The Impact of AI Chat Tools on Learning Communities: A Cognitive Science Perspective

Joshua K. Coleman

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2026

© 2026

Joshua K. Coleman

All Rights Reserved

Abstract

The Impact of AI Chat Tools on Learning Communities: A Cognitive Science Perspective

Joshua K. Coleman

This dissertation investigates how the emergence of large language models (LLMs), particularly ChatGPT, is transforming collaborative learning and public knowledge production in online communities. Focusing on five technical subforums within the Stack Exchange network, the study employs a quasi-experimental, pre-test/post-test design grounded in distributed cognition theory to examine linguistic and behavioral changes before and after the public release of ChatGPT (pp. 28-47). Quantitative findings indicate a statistically significant increase in question complexity post-release, as measured by higher Flesch-Kincaid Grade Levels across the dataset (pp. 55-59). No significant differences were observed in Lexical Density (proportion of content words), suggesting that syntactic complexity increased without a corresponding rise in semantic information load (pp. 56-59). Across the two nine-month sampling windows, overall question volume declined by 15.6% (75,923 pre vs. 64,072 post), with decreases concentrated in Computer Science, Data Science, and Mathematics, while Academia rose modestly and Artificial Intelligence remained near-flat. In the answered-question analytic sample (N = 97,513), user engagement metrics remained relatively stable overall, though some communities experienced localized declines in question resolution (accepted-answer rate) and response activity (number of answers per question) (pp. 48-56, Table 8). These patterns support the hypothesis that users offload simpler queries to ChatGPT and use public forums for more complex or ambiguous

problems (pp. 59-64). The final chapter contextualizes these findings within broader concerns about educational equity, platform governance, and the future of digital commons (pp. 72–76).

Keywords: artificial intelligence (AI), distributed cognition, large language models (LLM), question complexity, epistemic infrastructure, collaborative learning

Table of Contents

List of Tables iv

List of Figures v

Acknowledgments vi

Dedication vii

Chapter 1: Statement of the Problem 1

The Shifting Landscape of Knowledge in the Age of AI 1

The Problem in Context: AI, Distributed Cognition, and the Digital Commons 4

AI, Human Cognition, and Educational Implications 7

Significance of Study 10

Chapter 2: Literature Review 13

Distributed Cognition Theory: A Framework for Understanding Collaborative Intelligence 13

The Role of Mediating Artifacts in Cognition 14

Distributed Cognition in Digital Learning Environments 15

Frameworks for Evaluating Distributed Cognition 17

Prior Studies on AI and Communities of Learning 20

The Impact of AI Integration in Educational Contexts 21

Chapter 3: Methodology 24

Unit of Analysis 24

Research Design and Rationale 25

Phase 1: Pilot Study 26
Pilot Study Findings and Methodological Refinements for a Full Study 31
Implications for Future Study Design 33
Main Study Design: Full-Scale Investigation 34
Ethical Considerations and Limitations 50
Chapter 4: Findings & Results from Analysis of AI's Impact on Question Complexity in Online Learning Communities 52
Overview 52
Descriptive Results of Engagement and User Patterns 52
Community-Specific Time Period Patterns of Engagement 56
MANOVA Analysis: Impact of AI Chat Tools on Question Complexity 60
Univariate Follow-Up Tests: Decomposing the Multivariate Effect 61
Interaction Effects: Temporal Changes Across Communities 63
Answer to the Research Question 65
Chapter 5: Interpretation, Implications, and Future Directions 68
Interpretation of Key Findings 69
Analyzing the Decline in Participation 70
Theoretical Framing Through Distributed Cognition 72
Cognitive and Neural Effects of AI Chat Tools: Synthesis of Empirical and Theoretical Evidence 74
Broader Implications 76
Practical Strategies for Sustaining Online Q&A Platforms in the Age of AI 78

Limitations and Future Directions 81

Conclusion 84

References 85

Appendix A Glossary of Key Technical Terms 104

Appendix B Descriptive Statistics and Distribution Checks 106

Appendix C Comparison of Raw Data to Final Sample 109

Appendix D Participation Volume and Sample Retention 110

List of Tables

Table 1 Normality Assessment Results for Continuous Variables 30

Table 2 Variance Homogeneity Assessment and Test Selection 31

Table 3 Mann-Whitney U Tests: Pre- vs. Post-ChatGPT Question Characteristics 32

Table 4 Normality Assessment Results for Winsorized Continuous Variables 42

Table 5 Final Sample Sizes by Community and Time Period 45

Table 6 Descriptive Statistics by Time Period 53

Table 7 Descriptive Statistics by Community 55

Table 8 Descriptive Statistics by Community and Time Period 56

Table 9 Pearson Correlations Among Question Characteristics 58

Table 10 MANOVA: Effect of Time Period on Question Complexity 60

Table 11 Univariate Follow-Up Tests: Effect of Time Period 62

Table 12 Two-Way ANOVA for Flesch-Kincaid Grade Level by Time Period and Community 63

Table 13 Two-Way ANOVA for Lexical Density Level by Time Period and Community 65

List of Figures

Figure 1 Example data obtained from a typical question asked in the Stack Exchange Data Science community. 28

Figure 2 Sampling Periods for Pre- and Post-ChatGPT Analysis 36

Figure 3 Comparison of Original Corpus vs. Final Analysis Sample..... 38

Figure 4 Stability of Community Representation Before and After Data Cleaning 39

Figure 5 Stability of Time Period Distribution (Pre- vs. Post-ChatGPT)..... 40

Figure 6 Low Complexity Example: Mathematics (Pre-ChatGPT) 47

Figure 7 High Complexity Example: Mathematics (Pre-ChatGPT)..... 48

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Gary Natriello, for his patient guidance, enthusiastic encouragement, and invaluable critiques of this research. I would also like to thank Drs. James Corter, Youmi Suk, Ryan Baker, and Russell Neuman for their insightful feedback and for serving as members of my committee. My thanks also go to the staff in the Department of Human Development and the Office of Doctoral Studies for their technical and administrative support throughout this process. Finally, I wish to thank my family for their unwavering support and belief in me; This journey would not have been possible without them.

Dedication

To Christiana, for your unwavering belief and endless patience.

Chapter 1: Statement of the Problem

The Shifting Landscape of Knowledge in the Age of AI

The rapid adoption of Large Language Models (LLMs) such as ChatGPT has fundamentally altered the mechanics of information retrieval and creation. While these tools increase the speed and accessibility of problem-solving, they also disrupt established methods of public knowledge exchange. However, this technological revolution also introduces a critical and underexplored challenge: the potential erosion of publicly accessible, human-generated knowledge. As individuals increasingly turn to private interactions with AI for answers and assistance, the rich ecosystems of public knowledge-sharing, such as online forums and collaborative platforms, face a potential decline. This shift threatens to diminish the diversity and richness of the digital commons, a resource that has become fundamental to learning and innovation in the 21st century (del Rio-Chanona et al., 2023; del Rio-Chanona et al., 2024).

This dissertation investigates the impact of AI chat tools on these vital online learning communities through the theoretical lens of distributed cognition. The theory of distributed cognition posits that cognitive processes are not confined to the individual mind but are spread across individuals, tools, and artifacts within a given environment (Hutchins, 1995). Online question-and-answer (Q&A) platforms like the Stack Exchange network are prime examples of distributed cognition in action. On these platforms, a user's question initiates a collective cognitive process, with community members contributing pieces of a solution, thereby creating a public and persistent cognitive artifact that benefits countless future learners. The introduction of AI tools like ChatGPT brings a new, powerful cognitive artifact

into this system, but one that operates primarily in private. This privatization of knowledge-seeking behavior may fundamentally alter the dynamics of these communities, leading to a reduction in the public creation of shared knowledge. In this dissertation, the unit of analysis is the initiating question in each thread; the linguistic content of answers was not analyzed, although question-level metadata (e.g., number of answers received and whether an accepted answer exists) was used to characterize engagement and resolution.

The core of this research is to understand how the integration of AI tools is reshaping the nature of public inquiry and collaborative learning. Specifically, this study examines whether the availability of ChatGPT has led to a change in the complexity of questions asked on Stack Exchange and a corresponding shift in user engagement. By analyzing large-scale data from several Stack Exchange communities before and after the widespread adoption of ChatGPT, this research seeks to quantify these changes and interpret their implications for the future of online learning and knowledge dissemination. The central hypothesis is that as users offload simpler queries to AI, the questions that remain in the public sphere will become more complex, specialized, and technical. While such an evolution may signal a maturation of the discourse, it could also be accompanied by a decline in overall community participation, as fewer users engage in the collaborative process of asking and answering questions.

This study is not merely a technical analysis of platform metrics; it is an exploration of the evolving relationship between human and artificial intelligence in the context of learning. The findings aim to provide a broad overview of how AI is transforming the cognitive and social architecture of online knowledge-sharing environments. By examining both the potential benefits, such as increased efficiency and the ability to tackle more

complex problems, and the potential drawbacks, including the erosion of public knowledge and the creation of new barriers for novice learners, this research will contribute to a more informed and intentional approach to integrating AI into our educational ecosystems.

The central research question guiding this dissertation is:

- How does the introduction and utilization of AI chat tools within the Stack Exchange online community influence the complexity of user-generated questions, and how can changes in question complexity be interpreted through the lens of Distributed Cognition theory to understand the shifting quality of collaborative learning within the platform (Hutchins, 1995; Lave & Wenger, 1991)?

To address this primary question, the study will also investigate the following sub questions:

- Has the introduction of AI chat tools significantly impacted dispersed cognitive processes in collaborative learning environments on Stack Exchange (Loeckx, 2014)?
- To what extent have the temporal patterns of question complexity changed before and after the introduction of an AI chat tool?
- Has there been a significant effect of AI chat tools on learner engagement and the collaborative building of knowledge (Srba & Bielikova, 2014)?

By answering these questions, this dissertation has yielded critical insights for educators, platform designers, policymakers, and researchers. The results highlight the specific trade-offs inherent in AI adoption, pointing toward design strategies that allow automated tools to support, rather than replace, human collaboration. Ultimately, this research seeks to contribute to the development of digital learning environments that are not only technologically advanced but also inclusive, engaging, and conducive to the creation of a vibrant and enduring public sphere of knowledge.

The Problem in Context: AI, Distributed Cognition, and the Digital Commons

The internet has facilitated the expansion of extensive digital commons, where individuals freely exchange knowledge and collaboratively construct it. Platforms like Wikipedia, open-source software repositories, and Q&A forums such as Stack Exchange have become indispensable resources for learners and professionals worldwide. These platforms embody the principles of distributed cognition and the wealth of networks created through social production, where the cognitive load of problem-solving is shared, aligning with cognitive architecture theories that advocate for managing intrinsic and extraneous load (Sweller et al., 1998) across a network of human minds and technological artifacts (Benkler, 2006; Hollan et al., 2000). For instance, asking a question on Stack Overflow initiates a distributed cognitive event, not just a request for information. The subsequent answers, comments, and edits build a network of shared problem-solving, resulting in a durable, public artifact that future learners can reference and expand upon. This process of public knowledge creation has been a cornerstone of the internet's democratizing power.

The advent of powerful generative AI tools like ChatGPT introduces a profound disruption to this ecosystem. These models are designed to provide direct, immediate, and often highly accurate answers to a vast range of queries. For the individual user, this offers a significant increase in efficiency. A programmer struggling with a syntax error can get an instant solution from an AI, rather than asking a question on Stack Overflow and waiting for a human response. This alters the established dynamic between searching and asking (Chen & Xing, 2016). A student can ask an AI to explain a complex concept, receiving a tailored explanation without the need to sift through forum discussions. This "cognitive offloading" to AI, like the "Google effect" on memory (Sparrow et al., 2011), can free up mental

resources for more complex tasks, potentially leading to greater productivity and more sophisticated problem-solving (Gray & Tobin, 2010; Zhang & Norman, 1994).

However, this shift from public, many-to-many interactions to private, one-to-one interactions with an AI has significant and potentially detrimental consequences for the digital commons. Every query answered by ChatGPT is a query that is not asked on a public forum. An AI's solutions do not contribute to a shared knowledge base. This scenario leads to a potential "tragedy of the digital commons," where the collective resource is depleted because individuals, acting in their own rational self-interest (seeking the most efficient solution), cease to contribute to its upkeep (Aarset & Johannessen, 2022; Greco & Floridi, 2004; Meghmala, 2023; Salomon, 1993; Xue et al., 2023).

This study posits that this shift will manifest in two primary ways on platforms like Stack Exchange:

1. An Increase in Question Complexity: As users turn to AI for simpler, more foundational questions, the queries that are still asked on public forums will likely be those that are too complex, nuanced, or novel for current AI models to handle effectively. This will lead to an overall increase in the technical and difficulty of the questions on these platforms. The discourse may become more advanced, catering to an audience of experts and specialists.
2. A Decrease in User Engagement: The increase in efficiency offered by AI may lead to a decrease in the overall volume of activity on public forums. If users can get their answers from an AI, they have less incentive to participate in the community, either by asking or answering questions. This decrease in participation could lead to a decline in the number of answers per question, the number of votes, and other metrics

of community engagement. The platforms could transition from vibrant, interactive communities to static archives of expert-level problems.

These two trends are deeply intertwined and have significant implications for learning. An environment dominated by highly technical questions can be intimidating and inaccessible to novices, creating new barriers to entry for those seeking to learn a new skill or enter a new field. A decline in peer-to-peer interaction erodes the social fabric of learning communities, reducing opportunities for mentorship, collaborative problem-solving, and the development of a shared sense of identity and purpose.

From a distributed cognition perspective, the introduction of AI represents a fundamental restructuring of the cognitive system. The system is no longer primarily distributed among a network of humans; it is increasingly becoming a hybrid system distributed between a human and an AI (Cheng & Zhang, 2024). While this may enhance the cognitive capabilities of the individual, it may come at the cost of the collective intelligence of the community. This dissertation will explore this trade-off, examining how the architecture of distributed cognition is changing in the age of AI and what this means for the future of collaborative learning.

The problem, therefore, is not simply that people are using AI. The problem is that the way they are using AI may be undermining the public, collaborative systems of knowledge creation that have been so crucial to the internet's success. This study will provide data-driven insights into a trend that has, until now, been predominantly speculative. The findings will be essential for understanding how to navigate the transition to an AI-driven world in a way that preserves the invaluable resource of our shared digital commons.

AI, Human Cognition, and Educational Implications

As the theoretical foundation has established, it is crucial to explicitly connect the rapid evolution of AI chat tools with emerging concerns about cognitive processes, as these issues resonate throughout educational research and public discourse.

Large language models, or LLMs, are advanced artificial intelligence systems trained on extensive collections of text to produce coherent, human-like responses. Their growing use in education, particularly through tools like ChatGPT, presents the possibility of expanding access to information and increasing learning efficiency. However, their adoption also introduces significant cognitive challenges that require careful investigation. Empirical research shows a double effect. LLMs reduce the mental effort needed to find and process information, but they may also weaken critical thinking and discourage deep analytical engagement (Fan et al., 2025; Lee et al., 2025). These concerns are especially relevant in educational contexts where long-term cognitive development is essential.

Studies using electroencephalography, a method for measuring brain activity, have compared writing tasks completed with AI support, web search, or no assistance. Data show that AI-assisted writing reduces coordination across brain regions, signaling lower neural engagement. In contrast, unassisted writing increases activity in the alpha, beta, theta, and delta frequency bands (Kosmyna et al., 2025). These bands reflect different cognitive functions. Alpha waves support attentional control. Beta waves are associated with active reasoning. Theta waves aid memory and creative thinking. Delta waves enable deep cognitive integration. Together, these frequencies underlie executive function, conceptual synthesis, and original idea generation.

Recent findings from Kosmyna and colleagues demonstrate that not only the use of AI, but the order in which it is introduced, shapes cognitive performance. Participants who used ChatGPT from the outset showed lower neural activation and produced fewer creative essays over time. Their writing behavior also shifted toward copy-paste patterns, suggesting minimal internal processing. In contrast, participants who began with unaided writing and used ChatGPT only in later tasks showed stronger neural activation and more creative output. Those who started with ChatGPT and later lost access experienced a measurable drop in performance. These results indicate that early cognitive effort prepares the brain to benefit from AI augmentation, a finding consistent with brain-based learning principles (Akasheh et al., 2018). This parallels the evolution of Intelligent Tutoring Systems (VanLehn, 2011) and the use of AI teaching assistants (Goel & Joyner, 2017), pointing toward a future of hybrid human-AI intelligence (Luckin, 2018). Immediate reliance appears to inhibit this process and may reduce long-term learning potential.

These findings are consistent with distributed cognition theory, which conceptualizes thinking as extending across tools, individuals, and environments (Hutchins, 1995; Hollan et al., 2000). As users increasingly rely on AI to solve problems, mental effort shifts away from the individual. While this reduces task complexity, it may also limit engagement with the internal processes that support learning transfer and adaptive reasoning. These functions are central to adaptive expertise, the ability to apply knowledge flexibly in unfamiliar situations (Hatano & Inagaki, 1986).

Neural and behavioral data raise further concerns about cognitive offloading, where individuals use external systems to reduce mental effort. Frequent use of large language models has been associated with lower metacognitive engagement. Learners become less

inclined to plan, monitor, or evaluate their own thinking. They may also struggle to recall AI-generated content or explain it in their own terms (Fan et al., 2025). Even when prompted to build on earlier AI-assisted work, individuals who began with AI support show lower neural activation than those who wrote independently. This pattern suggests that AI use may affect not only attention in the moment, but also the ability to reengage with complex tasks in future settings.

To address these risks, educators and designers of learning technologies should support mixed approaches that maintain both efficiency and deep thinking. AI can be useful for simple tasks such as grammar correction, citation formatting, moving from "prompt to polished" drafts (Usher & Amzalag, 2025), and basic information gathering (Mai et al., 2024). However, it is important to pair these uses with tasks that require independent thinking. This supports the transition from passive to active-constructive-interactive learning activities, facilitating knowledge sharing across boundaries (Hwang et al., 2015). One option is to use AI to help students generate initial ideas, followed by structured assignments that ask them to analyze, evaluate, and expand those ideas without technological assistance. These strategies align with the cognitive apprenticeship model, which emphasize guided practice and gradual development of expert thinking through reflection and feedback (Collins et al., 1989).

Preserving this balance is necessary if schools and universities want to make use of AI tools without weakening essential intellectual skills. Higher order thinking abilities such as analysis, synthesis, and creative problem solving remain critical to lifelong learning. These skills, central to the upper levels of Bloom’s taxonomy, are necessary for adapting to

complex, unpredictable challenges in a knowledge-based society (Anderson & Krathwohl, 2001).

Significance of Study

This research holds significant implications for multiple fields, including education, computer science, sociology, and public policy. By providing a detailed, empirical analysis of the impact of AI on online learning communities, this study will make several key contributions.

First, for the field of education and learning sciences, this dissertation will offer a novel application of distributed cognition theory to the study of human-AI interaction. It will provide a framework for understanding how AI tools are integrated into existing cognitive systems and how they reshape the processes of learning and knowledge construction. The findings will have practical implications for educators and instructional designers, highlighting the need to create learning environments that leverage the power of AI while still fostering collaboration, critical thinking, and community engagement. The study will raise important questions about pedagogical strategies in an AI-rich world: How can we teach students to be effective and ethical users of AI? How can we design learning activities that encourage both individual exploration with AI and collaborative inquiry with peers?

