Balancing Bytes and Ethics: Stakeholder Implications of Private LLMs

Timothy R. McIlveene
University of West Florida

Stephen A. LeMay
University of West Florida

John Batchelor
University of West Florida

Andrya Allen
Pensacola

This research explores the ethical implications of private large language models (PLLMs) through the lens of stakeholder theory. Private LLMs, tailored for specific organizational needs, present unique privacy and data protection challenges. We examine the historical development of LLMs and their impact on stakeholders, including shareholders, employees, customers, and society. Our proposed framework balances stakeholder interests with ethical considerations, offering a comprehensive approach to the ethical development and deployment of PLLMs. This framework emphasizes transparency, accountability, and sustainable practices to ensure long-term value creation. Future research directions include developing regulatory frameworks, conducting detailed social impact assessments, and exploring strategies for effective human-AI collaboration. This study contributes to academic discourse by providing a multi-faceted approach to managing the ethical challenges posed by PLLMs, fostering best practices, and mitigating potential conflicts among stakeholders.

Keywords: artificial intelligence, private large language models, AI governance, data privacy, stakeholder theory, ethics

INTRODUCTION

The first chatbot, Eliza, emerged early in natural language processing (NLP). MIT researcher Joseph Weizenbaum designed Eliza to simulate human conversation based on predefined rules. Eliza marked the start of NLP research, laying the foundation for more sophisticated large language models (LLMs) (Weizenbaum, 1966). Long-Short Term Memory (LSTM) networks appeared in 1997. LSTMs fostered more extensive, more complex neural networks that could handle larger datasets, paving the way for more advanced language models (Hochreiter & Schmidhuber, 1997). Stanford's CoreNLP Suite introduced tools and algorithms that recognized named entities and permitted NLP tasks like sentiment analysis. This helped

researchers with real-world language challenges (Manning et al., 2014). However, the real revolution came later.

In 2018, Google introduced BERT (Bidirectional Encoder Representations from Transformers). BERT was pre-trained on an enormous volume of text, setting new standards for benchmarks like XXX. BERT was often tuned for specific tasks (Devlin et al., 2018). In 2019, OpenAI released GPT-2 (Generative Pre-trained Transformer 2). GPT-2 showed staggering language generation abilities with 1.5 billion parameters, but it has since been supplanted by GPT-3 (175 billion parameters) and GPT-4 (a speculative 1.7 trillion parameters) (Khobragade, 2023; Chu et al, 2024). This rapid progression underscores the transformative potential of LLMs in reshaping communication, automation, and decision-making processes across various sectors.

This history brings us to the moment's conversation: the conversation about ChatGPT-4, its relatives, its competitors, and its potential successors. The list has become long and immediately impactful: ChatGPT-4, Gemini, Claude, CoPilot, Intelligence, Perplexity, and many more. AIxploria lists the Top 100 most popular AIs, the Top 100 in weekly trends, and the Top 100 in 24-hour trends (https://www.aixploria.com/en/top-100-ai). The lists are far from identical.

These developments highlight where we stand —in the early yet pivotal LLM/AI development stages. Brands, models, and applications proliferate, but if they follow the pattern of technological history, they will then consolidate as winners absorb losers and small competitors (Hidalgo et al., 2007). Regardless, the AI genie has been released and will be with us for the foreseeable future.

In this research, we explore the impact of LLMs on society through the lens of stakeholder theory. Although the ideas might apply more broadly to other AIs and applications, we limit the scope by focusing on private LLMs (PLLMs), built for specific organizations with specific privacy and data protection needs. Our purpose is to examine the ethical implications of private LLMs, considering the perspectives of multiple stakeholders. Our thesis is that private LLMs' ethical considerations must balance the stakeholders' interests, including shareholders, employees, customers, and civil society.

This work contributes to the academic literature on this topic by developing an ethical framework that considers the perspectives of many stakeholders. This framework allows for ethical analysis to lead to best practices, unravel potential conflicts of interest among stakeholders, and foster regulatory insights. It takes an interdisciplinary perspective, offering case studies and practical examples. It also points to future research directions.

This article proceeds through the following steps. First, we review the literature and definitions of LLMs, defining private LLMs and their ethical implications. Second, we review the literature on stakeholder theory and its ethical considerations. Third, we combine the two conceptual analyses to form a framework that fosters our thesis. Fourth, we demonstrate the framework's value with specific cases and practical examples. Fifth, we suggest future theoretical and applied research directions in this arena.

LITERATURE REVIEW

LLMS and Private LLMS

We start with the literature's definitions of LLMs and private LLMs. We modified the definition from Naveed et al. (2024) and chose this: a large language model (LLM) is a complex mathematical representation of human communication based on high volumes of data (Naveed et al., 2024). A private LLM is customized to work within the boundaries of a specific organization. These boundaries may include use cases, privacy, legal compliance issues, and niche applications (Dialpad, 2024; Signity Solutions, 2024).

Private LLMs improve data control by customizing training and structure to fit organizational policies and privacy needs, ensure legal compliance, and minimize risks like data breaches. They reduce third-party access and protect sensitive data from unauthorized exposure. They can integrate with an organization's existing systems (*What Are Private LLMs? Running Large Language Models Privately - privateGPT and Beyond - Zilliz Blog*, n.d.).

