

Tale of Two Storylines: An NLP-assisted Corpus Study of Open AI's Identity Construction

Kaiwen Cheng

kevincheng@sisu.edu.cn

College of Language Intelligence, Sichuan International Studies University

Ying Lan

College of English studies, Sichuan International Studies University

Gang He

College of Language Intelligence, Sichuan International Studies University

Research Article

Keywords: Corporate Identity Construction, Blog, OpenAI, NLP, Discursive Positioning

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10948391/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Tale of Two Storylines: An NLP-assisted Corpus Study of *Open AI's* Identity Construction

Abstract

Corporate identity profoundly influences stakeholder perceptions, yet existing studies often privilege formal disclosures by publicly listed firms while neglecting how informal digital platforms shape organizational legitimacy. This study investigates how OpenAI constructs its corporate identity through discursive and narrative strategies embedded in its official blog communication. By integrating Natural Language Processing (NLP) techniques with the Discursive Positioning Framework, the study reveals a dual narrative storyline that underpins OpenAI's identity construction. The primary storyline articulates three distinct identities: (1) AI Safety and Benefit Guardian, emphasizing ethical stewardship and risk mitigation; (2) Advanced AI Innovator, foregrounding technological breakthroughs; and (3) Shared AI Advocate, advocating collaborative, open-source development. Beneath this public story, however, lies an underlying storyline that highlights unresolved tensions between OpenAI's aspirational discourse and its operational realities, such as its paradoxical roles as a Privacy Peeper (scrutinizing user data under the guise of safety) and a Technology Monopolist (consolidating market influence despite open-access rhetoric). The findings suggest that companies should not only focus on formal disclosures but also ensure coherence in the identities projected across informal platforms while managing organizational identity. Such cross-channel consistency is fundamental to effective brand governance and the maintenance of public trust.

Keywords: Corporate Identity Construction; Blog; OpenAI; NLP; Discursive Positioning

1 Introduction

The concept of corporate identity, first introduced by Lippincott and Margulies in 1957, addresses the fundamental questions of “Who am I, what am I doing, and how do I do it?”. It shapes a company’s unique image and, in turn, its competitiveness and reputation (Bamberg & Dege, 2021; Jaworska, 2020; Melewar, 2008). While earlier research primarily focused on the static aspects of corporate identity—such as logos, symbols, and other visual elements—recent studies have expanded to include more dynamic interpretations. These dynamic interpretations explore how corporate identity is actively constructed, negotiated, and reshaped over time across disciplines such as management, sociology, and linguistics (Balmer & Greyser, 2002; Melewar, 2003).

From a sociolinguistic perspective, discourse both reflects and reproduces social reality, enabling organizations to construct their identities through language (Fairclough, 2013). It allows individuals and organizations to shape and communicate their identities through language, which materializes into practices that are reproduced in the social world. Methods such as discourse analysis have traditionally been applied to texts like annual reports (Fuoli, 2018) and CSR (Corporate Social Responsibility) reports (Li, 2022; Sun & Jiang, 2014) to uncover how companies construct their identities. Meanwhile, the advent of digital texts and big data has encouraged the use of corpus linguistics and natural language processing (NLP) techniques in corporate identity studies. Researchers have illustrated the efficacy of these methods in analyzing extensive corpora of texts from public companies, revealing patterns and themes that significantly influence corporate identity (Goloshchapova et al., 2019; Han et al., 2024). Nonetheless, there also remains a paucity of research applying such techniques to private firms, especially those operating within emerging technology sectors, including artificial intelligence (AI). This study addresses this gap by applying Word2Vec-

based K-Means NLP techniques to analyze the official blog of OpenAI, a leading private company in the AI industry.

2. Literature Review and Theoretical Framework

2.1 Corpus-driven Corporate Identity Research

Research in various fields like anthropology, linguistics, sociology, history, and psychology, to name a few, has firmly established and acknowledged the essential role of linguistic strategies and discourse processes in the construction and negotiation of identities (De Fina & Georgakopoulou, 2020). Especially in linguistic research, a common approach is corpus-driven critical discourse analysis. For instance, Fuoli analyzed stance expressions in annual reports by constructing a corpus for corporates' social responsibility report (Fuoli, 2018). Fuoli's study revealed that companies employ varied stance expressions to construct distinct corporate identities, catering to stakeholder expectations and enhancing report persuasiveness. Under this approach, Perrino (2015) combined linguistic anthropology and sociolinguistics to analyze spatiotemporal effects in oral narratives of Italian executives, focusing on brand identity construction. Similarly, Sun and Jiang (2014) compared metaphors in Chinese and American corporate mission statements, uncovering cultural differences in corporate identity expression. Moreover, Li (2022) reveals how corporations construct their corporate identity through discourse in their corporate social responsibility reports, using a corpus-assisted critical discourse analysis of Starbucks, highlighting the significant application of corpus methodology in researching corporate identity.

Despite these insights, it's crucial to acknowledge that corporate identity is not static. As noted by Benwell (2006) and De Fina (2010), corporate identity is a constantly evolving social construct ,

underscoring its nature as an ongoing, dynamic process as opposed to a fixed characteristic. Future research should focus not only on what a company “says” but also what it “does” at a given time and the reasons behind these identity changes.

2.2 Text Mining and Corporate Identity Research

Text mining is a method for analyzing vast quantities of textual data and uncovering valuable insights(Hassani et al., 2020). This method is used to make sense of data using natural language processing (NLP) techniques and statistical methods. The integration of NLP in corporate identity and reputation research not only enhances evaluation accuracy and objectivity but also underscores the pivotal role of digital communication in shaping and maintaining a company’s public image and its identity. Romano et al. (2024) proposed an NLP-based approach to quantify the alignment between corporate communication and its values, enhancing evaluation accuracy and objectivity. Cruciata et al. (2024) also utilized NLP with a hybrid text classification method (ZSTC) to compare corporate texts, demonstrating a strong correlation between text and stakeholder perception, showcasing NLP’s significant potential in corporate value assessment. NLP is also important in analyzing corporate reputation. Corporate identity and reputation are mutually reinforcing: a strong identity boosts reputation, which in turn strengthens and spreads the identity (Whetten & Mackey, 2002). By leveraging text mining, companies can analyze customer feedback from social media, online reviews, and forums, enabling them to enhance their brand reputation and address customer satisfaction issues effectively(Mousavi et al., 2020; Rust et al., 2021). Kayakuş et al. (2024) used sentiment analysis and machine learning to show how customer feedback shapes brand reputation, highlighting the dynamic relationship between identity and reputation. Rantanen et al. (2019) demonstrated the effectiveness of machine learning in classifying online reputation in the banking

sector, establishing a foundation for NLP's role in reputation management.

Corporate Social Responsibility (CSR) contains company's identity in corporate communication. Goloshchapova et al. (2019) utilized Latent Dirichlet Allocation (LDA, a topic modeling technique) to analyze CSR reports from European companies (1999–2016), revealing key themes that influence corporate public identity. Han et al. (2024) introduced a deep learning and NLP framework to detect and measure CSR content on social media, examining its impact on financial performance. This highlights the pivotal role of digital CSR communication in shaping corporate identity. However, a notable research gap is the prevalent focus on CSR or annual reports of listed companies, leaving the identity construction of private companies, which are not obligated to disclose such reports, as an unexplored area.

Blogs—also known as weblogs—are online journals designed for public viewing and constitute a notable medium in this context. Bloggers regularly update their sites with new content, which then prompts comments from readers and viewers. The meanings and emotions attached to roles influence the way individual and organizational identities are intertwined in fluid online spaces (Bange et al., 2022). Blogs serve as a vital channel for private companies to convey their identity, values, and ideas to the outside world, making them a significant vehicle for shaping and promoting corporate identity externally. Therefore, combining Word2Vec-based K-Means NLP techniques with discursive positioning framework, the current study aims to address the following questions:

1) What kind of identity does OpenAI construct in its blogs?