Second, for computer science and platform design, the research will provide valuable data for the developers and managers of online communities. The findings will illuminate the changing behaviors and needs of users in the age of AI, offering insights that can inform the design of more resilient and adaptive platforms. For example, if platforms are indeed becoming more specialized and less accessible to novices, designers may need to implement new features to support and onboard new users, such as AI-powered tutorials, mentorship

programs, or tiered levels of discussion. The study will also highlight the challenges of content moderation in an era of hybrid human-AI content, pointing to the need for new tools and policies to ensure the integrity and transparency of online discourse (Cotton et al., 2024).

Third, for sociology and the study of digital communities, this dissertation will contribute to our understanding of the social dynamics of online interaction. It will provide a case study of how a major technological shift can transform the norms, structures, and social fabric of a community. The research will explore the potential for AI to create new forms of social stratification, with a divide between expert users who can navigate the increasingly complex discourse and novice users who are left behind. It will also touch on broader sociological questions about the future of work and expertise, as AI begins to automate tasks that were once the exclusive domain of human professionals, raising concerns about the risk of automation in the labor market (Alam, 2021; Arntz et al., 2016).

Finally, this study holds significant implications for public policy. The potential erosion of the digital commons is a matter of significant public concern. Publicly accessible knowledge is a vital resource for innovation, economic growth, and democratic participation. If this resource is diminished, it could have far-reaching negative consequences. The findings of this dissertation could inform policy discussions about how to protect and sustain the digital commons in the age of AI. This might include policies to incentivize public knowledge sharing, regulations to ensure the transparency and accountability of AI models, or public investments in open-source AI and open data initiatives (Dave, 2023).

This dissertation provides an empirical baseline for understanding a critical shift in the information ecosystem. As AI adoption accelerates, quantifying its immediate impact on public learning communities is essential for future policy and design. This study will provide

the data-driven analysis needed to understand this impact and to begin a conversation about how we can shape a future where AI and human intelligence can coexist and collaborate for the benefit of all. By bridging theory and empirical data, this research will offer a comprehensive perspective on one of the most important technological and social transformations of our time.

The results will be of interest to anyone who is concerned with the future of learning, collaboration, and knowledge in an increasingly digital world. AI is fundamentally reshaping educational environments by enabling new forms of collaboration, personalized learning, and knowledge construction. Recent studies highlight how the integration of AI-powered tools foster both individual and group engagement, allowing learners to access tailored resources, receive instant feedback, and participate in dynamic, technology-enhanced communities (Treve, 2024; Zawacki-Richter et al., 2019). However, the adoption of AI in education also raises important questions about digital equity, learner agency, and ethical considerations in data use. As institutions navigate these challenges, a balanced, evidence-based approach is needed to ensure that AI-supported learning environments promote inclusivity, critical thinking, and positive outcomes (UNESCO, 2025).

Chapter 2: Literature Review

This chapter provides the theoretical and empirical foundation for the present study by reviewing the existing literature on the impact of artificial intelligence (AI) on online learning communities. The review is structured around three key areas. First, it offers a comprehensive exploration of distributed cognition theory, the central theoretical framework guiding this dissertation. This section outlines the core principles of the theory, its evolution from traditional cognitive science, and its application to digital and AI-mediated environments. Second, it examines several established frameworks for evaluating distributed cognition, providing the analytical tools necessary to operationalize the theory for empirical research. Finally, the chapter reviews prior studies on AI and communities of learning, synthesizing existing research on the effects of AI integration in educational settings, with a particular focus on online question-and-answer (Q&A) platforms like Stack Exchange (Allamanis & Sutton, 2013; Choi et al., 2012). By integrating these three areas, this chapter establishes the context for the study, identifies the gaps in current knowledge, and justifies the research questions and methodology that follow.

Distributed Cognition Theory: A Framework for Understanding Collaborative Intelligence

The study of human cognition has traditionally focused on the internal mental processes of the individual. However, this perspective often overlooks the profound influence of the external world, the tools we use, the people we interact with, and the environments we inhabit on how we think, learn, and solve problems. Emerging as a powerful alternative, the theory of distributed cognition, pioneered by Edwin Hutchins,

proposes a radical shift in focus. It posits that cognition is not confined to the individual brain but is instead a distributed phenomenon, spread across a system of individuals, artifacts, and their social and physical environment (Argote, 2025; Hutchins, 1995; Stahl, 2006), often manifesting in knowledge-building communities (Scardamalia & Bereiter, 1994). This perspective aligns with recent research on group dynamics and organizational learning and moves the unit of analysis from the individual mind to the larger cognitive system, examining how coordinated activities and the use of tools and technologies enable complex cognitive achievements that would be impossible for an individual alone (Argote, 2025).

Since its inception, distributed cognition has transformed our understanding of human intelligence, offering a robust framework for analyzing complex socio-technical systems (Hollan et al., 2000). Early research applied the theory to settings like ship navigation and airline cockpits, demonstrating how cognition is shared and coordinated among team members and their technological instruments (Hutchins, 1995). These studies revealed that cognitive labor is dynamically allocated, with individuals and artifacts each playing specialized roles in a larger, integrated system. The focus is not just on what is happening inside people's heads, but on the interactions between people and the tools they use to externalize memory, process information, and represent knowledge.

The Role of Mediating Artifacts in Cognition

A central tenet of distributed cognition is the concept of mediating artifacts. These are external objects, symbols, or tools that shape and guide cognitive processes (Cole & Engeström, 1993). Artifacts are not merely passive aids; they actively structure thinking and enable new forms of cognitive activity. This idea has deep roots in Vygotsky's sociocultural

theory, which emphasized how cultural tools mediate psychological functioning (Vygotsky, 1978). In a distributed cognitive system, artifacts like checklists, diagrams, or computer software become integral components of the cognitive process itself.

A compelling illustration of this principle is found in the study by Morgan et al. (2008) on the "Copy and Paste" function. The researchers investigated this seemingly simple software function as a mediating artifact in learning. They found that a modified version of the tool, which embedded strategies for notetaking and summarization, significantly enhanced higher-level cognitive processing among learners compared to the standard version. The experimental group spent a far greater percentage of their time on high-level cognitive activities (27.65%) compared to the control group (6.67%). This study powerfully demonstrates that the design of a digital tool can fundamentally alter how individuals think and learn, underscoring the interplay between external artifacts and internal cognitive processes.

Distributed Cognition in Digital Learning Environments

With the rise of the internet and digital technologies, the theory of distributed cognition has become increasingly relevant. Researchers have extended the framework to explore how cognitive processes are distributed across digital platforms and online resources (Perry, 2003). Online learning communities, in particular, serve as prime examples of distributed cognition in action, reflecting principles of cooperative learning (Johnson & Johnson, 2009; Oxford, 1997) and communities of practice (Wenger, 1999). Anderson et al. (2012) demonstrated how this community activity generates significant value for the platform. A key area of this research involves estimating question difficulty, with models

evolving from competition-based metrics (Liu et al., 2013; Wang et al., 2014) to supervised learning approaches (Hassan et al., 2018; Raida et al., 2024).

In a groundbreaking study, Roschelle (2012) examined how high school students used a computer-based simulation tool to collaboratively learn physics concepts. The simulation acted as a shared cognitive artifact, allowing students to jointly manipulate variables, observe outcomes, and build a common understanding of complex principles like acceleration and vector addition. The study demonstrated that learning was not an individual act but a collective achievement, distributed across the students and the technological tool. Similarly, Suthers and Hundhausen (2003) showed that the way information was represented online, as a graph versus as text, significantly influenced students' cognitive activities during collaborative argumentation, highlighting the power of "representational guidance" in shaping distributed cognition.

The application of the theory has expanded to a wide range of digital learning contexts. Research by Zurita and Nussbaum (2004) on the use of handheld computers in the classroom found that these devices facilitated shared understanding and improved learning performance, demonstrating the benefits of distributing cognition in a mobile, digital environment. A study by Raes et al. (2012) examined how different communication tools (forums, wikis, chat) influenced students' knowledge construction. They found that the choice of tool shaped interaction patterns led to varying levels of knowledge building, suggesting that the design of the digital environment is a critical factor in the success of collaborative learning. Zheng et al. (2015) explored the distribution of cognition in Massive Open Online Courses (MOOCs), which involve vast, dispersed communities of learners. Through a quantitative analysis of forum posts, they found that active participation in

discussions was significantly related to course completion, emphasizing the role of distributed social interaction in supporting learning at scale.

Most recently, the theory has been extended to the study of artificial intelligence, examining how cognitive processes are distributed between humans and AI systems (Zhang & Patel, 2006). Cheng and Zhang (2024) describe this evolution as a shift from using AI as a tool to interacting with it as a teammate. As AI becomes more integrated into our daily lives, it acts as a new and powerful cognitive artifact, augmenting human capabilities but also fundamentally altering the structure of our cognitive systems. This dissertation is situated at this cutting edge, exploring how AI chat tools are reshaping the distribution of cognition in the established learning ecosystems of online Q&A forums.

Frameworks for Evaluating Distributed Cognition

To move from theoretical concepts to empirical analysis, researchers have developed several frameworks to evaluate, and measure distributed cognition in practice. These frameworks provide structured methods for analyzing how information is shared, how cognitive tasks are allocated, and how digital tools facilitate these processes within a cognitive system (Halverson, 2002).

- Distributed Cognition for Teamwork (DiCoT): Developed by Blandford and Furniss (2006), the DiCoT model is designed to analyze team-based work, both online and offline. It provides a structured approach by focusing on five key aspects of a cognitive system:
    - Physical Layout: The spatial arrangement of individuals and artifacts. In an online context, this translates to the user interface and the layout of the digital workspace.

- Artifacts: The tools, representations, and objects used in the cognitive process. This includes everything from software interfaces to the questions and answers themselves on a platform like Stack Exchange.
- Social Structure: The roles, relationships, and norms that govern interactions within the group. This includes community rules, reputation systems, and the unwritten social codes of a platform.
- Information Flow: The pathways through which information is propagated, transformed, and shared among the components of the system.
- Evolutionary Aspects: How the system changes and adapts over time in response to new challenges or the introduction of new tools.

The DiCoT framework is particularly useful for this study as it provides a holistic lens to analyze how the introduction of an AI chat tool (a new artifact) might impact the social structure, information flow, and overall evolution of the Stack Exchange community.

- Connection, Translation, Offloading, and Monitoring (CTOM): Presented by Martin (2012), the CTOM framework focuses specifically on the pedagogical functions of educational technologies. It views learning as an increase in coordination between cognitive systems, facilitated by technology through four key functions:
  - Connection: Technology facilitates the transfer of information between different systems (e.g., between a student and a database, or between two students).
  - Translation: Technology transforms information into more usable or understandable representations (e.g., visualizing data in a graph).

    - Offloading: Technology takes on tedious or repetitive cognitive tasks, freeing up human cognitive resources for higher-level thinking. This function is directly relevant to the central hypothesis of this dissertation, which posits that users are offloading simpler queries to ChatGPT.
    - Monitoring: Technology provides feedback on the quality of coordination, helping learners to assess their understanding and progress.

      The CTOM framework provides a powerful analytical tool for examining precisely how AI chat tools are changing the distribution of cognitive labor in learning.
- Cognitive Tutors: This framework utilizes AI to model the cognitive processes involved in problem-solving (Walker et al., 2006). By creating detailed computational models of how an expert solves problems in a specific domain, cognitive tutors can provide personalized, step-by-step guidance to learners. These intelligent tutoring systems are a prime example of distributed cognition, where the cognitive process is distributed between the learner and the AI system. AI provides scaffolding, tracks the learner's progress, and adapts to their individual needs. This framework informs the present study by providing a model for how AI can act as a cognitive partner, a role that ChatGPT is increasingly playing for users of platforms like Stack Exchange. It also highlights the way these systems can affect linguistic responses and the construction of knowledge within a learning environment (Chomsky, 1980; Stahl, 2015).

These frameworks provide the analytical scaffolding needed to investigate the impact of AI on learning communities. They allow us to move beyond simple observation and systematically analyze the changes in the cognitive architecture of these online spaces.

Prior Studies on AI and Communities of Learning

The integration of AI into education is not a new phenomenon, but the recent advent of powerful, publicly accessible generative AI like ChatGPT has accelerated its impact and raised new questions. This section reviews prior research on online learning communities, particularly Q&A forums, and the emerging body of work on the effects of AI on these environments.

Researchers have long used Stack Exchange data to understand developer behavior, analyzing questions by topic (Treude et al., 2011), quality (Squire & Funkhouser, 2014), the effectiveness of code examples (Nasehi et al., 2012), type, and code snippets (Allamanis & Sutton, 2013; Subramanian & Holmes, 2013). Online social Q&A forums like the Stack Exchange network have become invaluable resources for learning and knowledge exchange (Kang, 2022). They are living in ecosystems of distributed cognition, where individuals use natural language to ask questions, provide answers, and learn by observing the interactions of others.

Several key studies have illuminated the learning dynamics on these platforms. A study by Mamykina et al. (2011) on Stack Overflow, the largest forum in the Stack Exchange network, used mixed methods to identify different categories of learning behaviors. They found that users learn not only through the active participation of asking and answering but also through the passive observation of others' exchanges. Successful learning was associated

with a balance of these different forms of engagement, highlighting the importance of a diverse and active community.

Parnin and Treude (2011) further underscored the importance of observational learning. In a survey of Stack Overflow users, they found that more than half of the respondents reported often finding solutions to their problems simply by reading existing questions and their replies, without ever asking a question themselves. This phenomenon, often referred to as "lurking," demonstrates that the publicly accessible cognitive artifacts (the Q&A pairs) created by the community have immense value for a much larger audience than just the active participants.

Research by Yang et al. (2020) provided further evidence for the learning benefits of these platforms. Through a sophisticated analysis of user activity, they found that users' answers improved in quality after being exposed to the answers of others. This effect was particularly significant for less experienced users, suggesting that these forums can be powerful environments for novices to learn from experts and peers.

These studies collectively establish that online Q&A forums are not just repositories of information but dynamic environments for distributed, collaborative learning. The health of these ecosystems depends on a continuous cycle of public questioning, answering, and observation.

The Impact of AI Integration in Educational Contexts

The potential for AI to enhance learning has been explored in various contexts. Akasheh et al. (2018) conducted a notable study that explored the integration of cognitive neuroscience principles and AI tools in an engineering mechanics course. The development of an instructional teaching model called the "Knowledge and Curriculum Integration

Ecosystem" (KACIE), initially introduced by Solomon et al. (2017), broke down complex topics into smaller, manageable concepts and utilized AI tools to provide prompt feedback on students' work. The results were highly positive, with significant improvement in student engagement and learning outcomes. A follow-up study in a different course with a different instructor confirmed the transferability of the KACIE framework. The experimental group in the follow-up study scored, on average, 10 points higher on examinations than the control group ($p < .05$), and student surveys revealed strong agreement that the approach was more engaging and conducive to learning. This research demonstrates that when thoughtfully designed and integrated, AI can be a powerful tool for improving educational outcomes.

Other research has looked at the broader economic and social implications of AI. Studies by Arntz et al. (2016) and a White House (2016) report on the economy found that while AI is likely to automate many tasks, it is less likely to replace jobs that require complex social and emotional intelligence, such as teaching. They suggest that AI will more likely serve as a tool to supplement and enhance the work of human educators rather than replace them entirely. A review by Alam (2021) also found that AI has been widely embraced in educational institutions for administrative tasks, instruction, and creating personalized learning experiences, further highlighting the growing integration of AI into all facets of education.

However, the emergence of generative AI chat tools such as ChatGPT introduces a novel and potentially disruptive dynamic. These tools do not just supplement; they can also substitute for the very interactions that form the bedrock of platforms like Stack Exchange. Every question answered privately by ChatGPT is a question that does not become a public cognitive artifact on Stack Exchange. This dynamic leads to the central problem investigated

in this dissertation: the potential for a "tragedy of the digital commons," where a valuable collective resource is eroded because individuals, acting in their own self-interest (seeking the most efficient answer), no longer contribute to its maintenance and growth (del Rio-Chanona et al., 2023).

This tension represents the frontier of research on AI in education. While AI holds immense promise for personalizing learning and increasing individual efficiency, it also poses a threat to the public, collaborative knowledge-building ecosystems that have become so vital. This dissertation is one of the first major studies that looks closely at this issue, using the idea of distributed cognition to examine how ChatGPT affects the active and important learning communities within the Stack Exchange network. By doing so, it seeks to fill a critical gap in the literature and provide data-driven insights to inform the future of AI-enhanced learning.

Chapter 3: Methodology

This chapter details the methodological framework employed to investigate the impact of large language models (LLMs), specifically ChatGPT, on the complexity of user-generated questions and engagement patterns within online learning communities. The research is designed to address the central research question: How has the introduction and utilization of AI chat tools within the Stack Exchange online community influenced the complexity of user-generated questions, and how can changes in question complexity be interpreted through the lens of Distributed Cognition theory to understand the shifting quality of collaborative learning within the platform (Hutchins, 1995; Lave & Wenger, 1991)?

To answer this question and its sub-questions, this study adopts quasi-experimental, pre-test/post-test design, integrating quantitative and qualitative analytical techniques. This approach allows for a comprehensive examination of the phenomenon, combining large-scale statistical analysis of linguistic and engagement data with a theoretical interpretation grounded in distributed cognition, and adheres to principles of open science and transparency (Miguel et al., 2014). The research was conducted in two primary phases: a preliminary pilot study designed to test and refine the methodology, followed by a full-scale investigation across multiple online communities. This chapter provides a detailed account of the research design, data collection and processing procedures, the operationalization of key variables, and the statistical methods used for analysis.

Unit of Analysis

Each observation in the dataset represents a single original Stack Exchange question (i.e., the initiating question in a Q&A thread). All linguistic complexity measures (e.g.,

Flesch–Kincaid Grade Level and lexical density) were computed from the question title and question body text only. The study does not analyze the textual content of answers. Variables such as the number of answers received, time to first answer, upvotes, and accepted-answer status are treated as question-level outcomes/metadata describing how the community responded to each question.

Research Design and Rationale

The overall research design is a retrospective, observational study with a quasi-experimental, comparative component. This design is particularly well-suited for investigating the effects of a large-scale, naturally occurring event (i.e., the public release and widespread adoption of ChatGPT). Since it is impossible to randomly assign users to "ChatGPT" and "no-ChatGPT" groups in a real-world setting, this study leverages a pre-test/post-test approach, comparing data from periods before and after the intervention. The primary "treatment" is the public availability of ChatGPT, and the study observes its correlation with changes in user behavior on the Stack Exchange platform.

This design choice is underpinned by the theoretical framework of distributed cognition. The theory posits that cognitive processes are distributed across individuals, tools, and artifacts (Hollan et al., 2000). Stack Exchange communities are excellent examples of such distributed cognitive systems, where knowledge is co-constructed through public interaction. The introduction of ChatGPT represents the integration of a powerful new cognitive artifact into this system. An observational design allows for the examination of how this new artifact alters the existing dynamics of the system in its natural context, rather than in a controlled, artificial environment. This study can detect changes in the underlying distributed cognitive processes by analyzing question texts (titles and bodies) and question-

level metadata (e.g., number of answers received, time to first answer, upvotes, and accepted-answer status) before and after the introduction of ChatGPT. Answer text was not analyzed.

