When Artificial Intelligence Meets Human Intelligence: Topics and Sentiments about ChatGPT in Online Knowledge Sharing Communities

Yao Sun
Department of Social Sciences and Humanities
New Jersey Institute of Technology
yao.sun@njit.edu

Lijing Wang
Department of Data Science
New Jersey Institute of Technology
lijing.wang@njit.edu

Kevin Diggs
New Jersey Institute of Technology
ktd25@njit.edu

Avanish Kulkarni
New Jersey Institute of Technology
agk4@njit.edu

Kevin Steiger
New Jersey Institute of Technology
kns7@njit.edu

Tai Vu
New Jersey Institute of Technology
tdv@njit.edu

Vibha Venkataraman
New Jersey Institute of Technology
vv66@njit.edu

Abstract

OpenAI's release of ChatGPT has sparked worldwide discussion and debates about its potential implications. Knowledge sharing conversations in online Q&A communities, for example, demonstrate how ChatGPT is perceived when human collective intelligence is unleashed. We focus on textual data collected from the Stack Exchange community to investigate the major topics and sentiments generated when community members share their knowledge about ChatGPT. Different topic modeling and sentiment analysis algorithms were applied and compared to understand major perceptions and attitudes toward ChatGPT, and to illustrate the perceived opportunities and challenges surrounding ChatGPT and AI algorithms. Our study presents the differences in terms of ChatGPT-related sentiments and discussion themes across different types of subcommunities and highlights the critical roles of feature extraction and threshold selection in further improving the accuracy of topic models and precisely capturing the patterns of prevalent sentiments.

Keywords: ChatGPT, Artificial Intelligence (AI), Online Q&A Communities, Topic Modeling, Sentiment Analysis

1. Introduction

Recent years have witnessed the revolutionary impact of artificial intelligence (AI) on all aspects of our society. One of the most recent milestones of AI development occurred in November 2022, marked by the launch of ChatGPT, a generative AI tool that has attracted over 1 million registered users within the first five days of its release[1]. Generative AI refers to a class of artificial intelligence models that can utilize the patterns and structures learned from existing data to generate new data such as text, images, and audio files (Ali & Djalilian, 2023; Naumova, 2023; Rudolph et al., 2023; Zhou et al., 2023). Using a conversational structure, ChatGPT can challenge false assumptions or detect errors while being used for large-scale applications.

Research has demonstrated the great potential of ChatGPT to support individuals, communities, and organizations in making informed decisions in such areas as public health (Baclic et al., 2020; Panch et al., 2019), environmental sustainability (Biswas, 2023), journalism and media practice (Pavlik, 2023), business management (George & George, 2023), etc. For example, ChatGPT-based marketing platforms and bots can help marketers quickly create and deliver content,

[1] https://explodingtopics.com/blog/chatgpt-users

URI: https://hdl.handle.net/10125/108864
978-0-9981331-8-8

better understand customers, increase consumer engagement, as well as improve customer service and marketing efficiency (Paul et al., 2023; Rivas & Zhao, 2023). ChatGPT has also been found to positively influence organizational and managerial activities at the strategic, functional, and administrative levels (Korzynski et al., 2023). However, in the meantime, ChatGPT has raised concerns over its privacy, security, potential bias, and other ethical considerations. For instance, research on education indicated that both educators and learners will need to develop competencies and literacies necessary to understand the potential limitations and brittleness of such technology as ChatGPT (Kasneci et al., 2023; Tlili et al., 2023). A recent study (Rozado, 2023) on ChatGPT's political orientation found that the majority of ChatGPT-generated responses exhibited "left-leaning political viewpoints" (p.148), although ChatGPT "claimed to be politically neutral and unable to take a stance" (p.149). Another research studied the accuracy of ChatGPT, demonstrating that ChatGPT indeed produces errors and may not be used as a sole source especially if the user lacks sufficient background knowledge of a certain topic (Shoufan, 2023). In addition, the content generated by ChatGPT and other AI algorithms can pose a challenge to copyright and intellectual property (Vincent, 2022). They may also be related to emerging workplace issues, such as job replacements (Hern 2023; Roose, 2023) or algorithmic discrimination against disadvantaged persons in the hiring process (Yarger et al., 2020).