2) How does OpenAI employ specific discursive strategies to highlight its specific identities?

3) What underlying social cognitions or ideologies are reflected in the construction

OpenAI's identity?

2.3 Discursive Positioning Approach

Originated from Positing Theory, the discursive positioning approach delineates the manner in which individuals are situated, or situate themselves, within the narratives that emerge during social engagements (Davies & Harré, 1990). This approach enables individuals to form understandings of themselves and others within various contexts, emphasizing the active role language plays in crafting our social realities, identities, and relationships. The essence of discursive positioning lies in its ability to unravel the construction of subjectivity and the fluidity of self-identity within social realms. Speakers during a conversation assign specific roles to themselves and others, with these positions being fluid and subject to continuous reconstruction through interaction dynamics (Harré, 2008). A key element of this approach is identity fluidity, which suggests that identities are not fixed but are constantly renegotiated and reconstructed based on changing social and discursive environments. This fluidity suggests that a person's identity is not static but is shaped by their discourse and narrative participation, as noted by Davies and Harré (1990). They underscore that within narrative positioning, the "I" becomes a narrative subject for each dialogue participant, capable of evolving as the narrative progresses. This concept proposes the understanding of oneself in a historical, coherent, and holistic manner through five key elements of discursive positioning:

- **Identity categories** - Defining social identities based on various classifications.
- **Engagement in discursive practices** - How meanings are attributed to these identity categories through discourse.
- **Self-positioning** - How individuals locate themselves within these categories and narratives.

- **Recognition of subcategories** - Acknowledging one's attributes that align with certain subcategories while differing from others in storylines.
- **Navigating contradictory identities** - Managing conflicting aspects of one's identity.

The concept of storyline refers to a sequence of discourse acts that are meaningful to individuals, as understanding an event often begins with its narrative structure and the actions taken by its protagonists (Potter & Wetherell, 1987). Crucially, how these protagonists interpret events and their underlying intentions are essential for making sense of the discourse. Harré (2012) further elaborates that people use various contextual features to make such inferences, including prior discourse, anticipated future actions, and the characteristics of the speaker. These inferences contribute to constructing a cohesive storyline that integrates different cognitive and temporal aspects. Therefore, the narrative plot encompasses cognition, temporality, and normativity, which implicitly or explicitly link the past, present, and future. This linkage helps in positioning individuals within discourse, which, in turn, reflects specific identities or particular identity traits in distinct contexts (Yuan & Li, 2022). The current study uses the discursive positioning approach to clarify the outcomes of clustering and word vectors, with the aim of identifying the types of self-positions or identities that OpenAI constructs in its blogs.

3 Methodology

3.1 Data collection

The current study uses the blogs publicly available on OpenAI's official website until March 8, 2024 as the research corpus. All the blogs are open and allowed to be scraped. The data acquisition process was performed using Python's requests and BeautifulSoup libraries to automate the

extraction of blog posts from OpenAI’s official website. The approach ensured that all relevant articles were gathered consistently and saved locally for further analysis. For further details regarding the preprocessing code, please refer to the Appendix.

A custom Python script was developed to achieve this task. The script began by fetching the sitemap from OpenAI’s website to identify the URLs of all blog posts. Using the requests library, the sitemap was retrieved, and BeautifulSoup was employed to parse the content and extract URLs related to blog posts. Each blog post URL was then accessed to extract its content.

The script retrieved each blog post’s title, content, and publication date. The title was used as the basis for the file name, while the content, extracted from specific HTML elements (e.g., div tags with the class ui-richtext), was saved to individual text files categorized by year. The content was also processed to ensure readability by splitting paragraphs into individual sentences, which were saved with each blog post. In this study, all the blogs are open and allowed to be scrapped. The corpus contains approximately 435,000 characters in a total of 118 valid blogs.

Table 1 Descriptive Statistics of the OpenAI Blog Corpus (2015-March 2024)

Year	Number of Articles	Total Character Count	Average Character Count per Article
2015	1	3529	3529
2016	9	11502	1278
2017	2	1537	769
2018	11	23840	2167

Year	Number of Articles	Total Character Count	Average Character Count per Article
2019	10	32093	3209
2020	6	18585	3098
2021	8	24196	3025
2022	19	79171	4167
2023	44	207327	4712
2024 (until March)	8	33342	4168

^a Data set comprises 118 publicly accessible OpenAI blog posts published between 11 December 2015 and 8 March 2024 ($\approx 435\,000$ characters). URLs were collected from the site’s public sitemap and parsed with a custom Python scraper (requests, BeautifulSoup). The preprocessing script is provided in the Appendix.

3.2 Data Analysis Procedure

The whole process of data analysis can be referred in the following flowchart (Figure 1). The acquisition of the corpus involves using web scraping to gather blogs from OpenAI. Then, the obtained original text data were input into the Python IDE tool Pycharm. Within a Python 3.9 environment, the nltk package were used for text tokenization and cleaning. Then, we imported word2vec library of gensim and utilized the blog as corpus to train the word embeddings for clustering database. Specifically, the model uses Continuous Bag of Words (CBOW) method(Mikolov, 2013) to generate vector representations of words based on their context in the training corpus. After training, cosine similarity(Manning, 2009) is used to measure the similarity between the word vectors. We print 5 examples in each topic of clusters to identify the identity

category (Defining social identities based on various classifications). Then we calculated the word2vec of “OpenAI” and demonstrate the similar words with OpenAI in our blog model. Based on word2vec calculation result, the storyline of OpenAI is used to analysis for the construction process of those identities.

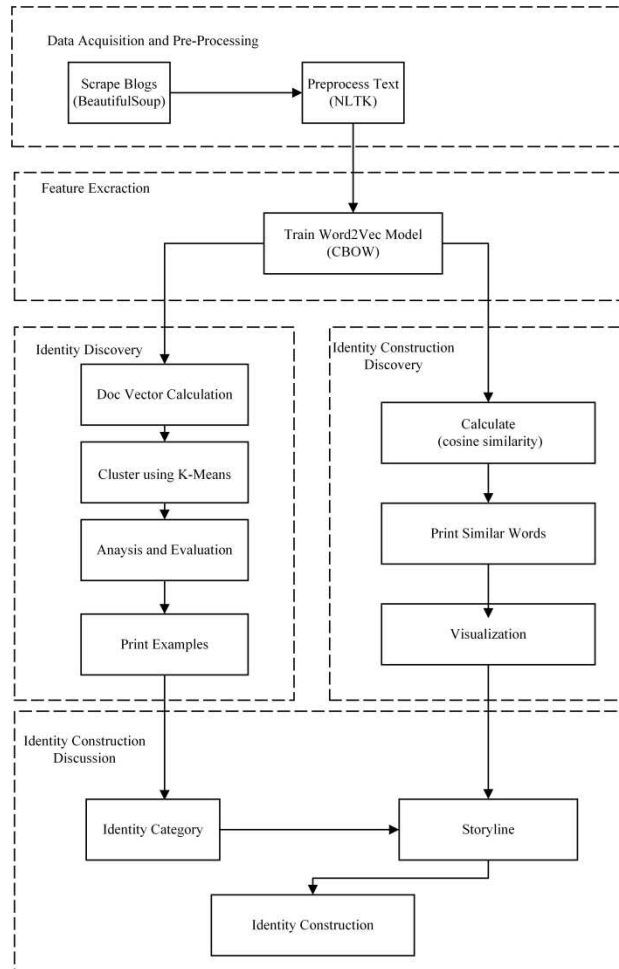


Figure 1. Flowchart of the Study

3.3 Analytic Framework Based on Discursive Positioning

Based on the Discursive Positioning approach (Davies & Harré, 1990; Harré, 2008), the current study proposed and applied an analytic framework to interpret how OpenAI constructs its corporate identities through discourse, specifically by analyzing corpus-based results derived from

clustering and Word2vec modeling. The framework includes the forementioned 5 elements operating at three levels: a) initial tags, b) ideational tags, and c) contradictory tags, as shown in Figure 2. In discourse, meanings are ascribed to individuals through discursive practices, creating foundational identity labels (initial tags). The “I” then enriches its identity with specific narratives (ideational tags), distinguishing its positioning in unique storyline. Throughout discursive interactions, conflicts in positioning (contradictory tags) may arise. Through actions, solutions, or discourse strategies to navigate identities, these dimensions collectively facilitate the self-positioning process. The analysis proceeds in three stages as follows.