The quasi-experimental approach further strengthens the research. The quantitative component provides robust, empirical evidence of changes in question complexity and engagement, allowing for the identification of statistically significant trends at scale. The qualitative component, informed by the theory of distributed cognition, provides the interpretive lens necessary to understand the meaning of these quantitative changes. This integration of methods ensures that the study moves beyond mere description to offer a theoretically grounded explanation of how AI is reshaping collaborative learning online.

Phase 1: Pilot Study

The primary purpose of the pilot was to test the feasibility of the proposed research design, validate the chosen metrics for question complexity, and refine the data collection and analysis pipeline. It served as a proof-of-concept, allowing for the identification of potential challenges and the optimization of the methodology for the main study.

The pilot study focused on the Stack Exchange *Data Science* subdomain. This community was selected for its relevance to the research topic and its active user base. Data was retrieved from a publicly available data dump accessed via the link (https://data.stackexchange.com). To ensure randomization and mitigate bias, a random number generator was used to select question IDs from the full population of eligible questions within the designated time frames for each group. First, all questions asked within the pre-ChatGPT period (January 1, 2022 - September 30, 2022) and the post-ChatGPT period (January 1, 2023 - September 30, 2023) were compiled into separate lists. Then, 50

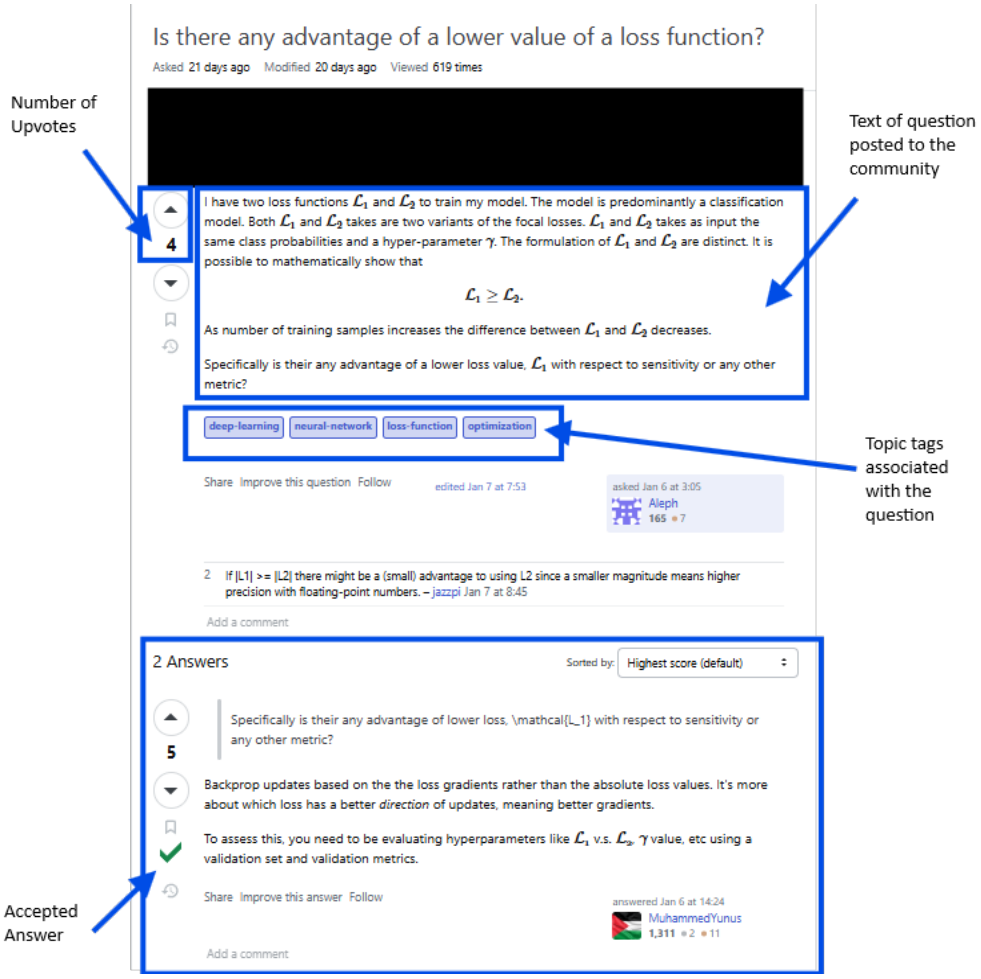
distinct questions were randomly selected from each period using this method, resulting in two equally sized, non-overlapping groups. A total of 100 questions were sampled. This approach aligns with best practices for random sampling in social science research and helps increase the validity and generalizability of the findings (Creswell & Creswell, 2018).

- Pre-ChatGPT group (*n*=50): Questions asked between January 1, 2022, and September 30, 2022.
- Post-ChatGPT group (*n*=50): Questions asked between January 1, 2023, and September 30, 2023.

This balanced, random sampling approach was designed to provide a preliminary, unbiased comparison between the two periods. Figure 1 shows an example of a typical question sampled from the data.

Figure 1

Example data obtained from a typical question asked in the Stack Exchange Data Science community.



Variables and Measures

Several engagement metrics were used as proxies for cognitive demand and community interaction (Mamykina et al., 2011; Yang et al., 2020). These included:

- *Number of Responses*: The total count of answers a question received.
- *Number of Upvotes*: A measure of community approval and perceived quality.

- *Time to First Response*: The time in seconds between when the question was created to when it received its first response, indicating the immediacy of community engagement.
- *Tagged Categories Count:* The number of topics the question is related to.
- *Total Word Count:* Total count of words contained in the title and body of the question.
- Has Accepted Answer: Coded as an independent binary variable (0 = No Answer, 1 = Accepted Answer) to represent if an accepted solution has been provided for the question.
- *Flesch-Kincaid Grade Level* (i.e., Question Complexity): This was operationalized using a widely recognized readability metric: the Flesch-Kincaid Grade Level (Flesch, 1948; Kincaid et al., 1975), chosen over code-specific readability metrics (Buse & Weimer, 2010) to capture natural language complexity. These scores were calculated using the *textstat* library in Python and served as proxies for the linguistic difficulty of the questions.
- *ChatGPT Period*: Coded as an independent binary dummy variable (0 = pre-ChatGPT, 1 = post-ChatGPT) to represent the treatment condition.

Statistical Assumptions Testing

Prior to conducting group comparisons, comprehensive assumption testing was performed to determine the appropriateness of parametric versus non-parametric statistical methods. Normality was assessed using *Shapiro-Wilk tests* (most powerful for samples $n < 50$) and descriptive statistics, including skewness and kurtosis. Significance levels were set at $p < .05$ for inferential tests and skewness < 2.0 and kurtosis < 7.0 for descriptive measures

(Table 1). The equal variances assumption was assessed using *Levene's test* (robust to non-normality), with the significance level set at $p < .05$.

Table 1

Normality Assessment Results for Continuous Variables

Variable	Shapiro-Wilk *W*		Skewness	
	Pre	**Post**	**Pre**	**Post**
Number of Upvotes	0.67***	0.81***	2.29	1.03
Tagged Categories	0.92**	0.89***	−0.05	−0.16
Number of Responses	0.56***	0.55***	1.77	1.09
Time to Response	0.83***	0.86***	1.08	0.68
Flesch-Kincaid Level	0.91***	0.95*	1.21	0.81
Total Word Count	0.92**	0.83***	0.88	1.80

Note. Pre = Pre-ChatGPT period; Post = Post-ChatGPT period. All variables violated normality assumptions, necessitating non-parametric tests.

**p < .05. **p < .01. ***p < .001.*

For variables where normality and homogeneity of variance assumptions were met, intended parametric tests included independent samples t-tests to compare means between the pre- and post-ChatGPT periods. However, thorough assumption testing revealed that the majority of variables exhibited violations of normality in both groups (Shapiro-Wilk $p < .05$ for nearly all variables), validating the use of non-parametric tests (primarily Mann-Whitney

U) for group comparisons (Table 2). Levene's test confirmed equality of variances was typically met (majority $p > .05$).

Table 2

Variance Homogeneity Assessment and Test Selection

Variable	Pre Var	Post Var	Levene *F*	*p*
Number of Upvotes	1.65	0.74	1.05	.307
Tagged Categories	1.44	1.67	0.59	.444
Number of Responses	0.34	0.20	0.15	.699
Time to Response	3151.23	3345.71	0.45	.504
Flesch-Kincaid Level	10.43	9.60	0.10	.755
Total Word Count	4186	9010	1.03	.313

Note. All variables met variance homogeneity (Levene $p > .05$) but violated normality (Table 1), requiring non-parametric tests.

Pilot Study Findings and Methodological Refinements for a Full Study

Due to systematic violations of normality assumptions across all continuous variables (see Parametric Validation section), Mann-Whitney U tests were conducted to compare question characteristics between pre-ChatGPT (n = 50) and post-ChatGPT (n = 50) periods. All analyses were two-tailed with $\alpha = .05$, and effect sizes were calculated using Cohen's *d*.

The pilot study yielded several important findings that informed the design of the full-scale investigation (Table 3).

Table 3

Mann-Whitney U Tests: Pre- vs. Post-ChatGPT Question Characteristics

Variable	Pre *M* (SD)	Post *M* (SD)	*U*	*p*	*d*
Upvotes	0.84 (1.28)	0.54 (0.86)	1376.50	.338	-0.27
Tags	3.10 (1.20)	3.38 (1.29)	1096.00	.277	0.22
Responses	1.30 (0.58)	1.26 (0.44)	1244.50	.964	-0.08
Time to First Response	418 (435)	456 (448)	1221.50	.847	0.09
FK Grade	9.92 (3.23)	10.45 (3.10)	1078.00	.237	0.17
Word Count	119 (65)	138 (95)	1151.50	.499	0.22

Note. N = 100 (50 per group). All tests two-tailed, α = .05. No results were statistically significant. Cohen's d: negligible (< .20), small (.20–.49), medium (.50–.79), large (≥ .80).

No statistically significant differences were observed between pre- and post-ChatGPT periods across any of the seven variables examined (all $p > .10$). Question engagement metrics showed stability across periods. Number of upvotes decreased from pre-ChatGPT ($M = 0.84$, $SD = 1.28$) to post-ChatGPT ($M = 0.54$, $SD = 0.86$), but this 36% reduction was not statistically significant, $U = 1376.5$, $p = .338$, $d = -0.27$. The number of responses remained virtually identical between periods (pre: $M = 1.30$, $SD = 0.58$; post: $M = 1.26$, $SD = 0.44$), $U = 1244.5$, $p = .964$, $d = -0.08$.

Question complexity measures revealed minimal change. Tagged categories count showed a slight increase from pre-ChatGPT ($M = 3.10$, $SD = 1.20$) to post-ChatGPT ($M = 3.38$, $SD = 1.29$), $U = 1096.0$, $p = .277$, $d = 0.22$. Flesch-Kincaid Grade Level remained stable (pre: $M = 9.92$, $SD = 3.23$; post: $M = 10.45$, $SD = 3.10$), $U = 1078.0$, $p = .237$, $d =$

0.17. Total word count increased modestly from pre-ChatGPT (M = 119.36, SD = 64.70) to post-ChatGPT (M = 137.52, SD = 94.92), but this 15% increase was not significant, U = 1151.5, p = .499, d = 0.22.

Community response patterns demonstrated remarkable consistency. Time to first response showed no meaningful change between pre-ChatGPT (M = 417.98, SD = 434.83) and post-ChatGPT (M = 456.19, SD = 448.04) periods, U = 1221.5, p = .847, d = 0.09. Additionally, question resolution rates showed a notable but non-significant decrease. Accepted answers occurred in 90% of pre-ChatGPT questions compared to 76% of post-ChatGPT questions, $\chi^2(1) = 2.55$, p = .110, representing a 14-percentage point decrease that approached statistical significance. Effect sizes across all measures ranged from negligible to small (d = -0.27 to d = 0.22), with the largest effects observed for number of upvotes (favoring pre-ChatGPT) and word count (favoring post-ChatGPT).

Implications for Future Study Design

The pilot study's systematic null findings provide crucial insights for designing the comprehensive investigation. This discovery led to several critical methodological refinements: The observed small effect sizes (d < 0.3) indicate that detecting meaningful differences would require substantially larger samples. The small sample size (N=100) likely lacked the statistical power to detect subtle effects of ChatGPT. Power analysis suggests approximately 800 questions asked per group would be needed to detect small effects (d = 0.2) with 80% power, compared to the pilot study's capacity to detect only medium to large effects. The full-scale investigation was therefore designed with a much larger dataset, spanning thousands of questions from multiple communities. The pilot study also showed that while readability scores are useful, they primarily capture syntactic complexity and may

not fully reflect the technical or conceptual difficulty of questions (Arora et al., 2015). This highlighted the need for a more nuanced measure of complexity. Consequently, Lexical Density (LD) analysis was incorporated into the main study to assess the technical specificity of the question content.

The pilot focused on a single community; its findings may not generalize to other Stack Overflow communities. To enhance the generalizability of the findings, the main study was expanded to include four diverse Stack Exchange communities, allowing for a comparative analysis of how the impact of AI might vary across different knowledge domains.

The pilot study was invaluable in shaping a more robust and comprehensive methodology for the main investigation. It validated the core research design while also revealing key areas for improvement, ultimately leading to a more rigorous and insightful full-scale analysis.

Main Study Design: Full-Scale Investigation

The main investigation built directly on the lessons learned from the pilot study, employing a more extensive and refined methodology to provide a comprehensive analysis of AI's impact on online learning communities.

Data Selection and Sampling

The scope of the full-scale study was expanded to include five distinct Stack Exchange communities: Data Science, Computer Science, Artificial Intelligence, Mathematics, and Academia. These communities were selected due to their strong orientation toward technical and scholarly inquiry, domains in which large language models (LLMs), such as ChatGPT, are likely to exert notable influence.

To facilitate a meaningful temporal comparison, a structured sampling strategy was employed to contrast community behavior before and after the public release of ChatGPT. OpenAI released ChatGPT to the public on November 30, 2022, marking a significant turning point in the availability and visibility of powerful generative AI tools. To clearly delineate pre- and post-ChatGPT conditions, the study introduced a three-month buffer around the release date. This buffer period (October 1, 2022 – December 31, 2022) was excluded from the analysis to account for early experimentation, uneven adoption rates, and initial public reaction, which could introduce noise into the data.

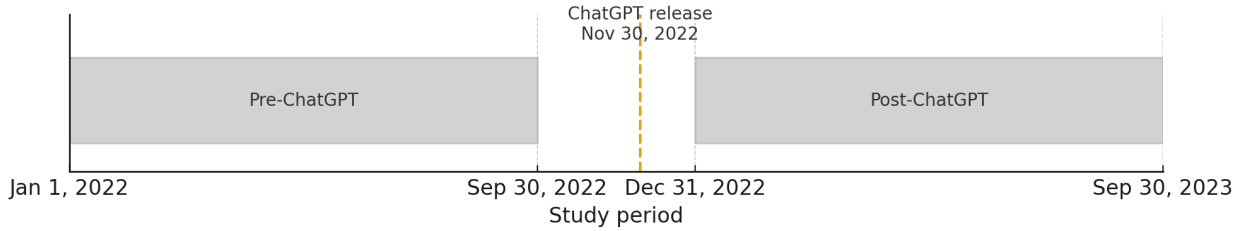
Two clearly defined sampling windows, consisting of the same three calendar months of two successive years to control for possible seasonality effects, were constructed around this buffer:

- Pre-ChatGPT period: January 1, 2022 - September 30, 2022
- Post-ChatGPT period: January 1, 2023 - September 30, 2023

The cutoff date of October 1, 2022, was selected as the end of the pre-period to ensure a full nine months of activity uninfluenced by the public release of ChatGPT. Similarly, the post-period began on January 1, 2023, allowing for widespread awareness and use of ChatGPT to have taken hold within the studied communities (Figure 2).

Figure 2

Sampling Periods for Pre- and Post-ChatGPT Analysis



Note. This timeline illustrates the two distinct sampling windows used in the study: the pre-ChatGPT sample (January 1, 2022 - September 30, 2022) and the post-ChatGPT sample (January 1, 2023 - September 30, 2023). A three-month buffer period (October 1, 2022 - December 31, 2022) was excluded from analysis to reduce adoption rate of treatment status during ChatGPT's initial rollout. The first 9 months of each year were selected to control for seasonality. The vertical marker indicates the public release of ChatGPT on November 30, 2022.

Prior to data preprocessing and cleaning, the initial dataset resulted in the following sample sizes:

- Academia: 1,976 (pre), 2,128 (post)
- Artificial Intelligence: 942 (pre), 953 (post)
- Computer Science: 2,243 (pre), 1,632 (post)
- Data Science: 2,525 (pre), 1,523 (post)
- Mathematics: 68,237 (pre), 57,836 (post)

As a direct behavioral indicator of participation, these pre/post counts reflect the volume of initiating questions asked during each sampling window. Aggregated across communities, questions declined from 75,923 (pre) to 64,072 (post), a 15.6% decrease. The

decline is concentrated in Computer Science (-27.2%), Data Science (-39.7%), and Mathematics (-15.2%), while Academia increased (+7.7%) and Artificial Intelligence remained essentially flat (+1.2%).

Preprocessing and Cleaning

The data preprocessing was conducted by applying three systematic filtering operations to ensure data integrity and analytical validity (Osborne, 2013). Each exclusion criterion was defined a priori based on operational requirements and theoretical considerations. Duplicate observations were identified through exact string matching on the Body variable (Müller & Freytag, 2005), which prevented artificial sample size inflation and maintained independence assumptions in subsequent analyses (Field, 2013). The first chronological occurrence was retained, eliminating 36 duplicate records (0.03% of the dataset).

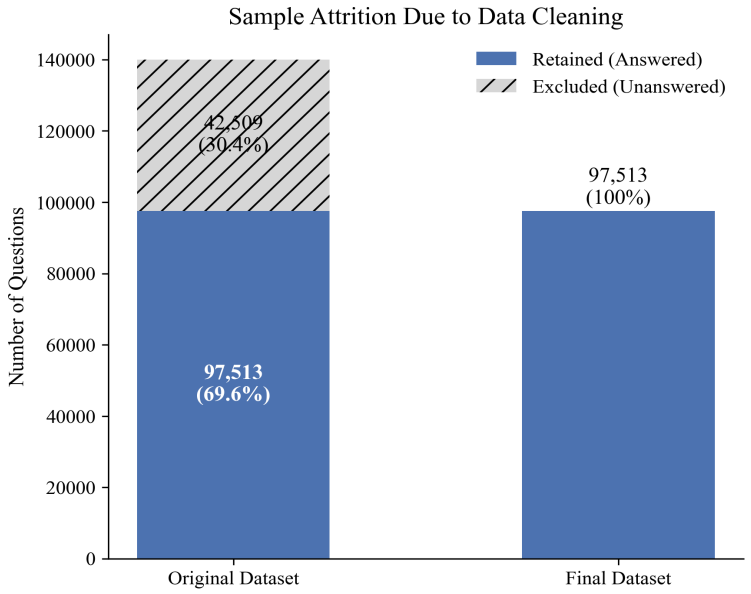
Complete case analysis was implemented, meaning that observations with null or empty *Body* values (Allison, 2001) were deleted. This listwise deletion was justified because the absence of question text made the computation of linguistic complexity metrics and readability indices impossible. Imputation was neither feasible nor conceptually meaningful, as Body serves as the source for multiple predictor variables (Little & Rubin, 2019). This operation excluded two observations (0.0014% of the dataset).

