanics of hijacking"— including structural reconstitution, selective disclosure, ecosystem lock-in, and narrative framing — illustrate how economic incentives and power dynamics influence control over AI innovation.

Structural Reconstitution is the first mechanism of control. By creating OpenAI LP in 2019, the organization shifted from a nonprofit to a capped-profit subsidiary. This hybrid structure allowed it to attract large-scale investment while maintaining a veneer of mission-oriented governance (TechCrunch, 2019; Microsoft, 2025). Political economy theory explains this as the prioritization of capital accumulation over collective knowledge dissemination: the organization's primary accountability moves toward investors and strategic partners, subordinating openness to financial and competitive incentives (Benkler, 2006).

Model Secrecy and Selective Disclosure further exemplify the concentration of scientific and economic power. With GPT-4, critical details such as architecture, training data, and compute resources were withheld, citing safety and competitive concerns (OpenAI, 2023; Xiang, 2023). Political economy theory interprets this as a strategic appropriation of intellectual capital: control over knowledge resources generates monopolistic advantages, limits reproducibility, and restricts access for smaller developers or academic institutions (Fuchs, 2021; Bengio, 2020).

Narrative Framing operates as a legitimizing mechanism. Public statements by Sam Altman and other executives frame proprietary restrictions as necessary for safety and responsible AI development (Tech Policy Press, 2023; OpenAI, 2023). However, this framing also serves as a strategic tool to consolidate market power under the guise of ethical governance—a classic example of how political and economic narratives are used to normalize concentration of control (West, 2023).

Ecosystem Lock-In and IP Exclusivity illustrate how technological platforms reinforce dependence and maintain competitive advantage. Long-term exclusive agreements with Microsoft, including Azure provisioning and intellectual property rights extending through 2032, create high switching costs for developers and embed proprietary AI within corporate infrastructures (Microsoft, 2025; TechCrunch, 2023). Political economy theory explains this as an institutional mechanism that converts technological adoption into economic leverage, concentrating value within a narrow set of actors while marginalizing alternative contributors (Benkler, 2006; Zuboff, 2019).

Finally, **Selective or Symbolic Openness** demonstrates the strategic performance of benevolence. OpenAI occasionally releases older models or limited components in response to external pressure, but frontier models remain tightly controlled (Nuñez, 2025). This aligns with political economy insights: symbolic openness creates legitimacy without relinquishing control over high-value knowledge assets, maintaining an economic moat while cultivating public trust (Fuchs, 2021; Xiang, 2023).

In sum, political economy theory provides a robust framework to understand the mechanics of hijacking in the AI domain. It highlights the interplay of economic incentives, institutional design, and governance narratives in transforming ostensibly open scientific resources into proprietary, revenue-generating assets. In the case of OpenAI, the combination of structural reconstitution, selective disclosure, ecosystem lock-in, and strategic narrative framing reflects a conscious alignment of AI development with investor interests, competitive advantage, and market concentration.

X. Assessing the Integrity of AI Applications in Higher Education

The "hijacking" of open-source AI by OpenAI, exemplified by ChatGPT's closed architecture, selective disclosure, and ecosystem lock-in (OpenAI, 2023; Xiang, 2023), poses unique challenges for HEIs. Assessing AI integrity requires institutions to implement structured evaluation mechanisms that address transparency, reliability, ethical alignment, and accountability.

Transparency and Documentation Review. HEIs should prioritize AI applications that provide clear documentation on their underlying architecture, training data, and operational constraints. In the case of ChatGPT and GPT-4/5, critical details are withheld under the guise of safety and competitive advantage (OpenAI, 2023; Xiang, 2023). Institutions should therefore require vendors to disclose:

- Data sources and preprocessing methods
- Model limitations and intended use cases
- Bias mitigation strategies
- Update and version history

Without transparency, faculty and students cannot verify outputs, evaluate risks, or ensure reproducibility in research, undermining academic integrity.

Verification of Performance and Safety Claims. Given the selective disclosure of model capabilities (Nuñez, 2025), HEIs should independently test AI tools rather than relying solely on vendor-provided benchmarks. This could involve:

- Running controlled evaluations on representative datasets
- Checking for consistency, accuracy, and fairness of outputs
- Simulating edge cases to assess safety and ethical risks

Independent verification ensures that AI outputs are trustworthy, particularly when applied in grading, research analysis, or decision-support systems.