The literature addresses a range of issues regarding private LLMs. These academic issues include privacy risks and data leakage, regulatory compliance, balancing utility and privacy, and mitigating

interference with intellectual property. PLLMs are less likely to run into some of the dangers associated with LLMs in general, such as the problem of size. Bender et al. (2021) refer to this problem as the stochastic parrot. However, there is a gap in the literature regarding the long-term implications of private LLMs on organizational culture and stakeholder trust. Our research seeks to address this gap by examining how private LLMs can enhance or erode stakeholders' trust, depending on their ethical deployment.

Next, we address privacy risks and data usage, regulatory compliance, and intellectual property violations as principal categories of ethical concern. These three categories cover most of the ethical issues in the literature.

Privacy Risks and Data Usage

Yao et al. (2024) addressed three ideas on data security and privacy in LLMs. They examined the positive impacts of LLMs on security, the threats that emerge from their use in that area, and the defense of LLM vulnerabilities. They found that LLMs often discover vulnerabilities in systems that might otherwise be overlooked. LLMs can improve code and data security, assure confidentiality, and leverage speed in uncovering system vulnerabilities. Simultaneously, LLMs can increase vulnerabilities because of their scale, openness to user-level attacks, and opacity derived from their scale. Their use may also raise new legal and regulatory issues. In their view, LLMs benefit privacy and security more than they harm it.

This finding seems positive for ethical considerations among stakeholders. However, the difficulties arise in the details. Evertz et al. (2024) cataloged potential attacks through PLLMs. These included malicious prompts, accidental leaks, lack of robustness, and vulnerabilities in integration with other internal systems. They recommend a similar catalog of potential defenses against such attacks, including access control, data encryption, security audits, prompt sanitization, and training (the LLM) for confidentiality awareness. An intriguing aspect of their analysis is the secret key game. In it, operators insert a secret string of code along with instructions not to reveal it. A 'white hat' attacker then attempts to get the LLM to reveal it. This helps with several of the attacks they listed.

The literature in this area proliferates at an astonishing rate and does so for good reason. Some researchers have found privacy gaps in LLMs like ChatGPT (Gupta et al., 2023). Consumers, patients, and other user-stakeholders expect organizations to protect their data. Their perspective is deontological: firms have an obligation to protect personal information.

Despite these comprehensive assessments, the literature lacks a thorough exploration of the ethical implications of these privacy risks, particularly how they affect stakeholder trust and organizational legitimacy. Our work aims to fill this gap by proposing a framework that explicitly addresses these ethical considerations and provides practical recommendations for organizations to manage these risks responsibly.

Regulatory Compliance

AIs can help firms maintain data privacy and security. They can do the same with regulatory compliance in the technological world and other regulatory environments (Ioannidis et al., 2023). AIs have impeccable memories, so they are likely to provide checklists for compliance that include regulations and compliance elements that a human might forget or overlook.

The other dimension also exists: Will they comply with regulations that cover them? That problem becomes more complex as software overlooks national and regional borders. The EU GDPR (General Data Protection Regulation) imposes stricter regulations on PLLMs than the U.S. It calls for greater transparency and concern for data accuracy (Stringhi, 2023) than do regimes in the U.S. and China.

Anderljung et al. (2023) outlined a regime for PLLM and LLM accountability. They stress the importance of access, independence, and expertise for regulators. They include appendices with detailed policy recommendations that include mandated access, scrutiny of all aspects of the AI's development, and scrutiny in proportion to the risks the AIs pose. Such policies have major implications for the ethical use of PLLMs, but they require an openness to regulation that some organizations resist steadfastly. The regulators and the regulated must trust one another. The regulators need independence and involvement. Andeljung et al. (2023) cite the Enron case as a situation where the regulators were too trusting and thus deceived.

While these studies provide important insights into regulatory challenges, there is a gap in understanding how different regulatory frameworks across regions can be harmonized to ensure consistent ethical standards for PLLMs. Our research addresses this gap by exploring the potential for international regulatory cooperation and the development of unified standards that protect stakeholders globally.

Intellectual Property Violations

This concept operates in two dimensions, especially with regard to PLLMs. First, companies want to avoid the legal issues of violating intellectual property rights. At the same time, they want to avoid having their own intellectual property rights violated. Picht et al. (2022) argue for a framework to develop policies that cover both issues. AI technologies affect the intellectual property system by reshaping how companies operate, innovate, and protect their assets. These technologies foster co-innovation networks, change innovation dynamics, and potentially shift firms towards trade secret protection (Drexl et al., 2021; Kappos & King, 2021). AI influences IP rights eligibility and protection criteria, potentially raising the bar for protectability (Picht et al., 2022; Makam, 2023). Existing human-centric intellectual property regimes may become irrelevant in an AI-driven environment (Lee et al., 2021). The Intersection of AI and intellectual property challenges patent, copyright, and trademark applicability (Singh & Singh, 2023). A sound framework should clarify and adapt current laws and carve out protection for AI innovations, as well as defend currently protected property rights from violation (Picht & Thouvenin, 2023). This relates to a flexible, adaptive regulatory framework (Ubaydullayeva, 2023). As a result, there's a growing need to clarify and adapt current IP laws and procedures. Legal frameworks may need to address issues such as AI creatorship, ownership of AI-generated output, and the allocation of neighboring rights. Policymakers and legal experts must consider these practical implications to ensure the IP system remains effective in an AI-driven future (Picht et al., 2022).