The most substantial filtering operation involved the systematic exclusion of observations lacking valid response times (*TimeToFirstResponse* = null), aligning with the study's focus on community engagement dynamics (Mamykina et al., 2011). Unanswered questions provide limited information regarding community responsiveness, rendering them less informative for engagement analyses (Anderson et al., 2012). This process removed

42,483 observations (30.3% of the dataset), consistent with unanswered question rates in online communities (Choi et al., 2012). The preprocessing enhanced data quality by eliminating pseudo replication (Hurlbert, 1984) and producing complete information across analytical variables (Schafer & Graham, 2002). The final dataset comprised 97,513 observations (69.65% retention), providing substantial statistical power (Cohen, 1988). Temporal distribution remained balanced (54.4% pre-ChatGPT, 45.6% post-ChatGPT), supporting difference-in-differences estimation (Angrist & Pischke, 2009). All preprocessing decisions were specified prior to analysis to minimize researcher degrees of freedom (Simmons et al., 2011). Figure 3 - 5 illustrate the stability of the dataset following the exclusion of unanswered questions (Appendix C).

Figure 3

Comparison of Original Corpus vs. Final Analysis Sample



Note. The reduction in sample size is almost entirely attributable to the removal of unanswered questions (n=42,483).

Figure 4

Stability of Community Representation Before and After Data Cleaning

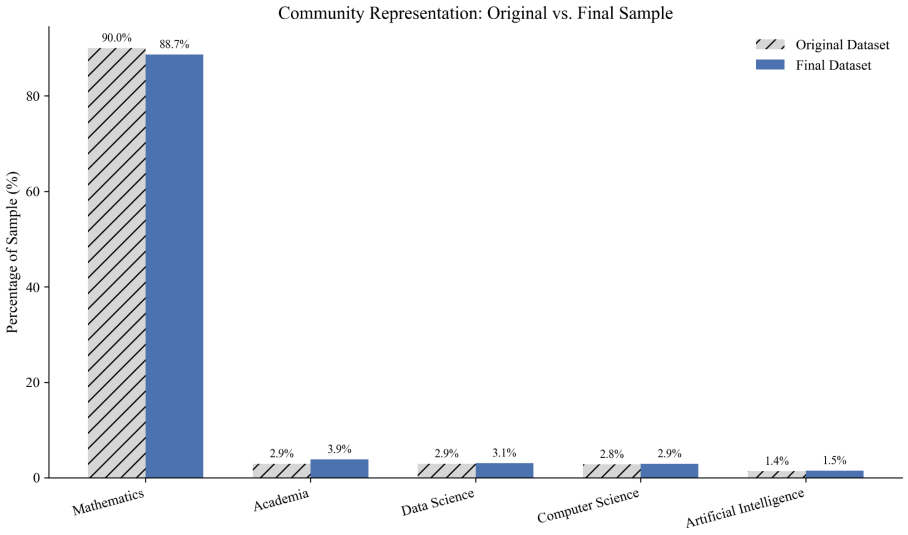
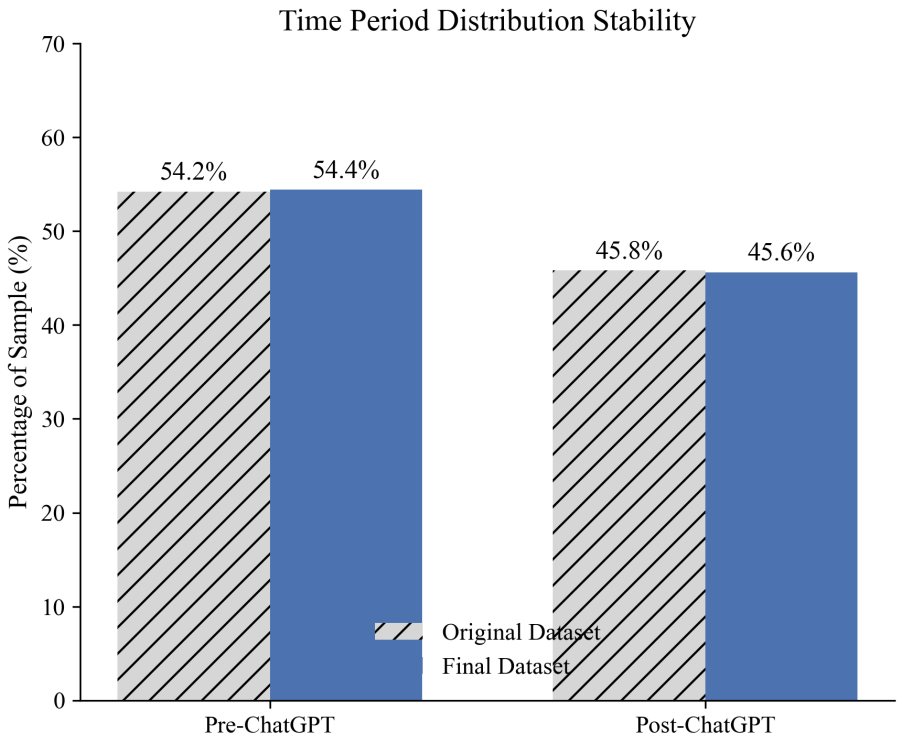


Figure 5

Stability of Time Period Distribution (Pre- vs. Post-ChatGPT)



The compositional integrity of the sample was preserved across both Communities and Time Periods. The filtering did not disproportionately affect any single community, and the relative proportions of each community in the sample remained consistent and the distribution shifted by less than 1.5%. Additionally, the proportion of questions between pre-ChatGPT and post-ChatGPT questions remains virtually unchanged ($\Delta < 0.2\%$). Data exclusions did not appear to bias the final sample composition and remained representative of the questions in the original corpus.

The dataset was also examined for the presence of extreme values (outliers) that could unduly influence results or violate statistical assumptions. Several continuous variables

exhibited substantial positive skewness and excess kurtosis, indicating the presence of extreme observations in the upper tails of their distributions. To address these distributional concerns while preserving the overall sample size and essential data structure, winsorization was applied to all continuous variables in the dataset.

Winsorization is a data transformation technique that addresses extreme values by replacing observations beyond specified percentiles with the values at those percentile thresholds (Dixon, 1960; Tukey, 1962). Unlike deletion methods that remove outliers entirely, winsorization retains all observations in the dataset, thereby maintaining statistical power and preventing the loss of potentially meaningful data points. This approach is particularly appropriate when extreme values may reflect genuine, albeit rare, occurrences rather than measurement error (Wilcox, 2012). In the context of online community data, extreme values, such as questions receiving exceptionally high upvote counts or unusually long response times, represent authentic community behaviors that merit inclusion, albeit with reduced influence on parameter estimates.

The winsorization procedure successfully reduced distributional skewness and kurtosis while preserving central tendency. The most substantial improvements occurred for variables with extreme initial skewness *Time to First Response* showed a 46.56% reduction in standard deviation, *Number of Upvotes* exhibited a 30.22% reduction, and *Total Word Count* showed a 17.36% reduction. Correspondingly, skewness values decreased from 10.74 to 7.36, 8.65 to 2.43, and 5.94 to 2.01 for these three variables, respectively. Winsorization had minimal impact on central tendency measures; mean values changed by less than 5% across all variables, and median values remained unchanged, reflecting the procedure's targeted focus on extreme tail values.

Because the sample size (N = 97,513) is larger than the maximum sample size for standard Shapiro-Wilk tests (n ≤ 5,000), a multi-method approach was used. This included D'Agostino-Pearson K^2 tests, which are suitable for large samples and combine skewness and kurtosis into one test, and Shapiro-Wilk tests on 10 independent random subsamples (n = 5,000 each) to determine the distributional properties while keeping the sample size in mind. Table 4 presents the results, which indicated departures from normality across all continuous variables (West et al., 1995).

Table 4

Normality Assessment Results for Winsorized Continuous Variables

Variable	Skewness	Kurtosis	*D'Agostino-Pearson K^2*	Shapiro-Wilk *W*
Number of Upvotes	2.43	8.56	51,014.30***	.76***
Tagged Categories	0.24	-0.92	15,603.11***	.91***
Number of Responses	2.36	6.23	47,082.84***	.57***
Time to Response	7.36	55.04	123,897.20***	.16***
Flesch-Kincaid Level	1.06	1.98	17,033.98***	.94***
Lexical Density	0.08	0.16	189.91***	.99***
Total Word Count	2.01	4.98	39,483.02***	.81***

*Note. Kurtosis values represent excess kurtosis (Fisher's definition). Shapiro-Wilk statistics represent mean values from 10 random subsamples (n = 5,000 each) due to the sample size constraints of the test. All variables were assessed following winsorization at the 1st and 99th percentiles. *p < .05. **p < .01. ***p < .001.*

D'Agostino-Pearson K^2 tests were highly significant for all variables ($p < .001$), indicating non-normal distributions. Shapiro-Wilk tests conducted on random subsamples corroborated these findings, with none of the 10 subsamples per variable demonstrating normal distributions ($p < .05$ for all).

Validation of Multivariate Assumptions and Robustness Checks

Preliminary diagnostic tests of the winsorized dataset revealed that significant skewness and kurtosis remained in several key dependent variables, notably *Time to Response* (Skewness = 7.36) and *Number of Upvote*s (Skewness = 2.43). However, the Central Limit Theorem suggests that, with large samples, parametric tests such as MANOVA are robust to violations of normality (Field, 2013). Here, the sample size ($N = 97{,}513$) is very large.

Nonetheless, to ensure the validity of the study's findings, a sensitivity analysis was conducted. A parallel transformed dataset was created by applying mathematical transformations to correct distributional irregularities. Specifically, logarithmic transformations ($\log_{10}(x+1)$ or $\log_{10}(x-x_{min}+1)$) were applied to count-based and time-based variables. For the variable *Tagged Categories*, the counts were relatively low, so a square root transformation was applied instead. The variable *Lexical Density* was found to be normally distributed and required no transformation.

Comparison of the MANOVA results between the original (winsorized) and transformed datasets showed that Wilks' Lambda and Pillai's Trace statistics remained nearly identical across both models (e.g., Time Period Effect on Complexity: Original $\Lambda = 0.999$, $p < .001$; Transformed $\Lambda = 0.999$, $p < .001$). Thus, the violation of normality assumptions did

not seem to inflate Type I error rates or alter the statistical conclusions. Consequently, for ease of interpretation and practical application, the original winsorized dataset is retained for all subsequent analyses reported in this study. Detailed descriptive comparisons and transformation diagnostics are provided in Appendix B.

Instrumentation and Operationalization of Constructs

The study's central construct, question complexity, was operationalized as a multidimensional concept using complementary metrics that capture distinct facets of complexity (Stenetorp et al., 2012). Syntactic complexity was quantified using the *Flesch-Kincaid Grade Level* (Flesch-K), computed via the *textstat* library in Python. This metric estimates the U.S. educational grade level required for text comprehension based on average sentence length and syllabic density (Flesch, 1948; Kincaid et al., 1975), providing a reliable indicator of syntactic and lexical complexity for large-scale analysis (DuBay, 2004).

Lexical Density (LD) was defined as the proportion of content words, nouns, proper nouns, main verbs, adjectives, adverbs relative to all tokens, following classic functional-linguistic accounts (Halliday et al., 2004) and subsequent corpus-based work on informational versus conversational registers (Biber, 1991). It was calculated using *the spaCy* Python library, after part-of-speech tagging with a modern English NLP pipeline, to quantify how much semantic information was packed into each question (title + body) (Graesser et al., 2011). In Stack Exchange data, a ‘post’ can refer to a question or an answer; this study restricts analysis to question posts (initiating questions) only.

Together, Flesch-K and LD capture complementary aspects of complexity. Flesch-K is more sensitive to sentence structure and orthography, whereas LD directly indexes the balance between content-bearing and function words. This dual-metric approach was

selected because it preserves continuity with existing readability work, incorporates a theoretically grounded measure of informational content, and avoids overreliance on high-dimensional vector-space models whose interpretation can be opaque in applied settings (Salton & Buckley, 1988).

Community engagement was operationalized at the question level using metadata extracted from the Stack Exchange API, including number of answers received (*ResponseCount*), *upvotes*, *time to first answer*, *tag count*, *accepted-answer indicator*, and question *word count* (Hanrahan et al., 2012). These measures describe how the community responded to each question but do not analyze the text of any answers. *ChatGPT* availability was coded as a binary indicator (0 = pre-ChatGPT release, 1 = post-ChatGPT release) to capture the temporal treatment condition and facilitate identification of shifts in question characteristics and engagement patterns. Table 5 below summarizes the final full sample sizes used in the study.

Table 5

Final Sample Sizes by Community and Time Period

Community	Pre	Post	Total
Academia	1,832	1,936	3,768
Artificial Intelligence	706	756	1,462
Computer Science	1,640	1,203	2,843
Data Science	1,794	1,190	2,984
Mathematics	47,118	39,338	86,456
Total	53,090	44,423	97,513

Note. Pre = January 1–September 30, 2022; Post = January 1–September 30, 2023. Sample includes only questions that receive at least one answer (i.e., answered questions). Pre/Post refers to the question creation date relative to ChatGPT's release.

Changes in Community Activity Levels

Beyond the characteristics of individual questions, the most striking behavioral shift observed following the release of ChatGPT is a substantial decline in the overall volume of new questions asked on the platform. As shown in Table 5, the total number of questions in the dataset (representing all answered questions in the sampled periods) fell approximately 16.3%, from 53,090 in the pre-ChatGPT period to 44,423 in the post-ChatGPT period.

This decline was not uniform across communities. The Data Science and Computer Science communities experienced the sharpest contractions, with question volumes dropping by 33.3% (1,794 to 1,190) and 26.6% (1,640 to 1,203), respectively. The Mathematics community, which constitutes the largest portion of the sample, also saw a significant reduction of 16.5% (47,118 to 39,338).

In contrast, the Academia sector saw a slight increase of 5.7% (1,832 to 1,936) and the Artificial Intelligence community remained relatively stable, rising slightly from 706 to 756 questions (+7.1%). Despite the global explosion of interest in AI during this period, the volume of questions on the AI forum did not skyrocket, suggesting that even in the domain of AI itself, users may be turning to generative tools rather than peer-to-peer forums for assistance.

Representative Question Examples: Low and High Complexity

To illustrate the types of questions analyzed in this study, representative examples are presented below. Questions were classified based on two complexity measures: Flesch-Kincaid Grade Level (linguistic complexity) and Lexical Density (LD) Low complexity questions exhibit lower Flesch-Kincaid Grade Level scores (indicating simpler sentence structure and vocabulary) and lower LD scores (indicating less specialized technical

terminology). These questions typically use straightforward language and common vocabulary accessible to broader audiences.

Figure 6

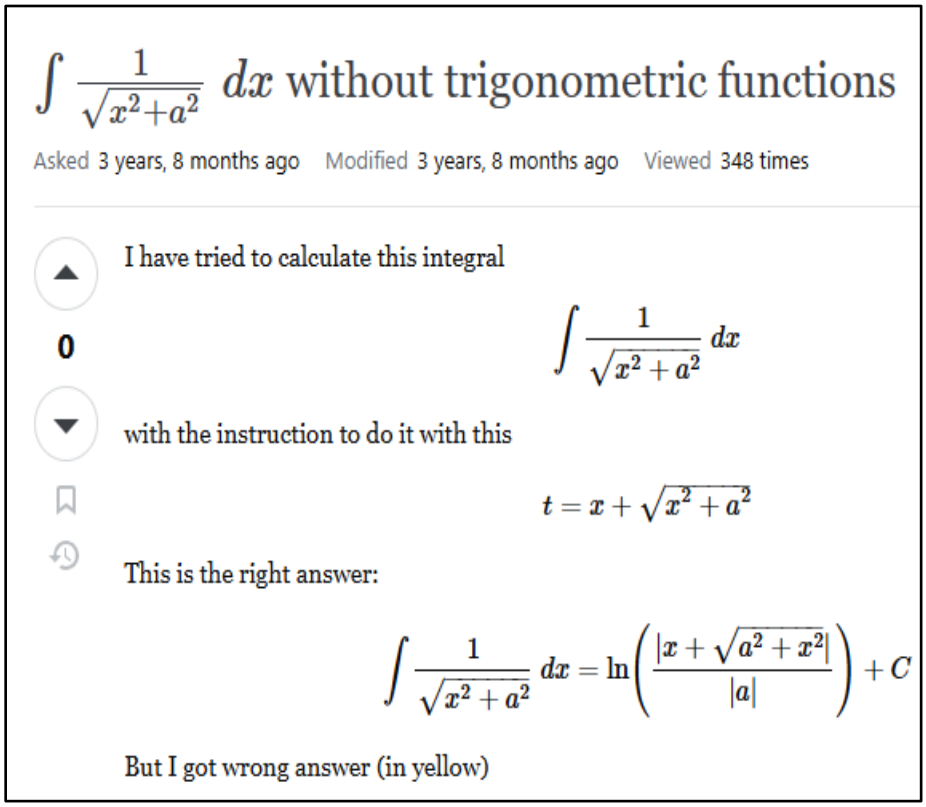
Low Complexity Example: Mathematics (Pre-ChatGPT)

Note. Flesch-Kincaid Grade Level = 4.24; Lexical Density = 0.483; Word Count = 118, Date Asked = 5/13/2022 07:02

High complexity questions exhibit higher Flesch-Kincaid Grade Level scores (indicating more complex sentence structure and advanced vocabulary) and higher LD scores (indicating greater use of specialized technical terminology). These questions typically employ domain-specific language and sophisticated linguistic constructions.

Figure 7

High Complexity Example: Mathematics (Pre-ChatGPT)



$\int \frac{1}{\sqrt{x^2+a^2}}\,dx$ without trigonometric functions

Asked 3 years, 8 months ago Modified 3 years, 8 months ago Viewed 348 times

0

I have tried to calculate this integral

$$\int \frac{1}{\sqrt{x^2+a^2}}\,dx$$

with the instruction to do it with this

$$t = x + \sqrt{x^2+a^2}$$

This is the right answer:

$$\int \frac{1}{\sqrt{x^2+a^2}}\,dx = \ln\left(\frac{|x+\sqrt{a^2+x^2}|}{|a|}\right) + C$$

But I got wrong answer (in yellow)

Note. Flesch-Kincaid Grade Level = 18.62; Lexical Density = 0.549; Word Count (including formulas) = 60; Date Asked = 2/25/2022 10:40.

Statistical Analysis Plan

The study employed a range of statistical techniques to analyze the data and test the research hypotheses. The analysis was conducted using Python with the SciPy, statsmodels, and scikit-learn libraries. The primary dependent variables for inferential testing were question-text complexity indices (Flesch–Kincaid Grade Level and lexical density) computed from question titles and bodies; engagement and resolution measures (e.g., number of

answers, time to first answer, accepted-answer rate) were analyzed as question-level outcomes rather than as features of answer text.

Descriptive statistics were computed to characterize the dataset across time periods and academic communities. Means and standard deviations were calculated for each variable stratified by time period, community, and the community-by-time period interaction. These statistics provide a comprehensive overview of question complexity, technical specificity, community engagement patterns, and question resolution success before and after the release of ChatGPT.