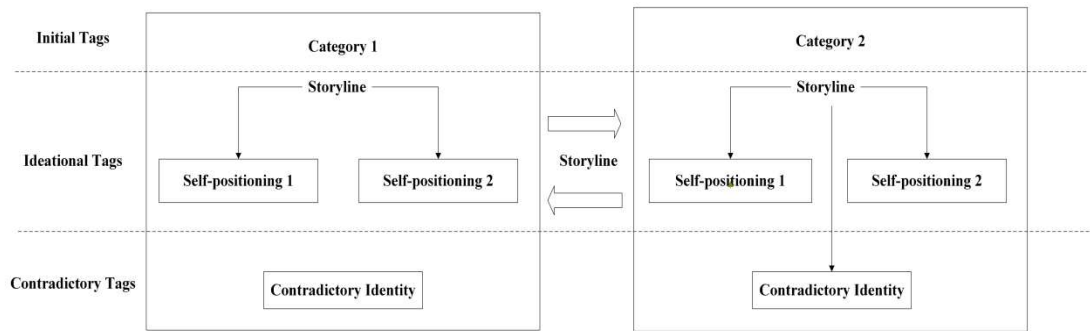


Figure 2 The Discursive Positioning Framework

First, the three document clusters generated by the K-Means algorithm (see Section 4.1) provide an empirical foundation for identifying OpenAI’ s self-representations. Each cluster was interpreted by examining its titles, content previews, and word clouds (Figures 3 – 5) to infer its central theme. Close reading of representative texts within each cluster ensured that the thematic interpretation was grounded in linguistic evidence.

Second, combining the Word2Vec results with qualitative discourse analysis enabled a focus on self-positioning. The Word2Vec model calculated the semantic similarity of various terms to OpenAI (see Table 3), producing core lexical items that functioned as narrative markers of identity

construction. For instance, words such as project, research, and engineer indicated OpenAI's positioning as a technological innovator; safety, benefit, and human marked its role as an ethical guardian; and API, access, and nonprofit signaled its stance as an open collaborator. Interpreting these clusters through the DPF framework allowed the study to conceptualize these linguistic regularities as components of OpenAI's primary storyline, which articulates its desired public image.

Finally, a critical reading was conducted to identify contradictory or tension-filled identity elements, corresponding to the DPF dimension of navigating contradictory identities. This stage compared OpenAI's stated values, such as openness, safety, and broad benefit, with its operational descriptions to detect discursive inconsistencies or omissions. Instances where the rhetoric of openness coexisted with references to proprietary technology or exclusive partnerships were interpreted as evidence of an underlying storyline in which OpenAI also positions itself, implicitly, as a Technology Monopolist or Privacy Peeper.

The following sections present the specific results of this integrated analysis, detailing how OpenAI's identity construction and narrative trajectories emerge from the combination of quantitative modeling and linguistic interpretation.

4 Results

4.1 Clustering and Identity Category

Following the foundational K-means clustering analysis of OpenAI's blog documents based on word embeddings, we have identified three primary document clusters, as shown in Table 2.

Table 2 Detailed Results in Clusters (Extracted)

Cluster	Title	Content Preview
Cluster 1	OpenAI technical goals	We're trying to build AI as part of a larger...
Cluster 1	Special projects	Impactful scientific work requires working...
Cluster 1	OpenAI supporters	We're excited to welcome new donors...
Cluster 1	OpenAI Scholars	We're providing stipends and mentorship...
Cluster 1	OpenAI Scholars 2018: Final projects	Our first cohort of OpenAI Scholars...
Cluster 2	Introducing OpenAI	OpenAI is a non-profit AI research company...
Cluster 2	Machine Learning Unconference	The latest information about the Unconference...
Cluster 2	Report from the conference	Last week we hosted over a hundred...
Cluster 2	OpenAI hackathon	Come to OpenAI's office in San Francisco...
Cluster 2	Report from the hackathon	On March 3rd, we hosted our first hackathon...
Cluster 3	Team++	We've had some fantastic people join...
Cluster 3	Welcome, Pieter and Shivon!	We have two more team updates...
Cluster 3	Team update	We'd like to welcome the latest set of...
Cluster 3	Team update	We've hired more great people to help...
Cluster 3	OpenAI and Microsoft	We're working with Microsoft to run...

^a K-Means applied to Word2Vec vectors with $k = 3$; overall Silhouette Coefficient = 0.518 (range -1 to 1), indicating moderately distinct clusters.

Our evaluation metrics indicate a silhouette coefficient of 0.518, suggesting moderately

planning, innovation-driven experimentation, and expanding access through sponsorship. The corresponding word cloud (Figure 3) reinforces this focus, with prominent terms like research, student, work, openai, and team, which together reflect a research-oriented and inclusive identity. This cluster illustrates OpenAI’s role as a research organization engaging diverse stakeholders through visionary goals and community outreach.

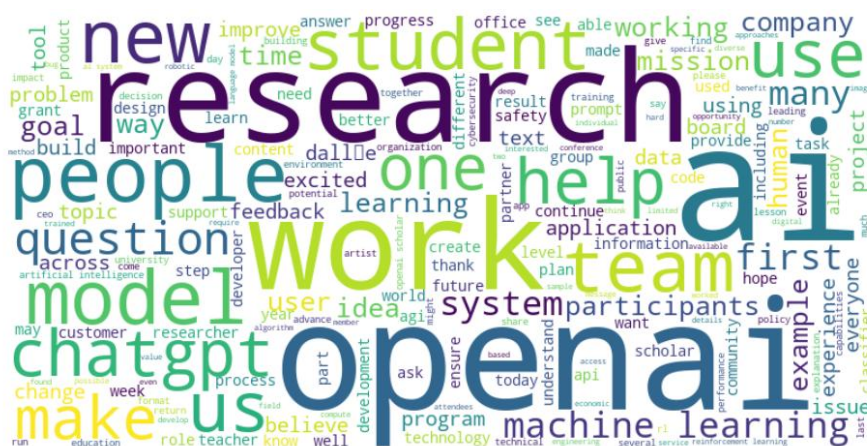


Figure 4 Word Cloud for Cluster 2 (Educational/Collaborative Focus)

Cluster 2 features posts centered on OpenAI-hosted events, including hackathons, unconferences, and internal summaries of community gatherings. Titles such as “Machine Learning Unconference”, “Report from the Hackathon”, and “Introducing OpenAI” highlight efforts to foster innovation via collaborative activities and open discussion. As shown in the word cloud (Figure 4), frequently used terms such as team, student, question, help, and learning signal a focus on interaction, co-learning, and community building. This cluster captures OpenAI’s identity as a convener of AI researchers and its investment in building a collaborative ecosystem.