Although the literature provides a robust analysis of intellectual property issues, it often overlooks the ethical implications of AI-driven innovations in this area, particularly how they impact the balance of power between large organizations and smaller entities or individual creators. Our research will address this oversight by exploring the ethical dimensions of intellectual property management in the context of PLLMs, particularly concerning equity and fairness among stakeholders.

Use Cases and Examples

PLLMs offer potential applications in most industries. Some industries suggest themselves as more vulnerable to high-damage data breaches, while others suggest themselves as points where AI applications will have the greatest financial impact. We limit our discussion to three broad use cases: customer service, finance, and healthcare. AI has ethical and practical implications in each of these industries. They include customer service as likely to benefit greatly from AI, while finance and healthcare have larger stakes in high-damage data breaches. These stakes stem from privacy and regulatory compliance.

Customer Service.

Firms have used chatbots in customer service for some time, more extensively in B2C than in B2B. Fotheringham & Wiles (2023) argue that B2B sellers would benefit most from adding more sophisticated chatbots because of this lag. Amazon has used chatbots in customer service for some time. They cite enhanced customer experience in the form of shorter wait times, 24/7 coverage, scalability, and cost-saving among the benefits of AI chatbots. Of course, Amazon also offers AI development services through AWS Lev (Hnatushenko et al., (2024), and Alexa functions as a chatbot for many services.

Although the literature on AI use in customer service is extensive, there is a gap in understanding how the implementation of PLLMs specifically affects customer trust and long-term customer relationships. Our research aims to fill this gap by exploring the ethical implications of using PLLMs in customer service, focusing on transparency, accountability, and the protection of customer data.

Financial Services

Financial services offer many possible use cases for AI. We will focus on a few peculiar to the industry. AI has roots in pattern recognition, so it can play a major role in fraud detection and prevention. It monitors and flags suspicious activities, like too frequent use of a credit card in unusual locations. It can alert a card issuer or cardholder to start further investigation. It detects anomalies that humans might miss, including behavioral signals like the use of different devices (Shoetan & Familoni, 2024).

AI offers better ways to score credit and loans. It can scan a borrower's credit history in real time. It can assess people with limited credit history based on behavior like paying rent and utilities on time, job stability, and other predictive modeling. It is likely to improve the outcomes of credit scoring, giving loans to more of the deserving and denying them to people who are too risky (Tmelkov & Svrtinov, 2024). It can speed up risk assessment, claims prediction, and dynamic pricing in insurance underwriting. However, these applications may increase the risk of ethical violations, even as they make assessments more precise, especially in health insurance (Kharlamova et al., 2024).

While the literature covers various AI applications in financial services, there is a notable gap in addressing the ethical challenges associated with AI-driven decision-making processes, particularly how these processes might perpetuate biases or exclude certain groups of customers. Our research addresses this gap by proposing a framework for the ethical deployment of PLLMs in financial services, with a focus on fairness, inclusivity, and transparency.

Healthcare

As with financial services, healthcare offers a legion of potential use cases for AI. These use cases include data imaging, drug development, personalized patient treatment, mental health support, and resource management (Sai et al., 2024). Mongan et al. (2020) developed a checklist for conducting research on medical imaging with AI. The Checklist for Artificial Intelligence in Medical Imaging (CLAIM) includes 42 items that remind users and researchers of key elements in AI use. However, like any other tool, it must be used to be useful. Kocak et al. (2024) found that CLAIM was neglected more than it was used.

AI assists healthcare providers in delivering personalized patient care. This delivery includes plain language explanations of treatment and advanced electronic record keeping. Patients have a better chance of understanding their treatments and conditions. Their records are more likely to be updated, helping to avoid maltreatment like drug interactions or the prescription of allergy-generating drugs (Nova, 2023). AI will help develop more complete and accessible treatment plans (Chintala, 2023), predictive analytics, and decision support systems (Rana & Shuford, 2024). Mental health support also can benefit from AI. AI can be especially useful in diagnosis and treatment (Talati, 2023). The likelihood of ethical risks arises when used in predictions (Tutun et al., 2023).

Although AI applications in healthcare are well-documented, the literature does not thoroughly explore the ethical implications of using PLLMs, particularly in terms of patient autonomy and informed consent. Our research contributes to this area by examining how PLLMs can be ethically integrated into healthcare settings, ensuring that patient rights and privacy are respected.

Final Note on Use Cases

Similar use cases exist in most industries. Each use case offers benefits, but those benefits are often accompanied by ethical risks. In the following sections of the paper, we address these risks from the stakeholders' perspectives. Our work seeks to highlight the ethical considerations that are often overlooked in the current literature, providing a comprehensive framework for addressing these concerns.

Stakeholder Theory

Stakeholder theory recognizes and describes the complex interaction between the organization and a broader group of constituencies (i.e., stakeholders) beyond just its owners (i.e., shareholders) (Freeman, 1984; Deng et al., 2013). Freeman (1984) articulated the current conception of stakeholder theory by describing stakeholders as individuals or groups who can impact or are impacted by the actions of an

organization. Other definitions characterize organizational stakeholders as "individuals and constituencies that contribute, either voluntarily or involuntarily, to its wealth-creating capacity and activities, and who are therefore its potential beneficiaries or risk bearers" (Post et al., 2002, p. 8).There are a multitude of constituencies that could be considered organizational stakeholders, but primary stakeholders include shareholders, employees, customers, suppliers, competitors, government, and environmentalists (Freeman, 1984).