Pearson product-moment correlations were calculated to examine bivariate relationships among key variables. Eight variables were included in the correlation matrix: Flesch-Kincaid Grade Level, Lexical Density Score, Total Word Count, Number of Responses, Number of Upvotes, Time to First Response (seconds), Number of Tagged Categories, and Accepted Answer Rate.

Statistical assumptions (e.g., normality of residuals, homogeneity of variances) were checked prior to conducting the inferential tests to ensure the validity of the results. To address research questions concerning how the introduction of AI chat tools influences question complexity and distributed cognitive processes in collaborative learning environments, a one-way multivariate analysis of variance (MANOVA) was conducted. This analysis examined whether the temporal period (pre-ChatGPT vs. post-ChatGPT) significantly affected question complexity, operationalized through two dependent variables: Flesch-Kincaid Grade Level and Lexical Density. Additionally, to examine whether the temporal effect varied across communities for each complexity measure, separate univariate

two-way analyses of variance (ANOVAs) were conducted with time-period and community as independent variables.

Cohen's *d* was calculated as a measure of effect size to assess the practical significance of the observed changes. This standardized metric provides a more intuitive understanding of the magnitude of the difference between the pre- and post-ChatGPT groups, independent of sample size (Cohen, 1988).

Ethical Considerations and Limitations

The research was designed and conducted with careful attention to ethical considerations. The study exclusively used publicly available, anonymized data from the Stack Exchange data dump (Internet Archive, 2024). As no personal identifiable information was collected or analyzed, the research did not require informed consent from individual users and posed no risk to their privacy.

Despite the rigorous methodology, the study has several limitations that must be acknowledged:

- The observational design of the study can only establish a strong correlation between the availability of ChatGPT and the observed changes in user behavior. It cannot definitively prove causation. Other confounding factors, such as the natural maturation of the online communities or broader shifts in the tech industry, could also have contributed to the observed trends.
- The study infers the influence of AI from macro-level changes in platform data. The study does not use any methods to identify specific questions that AI generated or assisted with. The analysis relies on the assumption that the widespread availability of ChatGPT is the primary driver of the observed changes in the post-treatment period.

- The findings are based on five specific, technically oriented Stack Exchange communities. While this offers a deep understanding of these particular environments, the results may not be generalizable to other types of online communities (e.g., social media platforms, less formal forums) with different norms and user bases.
- The metrics used for complexity and engagement are proxies for the underlying constructs. While they are well-established in literature, they are imperfect measures. LD, for instance, assesses the proportion of content words (information load) but not syntactic integration, discourse coherence, argument quality, or conceptual depth.
- 42,483 unanswered questions were removed during preprocessing to ensure complete cases for response-time analyses and focus on observable engagement. This choice can induce selection bias toward answered questions, understate disengagement, and overstate platform responsiveness. Since nonresponses are unlikely to be missing at random, this approach also depends on participation, which may obscure temporal shifts that would be evident if nonresponses were modeled. Unanswered questions are themselves theoretically informative: a rising share may signal strain in the epistemic infrastructure (e.g., answerer supply, attention allocation, problem tractability) of a distributed cognitive system. Future work should model nonresponse directly (e.g., censored event-time approaches or two-stage selection models) to estimate system-level responsiveness.

These limitations highlight valuable directions for future research, including longitudinal studies that track individual users, the integration of AI-detection models, and qualitative studies (such as surveys or interviews) to understand user motivations directly.

Chapter 4: Findings & Results from Analysis of AI's Impact on Question Complexity in Online Learning Communities

Overview

The rapid emergence of AI chat tools, most notably large language models like ChatGPT, in educational and collaborative environments has produced fundamental shifts in patterns of inquiry, learning, and public knowledge-sharing. To comprehensively understand these shifts, this section presents an integrated analysis of linguistic complexity, engagement metrics, and distributed cognition within selected Stack Exchange communities. Drawing on both established and emerging literature, these findings reveal the relationship between cognitive offloading, community structure, and evolving norms of digital learning.

Descriptive Results of Engagement and User Patterns

Descriptive statistics were calculated using the final dataset (N= 97,513) to summarize key variables (i.e., reading level, technical content, word count, number of answers, upvotes, time to first response, topic tags, and accepted answer rate) across different time periods (before and after ChatGPT's release) and academic communities. Means and standard deviations for each variable were analyzed by time-period, community, and their combination, providing an overview of question difficulty, technical content, community engagement, and answer success before and after November 2022.

Table 6 summarizes results for the pre-ChatGPT period (n = 53,090) and the post-ChatGPT period (n = 44,423) participation volume (answered questions). Even after restricting analysis to answered questions, the volume of initiating posts declined by 16.3% from pre-ChatGPT (n = 53,090) to post-ChatGPT (n = 44,423). This pattern mirrors the raw

pre/post counts of questions asked and is driven primarily by declines in Computer Science, Data Science, and Mathematics, alongside modest increases in Academia and Artificial Intelligence (Table 5). The average total word count rose from 157.26 (*SD* = 112.42) before ChatGPT to 167.20 (*SD* = 118.42) after its release, showing that users wrote longer, more detailed questions in the later period.

One clear trend is an increase in question complexity after the introduction of ChatGPT. The average Flesch-Kincaid Grade Level rose from 10.31 (*SD* = 4.67) before ChatGPT to 10.98 (*SD* = 5.21) afterwards, a difference of about two-thirds of a grade level. This shift suggests a modest movement toward more complex language in questions during the later period. In contrast, LD was virtually identical across time, with both periods exhibiting an average lexical density of approximately 0.48 (post: M = 0.48, SD = 0.06; pre: M = 0.48, SD = 0.06). With such similar scores and small standard deviations, it is possible these numbers reflect limits in how the measure was applied.

Table 6

Descriptive Statistics by Time Period

Variable	Pre-ChatGPT *M* (*SD*)	Post-ChatGPT *M* (*SD*)
Flesch-Kincaid Grade	10.31 (4.67)	10.98 (5.21)
Lexical Density	0.48 (0.06)	0.48 (0.06)
Word Count	157.26 (112.42)	167.20 (118.42)
Number of Responses	1.39 (0.75)	1.41 (0.78)
Number of Upvotes	1.45 (2.17)	1.52 (2.27)
Time to First Response (s)	9515.90 (56514.90)	8948.06 (52729.28)
Number of Tags	2.79 (1.25)	2.81 (1.25)
Accepted Answer (%)	63%	62%

Note. N = 97,513 (Pre: 53,090; Post: 44,423).

Looking at engagement, the number of responses and upvotes did not change much between periods. Users received about 1.39 answers per question before ChatGPT (*SD* = 0.75) and 1.41 after (*SD* = 0.78). Upvotes were also fairly stable, averaging 1.45 (*SD* = 2.17) before and 1.52 (*SD* = 2.27) after ChatGPT. Notably, the average time to first response fell by about nine minutes after ChatGPT launched, from 9,515 seconds (*SD* = 56,514.90) in the earlier period to 8,948 seconds (*SD* = 52,729) post-ChatGPT. The number of topic tags per question hardly changed, remaining near 2.8 in both periods.

Finally, the rate of accepted answers remained largely stable, shifting only marginally from 63% before ChatGPT's release to 62% afterwards. This minor difference suggests no substantial change in question resolution success following the introduction of ChatGPT.

Table 7 displays descriptive statistics by the five Stack Exchange communities, revealing substantial heterogeneity across platforms. The mathematics community comprised the vast majority of the sample (*n* = 86,456, 88.68%), with considerably smaller contributions from Academia (*n* = 3,768, 3.86%), Data Science (*n* = 2,984, 3.06%), Computer Science (*n* = 2,843, 2.92%), and Artificial Intelligence (*n* = 1,462, 1.50%). This unequal distribution reflects genuine differences in community size and activity levels on Stack Exchange.

Question complexity, reported by the Flesch-K dimension, showed notable variation across the five communities studied. Mathematics questions proved most complex linguistically (*M* = 10.71, *SD* = 5.12), while Data Science questions used simpler language (*M* = 9.42, *SD* = 3.78). Academia (*M* = 10.01, *SD* = 2.77), Artificial Intelligence (*M* = 9.83, *SD* = 3.38), and Computer Science (*M* = 10.22, *SD* = 4.32) occupied the middle ground.

Interestingly, Academia demonstrated the most consistent complexity levels, reflected in its smallest standard deviation. However, Lexical Density showed very little variation across each community (0.48 to 0.54).

When examining question length, Academia questions averaged the longest at 182.69 words (*SD* = 130.47), while Mathematics questions were shortest (*M* = 160.03, *SD* = 112.67) despite their linguistic complexity. This suggests mathematicians favor concise notation over lengthy explanations.

Table 7

Descriptive Statistics by Community

Variable	Academia (*n*=3768)	AI (*n*=1462)	CS (*n*=2843)	DS (*n*=2984)	Math (*n*=86456)
Flesch-Kincaid Grade	10.01	9.83	10.22	9.42	10.71
	(2.77)	(3.38)	(4.32)	(3.78)	(5.12)
Lexical Density	0.49	0.53	0.50	0.54	0.48
	(0.05)	(0.06)	(0.06)	(0.07)	(0.06)
Word Count	182.69	170.80	170.36	173.72	161.46
	(130.47)	(131.16)	(127.06)	(143.01)	(112.67)
Responses	2.24	1.31	1.37	1.26	1.37
	(1.37)	(0.63)	(0.73)	(0.57)	(0.71)
Upvotes	3.83	1.28	1.03	0.85	1.42
	(4.70)	(2.03)	(1.80)	(1.42)	(2.02)
Time to Response(s)	4869	16692	12106	19029	8893
	(40827)	(63191)	(62919)	(73652)	(54103)
Tags	2.87	3.26	2.68	3.20	2.78
	(1.21)	(1.27)	(1.25)	(1.31)	(1.25)
Accepted Answer(%)	40%	45%	53%	38%	65%

Note. AI = Artificial Intelligence; CS = Computer Science; DS = Data Science; Math = Mathematics. Values aggregate across pre- and post-periods within each community. Standard deviations in parentheses.

Engagement patterns differed substantially between communities. Academia led in both responses (M = 2.24, SD = 1.37) and upvotes (M = 3.83, SD = 4.70), while Data Science lagged behind with the fewest upvotes (M = 0.85, SD = 1.42). Response times also varied considerably; Academia questions received the quickest first response at roughly 1.4 hours (M = 4,869.43 seconds, SD = 40,827.30), whereas Data Science questions waited approximately 5.3 hours (M = 19,028.57 seconds, SD = 73,651.51).

Artificial Intelligence and Data Science employed more category tags per question (M = 3.26 and 3.20 respectively), potentially reflecting their interdisciplinary nature. Finally, answer acceptance rates ranged from 65% in Mathematics to just 36% in Data Science, possibly indicating different community norms or question difficulty levels across disciplines.

Community-Specific Time Period Patterns of Engagement

Table 8 presents the community by time-period interaction, revealing how each community's characteristics changed from the pre-ChatGPT to post-ChatGPT period.

Table 8

Descriptive Statistics by Community and Time Period

	Flesch-K *M* (*SD*)	LD *M* (*SD*)	Word *M* (*SD*)	Resp *M* (*SD*)	Upvt *M* (*SD*)	Time(s) *M* (*SD*)	Tags *M*(*SD*)	Acc%
Academia								
Pre	9.80(2.67)	.49(.05)	178(129)	2.26(1.38)	3.97(4.78)	5209(43552)	2.85(1.21)	42
Post	10.21(2.85)	.49(.05)	187(132)	2.21(1.37)	3.70(4.61)	4548(38079)	2.89(1.21)	38
AI								
Pre	9.94(3.30)	.53(.06))	175(130)	1.36(0.69)	1.44(2.29)	16882(64450)	3.07(1.32)	43
Post	9.72(3.44)	.531(.06))	167(123)	1.25(0.58)	1.11(1.71)	16214(61857)	3.47(1.18)	47
CS								
Pre	10.00(4.21)	.50(.06)	171(124)	1.36(0.72)	1.09(1.74)	12106(58065)	2.68(1.24)	53
Post	10.52(4.45)	.50(.06)	171(132)	1.38(0.75)	0.93(1.86)	16214(68658)	2.68(1.25)	49

DS								
Pre	9.25(3.72)	.54(.07)	173(142)	1.27(0.57)	0.82(1.37)	20906(77190)	3.18(1.31)	36
Post	9.67(3.87)	.54(.06)	175(144)	1.26(0.58)	0.76(1.49)	16188(67193)	3.24(1.31)	36
Math								
Pre	10.35(5.21)	.48(.06)	156(112)	1.37(0.72)	1.42(2.04)	8852(54625)	2.77(1.25)	66
Post	10.98(5.20)	.48(.06)	168(113)	1.38(0.71)	1.43(1.99)	9036(53564)	2.80(1.24)	63

Note. N = 97,513. Pre = Pre-ChatGPT; Post = Post-ChatGPT. AI = Artificial Intelligence; CS = Computer Science; DS = Data Science; Math = Mathematics. Flesch-K = Flesch-Kincaid Grade Level; LD = Lexical Density; Resp = Responses; Upvt = Upvotes; Acc% = Accepted Answer percentage.

In Academia, Flesch-Kincaid Grade Level rose from 9.80 (*SD* = 2.67) to 10.21 (*SD* = 2.85) post-ChatGPT, while Word Count increased from 178.18 to 186.96. Upvotes declined slightly (3.97 to 3.70), response times improved by 11 minutes, and Accepted Answer Rate dropped from 42% to 38%. The Artificial Intelligence community showed increased complexity (Flesch-Kincaid: 9.75 to 9.91) and longer questions (165.71 to 175.17 words). Upvotes increased from 1.11 (*SD* = 1.71) to 1.44 (*SD* = 2.29), response times grew by 10 minutes, tagged categories decreased (3.47 to 3.07), and acceptance rates fell from 47% to 43%. Computer Science experienced increased complexity (10.00 to 10.52) with stable word counts. Upvotes dropped (1.10 to 0.93), response times increased dramatically by 90 minutes, and Accepted Answer Rate declined 7 percentage points, from 56% to 49%, representing the largest decrease across communities. Data Science showed increased complexity (9.25 to 9.67) and length (172.70 to 175.24 words). Upvotes declined (0.92 to 0.76), response times improved by 79 minutes, and acceptance rates remained stable at 36%, the lowest across all communities. Mathematics, comprising 88.68% of the dataset, exhibited the largest complexity increase (10.42 to 11.17) and longer questions (155.28 to 165.73

words). Response times improved by 9 minutes, and acceptance rates remained perfectly stable at 65% in both periods. Given Mathematics' dominance in the dataset, these patterns largely drive the overall temporal trends observed in Table 7.

Bivariate Correlations

Table 9 displays the Pearson correlation matrix for all eight variables. The overall pattern reveals predominantly weak correlations, with the majority of relationships falling below $|r| = .20$, suggesting that the variables measure relatively distinct constructs with limited bivariate overlap. This finding is methodologically favorable for subsequent multivariate analyses, as it minimizes multicollinearity concerns that could compromise regression model stability.

Table 9

Pearson Correlations Among Question Characteristics

Variable	1	2	3	4	5	6	7	8
1. Flesch-Kincaid	—							
2. LD	−.07**	—						
3. Word Count	.00	.07**	—					
4. Responses	−.01	−.08**	−.01	—				
5. Upvotes	.00	.02**	.17**	.35**	—			
6. Time to Resp	.00	.05**	.05**	−.06**	.06**	—		
7. Tags	.00	.12**	.18**	−.01	.13**	.03**	—	
8. Accepted Ans	.01	−.01	.02	.00	.06**	−.10**	.03**	—

Note. N = 97,513. Correlations ≥ |.02| are significant at p < .001 given large sample size.

LD = Lexical Density

***Significant at p < .001.*

Several overarching patterns emerged from the correlation analysis. The predominance of weak associations ($|r| < .20$ for 27 of the 28 bivariate relationships) indicates that the variables capture largely distinct dimensions of question characteristics and community engagement. This independence is advantageous for subsequent regression analyses because it reduces the risk of multicollinearity.

The largest association was observed between the number of responses and the number of upvotes ($r = .35$). This relationship suggests that questions attracting more replies also tend to receive more positive evaluations from the community. However, the shared variance between these variables remains modest, with approximately 88% of the variance unexplained, indicating that they measure related but distinct aspects of engagement.

Question resolution success, measured by the accepted answer rate, exhibited weak or negligible associations with question complexity, length, engagement, and topical breadth. Its only notable relationship was with response timing ($r = -.10$). This pattern implies that factors not represented in the quantitative metrics may exert greater influence on whether a question receives an accepted answer.

Finally, the two complexity measures (Flesch-Kincaid Grade Level and Lexical Density) showed minimal relationships with engagement and resolution outcomes. These results suggest that question difficulty or information density do not substantially affect community responsiveness or resolution success at the bivariate level. Collectively, these findings establish baseline expectations for the relationships among variables and confirm that the constructs are sufficiently distinct to warrant separate analytical consideration in the subsequent multivariate analyses.

MANOVA Analysis: Impact of AI Chat Tools on Question Complexity

MANOVA was selected over separate univariate analyses because the two complexity measures, while distinct, jointly characterize question sophistication and may share common variance. Examining them with a multivariate provides a more comprehensive and statistically powerful test of temporal differences while controlling for Type I error inflation that would occur with multiple independent tests (Tabachnick & Fidell, 2019). The multivariate omnibus test employed Wilks' Λ as the primary test statistic, supplemented by Pillai's Trace for robustness. Given the large sample size, both statistics were expected to yield similar conclusions. Univariate follow-up tests were used to break down significant multivariate effects and find out which specific complexity dimensions changed after ChatGPT was added.

The one-way MANOVA yielded a statistically significant multivariate effect of Time Period (pre-ChatGPT vs. post-ChatGPT) on the combined complexity measures, as presented in Table 10. Wilks' $\Lambda = 0.920$, $F(2, 97{,}510) = 4256.40$, $p < .001$, partial $\eta^2 = 0.074$. Although the effect size is small by conventional benchmarks (Cohen, 1988), this result indicates that, considered together, the two complexity indices differed reliably before versus after the release of ChatGPT.

Table 10

MANOVA: Effect of Time Period on Question Complexity

Statistic	Value	F	df_1	df_2	η^2_p
Wilks' Λ	0.920	4256.40***	2	97510	.074
Pillai's Trace	0.080	4256.44***	2	97510	.074

Note. N = 97,513. Multivariate tests of Time Period (Pre vs. Post-ChatGPT) on combined complexity measures (Flesch-Kincaid Grade Level, Lexical Density).

****p < .001.*

This significant multivariate effect provides robust empirical evidence that ChatGPT's introduction altered the complexity profile of questions posed in Stack Exchange communities. The effect size magnitude (7.4% of variance explained) is particularly noteworthy given that the analysis compares aggregate temporal periods rather than individual-level interventions, suggesting that the technological shift represented a substantial perturbation to collaborative learning dynamics across these academic communities.

Univariate Follow-Up Tests: Decomposing the Multivariate Effect

To interpret the nature of the significant multivariate temporal effect, univariate follow-up tests were conducted for each complexity measure, with results presented in Table 11. For Flesch-Kincaid Grade Level, the pre-ChatGPT period (M = 10.31, SD = 4.67) differed significantly from the post-ChatGPT period (M = 10.98, SD = 5.21), $F(1, 97{,}511)$ = 242.00, $p < .001$, Cohen's d = 0.10. While statistically significant, the effect size falls below the conventional threshold for a 'small' effect (0.20), classifying this as a negligible practical effect according to Cohen's (1988) taxonomy.