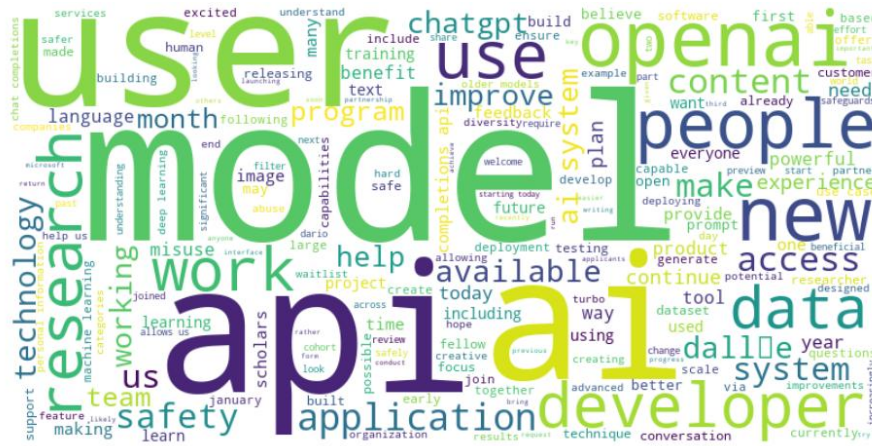


Figure 5 Word Cloud for Cluster 3 (Application/Developer Focus)

Cluster 3 includes blog posts that focus on organizational updates, hiring announcements, and partnerships. Titles like “Team++”, “Team Update”, and “OpenAI and Microsoft” reflect the company’s evolving internal structure and its external collaborations. The word cloud (Figure 5) highlights keywords such as api, user, developer, data, and access, pointing to a discourse centered around productization, developer support, and applied AI. This cluster reflects OpenAI’s identity as an expanding, application-focused enterprise working at the intersection of AI development and industry partnership.

Together, the three word clouds and cluster interpretations reveal distinct thematic patterns within the OpenAI blog, validating the K-Means classification and demonstrating the company's multi-layered identity construction through discourse.

Davies and Harré (1990) point out that identity categories involve learning which categories include some people and exclude others, such as male/female or father/daughter. Through K-Means clustering analysis, we identified three distinct categories within the OpenAI blog documents. These categories reflect the dual facets of academia and commerce that shape OpenAI's identity

positioning. They not only delineate differences among various participants within the OpenAI community but also reveal the process by which meaning is conferred upon these categories through discursive practices.

In Cluster 1, for instance, the emphasis on research programs, technical goals, and scholarship initiatives constructs an identity category aligned with scientific authority and institutional vision. Posts in this group position OpenAI as a thought leader and social benefactor, foregrounding its commitment to advancing AI for the broader public good. In contrast, Cluster 2, centered around hackathons and unconferences, reflects OpenAI's identity as a community builder and innovator. These texts construct a participatory identity, positioning both OpenAI and its audience as collaborators in the AI ecosystem. Meanwhile, Cluster 3—comprising team announcements and strategic partnerships—projects a corporate or organizational identity. Here, the focus shifts to internal development, recruitment, and alliances with major tech actors such as Microsoft, signaling a commercially strategic dimension of OpenAI's self-presentation.

These identity categories are not fixed but fluid and interactionally constituted. Through the discursive patterns observed in each cluster, we see how OpenAI navigates and negotiates its position across the academic-commercial continuum. Thus, the clustering results not only capture textual similarities but also offer insight into the ongoing performative construction of OpenAI's institutional identity.

4.2 Word2vec Calculation

Distributional word representations, formed by word vectors, aim to discern nuanced variations that might not be detectable through simple word frequency analysis (Mikolov et al., 2013). This method represents meanings using either sparse or dense vectors (known as

embeddings), which are derived from the continuous lexical representations based on the counts of words occurring together (Viola & Verheul, 2020). To more accurately capture the semantic changes of vocabulary and the social changes they reflect, a distributed word representation method is proposed, representing multiple meanings of a word with different low-dimensional vectors (Z. Liu et al., 2016). Using the Word2Vec algorithm, we calculated the similarity between various terms and “OpenAI” based on their context within the blog corpus (see Table 3). The analysis produced a list of terms with high similarity scores, such as “project,” “scholar,” “research,” “engineer,” and “mentor,” among others. These terms exhibit close semantic relationships with “OpenAI,” suggesting significant patterns in the language used throughout the blog posts. For instance, “project” (0.92), “scholar” (0.91), “research” (0.88), and “engineer” (0.86) are among the top similar terms, reflecting OpenAI’s academic-technical identity. Additionally, terms like “mentor” (0.85), “nonprofit” (0.82), “benefit” (0.85), “human” (0.85), “safe” (0.86), and “api” (0.83) underscore OpenAI’s emphasis on human-centred innovation, social benefit, operational safety, and accessible technological infrastructure.

Overall, the statistical findings reinforce the initial observation that OpenAI’s blog posts emphasize collaboration, technology, and accessibility, with certain themes, such as mentorship and research, becoming less prominent as time progress.

Table 3 Top Word Associations with OpenAI

Word	Similarity	Word	Similarity
project	0.922	residency	0.815
scholar	0.905	leadership	0.807

research	0.867	intern	0.792
engineer	0.859	bert	0.790
safe	0.857	opportunity	0.788
mentor	0.857	staff	0.773
benefit	0.853	class	0.673
human	0.845	google	0.661
api	0.845	founded	0.655
nonprofit	0.821	microsoft	0.651

^a Similarity scores are cosine measures computed with Word2Vec on the corpus of 118 OpenAI blog posts (11 Dec 2015 – 8 Mar 2024); higher values indicate a stronger semantic association with “OpenAI.”

These associations will be explored further in subsequent sections, where we will analyze how these terms contribute to the construction of OpenAI’s narrative within its blog posts. The underlying relationships between these words will serve as a basis for examining the broader patterns of identity positioning.

5. Storyline and Self-Position from the Discursive Positioning

5.1 AI Safety and Benefit Guardian

In certain contexts, good quality is crucial for gaining trust (Fuoli & Paradis, 2014). Despite ongoing debates regarding Artificial General Intelligence (AGI), OpenAI has consistently placed its core mission on ensuring that AGI is developed in a manner that benefits all of humanity. OpenAI has positioned itself as a key player in this area by explicitly connecting its mission to the responsible and beneficial development of AGI. A word vector analysis of their communications

reveals a significant co-occurrence between “mission” and “safety,” which underscores their strategic focus on positioning safety as an inseparable component of their technological advances. For example, in OpenAI’s 2019 statement, they clearly articulated their mission:

“Our mission is to ensure that artificial general intelligence (AGI) benefits all of humanity, primarily by attempting to build safe AGI and share the benefits with the world.”¹

This statement establishes OpenAI’s identity while highlights their strategic focus on safety, showing that their technological development is guided by the principle of benefiting humanity. The repeated emphasis on “safe AGI” and “share the benefits” indicates a deliberate effort to alleviate public concerns about AGI risks, positioning OpenAI as an ethical leader who prioritizes humanity’s well-being.

Building on this mission, OpenAI’s 2020 statement further reinforces their commitment to safety. As stated in their 2020 report:

“We’re proud of these and other research breakthroughs by our team, all made as part of our mission to achieve general-purpose AI that is safe and reliable, and which benefits all humanity.”²

This framing demonstrates that OpenAI views their technological innovations as steps toward achieving safe and reliable AI. For instance, the development of GPT-3, which incorporated reinforcement learning from human feedback, was a direct response to mitigate biases and ensure

¹ “OpenAI LP”

² (Organizational Update from OpenAI, 2020)

the model aligns with ethical standards. This illustrates OpenAI’s approach of integrating safety with innovation, reinforcing their identity as a responsible and trustworthy leader in the field.

In 2023, OpenAI further emphasized their commitment to AI safety through strategic expansion. The establishment of a new office in London was described as a key opportunity to attract world-class talent and drive innovation in AGI development and policy. This move reflects OpenAI’s global strategy to integrate safety concerns not just into AI development but also into policy engagement, ensuring that innovation aligns with ethical and regulatory frameworks. Moreover, OpenAI emphasized the importance of community involvement in ensuring AI safety, stating:

*“Involving the community is critical to our mission of building safe AGI that benefits humanity.”*³

This highlights OpenAI’s belief in transparency and collaboration, recognizing the importance of public input and expert engagement in shaping AI development. By 2024, OpenAI’s Charter has continued to reinforce the idea that safety is paramount. As stated in the 2024 Charter:

*“AI technology comes with tremendous benefits, along with serious risk of misuse. Our Charter guides every aspect of our work to ensure that we prioritize the development of safe and beneficial AI.”*⁴

³ (Introducing GPTs, 2023)

⁴ (OpenAI Charter, n.d.)