Stakeholder theory places the organization within an established and interconnected network of internal and external actors. To create wealth and ensure survival, organizations must successfully navigate these relationships by creating trust and mutual cooperation to satisfy their stakeholders (Post et al., 2002). Stakeholder theory also provides an ethical lens as it examines the organization's morals and values in managing its relationships with its stakeholders and creates effective corporate governance by addressing stakeholder, and not just shareholder, needs (Phillips et al., 2003; Stoelhorst & Vishwanathan, 2024).A major venue for satisfying organizational stakeholders and developing sound corporate governance is through successful corporate social responsibility (CSR) programs (Khalil & Rashed, 2023).

Despite its broad application, stakeholder theory in the context of AI and specifically PLLMs remains underexplored. While the literature covers general stakeholder relationships and the importance of CSR, there is a gap in understanding how stakeholder theory can be applied to manage the ethical challenges posed by PLLMs. Our research contributes to this emerging field by developing a framework that integrates stakeholder theory with the specific ethical considerations of PLLMs, focusing on how organizations can balance the competing interests of diverse stakeholders while ensuring ethical AI practices.

Corporate Social Responsibility (CSR)

Corporate social responsibility (CSR) is an acknowledgment that the organization is responsible to many stakeholders and not just to its shareholders (Duhaime et al., 2021). CSR is often conceptualized as a form of "enlightened self-interest" where the organization undertakes actions and policies that benefit itself and society (Mintzberg, 1983; Dutta et al., 2021).Carroll (1979, p. 500) provides a more formal definition and describes CSR as "the social responsibility of businesses that encompasses the economic, legal, ethical and discretionary expectations that society has of organizations at a given point in time."

CSR is a way for the organization to integrate, operationalize, and effectively respond to divergent stakeholder demands concerning social, environmental, and economic issues (Carroll, 1999). In particular, CSR is a mechanism of the organization's ability to create an environment that encourages sound corporate governance, sustainable practices, and accountability (Cai et al., 2012). These types of activities undertaken through CSR outreach help ensure corporate trust is built between the organization and its stakeholders. This happens as stakeholders are assured that the organization has not acted exploitatively and behaves in legal and ethical ways (Caruna & Chatzidakis, 2014).

However, the current literature on CSR does not fully address the complexities introduced by AI technologies like PLLMs. Specifically, there is a gap in understanding how CSR strategies can be adapted to address the unique ethical challenges posed by PLLMs, particularly concerning privacy, data security, and the potential for AI-driven inequalities. Our research aims to bridge this gap by proposing CSR strategies that are specifically tailored to the ethical management of PLLMs, ensuring that organizations can maintain stakeholder trust while innovating responsibly.

The reviewed literature provides a comprehensive understanding of the current state of research on PLLMs, highlighting critical areas such as privacy risks, regulatory compliance, and intellectual property issues. However, to move beyond theoretical considerations, examining how these issues manifest in real-world applications is essential. The following section will explore the ethical implications of PLLMs, focusing on how these technologies impact various stakeholders.

ETHICAL IMPLICATIONS OF PRIVATE LLMS FROM A STAKEHOLDER THEORY PERSPECTIVE

With the foundational understanding of PLLMs established, we now turn to the ethical implications that arise when these technologies are deployed within organizations. This section will analyze the consequences for key stakeholders, including employees, customers, and society at large. We offer insights into the practical challenges and ethical dilemmas that organizations must navigate.

As organizations begin to harness the incredible potential of developing and utilizing PLLMs, it is important to consider the ethical implications that invariably arise when revolutionary technology becomes mainstream (de Almeida et al., 2021). These implications include impacts on stakeholders like employees, customers, society, and the environment. The ethical challenges are multifaceted, requiring a balanced approach that takes into account the diverse needs and interests of all stakeholders. This section explores these ethical implications through the lens of stakeholder theory, emphasizing the importance of creating strategies that align with both organizational goals and ethical responsibilities.

Employees

For employees, private LLMs present both a threat and an opportunity. Some vocations will be completely remade, and the need for human workers will be dramatically reduced. Other vocations will foster new opportunities related to AI (Wang, 2023). Occupations most at risk for automation include vocations that largely consist of routine, predictable, and repetitive tasks. For example, Chui et al. (2016) predict that 78% of predictable physical work like welding, assembly line tasks, and food preparation could be automated with a minimal need for human interaction. To put this in context, repetitive tasks or predictable work represents about 20% of employee time in the United States (Chui et al., 2016). Specialized professions are also a rich target for employee elimination (Wang, 2023). AI is commonly used in white-collar professions, such as healthcare, legal research, and education, as organizations seek gains in efficiency and cost reduction (Chelliah, 2017; Wang, 2023).

From an ethical standpoint and a stakeholder perspective, organizations must consider the broader societal implications of these shifts. While cost savings and efficiency gains are significant, the displacement of workers can lead to social instability and a loss of trust in organizations. It is imperative for organizations to formulate strategies to upskill employees, ensuring they have the talent and skills necessary for both organizational survival and broader societal stability. This is an urgent issue, as the World Economic Forum predicted in 2020 that approximately 50% of employees across the globe will need re-skilling by the mid-2020s (Li, 2022). Specifically, in regards to private LLMs, employees will need a variety of skills in order to be successful. These include core skills around understanding and using AI, thinking skills that utilize creativity to solve complex and novel problems, self-management skills like effective time usage when using advanced technology like generative AI, and social and communication skills to ensure the use of AI is being used in the most effective manner by employees (Sofia et al., 2023). Organizations that neglect these considerations risk ethical breaches and long-term harm to their reputations and operational effectiveness.