Table 11

Univariate Follow-Up Tests: Effect of Time Period

Dependent Variable	M_{pre}	M_{post}	F	η^2_p
Flesch-Kincaid Grade Level	10.31 (4.67)	10.98 (5.21)	242.00***	.10
Lexical Density	0.48 (0.06)	0.48 (0.06)	0.01	<.0001

Note. N = 97,513. Standard deviations in parentheses. Univariate ANOVAs decompose the significant multivariate effect (Table 10).

****p < .001.*

In contrast, Lexical Density did not differ meaningfully across periods, $F(1, 97{,}511) = 0.01$, $p = .93$, partial $\eta^2 < .0001$, Cohen's $d = 0.00$. The proportion of content words remained effectively constant from the pre-ChatGPT to the post-ChatGPT period, suggesting that the informational "density" of questions, in terms of content versus function words, was stable over time even as Flesch-K rose modestly.

The divergent effect sizes between Flesch-Kincaid Grade Level ($d = 0.10$, negligible) and Lexical Density ($d = 0.00$) indicate that the significant multivariate effect documented in the omnibus MANOVA is primarily driven by changes in technical specificity rather than by changes in syntactic density. While questions became marginally more technically complex, they simultaneously maintained the same information density (i.e., the proportion of vocabulary used between content words and function words was the same). This pattern suggests a qualitative transformation in question formulation practices: users appeared to construct more elaborate, verbose questions that sacrifice technical precision for accessibility or comprehensiveness.

Interaction Effects: Temporal Changes Across Communities

Separate two-way analyses of variance (ANOVAs) were conducted for Flesch–Kincaid Grade Level (Flesch-K) and Lexical Density, with Time Period (pre-ChatGPT vs. post-ChatGPT) and Community (Academia, Artificial Intelligence, Computer Science, Data Science, Mathematics) entered as fixed factors. These models tested whether (a) question complexity differed across communities, (b) complexity changed from the pre- to the post-ChatGPT period, and (c) any temporal changes were moderated by community.

For Flesch-K, the two-way ANOVA revealed a significant main effect of Time Period, $F(1, 97{,}503) = 241.06$, $p < .001$, indicating that questions asked after ChatGPT's introduction were written at slightly higher estimated grade levels than those asked before, averaged across communities. There was also a significant main effect of Community, $F(4, 97{,}503) = 52.17$, $p < .001$, reflecting systematic differences in linguistic complexity across the five Stack Exchange sites. Communities such as Mathematics tended to show higher Flesch-K values than communities such as Data Science or Artificial Intelligence, consistent with their differing disciplinary conventions. The Time Period × Community interaction for Flesch-K was not significant, $F(4, 97{,}503) = 0.53$, $p = .71$, suggesting that the modest increase in Flesch-K from pre- to post-ChatGPT was similar across communities rather than being driven by a specific site discipline (Table 12).

Table 12

Two-Way ANOVA for Flesch-Kincaid Grade Level by Time Period and Community

Flesch-K Effect	***Sum of Squares***	***df***	***F***
Time period	5,731.25	1	241.06***

Community	4,961.89	4	52.17***
Time Period x Community	50.51	4	0.53
Residual	2,318,178.27	97,503	

Note. N = 97,513. Dependent variable is Flesch–Kincaid Grade Level (Flesch-K). Time period contrasts pre- vs. post-ChatGPT. Community refers to Stack Exchange sites (Academia, Artificial Intelligence, Computer Science, Data Science, Mathematics).

****$p < .001$*

For Lexical Density (LD), the two-way ANOVA showed no main effect of Time Period, $F(1, 97{,}503) = 0.36$, $p = .55$. Lexical Density was essentially unchanged before versus after ChatGPT's release when averaged across communities. In contrast, there was a very strong main effect of Community on LD, $F(4, 97{,}503) = 1{,}055.66$, $p < .001$, indicating that communities differed markedly in how densely they packed content words into questions. Some communities consistently produced lexically denser questions than others, reflecting stable disciplinary "signatures" in how information is packaged linguistically. The Time Period × Community interaction for LD was not significant, $F(4, 97{,}503) = 1.31$, $p = .26$, indicating that these community-specific LD profiles were highly stable over time; the availability of ChatGPT did not materially alter the relative lexical density patterns across sites (Table 13).

Taken together, the two-way ANOVA results indicate that the small temporal increase in Flesch-K is a global pattern shared across communities, whereas LD is primarily shaped by community membership and remains remarkably stable across the pre- and post-ChatGPT periods. The introduction of ChatGPT is associated with a slight overall rise in

estimated grade level across all communities, but it does not appreciably change the proportion of content words used within any given community.

Table 13

Two-Way ANOVA for Lexical Density Level by Time Period and Community

LD Effect	***Sum of Squares***	***df***	***F***
Time period	0.00137	1	0.36
Community	16.06	4	1055.66***
Time Period x Community	0.02	4	1.31
Residual	370.92	97503	

Note. N = 97,513. Dependent variable is lexical density (LD), computed as the proportion of content words (nouns, proper nouns, main verbs, adjectives, and adverbs) among all tokens. Time period contrasts pre- vs. post-ChatGPT. Community refers to Stack Exchange sites (Academia, Artificial Intelligence, Computer Science, Data Science, Mathematics).
***p < .001.

Answer to the Research Question

After ChatGPT's public release, question complexity on Stack Exchange showed statistically reliable but modest changes. A multivariate analysis of variance indicated that time period (pre-ChatGPT vs. post-ChatGPT) had a significant effect on the combined set of complexity indices (Flesch-Kincaid Grade Level and Lexical Density), Wilks' $\Lambda = 0.998$, $F(2, 97{,}510) = 121.00$, $p < .001$, partial $\eta^2 = .074$. Follow-up univariate tests showed that Flesch-Kincaid Grade Level was slightly higher in the post-ChatGPT period ($M = 10.71$, $SD = 5.05$) than in the pre-ChatGPT period ($M = 10.22$, $SD = 4.74$), $F(1, 97{,}511) = 242.00$, $p <$

.001, partial η^2 = .003, Cohen's d = 0.10, indicating a small but reliable increase in estimated reading grade level. In contrast, Lexical Density remained virtually unchanged across time (pre: M = 0.48, SD = 0.06; post: M = 0.48, SD = 0.06), F(1, 97,511) = 0.01, p = .93, partial η^2 < .0001, Cohen's d = 0.00.

Two-way ANOVAs further revealed that these patterns were consistent across communities. For Flesch-Kincaid Grade Level, there were significant main effects of time period, F(1, 97,503) = 241.06, p < .001, and community, F(4, 97,503) = 52.17, p < .001, but no Time Period × Community interaction, F(4, 97,503) = 0.53, p = .71, indicating that the small increase in grade level occurred in all five communities rather than being driven by a single site. For Lexical Density, the main effect of time period was nonsignificant, F(1, 97,503) = 0.36, p = .55, whereas the main effect of community was large, F(4, 97,503) = 1,055.66, p < .001, with no significant interaction, F(4, 97,503) = 1.31, p = .26. Thus, lexical density appeared to be a stable, community-specific property that did not change meaningfully following ChatGPT's introduction.

Taken together, these findings provide empirical evidence that the launch of ChatGPT coincided with a small, broadly distributed increase in the syntactic and lexical difficulty of questions (as indexed by Flesch-Kincaid Grade Level), while the informational density of questions (as indexed by Lexical Density) remained remarkably stable and primarily shaped by community norms. From a Distributed Cognition perspective (Hutchins, 1995), this transformation reflects a reorganization of cognitive labor between human questioners and anticipated AI assistance.

The aggregated temporal effects reveal that this impact varies substantially across disciplinary communities. The Mathematics community's pronounced shift (0.49 grade level

Flesch-Kincaid increase) contrasts sharply with Academia's stability and Computer Science's modest increases. This heterogeneity indicates that distributed cognitive systems, comprising human participants, technological tools, and community norms, responded differentially to the availability of AI based on existing epistemic cultures, consistent with situated cognition perspectives (Lave & Wenger, 1991). The findings suggest that ChatGPT's integration into collaborative learning environments does not follow a one-size-fits-all pattern but rather interacts with disciplinary norms and community-specific practices to produce contextually variable outcomes.

Chapter 5: Interpretation, Implications, and Future Directions

This chapter synthesizes the empirical results presented in the preceding chapters and interprets them through the conceptual lens of distributed cognition. The primary objective of this study was to assess how the emergence of artificial intelligence (AI) chat tools, particularly ChatGPT, influenced the complexity of public questions and engagement patterns within online learning communities. Through a quasi-experimental, pre-test/post-test design framework that combined large-scale quantitative analysis with theory-informed interpretation, this dissertation contributes to a growing body of work concerned with the cognitive, social, and infrastructural consequences of generative AI technologies.

To guide this inquiry, three research questions were posed. First, does the introduction of ChatGPT correspond with a measurable shift in the linguistic complexity of user-generated questions in technical forums such as Stack Exchange? Second, are there observable changes in engagement behaviors, including the number of answers, acceptance rates, and user response patterns? Third, how might these changes be explained through the framework of distributed cognition, particularly with respect to the evolving role of cognitive artifacts in online knowledge systems?

This chapter is organized into four sections. The first provides an interpretation of the key findings in light of the research questions. The second engages these results through the conceptual apparatus of distributed cognition theory. The third addresses the broader implications of these findings for education, digital platform governance, and the future of public knowledge. The final section outlines key limitations of the study and proposes directions for future research.

Interpretation of Key Findings

The results presented in Chapters 3 and 4 offer partial yet meaningful answers to the research questions. Concerning the first question, the analysis demonstrated a statistically significant increase in question complexity following the release of ChatGPT. This change was measured using the Flesch-Kincaid Grade Level, which rose from an average of 10.31 in the pre-ChatGPT period to 10.98 in the post-release phase. While the magnitude of this shift was modest (Cohen's d between 0.2 and 0.4), the trend was consistent across multiple communities, including Computer Science, Mathematics, and Artificial Intelligence.

These results suggest the presence of a "filtering effect," in which simpler or lower-effort questions are increasingly resolved through private interactions with AI, leaving more complex or ambiguous problems to be posed publicly. Users appear to offload basic information retrieval to AI tools, while continuing to rely on community platforms for tasks requiring context-specific judgment, subjective interpretation, or domain-specific expertise. Although the increase in complexity is not drastic, it may indicate the early stages of a broader cognitive reconfiguration in digital learning spaces.

In response to the second research question, user engagement metrics presented a mixed picture. The average number of answers per question increased slightly (from 1.39 to 1.41), while the acceptance rate of answers declined marginally (from 63% to 62%). More significantly, some communities, particularly Data Science and Artificial Intelligence, exhibited sharper reduction in response quality as measured by upvotes and accepted answers. These trends suggest that while platform participation has not collapsed, it may be evolving into a more specialized or uneven landscape.

These findings point toward the emergence of what might be termed the "super-user effect." As novice users rely more heavily on AI tools for entry-level questions, advanced contributors increasingly dominate high-complexity threads. This shift may stratify knowledge production within the community, reinforcing existing hierarchies of expertise and reducing the availability of accessible entry points for new participants. Recent sentiment analyses suggest users are ambivalent about this shift (Sun et al., 2025; Shoufan, 2023), identifying risks to the skill gap and intellectual diversity (Riedl & Bogert, 2024). It also raises concerns about sustainability, as fewer novice users engage in learning through observation and peripheral participation.

One unanticipated result involved the analysis of Lexical Density scores. These were employed to assess changes in domain-specific information load. However, the analysis revealed no significant differences between the pre- and post-ChatGPT periods. This null result implies that, although readability complexity difficulty increased, it did not correspond with an increase in content density. One plausible interpretation is that the observed complexity is more rhetorical than conceptual. Users may be adopting a more formal tone or producing longer sentences without necessarily introducing more advanced or structurally different subject-matter content. Alternatively, the limitation may lie in the Lexical Density metric itself, which is known to struggle with detecting semantic depth or conceptual abstraction. Future studies might employ more advanced semantic models to better capture cognitive complexity in discourse.

Analyzing the Decline in Participation

Beyond shifts in question complexity, the most direct behavioral indicator of change is the decline in question posting volume across three of the five communities. From a

distributed cognition perspective, fewer initiating questions implies fewer public ‘distributed cognitive events and fewer durable knowledge artifacts produced through community interaction. The declines are largest in domains where LLMs plausibly substitute for routine troubleshooting and procedural explanation (Computer Science, Data Science, and Mathematics), a pattern consistent with cognitive offloading and the privatization of epistemic activity described in Chapter 1 (Burtch et al., 2024).

Consequently, the stability of engagement metrics per question (answers each question received remaining ~1.4) must be interpreted with caution. While the community effectively services the questions that are asked, the community itself is shrinking. We are observing a "survivorship bias" in the public square: the only questions remaining are those that users felt could not be satisfactorily answered by an AI, or those asked by users deeply committed to the social norms of the platform.

The 16% decline in question volume observed in this study is not an isolated phenomenon but part of a broader "exodus" from public technical forums documented across the web. Industry reports indicate that Stack Overflow's traffic dropped by approximately 14% in March 2023 alone, shortly after the release of GPT-4. By 2024, some estimates suggested a year-over-year question volume decrease of nearly 78% in certain sectors, effectively erasing fifteen years of growth (Byteiota, 2025).

This aligns with findings from the 2023 Stack Overflow Developer Survey, where 83% of respondents reported using ChatGPT, often prioritizing the "instant" nature of AI responses over the "verified" nature of community answers. As noted by researchers, we are witnessing a "Great Decoupling" where developers still seek answers but no longer visit the sites that host them (Stack Overflow, 2023).

While industry analysts focus on this decline primarily as a loss of web traffic or ad revenue, this dissertation argues that the primary loss is cognitive. Previous analyses have largely treated this drop as a simple substitution effect: users are swapping one tool (forums) for another (LLMs) (Burtch et al., 2024). However, my analysis of the *remaining* questions reveals that this substitution is not random. The "simple" questions, those that previously allowed novices to practice articulating their problems and receive social corrections, have vanished.

What remains is a "skewed commons" where entry-level scaffolding is missing. Novices are no longer asking the "stupid" questions that lead to community mentorship. The "Knowledge Library" is stagnant. As users stop asking questions, the public repository of knowledge stops growing. AI models become much harder to train as the availability of human-verified data becomes scarcer. Social friction is removed. The "friction" of asking a question on Stack Overflow, which often requires formatting it correctly and checking for duplication in prior submissions, was arguably a learning mechanism itself. By bypassing this, users may be gaining speed but losing the metacognitive benefits of formulating a clear technical inquiry.

Theoretical Framing Through Distributed Cognition

Distributed cognition provides a robust framework for interpreting these platform-level changes. As theorized by Hutchins (1995), cognitive processes are not confined to individual minds but are extended across artifacts, environments, and social systems. In this view, tools such as ChatGPT are not external aids to cognition but integral components of a distributed cognitive system. They mediate, transform, and reallocate cognitive labor, thereby reshaping the structure of collective inquiry.

From this perspective, ChatGPT functions as a novel cognitive artifact. It is generative, conversational, and context-sensitive, distinguishing it from static resources such as documentation or search engines. The emergence of this tool introduces a shift in cognitive geography, redistributing tasks that were once conducted in public, collaborative spaces into private, AI-mediated exchanges. This is particularly evident in the offloading of low-level problem-solving, such as code debugging or definitional queries, which are now frequently resolved without the need for community engagement.

This reallocation of effort aligns with the offloading mechanism described in Martin's (2012) CTOM model, which outlines how individuals delegate cognitive tasks to external systems. However, it also raises new theoretical concerns. Unlike traditional cognitive artifacts that merely store or retrieve information, generative AI tools produce new knowledge representations, often without clear attribution or transparency. This complicates the monitoring and translation functions of distributed cognition, where community members traditionally evaluate, validate, and refine knowledge through social interaction.

One particularly salient concern is the privatization of epistemic activity. In traditional distributed systems, such as online Q&A forums, knowledge is produced publicly and archived for future access. By contrast, AI interactions are ephemeral, individualized, and non-replicable. The shift from many-to-many to one-to-one interaction models represents not only a functional change but a structural reconfiguration of how collective cognition is instantiated. This may weaken the community's role as a site of durable knowledge construction and social learning.

Cognitive and Neural Effects of AI Chat Tools: Synthesis of Empirical and Theoretical Evidence

Building on the descriptive and inferential analysis presented above, it is essential to situate these results within the broader scholarly context and explore their implications for cognitive engagement and learning outcomes.

Recent studies confirm that integrating AI chat tools into collaborative learning environments produces measurable changes in both cognitive processes and neural activity. In a controlled experiment, Kosmyna et al. (2025) assigned participants to complete essay writing tasks using either AI assistance, web search tools, or no external support. Across conditions, participants who used a large language model showed significant reductions in neural connectivity, particularly in the alpha, beta, theta, and delta frequency bands. Independent work fosters deeper cognitive involvement and creativity, whereas reliance on AI/web tools favors speed but limits depth and originality. These frequencies are associated with core cognitive functions, including strategic planning, creative thinking, memory formation, and sustained attention (Fan et al., 2025; Lee et al., 2025).

The same study showed that participants who started with unaided writing and then used ChatGPT performed better than those who used AI from the beginning. Initial independent effort appeared to prime the brain for more effective use of AI, while early reliance reduced neural engagement and weakened subsequent performance. However, distinct ethical and ontological questions remain regarding whether LLMs effectively "understand" content (Mitchell & Krakauer, 2023) or can be considered authors (Thorp, 2023)

Evidence from literature reviews supports these experimental findings. While AI-supported learning can increase efficiency and short-term performance, especially for fact-based or procedural tasks, repeated reliance on external tools may impair the development of creative thinking and metacognitive skills (Treve, 2024; Zawacki-Richter et al., 2019). Studies report that learners who regularly use AI tools often feel less ownership over their work and struggle to recall AI-generated content accurately. These outcomes raise concerns about the long-term impact of cognitive offloading in educational contexts (Fan et al., 2025).

Linguistic analysis provides further insight into these cognitive effects. Essays written with AI support tend to follow more uniform patterns in sentence structure and vocabulary. This suggests a tendency to reproduce output formats favored by the underlying language model (Niloy et al., 2024). In contrast, essays written without external tools exhibit more variation, reflective reasoning, and emotional nuance. These characteristics reflect deeper engagement with content and greater expressive range. The observed reduction in originality and diversity among AI-assisted texts supports concerns about potential constraints on creativity and intellectual autonomy (Herbold et al., 2023).

Taken together, findings from both empirical studies and theoretical literature indicate a clear tension. AI chat tools offer significant support for learning and productivity, but excessive dependence may limit cognitive depth, originality, and sustained intellectual growth. Educators and designers of learning technologies must manage this trade-off carefully. Intentional strategies are needed to preserve opportunities for independent inquiry, critical reflection, and long-term skill development. These findings contribute to the broader discussion on distributed cognition and highlight the importance of continued empirical

evaluation as AI becomes increasingly embedded in the infrastructure of collaborative and digital learning environments.

Broader Implications

The findings of this study carry implications for several key domains, including education, platform governance, and public policy. In educational contexts, the integration of AI into student workflows demands a rethinking of instructional goals and assessment practices. Learners must develop skills in interacting with AI responsibly and critically, including the ability to evaluate the accuracy, bias, and relevance of AI-generated responses. Educational institutions should emphasize the development of "AI literacy" as a foundational component of digital competence.