Ethical Alignment Assessment. HEIs must examine whether AI applications align with institutional values and research ethics. The commercialization-driven motives of OpenAI demonstrate how profit imperatives can conflict with openness, accountability, and equity (West, 2023; Bengio, 2020). Ethical assessment should consider:

- Potential for reinforcing bias or misinformation
- Impact on student learning and equitable access
- Compliance with privacy and data protection standards

AI tools that fail ethical scrutiny should be restricted or accompanied by user training to mitigate harm.

Auditability and Traceability. The closed nature of ChatGPT prevents external audits and independent verification (Xiang, 2023). HEIs should insist on AI applications that allow for:

- Logging interactions and outputs for accountability
- Mechanisms for error reporting and correction
- Documentation of model decisions in high-stakes applications (e.g., research, grading, administrative decisions)

Auditability protects academic and institutional integrity by enabling verification and traceability of AI-generated outputs.

Risk Assessment and Regulatory Compliance. Given the legal and societal concerns raised by proprietary AI (Whittlestone et al., 2019; Zuboff, 2019), HEIs should perform risk assessments that consider:

- Dependence on single-vendor ecosystems (e.g., Microsoft-Azure integration)
- Compliance with local, national, and international AI regulations
- Long-term sustainability and vendor stability

Risk-aware policies prevent over-reliance on AI platforms that may restrict access, raise costs, or embed monopolistic control.

Encouraging Critical Use and Digital Literacy. Finally, integrity assessment must include human oversight. Students and faculty should be trained to:

- Critically evaluate AI outputs
- Recognize limitations and potential manipulations of proprietary AI
- Attribute AI-assisted work properly in research and academic submissions

By embedding AI literacy, HEIs reduce the risk of blind reliance on potentially opaque AI systems, promoting responsible and informed use.

XI. HEI AI Integrity Assessment Checklist

Below is the proposed AI Integrity Assessment Checklist for HEIs:

1. Transparency and Documentation

- Does the AI provider disclose model architecture, training data sources, and system limitations?
- Are model updates, version histories, and known biases clearly documented?
- Are usage guidelines, intended applications, and output interpretation instructions provided?

2. Performance Verification

- Has the AI been tested on institution-specific tasks (e.g., research summarization, grading support)?
- Are accuracy, reliability, and consistency of outputs independently verified?

- Are edge cases or high-stakes scenarios evaluated to detect failure modes?

3. Ethical Alignment

- Does the AI adhere to the institution's ethical and academic standards?
- Are potential biases, discriminatory outputs, or misinformation risks assessed?
- Does the AI safeguard user privacy and comply with data protection regulations (e.g., GDPR, FERPA)?

4. Auditability and Traceability

- Can outputs be logged for review, verification, or dispute resolution?
- Are decisions made by the AI traceable to specific inputs and processes?
- Are error reporting and correction mechanisms in place?

5. Vendor Dependence and Ecosystem Lock-In

- Does use of the AI create exclusive reliance on a single provider (e.g., cloud platform, APIs)?
- Are there alternatives or contingency plans if the vendor changes terms or access?
- Are licensing agreements clear, fair, and aligned with institutional goals?

6. Regulatory and Compliance Review

- Does the AI comply with relevant local, national, and international regulations?
- Are safety, security, and accessibility standards met?
- Is there an ongoing review process to ensure continued compliance?

7. Critical Use and Digital Literacy

- Are students, faculty, and staff trained to critically evaluate AI outputs?
- Are guidelines for responsible citation and attribution of AI-assisted work established?
- Are users informed about AI limitations and potential manipulations?

8. Continuous Monitoring

- Is there a process to periodically reassess AI integrity as the model or vendor evolves?
- Are incident response protocols established for harmful or misleading AI outputs?
- Are feedback mechanisms in place to collect user experiences and improve policy?

XII. Conclusion

In light of OpenAI's strategic hijacking of open-source AI, HEIs cannot assume inherent trust in AI applications like ChatGPT. Assessing integrity requires a **multi-dimensional framework** encompassing transparency, independent verification, ethical alignment, auditability, regulatory compliance, and critical literacy. Institutions must treat proprietary AI tools not as neutral assistants but as commercial artifacts whose outputs require careful scrutiny to protect academic standards, equity, and societal trust.

References

[1] American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). APA.