Customer Privacy and Trust

Organizations must also consider the ethical implications of private LLMs for customers, who are an obviously extremely important group of stakeholders. First, organizations must implement systems and safeguards to protect user data. Users of private LLMs have the potential to share information that would be considered sensitive or even confidential and could be used for nefarious purposes in the wrong hands (Yao et al., 2024). Organizations can take meaningful steps to protect customers. These actions include implementing security measures that limit who can access and use data entered into the LLM, removing or disguising personally identifiable data, and implementing robust encryption (Villegas-Ch & Garcia-Ortiz, 2023).

Second, organizations must be transparent and gain consent from users of their private LLMs. The European Union, through its General Data Protection Regulation (GDPR), asserts that an individual's

personal information is a fundamental right and accords a mechanism for controlling how organizations use their information (Felzmann et al., 2019). This includes informing individuals accessing AI systems about what personal data is being utilized and any potential implications (Wulf & Seizov, 2022). Transparency is not just a regulatory requirement but an ethical imperative. We propose that stakeholders would be well served by organizations that provide adequate transparency about how personal data could be used and then seek user consent so stakeholders can make informed decisions. Failure to do so could result in a significant erosion of trust, leading to reputational damage and potential legal liabilities.

Social & Environmental Impact

Finally, the ethical implications of private LLMs regarding social and environmental impacts must be factored into organizational stakeholder considerations. Floridi et al. (2018) provide a six-pronged guide to developing an ethical framework for AI that benefits society. These principles are: (1) creating technology that promotes the well-being of society, (2) does not harm society, (3) providing decision-making autonomy to users, (4) promoting common prosperity, (5) understanding how the technology works, and its creators are held accountable.

In addition to these principles, organizations must also consider the environmental impact of PLLMs. The computational resources required to train and operate large language models can be substantial, leading to significant energy consumption and carbon emissions. As stakeholders become more environmentally conscious, the pressure on organizations to adopt sustainable practices will increase. In this context, organizations should design systems that promote long-term sustainability practices as they promote stakeholder engagement and firm success (Eccles et al., 2014). This includes exploring energy-efficient algorithms, offsetting carbon emissions, and being transparent about the environmental costs of AI initiatives. By doing so, organizations can align their ethical responsibilities with stakeholder expectations, fostering a more sustainable and equitable approach to AI development.

The ethical challenges PLLMs present are complex and multifaceted, affecting a wide range of stakeholders in different ways. These challenges underscore the need for a structured approach to ensure organizations can balance their technological ambitions with their ethical responsibilities. The following section introduces a framework designed to help organizations navigate these challenges by aligning ethical considerations with stakeholder value.

Balancing Ethical Considerations With Stakeholder Value

Given the ethical challenges outlined in the previous section, a proactive approach is required to balance the interests of all stakeholders. This section presents a framework for integrating ethical considerations into organizations' decision-making processes, ensuring that PLLMs are developed and deployed responsibly and sustainably.

We propose a framework that enables organizations to create stakeholder value via private LLMS while acting ethically, responsibly, and sustainably. This framework is grounded in the principles of responsible innovation, stakeholder engagement, and long-term value creation. By aligning these principles with organizational objectives, firms can navigate the ethical complexities of PLLMs while ensuring that all stakeholders' interests are adequately considered.

Responsible Innovation

We believe it is imperative for organizations to act in a responsible manner when creating private LLMs, lest they inadvertently harm stakeholders. Irresponsible development creates conditions where organizations violate existing laws and regulations, such as allowing conditions to exist that create bias when evaluating resumes during the hiring process (Dastin, 2018) or when it is unclear how the algorithm is making the decision in the first place (Drage et al., 2024). Another potential hazard of irresponsible innovation around AI is damage to the organization's brand when users are allowed to leverage its technology for offensive and inflammatory purposes, such as training a chatbot on racist and discriminatory language (Hunt, 2019).

Eitel-Porter (2020) provides key principles for developing responsible AI and an implementation approach that ensures these principles are followed. His key principles include creating conditions for fairness in the model, accountability for prejudiced or incorrect information, transparency about the model's inner workings, explainability for how the model returns output, and ensuring user privacy. Eitel-Porter recommends a robust governance structure where these principles are operationalized. These recommendations include starting with the founding principles summarized above, establishing an ethics board to deal with the thornier issues that arise with AI development, organization-wide training on the organization's AI governance policies, creating the conditions that allow employees to challenge these policies, establishing metrics to ensure the guiding principles are adhered to, and finally establishing an environment where concerns can be aired.

While these principles provide a solid foundation, our framework extends beyond them by incorporating stakeholder-specific considerations. For example, responsible innovation should address fairness and accountability and ensure that the interests of marginalized or vulnerable stakeholders are protected. This includes considering the broader societal impacts of PLLMs, such as potential job displacement or exacerbation of digital divides, and developing strategies to mitigate these risks. By integrating these considerations into the innovation process, organizations can ensure that their AI initiatives are not only responsible but also equitable and inclusive.