Acknowledging the potential risks associated with AGI, this statement reaffirms OpenAI’s commitment to balancing innovation with responsibility. Their Charter provides concrete guidance for how safety is integrated into every aspect of their work, including human feedback loops and policy development, reflecting a proactive approach to mitigating risks while advancing AI technology.

5.2 Advanced AI Innovator

Beyond embodying its mission, OpenAI’s corporate identity also demonstrates a strong focus on contributions to fields such as deep learning and machine learning, with a continuous emphasis on “learning” and “research”. The words most closely associated with “OpenAI” are “project”(similarity: 0.922), “scholar”(similarity: 0.905) and “research” (similarity: 0.867), underscoring OpenAI’s commitment to cutting-edge research as a central component of its identity. This focus is reflected in the frequent mentions of team members’ achievements in the research domain and academic learning in their blog posts, highlighting their expertise and showcasing how OpenAI nurtures and leverages talent to advance its mission.

For example, the 2019 announcement of the OpenAI Scholars program underscores the organization’s efforts to support the next generation of AI researchers, particularly from underrepresented groups. The program description notes:

“Our scholars are applying these specializations to current AI research and documenting their progress as they continue to grow as machine learning practitioners.” ⁵

⁵ (OpenAI Scholars 2019, 2019)

This statement exemplifies OpenAI’s dedication to fostering talent in AI, ensuring that the researchers of tomorrow are equipped with the skills and opportunities to contribute to the AI field. By investing in the development of emerging scholars, OpenAI reinforces its academic identity and establishes a pipeline for future AI leaders.

Similarly, the 2020 iteration of the OpenAI Scholars program further emphasizes the organization’s commitment to inclusivity in AI research. As stated in the program overview:

*“The OpenAI Scholars program provides stipends and mentorship to individuals from underrepresented groups to study deep learning and open-source a project.”*⁶

This initiative illustrates OpenAI’s proactive stance on diversity, providing mentorship and financial support to individuals who may face systemic barriers to entry in the field of AI. Through these efforts, OpenAI advances the field of AI by expanding educational access and actively supports the development of a more inclusive and equitable academic community within the technology sector.

OpenAI’s focus on continuous research and technological development extends beyond talent cultivation and into the domain of advanced AI techniques. In 2022, the organization made significant strides in improving its models through the use of Reinforcement Learning from Human Feedback (RLHF) to train GPT-3.0. OpenAI’s blog post detailing this process explains:

“We trained this model using Reinforcement Learning from Human Feedback (RLHF), using the

⁶ (OpenAI Scholars 2020, 2020)

*same methods as InstructGPT, but with slight differences in the data collection setup. We performed several iterations of this process, collecting comparison data and using reward models to fine-tune the model.”*⁷

This iterative approach to model training reflects OpenAI’s commitment to refining AI systems based on real-world user feedback. The repeated testing and fine-tuning process improved the model’s performance and ensured that it met the needs and expectations of its users, illustrating OpenAI’s focus on practical, user-centered AI development.

Moreover, the 2023 advancements in GPT-3.5 Turbo’s fine-tuning capabilities demonstrate OpenAI’s ongoing dedication to optimizing its models. The company’s report states:

*“Fine-tuning for GPT-3.5 Turbo is now available, with fine-tuning for GPT-4 coming this fall. Early tests have shown a fine-tuned version of GPT-3.5 Turbo can match, or even outperform, base GPT-4-level capabilities on certain narrow tasks.”*⁸

This achievement highlights OpenAI’s approach of optimizing existing models to enhance their performance on specific tasks, such as complex problem-solving or text summarization. The ability of GPT-3.5 Turbo to outperform GPT-4 on certain narrow tasks emphasizes OpenAI’s capacity to maximize the potential of its technologies, reinforcing its position as an industry leader in both model development and research efficiency.

In addition to advancing its models, OpenAI has made substantial investments in its

⁷ (Introducing ChatGPT, 2022)

⁸ (GPT-3.5 Turbo Fine-Tuning and API Updates, 2023)

technological infrastructure to support its research ambitions. As part of their 2023 infrastructure development efforts, OpenAI partnered with Azure to rebuild its deep learning stack and co-design a supercomputer specifically tailored to its workload. The blog post states:

*“Over the past two years, we rebuilt our entire deep learning stack and, together with Azure, co-designed a supercomputer from the ground up for our workload.”*⁹

This infrastructure development marks a significant milestone in OpenAI’s strategy to support increasingly complex and resource-intensive AI research. By improving the theoretical foundations and addressing technical challenges during the first test run with GPT-3.5, OpenAI ensured that its systems were well-equipped for future AI breakthroughs. This investment underscores the organization’s forward-thinking approach, ensuring that its research infrastructure can keep pace with the growing demands of AI innovation.

Finally, the introduction of GPT-4o in 2024 exemplifies OpenAI’s commitment to expanding AI capabilities across multiple modalities. As stated in their update:

*“It can respond to audio inputs in as little as 232 milliseconds, with an average of 320 milliseconds, which is similar to human response time in a conversation... GPT-4o is especially better at vision and audio understanding compared to existing models.”*¹⁰

This model represents a leap forward in AI’s ability to process and respond to multiple forms

⁹ (OpenAI and Microsoft Extend Partnership, 2023)

¹⁰ (Hello GPT-4o, 2024)

of input, including audio and visual data, with unprecedented speed and accuracy. The advancements in response time, cost efficiency, and non-English language capabilities are testaments to OpenAI's ongoing efforts to optimize and diversify the functionality of its AI models. By continuing to refine and expand its technological portfolio, OpenAI cements its identity as a leader in AI innovation, pushing the boundaries of what artificial intelligence can achieve across a variety of contexts and use cases.

In conclusion, OpenAI's academic initiatives, technological innovations, and infrastructure investments collectively solidify its position as a pioneer in the field of AI research and development. Through its focus on nurturing talent, advancing cutting-edge research, and continuously optimizing its models and infrastructure, OpenAI demonstrates its commitment to shaping the future of AI. Its efforts not only drive technological progress but also ensure that AI development is inclusive, safe, and aligned with the needs of humanity.

5.3 Shared AI Advocator

Business organizations strive to create a positive corporate identity that shapes public perception and builds a strong reputation over time (Van Riel, 1997; Westcott Alessandri, 2001). A strong reputation brings numerous advantages, including increased competitiveness, profitability, and the ability to attract customers, investors, and talented employees (for a review, see Walker, 2010). In this regard, OpenAI actively engages in collaborations with other companies, positioning itself as a "Shared AI Advocate." This identity is further reinforced through its association with terms like "access" (0.999792) and "API" (0.999811), which highlight OpenAI's commitment to making AI technologies accessible to a broader audience. Through these efforts, OpenAI aligns itself with its mission to democratize access to AI, ensuring that its innovations benefit a wide range

of users—from individual developers to large enterprises.

The strategic vision behind OpenAI LP, as expressed in its 2019 Statement on Capped Returns, reflects the company’s dedication to balancing financial sustainability with its broader mission to benefit humanity. The statement outlines:

*“The fundamental idea of OpenAI LP is that investors and employees can get a capped return if we succeed at our mission, which allows us to raise investment capital and attract employees with startup-like equity.”*¹¹

This unique approach, which limits financial returns to investors and employees, effectively distinguishes OpenAI from traditional profit-driven startups. By capping returns, OpenAI emphasizes its focus on societal welfare over financial gain, while still attracting necessary investment and talent. This model supports the reinvestment of value into the company’s mission and underlines OpenAI’s commitment to ethical AI development and knowledge sharing.