A large reservoir of unanswered queries (~30%) is not a data artifact but a signal of digital commons health (Frischmann et al., 2014; Ostrom, 1990). Nonresponses thus function as a leading indicator of a potential tragedy of the digital commons, where individual efficiency gains can coincide with a collective shortfall of shared artifacts, constraining observational learning and newcomer onboarding (Halfaker et al., 2013; Kraut & Resnick, 2012). Accordingly, platform design and policy should treat unanswered questions as maintainable commons infrastructure (e.g., surfacing orphaned questions, incentives for first responses, and workflows that reintegrate AI-assisted drafts into public threads) (Anderson et al., 2012; Kraut & Resnick, 2012; Mamykina et al., 2011).

Curriculum designers may also consider hybrid pedagogies that incorporate AI tools while maintaining collaborative learning environments. For instance, AI could be used for preliminary exploration, followed by peer discussion or instructor-guided synthesis. Such models preserve the scaffolding and social dimension of learning that AI alone cannot

replicate. Without intentional design, there is a risk that educational spaces will become fragmented, with students relying on AI in isolation and missing opportunities for deeper engagement.

For platform designers and moderators, the results underscore the importance of preserving epistemic inclusivity. As complexity levels rise and participation narrows, novice users may find it increasingly difficult to enter community knowledge practices. Interventions such as guided onboarding, tiered question spaces, or mentorship programs may help mitigate these effects. Equally important are mechanisms to detect and incorporate AI-generated insights into public discourse, ensuring that private knowledge production does not come at the expense of communal learning.

At the policy level, the study points to the urgent need for public stewardship of digital knowledge systems. Platforms like Stack Exchange serve as critical epistemic infrastructures, supporting informal learning, professional development, and collective intelligence. As more cognitive labor shifts to proprietary, closed-source systems, the sustainability of these public knowledge commons is at risk. Policymakers should consider funding models, content archiving strategies, and transparency standards that support open knowledge ecosystems. This may include investments in open-source AI models, collaborative moderation tools, and data accessibility initiatives.

The emergence of AI also presents a challenge to the social contract of knowledge production. If users become accustomed to resolving questions privately, without contributing back to the community, the feedback loop that sustains collective intelligence may degrade. In this light, maintaining community participation is not merely a technical or managerial challenge but a normative one. It calls for a renewed commitment to knowledge

as a public good, requiring cooperation among users, developers, educators, and policymakers.

Practical Strategies for Sustaining Online Q&A Platforms in the Age of AI

While recent findings clarify how question complexity and user engagement are evolving with the integration of artificial intelligence, these insights must be translated into concrete strategies to ensure the continued relevance and effectiveness of online question-and-answer platforms and address the sustainability challenges identified in community Q&A (Srba & Bielikova, 2016). This requires managing collective attention toward novel content (Wu & Huberman, 2007).

The widespread use of AI chat tools, particularly large language models such as ChatGPT, has significant implications for the design, governance, and long-term sustainability of Q&A sites like Stack Exchange. Empirical studies indicate a growing trend where users rely on AI to resolve simple, factual questions in private settings. As a result, public Q&A forums are increasingly populated with more complex, ambiguous, or highly specialized queries (Fan et al., 2025; Kosmyna et al., 2025). This shift requires careful rethinking of platform design, user engagement policies, and community norms to support both expert and novice users.

Supporting Novice Users Through Accessible Entry Points

As AI tools increasingly handle routine inquiries, the public discourse on Q&A platforms becomes more advanced and specialized. To maintain accessibility for newer users, platforms should develop clear onboarding processes that include guided tutorials, mentorship initiatives, and question refinement tools. AI-powered systems can assist users in

phrasing meaningful and appropriately scoped questions, lowering the entry barrier and promoting balanced participation (Zawacki-Richter et al., 2019).

Encouraging Public Knowledge Contributions

Migration of simple questions to private AI interactions weakens the shared knowledge base of public platforms. Q&A sites should encourage users to share AI-assisted answers publicly, particularly when addressing ambiguous or non-standard problems. Possible measures include:

- Enabling users to contribute AI-generated solutions to public threads when relevant.
- Introducing community-driven incentives, such as bounties for complex or unanswered questions, to attract expert input (Srba & Bielikova, 2014).
- Creating metadata systems, including tags or classifications, for AI-assisted contributions to track their influence on discourse and highlight valuable collaborations.

Enhancing Collaboration Between Humans and AI

Despite advances in automation, human insight remains essential for resolving difficult and context-sensitive problems. Platforms should promote collaborative workflows that integrate both human judgment and AI-generated input. These might include co-authored answers with clearly attributed sources, or interactive formats such as debate threads where users can evaluate and respond to multiple perspectives, including those generated by AI (Hutchins, 1995).

Findings from Kosmyna et al. underscore the importance of timing and sequencing in AI tool integration. Participants who began their writing tasks with immediate access to ChatGPT demonstrated reduced neural engagement and lower creativity, along with a

tendency to copy AI-generated content rather than develop original ideas. By contrast, individuals who first engaged in unaided writing and only later incorporated ChatGPT showed higher cognitive activation and produced more creative work. Those who relied on AI initially but were later asked to work without it exhibited a decline in performance, suggesting that early cognitive effort may condition the brain to use AI more effectively, whereas premature dependence can undermine long-term outcomes.

These findings have direct implications for Q&A platform design. Rather than promoting continuous AI reliance, platforms should consider phased assistance models that encourage initial independent contribution. Prompting users to formulate their own ideas before offering AI support may enhance participation quality while sustaining the cognitive processes essential for creativity, memory, and problem solving (Gupta, 2017).

Safeguarding Quality and Transparency

As AI-generated content becomes more common, Q&A platforms must develop rigorous content oversight policies. Editors and moderators should implement clear standards for reviewing AI-assisted answers, including disclosure of AI use, provenance tracking, and mechanisms for peer validation. These steps can help maintain trust in the accuracy and authenticity of shared knowledge (Herbold et al., 2023).

Evolving Reputation and Reward Mechanisms

Current reputation systems often prioritize speed, frequency, or volume of responses. However, these metrics may overlook the value of thoughtful analysis and critical engagement. Scoring algorithms should be revised to recognize contributions that demonstrate synthesis, depth, and collaboration. Additionally, users who validate or critique

AI-generated answers should receive recognition, reinforcing practices that support careful reasoning and reflective discourse (Mamykina et al., 2011).

Promoting Continued Learning and Diverse Participation

As more technical content dominates public forums, casual learners and observers may find fewer opportunities for incremental participation. To address this, Q&A platforms should develop learning pathways tailored to users at different levels of expertise. These could include beginner-friendly walkthroughs, modular topic guides, and featured threads on core concepts. Such efforts can help sustain a diverse and engaged user base over time (Yang et al., 2020).

Limitations and Future Directions

Like all empirical research, this study has several limitations. Notably, it did not take place under controlled laboratory conditions but instead draws on the large-scale, exogenous introduction of ChatGPT as a natural experiment. This context offered a rare opportunity to examine the impact of a transformative AI tool on collaborative learning environments, enabling quasi-experimental comparisons across periods before and after its release (Fischer et al., 2004).

First, the study relies on correlational research design. Although the temporal alignment between ChatGPT's emergence and observed shifts in user behavior is striking, causal claims remain limited without longitudinal or experimental data. That said, the abrupt and widespread adoption of the tool lends credibility to the interpretation of these changes as consequences of the intervention, which is similar in logic to interrupted time series or other quasi-experimental methods (Andersson Hagiwara et al., 2016; Shadish et al., 2002).

Second, the analysis did not employ AI-detection tools capable of identifying AI-generated or AI-assisted content at the individual level. As a result, conclusions about AI's role are necessarily indirect, based on patterns in aggregate data rather than confirmed usage. While this limits precision, it reflects the constraints faced by practitioners and policymakers who also lack access to granular indicators. The observed shifts at the system level nonetheless provide meaningful evidence of structural transformation.

Third, the focus on technical communities within Stack Exchange constrains generalizability. The platform's openness and structured norms support robust analysis, but findings may not transfer directly to other online spaces, particularly those with more fluid interaction norms or broader topical scopes. Even so, Stack Exchange offers a particularly suitable setting for detecting ecosystem-level responses to technological disruptions.

Fourth, because analysis was conditioned on questions that received at least one answer (excluding closed or unfit questions; Correa & Sureka, 2013), results capture engagement among answered questions and may understate system-wide changes in responsiveness. Incorporating unanswered questions with event-time or selection models would enable inferences about platform-level attention dynamics and test whether the unanswered share changes across periods. These findings would relate directly to the vitality of the digital commons.

Future work could address these limitations in several directions, aligning with the five priorities for research on ChatGPT (Van Dis et al., 2023), potentially utilizing advanced NLP embeddings like Word2Vec (Church, 2017; Jurafsky & Martin (2019) to track semantic shifts. Longitudinal user-level studies would shed light on how engagement patterns evolve in response to AI integration. Incorporating AI-detection methods would allow for more

granular assessments of hybrid authorship and its implications for discourse and collaboration. Comparative research across platforms like Reddit, Quora, or Discord could reveal how differing affordances shape the adoption and impact of AI in peer-driven knowledge communities.

Experimental designs may also play a key role. For instance, interventions could prompt users to externalize AI-assisted insights in public discussions, exploring how private cognitive tools can be reintegrated into collective knowledge-building processes. Other studies might examine how different prompting techniques influence the depth, clarity, or creativity of user contributions.

Beyond individual behavioral metrics, future research should investigate the ecological factors governing the sustainability of online knowledge commons. As generative AI becomes a primary interface for information retrieval, the broader 'information ecology' undergoes a transformation that necessitates longitudinal study to observe potential 'ecological succession' within these communities. Such research should employ user-path analysis to determine how cross-platform information foraging, where learners move between AI-driven chats and peer-to-peer forums, alters the sequence of inquiry.

Of particular concern is the potential for 'environmental tightening' of the digital commons. If AI tools effectively siphon off lower-complexity inquiries, the resulting rise in average question difficulty may create socio-technical barriers to entry for novices. Future studies should evaluate the impact of this shift on newcomer retention and the long-term 'carrying capacity' of these forums to generate the high-quality, human-verified data upon which future AI models depend.

Framing this study as a natural experiment enhances its ecological validity and captures dynamic responses to technological change that are difficult to observe in more controlled settings. Despite its limitations, the research offers actionable insights into how distributed cognitive systems adapt in real-world conditions, or insights that are critical for understanding the evolving relationship between AI and collective intelligence.

Conclusion

This dissertation has presented a data-driven and theory-informed analysis of how generative AI tools, particularly ChatGPT, are beginning to reshape distributed cognitive systems within online learning environments. Through statistical modeling and conceptual interpretation, the study demonstrates that the public complexity of questions has increased, while participation remains unevenly distributed. These patterns signal a subtle but meaningful shift in the architecture of digital knowledge systems.

By articulating how AI tools alter cognitive labor, reshape participation structures, and challenge traditional forms of epistemic transparency, this study contributes to both the cognitive science and education technology literatures. It extends distributed cognition theory into the realm of generative AI, documents early indicators of epistemic privatization, and underscores the ethical and infrastructural stakes of AI integration.

Preserving distributed cognition as a public resource requires more than just technical adjustments, akin to employee retention strategies (Bode et al., 2015); it demands distinct policy choices regarding platform governance. As these environments become more automated, the durability of collective learning will rely on specific interventions that prioritize human participation and open inquiry over efficiency.

References

Aarset, M. V., & Johannessen, L. K. (2022). On distributed cognition while designing an AI system for adapted learning. *Frontiers in Artificial Intelligence, 5*, Article 910630. https://doi.org/10.3389/frai.2022.910630

Akasheh, F., Solomon, J. T., Hamilton, E., Nayak, C. R., & Viswanathan, V. K. (2018, June). *Application of brain-based learning principles to engineering mechanics education: Implementation and preliminary analysis of connections between employed strategies and improved student engagement* [Conference paper]. The 2018 ASEE Annual Conference & Exposition, Salt Lake City, UT, United States.

Alam, A. (2021). Should robots replace teachers? Mobilisation of AI and learning analytics in education. In *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)* (pp. 1–12). IEEE. https://doi.org/10.1109/ICAC353642.2021.9697300

Allamanis, M., & Sutton, C. (2013). Why, when, and what: Analyzing stack overflow questions by topic, type, and code. In *2013 10th Working Conference on Mining Software Repositories (MSR)* (pp. 53–56). IEEE. https://doi.org/10.1109/MSR.2013.6624004

Allison, P. D. (2001). *Missing data*. Sage Publications.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2012). Discovering value from community activity on focused question answering sites: A case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 850–858). Association for Computing Machinery. https://doi.org/10.1145/2339530.2339665

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.

Andersson Hagiwara, M., Andersson Gäre, B., & Elg, M. (2016). Interrupted time series versus statistical process control in quality improvement projects. *Journal of Nursing Care Quality, 31*(1), E1–E8. https://doi.org/10.1097/NCQ.0000000000000130

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Argote, L. (2025). Group dynamics meets organizational learning: Reflections on research. *Small Group Research, 56*(3), 524–552. https://doi.org/10.1177/10464964241263117

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The risk of automation for jobs in OECD countries: A comparative analysis* (OECD Social, Employment and Migration Working Papers No. 189). OECD Publishing. https://doi.org/10.1787/5jlz9h56dvq7-en

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Arora, P., Ganguly, D., & Jones, G. J. (2015, August). The good, the bad and their kins: Identifying questions with negative scores in stackoverflow. In *Proceedings of the 2015 ieee/acm international conference on advances in social networks analysis and mining 2015* (pp. 1232-1239).

Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.

Blandford, A., & Furniss, D. (2006). DiCoT: A methodology for applying distributed cognition to the design of teamworking systems. In S. W. Gilroy & M. D. Harrison (Eds.), *Interactive systems. Design, specification, and verification (DSV-IS 2005)* (Vol. 3941, pp. 26–38). Springer. https://doi.org/10.1007/11752707_3

Bode, C., Singh, J., & Rogan, M. (2015). Corporate social initiatives and employee retention. *Organization Science, 26*(6), 1702–1720. https://doi.org/10.1287/orsc.2015.1006

Burtch, G., Lee, D., & Chen, Z. (2024). The consequences of generative AI for online knowledge communities. *Scientific Reports*, 14(1), Article 10074.

Buse, R. P., & Weimer, W. R. (2010). Learning a metric for code readability. *IEEE Transactions on Software Engineering, 36*(4), 546–558. https://doi.org/10.1109/TSE.2009.70

Byteiota. (2025). Stack Overflow traffic collapses 75% as AI tools win. *Byteiota*.

Chen, C., & Xing, Z. (2016). Towards correlating search on Google and asking on Stack Overflow. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 83–92). IEEE. https://doi.org/10.1109/COMPSAC.2016.210

Cheng, X., & Zhang, S. (2024). Tool, teammate, superintelligence: Identification of ChatGPT-enabled collaboration patterns and their benefits and risks in mutual learning. In *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)* (pp. 158–167). https://doi.org/10.24251/HICSS.2024.020

Choi, E., Kitzie, V., & Shah, C. (2012). Developing a typology of online Q&A models and recommending the right model for each question type. *Proceedings of the Association for Information Science and Technology, 49*(1), 1–4. https://doi.org/10.1002/meet.14504901302

Chomsky, N. (1980). A review of B. F. Skinner's *Verbal Behavior.* In N. Block (Ed.), *Readings in philosophy of psychology* (Vol. 1, pp. 48–63). Harvard University Press.

Church, K. W. (2017). Word2Vec. *Natural Language Engineering, 23*(1), 155–162. https://doi.org/10.1017/S1351324916000334

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 1-46). Cambridge University Press.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315044408-14

Correa, D., & Sureka, A. (2013). Fit or unfit: Analysis and prediction of "closed questions" on Stack Overflow. In *Proceedings of the First ACM Conference on Online Social Networks (COSN '13)* (pp. 201–212). Association for Computing Machinery. https://doi.org/10.1145/2512938.2512954

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dave, P. (2023). *Stack Overflow will charge AI giants for training data.* Wired. https://www.wired.com/story/stack-overflow-will-charge-ai-giants-for-training-data/

del Rio-Chanona, M., Laurentsyeva, N., & Wachs, J. (2023). *Are large language models a threat to digital public goods? Evidence from activity on Stack Overflow*. arXiv. https://arxiv.org/abs/2307.07367

del Rio-Chanona, R. M., Laurentsyeva, N., & Wachs, J. (2024). Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus, 3*(9), 400. https://doi.org/10.1093/pnasnexus/pgae400

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics, 31*(2), 385–391. https://doi.org/10.1214/aoms/1177705900

DuBay, W. H. (2004). *The principles of readability*. Impact Information.

Elkins, A. (2021). *An evaluative pilot study of an online facilitator professional development program* [Doctoral dissertation, University of South Florida].

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Fischer, G., Arias, E., Carmien, S., Eden, H., Gorman, A., Konomi, S. I., & Sullivan, J. (2004). *Supporting collaboration and distributed cognition in context-aware pervasive computing environments* [Conference paper]. The Human Computer Interaction Consortium Winter Workshop, Winter Park, CO, United States.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233. https://doi.org/10.1037/h0057532

Frischmann, B. M., Madison, M. J., & Strandburg, K. J. (Eds.). (2014). *Governing knowledge commons*. Oxford University Press.

Goel, A. K., & Joyner, D. A. (2017). Using AI to teach AI: Lessons from an online AI class. *AI Magazine, 38*(2), 48–59. https://doi.org/10.1609/aimag.v38i2.2732

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Gray, K., & Tobin, J. (2010). Introducing an online community into a clinical education setting: A pilot study of student and staff engagement and outcomes using blended learning. *BMC Medical Education, 10*, Article 6. https://doi.org/10.1186/1472-6920-10-6

Greco, G. M., & Floridi, L. (2004). The tragedy of the digital commons. *Ethics and Information Technology, 6*(2), 73–81. https://doi.org/10.1007/s10676-004-2895-2

Gupta, R., & Reddy, P. K. (2017). Towards question improvement on knowledge sharing platforms: A Stack Overflow case study. In *2017 IEEE International Conference on Big Knowledge (ICBK)* (pp. 41–48). IEEE. https://doi.org/10.1109/ICBK.2017.25

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist, 57*(5), 664–688. https://doi.org/10.1177/0002764212469365

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Arnold.

Halverson, C. A. (2002). Activity theory and distributed cognition: Or what does CSCW need to do with theories? *Computer Supported Cooperative Work, 11*, 243–267. https://doi.org/10.1023/A:1015298005381

Hanrahan, B. V., Convertino, G., & Nelson, L. (2012). Modeling problem difficulty and expertise in Stack Overflow. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (pp. 91–94). Association for Computing Machinery. https://doi.org/10.1145/2141512.2141550

Hassan, S. A., Das, D., Iqbal, A., Bosu, A., Shahriyar, R., & Ahmed, T. (2018). SOQDE: A supervised learning based question difficulty estimation model for Stack Overflow. In *2018 25th Asia-Pacific Software Engineering Conference (APSEC)* (pp. 445–454). IEEE. https://doi.org/10.1109/APSEC.2018.00059

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). W. H. Freeman.