Stakeholder Engagement

As outlined in the literature review section of this work, it is critical for organizations to engage with their stakeholders in an effort to better meet their needs so as to create trust and mutual cooperation (Freeman, 1984; Post et al., 2002). This is particularly true in an environment where private LLMs, in part, guide organizational decision-making that affects stakeholders (Mills et al., 2023). Thus, engaging with stakeholders to understand broader societal impacts is essential. For example, ChatGPT utilizes reinforcement learning from human feedback (RLHF), where the LLM is given feedback from a human perspective (i.e., users) to reinforce the desired model for optimization, which is incorporated into the model's future behavior (Santhosh, 2023).

However, effective stakeholder engagement requires more than just gathering feedback; it involves a continuous dialogue with stakeholders, ensuring their concerns are addressed and their insights are integrated into decision-making processes. Our framework emphasizes the importance of transparency in this engagement process, ensuring that stakeholders are informed about how their feedback influences the development and deployment of PLLMs. Moreover, organizations should actively seek out the perspectives of underrepresented or marginalized groups, whose voices might otherwise be overlooked, to ensure that the benefits of AI are distributed equitably.

Long-Term Value Creation

Finally, we recommend that organizations focus on long-term value creation by investing in sustainable and ethical practices in their private LLM developments. The principles of shared values provide a good context for this type of framework. According to these principles, organizations can create economic value while also creating value for society by tackling difficult and often intractable societal issues (Porter, 2023). Google's AI for Social Good project is a prime example of shared values in the AI/ LLM space. Google's AI technology is being leveraged to improve healthcare in the less developed world, forecast floods, track other natural disasters like wildfires, and create greener and more sustainable cities (Google AI, n.d.). Through outreach efforts like this, organizations are able to better meet the needs of their broader stakeholder group, which engenders trust and cooperation and casts the organization in a favorable light (Mayer et al., 1995).

Our framework builds on the concept of shared value by emphasizing the importance of sustainability not just as a corporate responsibility but as a core component of long-term value creation. This includes considering the environmental impact of PLLMs, such as their energy consumption and carbon footprint, and implementing strategies to minimize these effects. Organizations should also consider the social implications of their AI initiatives, such as their impact on employment, inequality, and community well-

being, and take proactive steps to address these challenges. By doing so, organizations can ensure that their AI innovations contribute to a more sustainable and equitable future, benefiting not just their shareholders but society as a whole.

By aligning ethical considerations with stakeholder value, organizations can create a balanced approach to the development and deployment of PLLMs. However, achieving this balance requires more than just a theoretical framework; it necessitates the implementation of concrete policies and practices that can guide organizational behavior. The following section outlines policy recommendations that operationalize the principles discussed, providing practical steps for organizations to follow.

Policy Recommendations

Building on the framework for balancing ethical considerations with stakeholder value, this section provides actionable policy recommendations for organizations. These policies are designed to operationalize the principles of responsible innovation, stakeholder engagement, and long-term value creation. This is important to ensuring that PLLMs are used ethically and effectively within corporate environments.

As organizations develop PLLMs, it is crucial to establish clear governance policies. Policy builds trust in applying and using PLLMs while ensuring ethical, safe corporate usage (Meltzer, 2023). Comprehensive policies address structures for enabling opportunity, managing risk, enhancing security, privacy, misinformation controls, and addressing copyright infringement. These policies should be dynamic and adaptable, reflecting the evolving nature of AI technologies and the regulatory landscape.

TABLE 1
CORPORATE POLICY GOVERNANCE FOR PLLMS

Enable Opportunity	
Develop Transparency and Trust	Share PLLM foundational information, including training methods, a summary of training data, and how the system is maintained with stakeholders.
Develop PLLM Standards	Create standards that ensure PLLM systems are explainable and interpretable for stakeholders.
Document PLLM systems and processes	Establish consistent documentation processes to share PLLM development and access.
Increase Access to PLLM Products and Services	Expand the availability and adoption of PLLM-driven products and services within the corporation and its customer base.
Manage Risks	
Protect Against Discrimination, Exclusion, and Toxicity	Implement and adhere to corporate privacy and anti-discrimination policies to safeguard against PLLM biases and misuse.
Commit to PLLM Ethical Principles and Cooperation Standards	Align with internationally recognized ethical guidelines for PLLM development and deployment within corporate policies. Adhere to international PLLM standards relevant to stakeholder usage.
Adopt PLLM Risk Management Framework	Adopt frameworks like the NIST AI Risk Management Framework to guide corporate PLLM policies and practices.

Increase Sharing on Governance	Share knowledge and experiences related to PLLM governance with other organizations to improve corporate standards.
Establish and Implement Code of Conduct	Establish and implement a Code of Conduct for AI into corporate governance policies. Corporations can adopt adherence to existing standards like the G7 Code of Conduct for AI.
Apply Data Governance Best Practices	Share and implement best practices for managing data responsibly
Enhance Security, Privacy, and Misinformation Controls	
Personal Data Access	Establish principles for stakeholder access to personal data and apply minimum standards like the OECD principles for corporate access to personal data.
Privacy Regulations	Implement appropriate corporate privacy measures to protect user data and comply with international standards.
Implement Governance Best Practices	Exchange best practices for PLLM governance with other corporations and institutions to improve overall industry standards.
Cooperation and Reporting on Misinformation	Work together with other corporations to combat misinformation and disinformation. Report instances of known misinformation with public notice.
Address Copyright Infringement	
Implement Copyright Safeguards	Apply safeguards for intentional copyright misuse to protect stakeholders from purposeful copyright infringement.
Monitor Copyright Laws	Stay informed and adapt to changes in copyright laws affecting AI to ensure corporate compliance.