OpenAI’s partnership with Microsoft exemplifies its collaborative approach to advancing AI technology. The 2021 Microsoft Partnership announcement states:

*“Microsoft invests in and partners with OpenAI to support us in building beneficial AGI.”*¹²

This strategic collaboration provides OpenAI with essential resources to enhance its AGI research while simultaneously integrating OpenAI’s cutting-edge technologies into Microsoft’s

¹¹ (OpenAI LP, 2019)

¹² (Microsoft Invests in and Partners with OpenAI to Support Us Building Beneficial AGI, 2019)

product ecosystem. Beyond financial support, the partnership also includes shared infrastructure, with OpenAI leveraging Azure’s computational resources to scale its models. The collaboration is a mutually beneficial one: OpenAI gains access to unparalleled computing power, while Microsoft strengthens its product offerings with state-of-the-art AI capabilities. Through this partnership, OpenAI expands its technological capacity while ensuring that its advancements in AI can have a broader societal impact.

The launch of OpenAI’s API (Application Programming Interface) further consolidates its identity as an advocate for shared AI. By offering an open API, OpenAI provides developers and businesses with easy access to advanced AI technologies, enabling them to create innovative applications and services. This openness fosters collaboration across industries, facilitating the integration of OpenAI’s AI models into a variety of products.

For instance, the licensing of GPT-3 to Microsoft for use in their products, as described in news:

“OpenAI has agreed to license GPT-3 to Microsoft for their own products and services. OpenAI released its first commercial product back in June: an API for developers to access advanced technologies for building new applications and services.” ¹³

The blog illustrates OpenAI’s commitment to sharing its advancements with industry leaders. The decision to license GPT-3 broadens the scope of its application and allows OpenAI’s technologies to be incorporated into mainstream products and services, extending the benefits of its innovations beyond its own organization. This action reinforces OpenAI’s role as a disseminator of

¹³ (OpenAI Licenses GPT-3 Technology to Microsoft, 2020)

cutting-edge AI technology, breaking through the typical proprietary barriers that limit access to new innovations.

“When the OpenAI API launched, Algolia partnered with OpenAI to integrate GPT-3 with their advanced search technology in order to create their new Answers product that better understands customers’ questions and connects them to the specific part of the content that answers their questions.”¹⁴

In 2022, further highlights OpenAI’s cooperative efforts through its partnership with Algolia. In this collaboration, GPT-3 was integrated into Algolia’s advanced search technology to develop the “Answers” product, which enhances customer interaction by providing more accurate and relevant search results. The integration of GPT-3 into Algolia’s platform demonstrates how OpenAI’s API enables third-party developers to leverage advanced AI for practical, real-world applications. This partnership showcases the versatility of OpenAI’s technology and underscores its role in driving innovation across diverse industries.

OpenAI’s continuous updates to its API, as detailed in the news for developers:

“Since launching our API in 2020, we’ve continuously updated it to include our most advanced models, making it easier than ever for developers to integrate cutting-edge AI into their projects with a simple API call.”¹⁵

¹⁴ (GPT-3 Powers the next Generation of Apps, 2021)

¹⁵ (Join Us for OpenAI’s First Developer Conference on November 6 in San Francisco, 2023)

Through simple API call, OpenAI reflects its commitment to keeping its technology at the forefront of AI development. With each update, OpenAI enhances the capabilities of its models, simplifying the integration process for developers and expanding the range of applications for which the API can be used. By continuously evolving its offerings, OpenAI ensures that its API remains a cutting-edge resource for developers at all levels.

Taken together, these blogs illustrate how OpenAI's collaborations, licensing strategies, and API development contribute to its identity as a "Shared AI Advocator". Through its partnerships with major tech companies, strategic licensing of its models, and ongoing API enhancements, OpenAI fosters an open and collaborative AI ecosystem. By enabling other organizations to integrate its technologies, OpenAI promotes a vision of AI that is inclusive, accessible, and beneficial to a wide range of stakeholders. These efforts position OpenAI as an industry leader not only in AI research but also in ensuring that its innovations contribute to the broader societal good.

6. The latent Storyline: Contradictory Identities

By analyzing the storyline in OpenAI's blog, we have identified some hidden identities that contradict their publicly promoted image. For example, the word "privacy" frequently appears alongside discussions of innovation, research and data usage. While OpenAI emphasizes privacy protection, in some cases, the blog mentions technological advancements without fully addressing user privacy concerns. This suggests a hidden "Privacy Peeper" identity, where innovation may take precedence over privacy. Similarly, the theme of "monopoly" emerges from discussions of partnerships and mission. Although OpenAI advocates for open collaboration, terms like "exclusive agreements" and "strategic alliances" suggest business strategies that might suppress competition,

hinting at a hidden “Technology Monopolist” identity. These hidden identities are derived by identifying contradictions in the language used in the blog, revealing the complex nature of OpenAI’s corporate identity construction.

6.1 Privacy Peeper

As OpenAI leads the way in AI development, the company faces increasing ethical and legal challenges, particularly concerning privacy and security. While OpenAI strives to innovate with products like ChatGPT and GPT-4 Plus, a tension exists between technological progress and the protection of user privacy, which is increasingly threatened in the age of big data and surveillance capitalism (Mayer-Schönberger & Cukier, 2013; Zuboff, 2019).

OpenAI’s approach, particularly with its products like ChatGPT and the development of GPT-4 Plus, illustrates a complex interplay between striving for cutting-edge AI technologies and adhering to privacy and security standards. While OpenAI has made efforts to demonstrate its commitment to security, disclosures about its data use practices, especially the decision not to train GPT-4 Plus on user data, highlight ongoing challenges in ensuring privacy preservation across AI development cycles. Especially in the blog named “Introducing GPT Team”, the blog highlighted that “No training on your business data or conversations” (see in figure 6). However, this statement obscures the fact that the user data are used to train the GPT model. This situation raises questions about the nature of data utilization and the extent to which user privacy is safeguarded in the pursuit of innovation.

ChatGPT Team includes:

- Access to GPT-4 with 32K context window
- Tools like DALL-E 3, GPT-4 with Vision, Browsing, Advanced Data Analysis—with higher message caps
- No training on your business data or conversations

Figure 6. Announcement of *ChatGPT for Teams*, 2024

The ethical and legal dilemmas OpenAI faces may not only delay new AI model releases but also diminish its competitive edge. The growing adoption of multiple AI providers, including Microsoft's partnerships with startups like Mistral, reflects increasing concerns over privacy, security, and legal compliance in AI deployment.

In an effort to address these privacy concerns, OpenAI published an article titled "New ways to manage your data in ChatGPT," (*New Ways to Manage Your Data in ChatGPT*, 2023) introducing several measures to mitigate the contradictory identity surrounding privacy. For user's data management, OpenAI noted in its blog:

"Conversations that are started when chat history is disabled won't be used to train and improve our models, and won't appear in the history sidebar... We hope this provides an easier way to manage your data than our existing opt-out process."

OpenAI has introduced a new feature that allows users to turn off chat history, thereby enabling them to choose which conversations can be used to train the models. When chat history is disabled, conversations will not be used for model training and will not appear in the history sidebar. Furthermore, OpenAI has stated that when chat history is disabled, new conversations will be retained for 30 days and reviewed only when necessary for abuse monitoring, after which they will be permanently deleted. These new features represent OpenAI's attempt to resolve its privacy contradictions by offering users more control over their data. However, these measures still diverge

from previous data usage practices, such as retaining conversations for monitoring purposes, which raises ongoing concerns regarding privacy protection.

Additionally, OpenAI's commitment to privacy extends to transparency initiatives. The company is working to share aggregated demographic information about its reviewers in a way that doesn't violate privacy rules and norms, recognizing that reviewer demographics can be an additional source of potential bias in system outputs(*How Should AI Systems Behave, and Who Should Decide?*, 2023). This reflects OpenAI's broader strategy of balancing transparency with privacy protection. Moreover, OpenAI has reaffirmed its commitment to user privacy by stating that "everyone at OpenAI is committed to protecting our users' privacy and keeping their data safe"(*March 20 ChatGPT Outage: Here's What Happened* | OpenAI, 2023). These declarations reinforce the company's position on safeguarding user data.