[2] AOL/Goldman, S. (2025, May 8). Sam Altman urges lawmakers against regulations that could "slow down" U.S. in AI race against China. *AOL.* https://www.aol.com/finance/sam-altman-urges-lawmakers-against-212500170.html

[3] Bengio, Y. (2020). The role of openness in AI research. *Communications of the ACM, 63*(8), 54–61. https://doi.org/10.1145/3381831
[4] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). *On the opportunities and risks of foundation models* (arXiv:2108.07258). https://arxiv.org/abs/2108.07258
[5] Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE Publications.
[6] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[7] Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
[8] Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.
[9] Feller, J., & Fitzgerald, B. (2002). *Understanding open source software development*. Addison-Wesley.
[10] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1
[11] Frazier, K. (2025, May 15). Why OpenAI's corporate structure matters to AI development. *Lawfare*. https://www.lawfaremedia.org/article/why-openai-s-corporate-structure-matters-to-ai-development
[12] Gilpin, R. (2001). *Global political economy: Understanding the international economic order*. Princeton University Press.
[13] Harvey, D. (2005). *A brief history of neoliberalism*. Oxford University Press.
[14] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
[15] Marx, K. (1990). *Capital: Volume I* (B. Fowkes, Trans.). Penguin Classics. (Original work published 1867)
[16] Microsoft. (2023, January 23). Microsoft invests billions more dollars in OpenAI, extends partnership. *TechCrunch*. https://techcrunch.com/2023/01/23/microsoft-invests-billions-more-dollars-in-openai-extends-partnership/
[17] Microsoft. (2025, October 28). *The next chapter of the Microsoft–OpenAI partnership*. Microsoft Corporate Blogs. https://blogs.microsoft.com/blog/2025/10/28/the-next-chapter-of-the-microsoft-openai-partnership/
[18] Miller, R. (2023, November 21). OpenAI mess exposes the dangers of vendor lock-in for startups. *TechCrunch*. https://techcrunch.com/2023/11/21/openai-dangers-vendor-lock-in/
[19] Mosco, V. (2009). *The political economy of communication* (2nd ed.). SAGE Publications.
[20] Nuñez, M. (2025, January 31). Sam Altman admits OpenAI was "on the wrong side of history" in open source debate. *VentureBeat*. https://venturebeat.com/ai/sam-altman-admits-openai-was-on-the-wrong-side-of-history-in-open-source-debate/
[21] OpenAI. (2015, December 11). *Introducing OpenAI*. https://openai.com/index/introducing-openai/
[22] OpenAI. (2023). *GPT-4 technical report* (arXiv:2303.08774). https://arxiv.org/abs/2303.08774
[23] OpenAI. (2025, October 28). *Our structure*. https://openai.com/our-structure/
[24] Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y
[25] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI Blog. https://openai.com/research/gpt-2
[26] Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly Media.
[27] Resnik, D. B. (2020). The ethics of research. In *The ethics of science: An introduction* (pp. 85–108). Routledge. https://doi.org/10.4324/9780429266569
[28] Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118–137). SAGE Publications.

[29] Smith, A. (1994). *An inquiry into the nature and causes of the wealth of nations*. Modern Library. (Original work published 1776)

[30] Stallman, R. (2002). *Free software, free society: Selected essays of Richard M. Stallman*. GNU Press.

[31] Stiglitz, J. E. (2002). *Globalization and its discontents*. W. W. Norton & Company.

[32] Tech Policy Press (Hendrix, J.). (2023, May 16). Transcript: Senate Judiciary Subcommittee hearing on oversight of AI. *Tech Policy Press*. https://www.techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai/

[33] TechCrunch (Coldewey, D.). (2019, March 11). OpenAI shifts from nonprofit to "capped-profit" to attract capital. *TechCrunch*. https://techcrunch.com/2019/03/11/openai-shifts-from-nonprofit-to-capped-profit-to-attract-capital/

[34] West, D. M. (2023, May 17). Senate hearing highlights AI harms and need for tougher regulation. *Brookings Institution*. https://www.brookings.edu/articles/senate-hearing-highlights-ai-harms-and-need-for-tougher-regulation/

[35] Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 195–200. https://doi.org/10.1145/3306618.3314289

[36] Willis, J. W. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. SAGE Publications.

[37] Xiang, C. (2023, March 16). OpenAI's GPT-4 is closed source and shrouded in secrecy. *VICE (Motherboard)*. https://www.vice.com/en/article/openais-gpt-4-is-closed-source-and-shrouded-in-secrecy

[38] Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

[39] Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.