The recommendations provided here form a framework for ensuring that PLLMs are used ethically and responsibly, balancing stakeholder interests while fostering innovation. By adopting these policies, organizations can mitigate risks, enhance trust, and ensure that their use of AI technologies aligns with ethical standards and stakeholder expectations. As the AI landscape continues to evolve, these policies should be regularly reviewed and updated to remain effective and relevant.

CONCLUSION

This research delves into the ethical considerations surrounding private large language models (PLLMs) through the lens of stakeholder theory. In an era where AI technologies are rapidly transforming industries and societies, developing and deploying PLLMs present both opportunities and challenges. Our analysis underscores the importance of balancing stakeholder interests with ethical considerations, emphasizing that responsible innovation is not merely a corporate obligation but a societal imperative.

We have proposed a comprehensive framework that guides organizations in ethically navigating the complexities of PLLMs. This framework promotes transparency, accountability, and sustainability—principles crucial for fostering trust and long-term value creation. By integrating responsible innovation,

continuous stakeholder engagement, and a commitment to long-term value, organizations can ensure that their AI initiatives contribute positively to their bottom line and the broader society.

However, the journey toward ethical AI is ongoing. As technologies evolve, so must our governance, regulation, and stakeholder engagement approaches. Future research should focus on developing regulatory frameworks that foster innovation while protecting stakeholders, conducting detailed social impact assessments to identify emerging risks and opportunities, and exploring strategies for fostering effective human-AI collaboration to ensure human oversight remains central in AI-driven decision-making processes.

Ultimately, the success of PLLMs will depend not only on their technical capabilities but also on the ethical frameworks within which they operate. By continuing this conversation and conducting further research, we can ensure that PLLMs are developed and used in ethical, equitable, and beneficial ways to all stakeholders. The future of AI is not just about advancing technology; it's about advancing humanity—and it is incumbent upon us to guide this powerful tool toward the greater good.

REFERENCES

Anderljung, M., Smith, E., O'Brien, J., Soder, L., Bucknall, B., Bluemke, E., . . . Chowdhury, R. (2023, November). Towards publicly accountable frontier LLMs. In *Socially Responsible Language Modelling Research*.

Cai, Y., Jo, H., & Pan, C. (2012). Doing well while doing bad? CSR in controversial industry sectors. *Journal of Business Ethics*, *108*, 467–480.

Carroll, A.B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of Management Review*, *4*(4), 497–505.

Carroll, A.B. (1999). Corporate social responsibility: Evolution of a definitional construct. *Business & Society*, *38*(3), 268–295.

Caruana, R., & Chatzidakis, A. (2014). Consumer social responsibility (CnSR): Toward a multi-level, multi-agent conceptualization of the "other CSR." *Journal of Business Ethics*, *121*, 577–592.

Chelliah, J. (2017). Will artificial intelligence usurp white-collar jobs? *Human Resource Management International Digest*, *25*(3), 1–3.

Chintala, S. (2023). AI-driven personalised treatment plans: The future of precision medicine. *Machine Intelligence Research*, *17*(2), 9718–9728.

Chui, M., Manyika, J., & Miremadi, M. (2016). Where machines could replace humans—and where they can't (yet). *The McKinsey Quarterly*, pp. 1–12.

Dastin, J. (2018, October 10). *INSIGHT - Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.com. Retrieved July 10, 2024, from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

de Almeida, P.G.R., dos Santos, C.D., & Farias, J.S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, *23*(3), 505–525.

Deng, X., Kang, J.-k., & Low, B.S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, *110*(1), 87–109.

Drage, E., McInerney, K., & Browne, J. (2024). Engineers on responsibility: Feminist approaches to who's responsible for ethical AI. *Ethics and Information Technology*, *26*(1), 4.

Duhaime, I.M., Hitt, M.A., & Lyles, M.A. (Eds.). (2021). *Strategic management: State of the field and its future*. Oxford University Press.

Dutta, K., & Ring, K. (2021). Do do-gooders do well? Corporate social responsibility, business models and IPO performance. *Journal of Applied Business and Economics*, *23*(2).

Eccles, R.G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835–2857.

Evertz, J., Chlosta, M., Schönherr, L., & Eisenhofer, T. (2024). Whispers in the machine: Confidentiality in LLM-integrated systems. *arXiv preprint*. arXiv:2402.06922.

Felzmann, H., Villaronga, E.F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 2053951719860542.

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*, 689–707.

Fotheringham, D., & Wiles, M.A. (2023). The effect of implementing chatbot customer service on stock returns: An event study analysis. *Journal of the Academy of Marketing Science*, *51*(4), 802–822.

Freeman, R.E. (1984). *Strategic management: A stakeholder approach.* Cambridge University Press.

Google AI. (n.d.). *Google AI and social good*. Retrieved from https://ai.google/responsibility/social-good/

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access.*

Hnatushenko, V., Ostrovska, K., & Nosov, V. (2024). Development and research of a chatbot using the linguistic core of Amazon Lex V2. In *COLINS* (Issue 3, pp. 50–62).