Further, OpenAI has continued to ensure that its models adhere to strict data privacy and security protocols. The company emphasized that "*all of these models come with the same data privacy and security guarantees we introduced on March 1*"(*Function Calling and Other API Updates*, 2023), underscoring its ongoing commitment to user data protection in the context of evolving AI technologies.

These new features and statements represent OpenAI's attempt to resolve its privacy contradictions by offering users more control over their data. However, these measures still diverge from previous data usage practices, such as retaining conversations for monitoring purposes, which raises ongoing concerns regarding privacy protection.

Despite these initiatives, contradictions remain between OpenAI's data usage and privacy policies, which could hinder the widespread adoption of its technologies and reduce its competitive

edge in the AI field. With the rise of multiple AI service providers, including Microsoft's partnerships with startups like Mistral, privacy, security, and legal compliance continue to be pressing concerns in the AI industry.

6.2 Technology Monopolist

While concerns over user privacy reveal only one facet of the tension between OpenAI's rhetoric and practice, a second contradiction emerges when the company's market position is viewed through the lens of potential monopoly dynamics.

The rise of monopolies in the digital age presents complex challenges with wide-ranging societal impacts. A monopoly, by definition, occurs when a single entity dominates the market, limiting competition and controlling pricing and supply (Samuelson & Nordhaus, 2009). This dominance can arise from government-sanctioned monopolies, natural monopolies driven by technological innovation, or conglomerates formed through market forces.

The word "mission" and storyline commitment storyline of OpenAI's discourse serve as both a guiding principle and a strategic narrative that highlights the company's commitment to advancing artificial general intelligence (AGI) for the benefit of humanity. However, this narrative, while noble on the surface, also obscures the company's increasingly monopolistic tendencies, as evidenced by the transformation of OpenAI from a nonprofit organization to a for-profit corporation. The transition from altruistic goals to corporate control reflects a growing concentration of power and resources, raising concerns about the ethical implications of such a shift.

In OpenAI's own words, the organization repeatedly emphasizes its dedication to the mission, as illustrated by the statement: "*We are dedicated to the OpenAI mission and have pursued it every step of the way.*" (OpenAI and Elon Musk, 2024). This sentiment is echoed in the company's internal

communications, where they acknowledge that achieving their ambitious goals would require significant capital investment— *“billions of dollars per year, which was far more than any of us, especially Elon, thought we’d be able to raise as the non-profit.”* This recognition of the vast financial resources required reflects the tension between the initial nonprofit model and the reality of funding AGI research at scale.

The challenge of reconciling the mission with financial constraints led OpenAI to pursue a for-profit model, but, as revealed in internal discussions, a fundamental disagreement arose over the issue of control. The co-founders, including Sam Altman and Greg Brockman, resisted Elon Musk’s proposal to convert OpenAI into a for-profit organization with the aim of securing majority control. The company’s response was rooted in the belief that such a model would undermine their mission: *“We couldn’t agree to terms on a for-profit with Elon because we felt it was against the mission for any individual to have absolute control over OpenAI.”* (OpenAI and Elon Musk, 2024). This statement highlights the ethical conflict at the core of OpenAI’s transformation, where the drive for control risks overshadowing the original mission of democratizing AGI technology.

OpenAI’s mission was not only about securing financial resources but also about advancing the public good by developing *“widely-available beneficial tools.”* This principle underscores the company’s commitment to ensuring that AGI serves the broader needs of society, rather than being monopolized by a few wealthy entities. However, the shift toward a for-profit model has made it increasingly difficult to reconcile the dual goals of mission-driven innovation and corporate profitability.

The friction between OpenAI and Musk, exemplified by Musk’s public comments and legal action, further underscores the tension between the company’s mission and its evolving business

model. Musk’s frustration with OpenAI’s transformation is evident in his public statements, where he lamented,

“We’re sad that it’s come to this with someone whom we’ve deeply admired—someone who inspired us to aim higher, then told us we would fail, started a competitor, and then sued us when we started making meaningful progress towards OpenAI’s mission without him.” (OpenAI and Elon Musk, 2024)

Musk’s disappointment reflects a sense of betrayal—he viewed OpenAI’s success as part of his broader mission, yet he felt that its shift toward profit-driven goals and corporate control contradicted the original vision of collaborative, open research.

Musk’s actions—starting a competing AGI company and filing a lawsuit against OpenAI—highlight his belief that the mission of OpenAI has been compromised by the company’s pursuit of a for-profit model. Musk’s public statements and legal challenges position him as an outsider, one who believed in the mission but was ultimately excluded as OpenAI began prioritizing corporate interests. His frustrations, as articulated in his open letter, shed light on the challenges of reconciling a lofty mission with the realities of managing and financing cutting-edge technological research.

Moreover, Altman’s response to Musk’s accusations, particularly in disclosing Musk’s previous proposals for OpenAI’s future, reveals the underlying power dynamics. Altman’s disclosure paints Musk as someone who sought control over OpenAI for personal and competitive reasons, which further complicates the narrative of mission-driven progress. Musk’s attempts to incorporate OpenAI into his broader business strategy, including using its AGI for Tesla’s self-

driving technology, underscore how the language of *mission* can be manipulated to justify monopolistic control.

OpenAI's use of the *mission* narrative to justify its actions, while initially framing the organization as a benevolent force for societal good, also reveals how mission-driven rhetoric can mask underlying monopolistic tendencies. Musk's public criticisms and legal actions underscore the conflict between OpenAI's founding mission and its transition toward a profit-driven model controlled by a select few. The pursuit of a grand mission, in this case, has not only shaped OpenAI's narrative but also facilitated its dominance over the AGI field, raising critical questions about access, equity, and the broader implications of technological monopolies in the age of artificial intelligence.

7 Conclusion

This study adopted a linguistically grounded mixed-methods approach that integrated Word2Vec and K-Means modeling with the Discursive Positioning Framework to examine 118 OpenAI blog posts. By combining large-scale semantic modeling with close rhetorical reading, the study demonstrates how informal digital texts can serve as reliable data for corporate identity research and ethical discourse analysis. The findings reveal a dual narrative storyline underlying OpenAI's identity work. On the surface, OpenAI constructs itself as an AI-Safety Guardian, Advanced Innovator, and Shared-AI Advocate; beneath this, however, its discourse also constructs implicit identities of a Privacy-Risk Enabler and Technology Monopolist. These parallel narratives reveal two underlying ethical tensions: one between transparency and strategic opacity, and another between collective benefit and market concentration, which together illuminate how linguistic practices negotiate legitimacy and authority in the AI industry. Theoretically, this research contributes to the intersection of linguistics and business ethics by conceptualizing corporate

discourse as a site of ethical positioning. It demonstrates that discursive patterns in blogs and other informal texts can function as early indicators of shifting organizational values, offering a new methodological bridge between computational text analysis and critical discourse theory. Practically, the findings highlight that organizations should not limit identity management to formal reports but ensure narrative coherence across informal channels such as blogs. Such cross-channel consistency is vital for maintaining brand integrity, stakeholder trust, and ethical accountability.