Herbold, S., Hautli,Janisz, A., Heuer, U., Vogel, L., Müller, R., & Brandt, M. (2023). A large-scale comparison of humans, written versus ChatGPT, generated essays. *Scientific Reports, 13*, Article 18617. https://doi.org/10.1038/s41598-023-45644-9

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs, 54*(2), 187–211. https://doi.org/10.2307/1942661

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science, 26*(6), 1593–1611. https://doi.org/10.1287/orsc.2015.1006

Internet Archive. (2024). Stack Exchange Data Science Subdomain (September 2024) dataset [Data set]. https://archive.org/download/stackexchange_20240930/stackexchange_20240930/datascience.stackexchange.com.7z

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Jurafsky, D., & Martin, J. H. (2019). Vector semantics and embeddings. In *Speech and Language Processing* (pp. 1–31). Stanford University.

Kang, M. (2022). Motivational affordances and survival of new askers on social Q&A sites: The case of Stack Exchange network. *Journal of the Association for Information Science and Technology, 73*(1), 90–103. https://doi.org/10.1002/asi.24548

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, Fog count and Flesch reading ease formula) for navy enlisted personnel* (Research Branch Report 8-75). Naval Technical Training, U.S. Naval Air Station.

Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.H., Beresnitzky, A.V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an ai assistant for essay writing task.* arXiv. https://doi.org/10.48550/arXiv.2506.08872

Kraut, R. E., & Resnick, P. (2012). *Building successful online communities: Evidence-based social design*. MIT Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (Article 1121, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons.

Liu, J., Wang, Q., Lin, C. Y., & Hon, H. W. (2013, October). Question difficulty estimation in community question answering services. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 85–90). Association for Computational Linguistics. https://doi.org/10.18653/v1/D13-1009

Loeckx, J. (2014). Mining for evidence of collaborative learning in question & answering systems. In *Proceedings of the Workshop on Feedback from Multimodal Interactions in Learning Management Systems (FFMI 2014)* (pp. 1–5). CEUR.

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL Institute of Education Press.

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. In *Frontiers in Education* (Vol. 9, Article 1328769). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1328769

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 2857–2866). Association for Computing Machinery. https://doi.org/10.1145/1978942.1979366

Martin, L. (2012). Connection, translation, offloading, and monitoring: A framework for characterizing the pedagogical functions of educational technologies. *Technology, Knowledge and Learning, 17*, 87–107. https://doi.org/10.1007/s10758-012-9193-6

Meghmala. (2023). Will ChatGPT be the Stack Overflow killer? Retrieved February 12, 2024, from https://www.analyticsinsight.net/will-chatgpt-be-the-stack-overflow-killer/

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30–31. https://doi.org/10.1126/science.1245317

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences (120:13)*, e2215907120. https://doi.org/10.1073/pnas.2215907120

Morgan, M., Brickell, G., & Harper, B. (2008). Applying distributed cognition theory to the redesign of the "Copy and Paste" function in order to promote appropriate learning outcomes. *Computers & Education, 50*(1), 125–147. https://doi.org/10.1016/j.compedu.2006.04.006

Müller, H., & Freytag, J. C. (2005). *Problems, methods, and challenges in comprehensive data cleansing* (Technical Report HUB-IB-164). Humboldt-Universität zu Berlin.

Nasehi, S. M., Sillito, J., Maurer, F., & Burns, C. (2012). What makes a good code example?: A study of programming Q&A in Stack Overflow. In *2012 28th IEEE International Conference on Software Maintenance (ICSM)* (pp. 25–34). IEEE. https://doi.org/10.1109/ICSM.2012.6405249

Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning, 40*(2), 919–930. https://doi.org/10.1111/jcal.12929

Osborne, J. W. (2013). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Sage Publications.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal, 81*(4), 443–456. https://doi.org/10.1111/j.1540-4781.1997.tb05510.x

Parnin, C., & Treude, C. (2011). Measuring API documentation on the web. In *Proceedings of the 2nd International Workshop on Web 2.0 for Software Engineering* (pp. 25–30). Association for Computing Machinery. https://doi.org/10.1145/1984701.1984706

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Perry, M. (2003). Distributed cognition. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks: Toward a multidisciplinary science* (pp. 193–223). Morgan Kaufmann. https://doi.org/10.1016/B978-155860808-5/50008-3

Raes, A., Schellens, T., De Wever, B., & Benoit, D. F. (2012). Scaffolding information problem-solving in web-based collaborative inquiry learning. *Computers & Education, 59*(1), 82–94. https://doi.org/10.1016/j.compedu.2011.11.010

Raida, M. N., Sristy, Z. N., Ulfat, N., Monisha, S. M. A., Mostafa, M. J. I., & Haque, M. N. (2024). A study on classifying Stack Overflow questions based on difficulty by utilizing contextual features. *Journal of Systems and Software, 208*, 111903. https://doi.org/10.1016/j.jss.2023.111884

Riedl, C., & Bogert, E. (2024). *Effects of AI feedback on learning, the skill gap, and intellectual diversity.* arXiv. https://arxiv.org/abs/2409.18660

Roschelle, J. (2012). Learning by collaborating: Convergent conceptual change. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 209–248). Routledge.

Salomon, G. (Ed.). (1993). *Distributed cognitions: Psychological and educational considerations*. Cambridge University Press.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265–283. https://doi.org/10.1207/s15327809jls0303_3

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Solomon, J. T., Viswanathan, V. K., Hamilton, E., & Nayak, C. R. (2017, June). *Improving student engagement in engineering using brain-based learning principles as instructional delivery protocols* [Conference paper]. The 2017 ASEE Annual Conference & Exposition, Columbus, OH, United States.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Squire, M., & Funkhouser, C. (2014). "A bit of code": How the Stack Overflow community creates quality postings. In *2014 47th Hawaii International Conference on System Sciences* (pp. 1425–1434). IEEE. https://doi.org/10.1109/HICSS.2014.185

Srba, I., & Bielikova, M. (2014). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies, 8*(2), 173–186. https://doi.org/10.1109/TLT.2014.2373374

Srba, I., & Bielikova, M. (2016). Why is Stack Overflow failing? Preserving sustainability in community question answering. *IEEE Software, 33*(4), 80–89. https://doi.org/10.1109/MS.2016.34

Stack Overflow. (2023). 2023 Developer survey. *Stack Overflow*.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge (acting with technology)*. MIT Press.

Stahl, G. (2015). A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 337–344. https://doi.org/10.1007/s11412-015-9222-2

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). BRAT: A web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 102–107). Association for Computational Linguistics.

Subramanian, S., & Holmes, R. (2013). Making sense of online code snippets. In *2013 10th Working Conference on Mining Software Repositories (MSR)* (pp. 85–88). IEEE. https://doi.org/10.1109/MSR.2013.6624012

Sun, Y., Wang, L., Diggs, K., Kulkarni, A., Steiger, K., Vu, T., & Venkataraman, V. (2025). When artificial intelligence meets human intelligence: Topics and sentiments about ChatGPT in online knowledge sharing communities. In *Proceedings of the 58th Hawaii International Conference on System Sciences (HICSS)* (pp. 233–242). https://doi.org/10.24251/HICSS.2025.028

Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *The Journal of the Learning Sciences, 12*(2), 183–219. https://doi.org/10.1207/S15327809JLS1202_2

Sweller, J., van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313–313. https://doi.org/10.1126/science.adg7879

Treude, C., Barzilay, O., & Storey, M. A. (2011). How do programmers ask and answer questions on the web? (NIER track). In *Proceedings of the 33rd International Conference on Software Engineering* (pp. 804–807). University of Hawaii. https://doi.org/10.1145/1985793.1985907

Treve, M. (2024). Integrating artificial intelligence in education: Impacts on student learning and innovation. *International Journal of Vocational Education and Training Research*, 10(2), 61–69. https://doi.org/10.11648/j.ijvetr.20241002.14

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics, 33*(1), 1–67. https://doi.org/10.1214/aoms/1177704711

UNESCO. (2025). *Artificial intelligence in education: Disruptions, dilemmas and directions*. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/digital-education/artificial-intelligence

Usher, M., & Amzalag, M. (2025). From prompt to polished: Exploring student-chatbot interactions for academic writing assistance. *Education Sciences, 15*(3), 329. https://doi.org/10.3390/educsci15030329

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Walker, E., Koedinger, K., McLaren, B., & Rummel, N. (2006). Cognitive tutors as research platforms: Extending an established tutoring system for collaborative and metacognitive experimentation. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 207–216). Springer Berlin Heidelberg. https://doi.org/10.1007/11774303_21

Wang, Q., Liu, J., Wang, B., & Guo, L. (2014). A regularized competition model for question difficulty estimation in community question answering services. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1115–1126). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1118

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage Publications.

White House. (2016). *Artificial intelligence, automation, and the economy*. Executive Office of the President. https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Academic Press.

Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences of the United States of America, 104*(45), 17599–17601. https://doi.org/10.1073/pnas.0704916104

Xue, J., Wang, L., Zheng, J., Li, Y., & Tan, Y. (2023). ChatGPT is a user-generated knowledge-sharing killer. *ICIS 2023 Proceedings*. In *ICIS 2023 Proceedings* (Paper 8). Association for Information Systems.

Yang, D., Kraut, R., & Smith, T. (2020). Does watching others help people learn? Social learning on a question-and-answer site for MOOCs. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW2), 1–24.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*(1), 87–122. https://doi.org/10.1207/s15516709cog1801_3

Zhang, J., & Patel, V. L. (2006). Distributed cognition, representation, and affordance. *Pragmatics & Cognition, 14*(2), 333–341. https://doi.org/10.1075/pc.14.2.12zha

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Understanding student motivation, behaviors, and perceptions in MOOCs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1882–1895). Association for Computing Machinery. https://doi.org/10.1145/2675133.2675217.

Zurita, G., & Nussbaum, M. (2004). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers & Education, 42*(3), 289–314. https://doi.org/10.1016/j.compedu.2003.08.005

Appendix A

Glossary of Key Technical Terms

- **Lexical Density (LD):** A text-level indicator of information load defined as the proportion of content words (typically nouns, proper nouns, main (lexical) verbs, adjectives, and adverbs) to all tokens in a text. Higher LD suggests more informational/technical language; lower LD reflects more conversational or function-word–heavy prose.
- **Flesch–Kincaid Grade Level (Flesch-K).** A readability index estimating the U.S. school grade level required to comprehend a text, based on average sentence length and average syllables per word. Higher scores indicate greater reading difficulty (e.g., 12 ≈ twelfth-grade level).
- **Large Language Models (LLMs)**: Artificial intelligence systems trained on large text datasets to generate language-based outputs.
- **Cognitive Load:** The amount of mental effort required to understand and process information.
- **Electroencephalography (EEG)**: A technique that records brain activity using sensors placed on the scalp.
- **Neural Network Coupling:** The coordination between different brain regions during cognitive tasks.

- **Alpha, Beta, Theta, Delta Bands:** Brainwave frequencies that correspond to different cognitive processes.
- **Executive Function:** Mental skills involved in managing attention, memory, and problem solving.
- **Distributed Cognition:** A theory that views thinking as spread across people, tools, and environments.
- **Cognitive Offloading:** The act of using external aids to reduce mental workload.
- **Metacognition:** Awareness and control of one's own learning and thinking processes.
- **Adaptive Expertise:** The ability to apply knowledge in flexible and innovative ways in new situations.
- **Higher-Order Cognitive Skills:** Complex thinking abilities including analysis, evaluation, and creation.
- **Question** (Stack Exchange): The initiating entry in a Q&A thread. In this study, each observation corresponds to one question (title + body).
- **Post** (Stack Exchange): A database object that can be either a question or an answer. This dissertation analyzes question posts only and does not analyze answer text.

Appendix B

Descriptive Statistics and Distribution Checks

Table B1

Data Transformation Rationale by Variable

Variable	Data Type	Skew	Transformation	Rationale
Flesch-Kincaid Grade Level	Continuous (Readability)	1.06	$\log_{10}(x - x_{min} + 1)$	Addressed the "floor" effect and potential non-positive values. The shift ensured all values were positive before logging, successfully pulling in high-grade outliers.
Total Word Count	Count (Discrete)	2.01	$\log_{10}(x + 1)$	Compresses the "power law" distribution typical of text lengths. This linearizes the magnitude differences (e.g., 10 vs. 1000 words).
Number of Responses	Count (Zero-inflated)	2.36	$\log_{10}(x + 1)$	Addresses zero-inflation (since logarithm of 0 is undefined). Adding 1 handles zeros while normalizing the long tail of highly active discussions.
Number of Upvotes	Count (Net Score)	2.43	$\log_{10}(x - x_{min} + 1)$	"Upvotes" are net scores that can be negative. The shift subtracts the minimum value to ensure all inputs are positive, allowing the log function to normalize the skew.
Time to First Response	Continuous (Time)	7.36	$\log_{10}(x + 1)$	Standard remedy for reaction-time data which follows an exponential curve. Converts the extreme right skew into a normalized "bell curve" shape.
Tagged Categories Count	Count (Small Range)	0.24	$\sqrt{x}$	A gentler transformation recommended for small-range count data (1-5 tags). Stabilizes variance without "over-correcting" the mild skew like a log function would.

Lexical Density	Continuous (Ratio)	0.08	None	The distribution was already nearly perfectly symmetric.

Table B2

Period Effect (Pre- vs Post-ChatGPT)

Model	**Pillai's Trace**	***F***
Original Data	0.0044	71.42***
Transformed Data	0.0043	70.16***

Note: Degrees of Freedom for both models were (6, 97,502).

Table B3

Descriptive Statistics Comparing Original vs. Transformed Distributions

	Original Data (Winsorized)				**Transformed Data**			
Variable	***M***	***SD***	**Skew**	**Kurt**	***M***	***SD***	**Skew**	**Kurt**
Flesch-Kincaid Grade Level	10.44	4.89	1.06	1.98	0.96	0.23	-1.14	3.05
Lexical Density	0.48	0.06	0.08	0.16	0.48	0.06	0.08	0.16
Total Word Count	161.79	115.30	2.01	4.98	2.12	0.27	0.18	-0.18
Number of Responses	1.40	0.76	2.36	6.23	0.36	0.11	1.69	2.16
Number of Upvotes	1.48	2.22	2.43	8.56	0.61	0.18	0.12	2.06
Time to First Response	9257.22	54823.25	7.36	55.04	2.16	1.03	1.07	1.13
Tagged Categories	2.80	1.25	0.24	-0.92	1.63	0.39	-0.13	-0.93

Table B4

MANOVA Comparison: Wilks' Lambda (Λ)

Effect	Original Data (Non-Normal)			**Transformed Data (Normalized)**		
	Wilks' Λ	***F***	Sig.	**Wilks' Λ**	**F**	**Sig.**
Community	0.9315	247.17	$p < .001$	0.9373	225.71	$p < .001$
Period	0.9997	4.76	$p < .001$	0.9997	3.81	$p < .001$
Interaction (Community x Period)	0.9987	4.39	$p < .001$	0.9986	4.8	$p < .001$

Table B5

Univariate Two-Way ANOVA results:

	Variable	Original Data (Non-Normal)	Transformed Data (Normalized)
Flesch-Kincaid	Community	< .001	< .001
	Period	< .001	< .001
	Interaction	0.713	0.617
Lexical Density	Community	< .001	< .001
	Period	.549	.549
	Interaction	.264	.264

Appendix C

Comparison of Raw Data to Final Sample

Table C1

Sample Composition by Community (Original vs. Final)

Community	Original N	Original %	Final N	Final %	Difference (% pts)
Mathematics	126,073	90.0%	86,456	88.7%	-1.4
Academia	4,115	2.9%	3,768	3.9%	+0.9
Data Science	4,056	2.9%	2,984	3.1%	+0.2
Computer Science	3,882	2.8%	2,843	2.9%	+0.1
Artificial Intelligence	1,896	1.4%	1,462	1.5%	+0.1

Note. Mathematics has the lowest retention rate (68.6%) because math questions on Stack Exchange receive zero answers more frequently than discussion-based communities like Academia (91.6% retention).

Appendix D

Participation Volume and Sample Retention

Table D1

Raw Question Volume by Community and Time Period (All Sampled Questions)

Community	Pre (*n*)	Post (*n*)	Δ*n*	%Δ	Post/Pre (Index)
Academia	1,976	2,128	152	7.70%	107.7
Artificial Intelligence	942	953	11	1.20%	101.2
Computer Science	2,243	1,632	-611	-27.20%	72.8
Data Science	2,525	1,523	-1,002	-39.70%	60.3
Mathematics	68,237	57,836	-10,401	-15.20%	84.8
Total	75,923	64,072	-11,851	-15.6%	84.4

Note. Pre = Jan 1–Sep 30, 2022; Post = Jan 1–Sep 30, 2023. Counts reflect the initial sampled corpus (prior to filtering out unanswered questions). Index = (Post / Pre) × 100.

Table D2

Answered-Question Volume by Community and Time Period (Final Analytic Sample)

Community	Pre (*n*)	Post (*n*)	Δ*n*	%Δ	Post/Pre (Index)
Academia	1,832	1,936	104	5.70%	105.7
Artificial Intelligence	706	756	50	7.10%	107.1

Computer Science	1,640	1,203	-437	-26.70%	73.4
Data Science	1,794	1,190	-604	-33.30%	66.7
Mathematics	47,118	39,338	-7,780	-16.50%	83.5
Total	75,923	64,072	-11,851	-15.6%	84.4

Note. Sample includes only questions with at least one answer (answered questions). Index = (Post / Pre) × 100.

Table D3

Answered-Question Rate by Community and Period (Answered ÷ Raw)

Community	Answered rate (Pre)	Answered rate (Post)	Δ (percentage points)	Community	Answered rate (Pre)
Academia	92.70%	91.00%	-1.7	Academia	92.70%
Artificial Intelligence	74.90%	79.30%	4.4	Artificial Intelligence	74.90%
Computer Science	73.10%	73.70%	0.6	Computer Science	73.10%
Data Science	70.70%	78.10%	7.5	Data Science	70.70%
Mathematics	69.10%	68.00%	-1	Mathematics	69.10%
Total	69.90%	69.30%	-0.6	Total	69.90%

Note. This table helps demonstrate that the post-period volume declines (Table D1/D2) are not simply an artifact of a large shift in answerability; overall answered-rate changes are small at the aggregate level.

Table D4

Pre/Post Volume Change Across Analytical Stages

Analytics Stage	Answered rate (Pre)	Answered rate (Post)	Δ (percentage points)	Community	Answered rate (Pre)
Raw corpus (all sample questions)	75,923	64,072	139,995	-11,851	-15.60%
Final analytic sample (answered questions only)	53,090	44,423	97,513	-8,667	-16.30%

Note. The decline in posting volume is visible both in the raw sampled corpus and after applying the answered-only analytic restriction.

Table D5

Data Preprocessing Exclusions Summary (Overall)

Exclusion/Filter	*N* Removed	% of dataset	Rationale (brief)
Duplicate questions (exact Body match)	36	0.03%	Prevent pseudo-replication / inflated N
Null or empty Body text	2	0.00%	Required for computing text-based measures
Unanswered questions (missing valid response time)	42,483	30.30%	Focus on community engagement dynamics
Final analytic sample (retained)	97,513	69.65%	Answered questions only

Note. This supports the claim that sample reduction is driven almost entirely by excluding unanswered questions and that the exclusion is methodologically motivated (engagement-focused).