Hunt, E. (2019, September 9). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

Ioannidis, J., Harper, J., Quah, M.S., & Hunter, D. (2023, June). Gracenote.ai: Legal generative AI for regulatory compliance. In *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023).*

Khalil, M., & Rashed, A. (2023). The impact of female directors on the relationship between corporate social responsibility and capital structure: Evidence from Egypt. *Journal of Applied Business and Economics*, *25*(2).

Kharlamova, A., Kruglov, A., & Succi, G. (2024, May). State-of-the-art review of life insurtech: Machine learning for underwriting decisions and a shift toward data-driven, society-oriented environment. In *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1–12). IEEE.

Kocak, B., Keles, A., & Akinci D'Antonoli, T. (2024). Self-reporting with checklists in artificial intelligence research on medical imaging: A systematic review based on citations of CLAIM. *European Radiology*, *34*(4), 2805–2815.

Li, L. (2022). Reskilling and upskilling the future-ready workforce for Industry 4.0 and beyond. *Information Systems Frontiers*, pp. 1–16.

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

Meltzer, J.P. (2023, November 1). Toward international cooperation on foundational AI models: An expanded role for trade agreements and international economic policy. *SSRN.* Retrieved from https://ssrn.com/abstract=4685309

Mills, S., Sampanthar, K., & Dardaman, E. (2023, February 5). *Getting stakeholder engagement right in responsible AI*. VentureBeat.com. Retrieved July 10, 2024, from https://venturebeat.com/ai/getting-stakeholder-engagement-right-in-responsible-ai/

Mintzberg, H. (1983). The case for corporate social responsibility. *Journal of Business Strategy*, *4*(2), 3–15.

Mongan, J., Moy, L., & Kahn, Jr., C.E. (2020). Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiology: Artificial Intelligence*, *2*(2), e200029.

Nova, K. (2023). Generative AI in healthcare: Advancements in electronic health records, facilitating medical languages, and personalized patient care. *Journal of Advanced Analytics in Healthcare Management*, *7*(1), 115–131.

Phillips, R.A. (1997). Stakeholder theory and a principle of fairness. *Business Ethics Quarterly*, *7*(1), 51–66.

Porter, M.E. (2023, April 4). Creating shared value. *Harvard Business Review*. Retrieved from https://hbr.org/2011/01/the-big-idea-creating-shared-value

Post, J.E., Preston, L.E., & Sachs, S. (2002). Managing the extended enterprise: The new stakeholder view. *California Management Review*, *45*(1), 6–28.

Rana, M.S., & Shuford, J. (2024). AI in healthcare: Transforming patient care through predictive analytics and decision support systems. *Journal of Artificial Intelligence General Science (JAIGS)*, *1*(1).

Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J.J. (2024). Generative AI for transformative healthcare: A comprehensive study of emerging models, applications, case studies, and limitations. *IEEE Access*.

Santhosh, S. (2023, January 15). Reinforcement learning from human feedback (RLHF)–ChatGPT. *Medium*. Retrieved from https://medium.com/@sthanikamsanthosh1994/reinforcement-learning-from-human-feedback-rlhf-532e014fb4ae#:~:text=Reinforcement%20learning%20from%20human%20feedback%20(RLHF)%20has%20gained%20popularity%20with,language%20model%20with%20reinforcement%20learning

Shoetan, P.O., & Familoni, B.T. (2024). Transforming fintech fraud detection with advanced artificial intelligence algorithms. *Finance & Accounting Research Journal*, *6*(4), 602–625.

Sofia, M., Fraboni, F., DeAngelis, M., Puzzo, G., Giusino, D., & Pietrantoni, L. (2023). The impact of artificial intelligence on workers' skills: Upskilling and reskilling in organisations. *Informing Science: The International Journal of an Emerging Transdiscipline*, *26*, 39–68.

Stoelhorst, J.W., & Vishwanathan, P. (2024). Beyond primacy: A stakeholder theory of corporate governance. *Academy of Management Review*, *49*(1), 107–134.

Talati, D. (2023). Artificial intelligence (AI) in mental health diagnosis and treatment. *Journal of Knowledge Learning and Science Technology*, *2*(3), 251–253.

Temelkov, Z., & Georgieva Svrtinov, V. (2024). AI impact on traditional credit scoring models. *Journal of Economics*, *9*(1), 1–9.

Tutun, S., Johnson, M.E., Ahmed, A., Albizri, A., Irgil, S., Yesilkaya, I., . . . Harfouche, A. (2023). An AI-based decision support system for predicting mental health disorders. *Information Systems Frontiers*, *25*(3), 1261–1276.

Villegas-Ch, W., & García-Ortiz, J. (2023). Toward a comprehensive framework for ensuring security and privacy in artificial intelligence. *Electronics*, *12*(18), 3786.

Wang, Y. (2023). The large language model (LLM) paradox: Job creation and loss in the age of advanced AI. *Authorea Preprints*.

Wulf, A.J., & Seizov, O. (2024). "Please understand we cannot provide further information": Evaluating content and transparency of GDPR-mandated AI disclosures. *AI & Society*, *39*(1), 235–256.

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, *4*(2), 100211.

Zilliz. (2024, April 11). *What are private LLMs? Running large language models privately - privateGPT and beyond*. Retrieved from https://zilliz.com/learn/what-are-private-llms