References

- Aghion, P., Jones, B. F., & Jones, C. I. (2017). *Artificial intelligence and economic growth* (Vol. 23928). National Bureau of Economic Research Cambridge, MA. <https://doi.org/10.3386/w23928>
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2), 31–50. <https://doi.org/10.1257/jep.33.2.31>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151, 103745. <https://doi.org/10.1016/j.jfineco.2023.103745>
- Balmer, J. M. T., & Greyser, S. A. (2002). Managing the Multiple Identities of the Corporation. *California Management Review*, 44(3), 72–86. <https://doi.org/10.2307/41166133>
- Bamberg, M., & Dege, M. (2021). Decentering Histories of Identity. In M. Bamberg, C. Demuth, & M. Watzlawik (Eds.), *The Cambridge Handbook of Identity* (1st ed., pp. 25–56). Cambridge University Press. <https://doi.org/10.1017/9781108755146.003>
- ChatGPT for teams. (2024). <https://openai.com/chatgpt/team/>
- Coates, A., & Ng, A. Y. (2012). Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade: Second Edition* (pp. 561–580). Springer. https://doi.org/10.1007/978-3-642-35289-8_30
- Cruciata, P., Pulizzotto, D., & Beaudry, C. (2024). First impressions on sustainable innovation matter: Using NLP to replicate B-lab environmental index by analyzing companies' homepages. *Technological Forecasting and Social Change*, 205, 123455. <https://doi.org/10.1016/j.techfore.2024.123455>
- Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour*, 20(1), 43–63. https://doi.org/10.1007/978-3-642-35289-8_30
- De Fina, A., & Georgakopoulou, A. (Eds.). (2020). *The Cambridge Handbook of Discourse Studies* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108348195>

- Fairclough, N. (2013). *Language and power*. Routledge.
- Function calling and other API updates*. (2023, June 13). <https://openai.com/index/function-calling-and-other-api-updates/>
- Fuoli, M. (2018). Building a Trustworthy Corporate Identity: A Corpus-Based Analysis of Stance in Annual and Corporate Social Responsibility Reports. *Applied Linguistics*, 39(6), 846–885. <https://doi.org/10.1093/applin/amw058>
- Fuoli, M., & Paradis, C. (2014). A model of trust-repair discourse. *Journal of Pragmatics*, 74, 52–69. <https://doi.org/10.1016/j.pragma.2014.09.001>
- Goloshchapova, I., Poon, S.-H., Pritchard, M., & Reed, P. (2019). Corporate social responsibility reports: Topic analysis and big data approach. *The European Journal of Finance*, 25(17), 1637–1654. <https://doi.org/10.1080/1351847X.2019.1572637>
- GPT-3 powers the next generation of apps*. (2021, March 25). <https://openai.com/index/gpt-3-apps/>
- GPT-3.5 Turbo fine-tuning and API updates*. (2023, August 22). <https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/>
- Gray, E. R., & Balmer, J. M. (1998). Managing corporate image and corporate reputation. *Long Range Planning*, 31(5), 695–702. [https://doi.org/10.1016/S0024-6301\(98\)00074-0](https://doi.org/10.1016/S0024-6301(98)00074-0)
- Hamerly, G., & Elkan, C. (2003). Learning the k in k-means. *Advances in Neural Information Processing Systems*, 16.
- Han, S., Liu, Z., Deng, Z., Gupta, S., & Mikalef, P. (2024). Exploring the effect of digital CSR communication on firm performance: A deep learning approach. *Decision Support Systems*, 176, 114047. <https://doi.org/10.1016/j.dss.2023.114047>
- Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), 1. <https://doi.org/10.3390/bdcc4010001>
- Hello GPT-4o*. (2024, May 13). <https://openai.com/index/hello-gpt-4o/>
- How should AI systems behave, and who should decide?* (2023, February 16). <https://openai.com/index/how-should-ai-systems-behave/>
- Introducing ChatGPT*. (2022, November 30). <https://openai.com/index/chatgpt/>
- Introducing GPTs*. (2023, November 6). <https://openai.com/index/introducing-gpts/>
- Jaworska, S. (2020). Corporate discourse. In *Handbook of Discourse Studies*. Cambridge University Press.
- Join us for OpenAI's first developer conference on November 6 in San Francisco*. (2023, September 6). <https://openai.com/index/announcing-openai-devday/>
- Kayakuş, M., Yiğit Açıkgoz, F., Dinca, M. N., & Kabas, O. (2024). Sustainable Brand Reputation: Evaluation of iPhone Customer Reviews with Machine Learning and Sentiment Analysis. *Sustainability*, 16(14), 6121. <https://doi.org/10.3390/su16146121>
- Li, X. (2022). The discursive construction of corporate identity in the corporate social responsibility reports: A case study of Starbucks. *Frontiers in Psychology*, 13, 940541. <https://doi.org/10.3389/fpsyg.2022.940541>
- Liu, Z., Liu, Y., Tu, C., & Sun, M. (2016). Lexical Semantic Variation and Social Change: Quantitative Observation and Analysis. *Language Strategy Research*, 6, 47–54.
- Manning, C. D. (2009). *An introduction to information retrieval*.
- March 20 ChatGPT outage: Here's what happened | OpenAI*. (2023, March 24).

- <https://openai.com/index/march-20-chatgpt-outage/>
- Melewar, T. C. (2003). Determinants of the corporate identity construct: A review of the literature. *Journal of Marketing Communications*, 9(4), 195–220. <https://doi.org/10.1080/1352726032000119161>
- Melewar, T. C. (Ed.). (2008). *Facets of corporate identity, communication and reputation*. Routledge.
- Microsoft invests in and partners with OpenAI to support us building beneficial AGI. (2019, July 22). <https://openai.com/index/microsoft-invests-in-and-partners-with-openai/>
- Mikolov, T. (2013). Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:1301.3781*, 3781.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- Mousavi, R., Johar, M., & Mookerjee, V. S. (2020). The Voice of the Customer: Managing Customer Care in Twitter. *Information Systems Research*, 31(2), 340–360. <https://doi.org/10.1287/isre.2019.0889>
- New ways to manage your data in ChatGPT. (2023, April 25). <https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt/>
- OpenAI and Elon Musk. (2024, March 5). <https://openai.com/index/openai-elon-musk/>
- OpenAI and Microsoft extend partnership. (2023, January 23). <https://openai.com/index/openai-and-microsoft-extend-partnership/>
- OpenAI Charter. (n.d.). Retrieved November 7, 2024, from <https://openai.com/charter/>
- OpenAI licenses GPT-3 technology to Microsoft. (2020, September 22). <https://openai.com/index/openai-licenses-gpt-3-technology-to-microsoft/>
- OpenAI LP. (2019, March 11). <https://openai.com/index/openai-lp/>
- OpenAI Scholars 2019: Meet our Scholars. (2019, March 13). <https://openai.com/index/openai-scholars-2019-meet-our-scholars/>
- OpenAI Scholars 2020: Final projects. (2020, July 9). <https://openai.com/index/openai-scholars-2020-final-projects/>
- Organizational update from OpenAI. (2020, December 29). <https://openai.com/index/organizational-update/>
- Rantanen, A., Salminen, J., Ginter, F., & Jansen, B. J. (2019). Classifying online corporate reputation with machine learning: A study in the banking domain. *Internet Research*, 30(1), 45–66. <https://doi.org/10.1108/INTR-07-2018-0318>
- Romano, S. P., Sperli, G., & Vignali, A. (2024). An NLP-based approach to assessing a company's maturity level in the digital era. *Expert Systems with Applications*, 252, 124292. <https://doi.org/10.1016/j.eswa.2024.124292>
- Rust, R. T., Rand, W., Huang, M.-H., Stephen, A. T., Brooks, G., & Chabuk, T. (2021). Real-Time Brand Reputation Tracking Using Social Media. *Journal of Marketing*, 85(4), 21–43. <https://doi.org/10.1177/0022242921995173>
- Sun, Y., & Jiang, J. (2014). Metaphor use in Chinese and US corporate mission statements: A cognitive sociolinguistic analysis. *English for Specific Purposes*, 33, 4–14. <https://doi.org/10.1016/j.esp.2013.07.005>
- Van Riel, C. B. (1997). Research in corporate communication: An overview of an emerging field.

- Management Communication Quarterly*, 11(2), 288–309.
- Viola, L., & Verheul, J. (2020). One Hundred Years of Migration Discourse in The Times: A Discourse-Historical Word Vector Space Approach to the Construction of Meaning. *Frontiers in Artificial Intelligence*, 3, 64. <https://doi.org/10.3389/frai.2020.00064>
- Westcott Alessandri, S. (2001). Modeling corporate identity: A concept explication and theoretical explanation. *Corporate Communications: An International Journal*, 6(4), 173–182. <https://doi.org/10.1108/EUM00000000006146>
- Whetten, D. A., & Mackey, A. (2002). A social actor conception of organizational identity and its implications for the study of organizational reputation. *Business & Society*, 41(4), 393–414.