The rise of ChatGPT has stimulated lots of discussion in online question-and-answer (Q&A) communities. Such kind of online communities acts as a platform for individuals to freely ask, answer, and discuss questions based on their common interests (Gazan, 2011; Liu et al., 2021). As one of the world's largest and most successful online Q&A community, Stack Exchange (https://stackexchange.com/) features over 418.8 million monthly visits, as well as 3.1 million questions and 3.5 million answers contributed every year[2]. Users involved in the Stack Exchange community are very active in discussing a wide variety of topics ranging from computer programming to arts and philosophy and contribute to either technical topics largely focused on programming and technology development or non-technical topics discussing personal interests and hobbies (Lampe et al., 2010). Despite the burgeoning scholarly interest in ChatGPT, however, little research has been focused on the diverse public opinion and perspectives toward ChatGPT generated in online Q&A communities. Therefore, our present study seeks to fill this research gap by mining and comparing the major themes and sentiments generated in the Q&A website Stack Exchange. We ask two specific research questions: 1) What are the major topics and sentiments regarding ChatGPT generated in the online Q&A community and 2) How do these topics and sentiments vary across different sub-communities? Using data collected from Stack Exchange, we applied topic modeling and sentiment analysis algorithms to examine meaningful themes and salient sentiments, as well as compare different subcommunities in terms of the themes and sentiments generated about ChatGPT. Our research is one of the first studies examining the social impact of ChatGPT and providing insights into understanding the major opportunities and challenges surrounding ChatGPT.

The rest of the paper is organized as follows. We first review previous research related to online Q&A communities and how computational tools have been used to understand the content generated by Q&A community members. We then introduce our research methods, followed by the analytical results as well as the conclusion and discussion of our research findings.

2. Related Work

As a typical knowledge sharing community, online Q&A community is often considered as a social gathering platform where members express their opinion, exchange their knowledge, and form online relationships through ask-answer activities (Rheingold, 2000). Ardichvili et al. (2003) suggested that community members are willing to contribute their knowledge because of their personal interests and goals, the community culture, and their trust toward each other. Such knowledge sharing activities can further contribute to the construction of the communal knowledge within the community (Ng'ambi, 2013). In the community, traditional search-based knowledge service evolves into a two-way knowledge exchange that benefits the generation of collective intelligence (Qi et al., 2018). Such interactive process in which individuals exchange information to build communal knowledge through social communication is a manifestation of human collective intelligence (Kaplan, 2001; Krause et al., 2010; Leimeister, 2010).

When a new technology emerges, both professional developers and lay users go to Q&A communities such as Stack Exchange to exchange knowledge, information, and support, allowing this online community to offer profound opportunities for researchers to examine how collective human intelligence responds to new technologies such as AI. For example, Almansoury et al. (2022) studied the discussion among video-game developers on Stack

[2] https://stackexchange.com/about

Exchange and found that the pairing or a mix of different technologies often tends to generate more challenging but meaningful comments within the community. Comparing humanities- versus STEM-focused subcommunities of Stack Exchange, Santos et al. (2019) demonstrated that unlike STEM-focused conversations, discussion on topics in the humanities are more likely to be driven by active participants' peer influence in this community. Research on quantum computing systems-related discussion also showed that participating in Stack Exchange conversations can facilitate the explanation of quantum theories, the interpretation of quantum program outputs, and the understanding of the connection between quantum computing and classical computing (El aoun et al., 2021).

Previous research has employed various computational tools to understand major themes and sentiments generated on social media platforms (Choudrie et al., 2021; Wedel et al., 2021), especially in online Q&A communities. Novielli et al. (2015), for example, used an innovative sentiment analysis method to identify and confirm the polarities of sentiments demonstrated in the Stack Overflow community. Similarly, a textual analysis of software engineering knowledge shared in the Stack Overflow community has demonstrated that examining the tags used by knowledge contributors can facilitate scholarly understanding of the development of collective intelligence in this community (Ye et al., 2016). Employing semantic analysis method named WordSimSE DB, Tian et al (2014) calculated the weighted word similarity of the knowledge shared in the Stack Overflow community and successfully unearthed the most discussed topics about software development. Papoutsoglou et al. (2022), likewise, studied open-source software licensing questions in Stack Exchange community using topic modeling approaches and found that compared to basic questions, questions particularly focused on warranties, derivative, and creative commons tend to yield answers with a longer period of time. Taken together, the literature research carried out in this direction demonstrates that applying computational tools such as topic modeling and sentiment analysis can help deepen our understanding of how individuals perceive ChatGPT and artificial intelligence algorithms in general.

3. Methods

3.1. Data Collection and Preprocessing

The dataset used in this study was collected from a total of 173 subsites of Stack Exchange (https://stackexchange.com/). Stack Exchange is known for its rigor in question-and-answer quality thanks to its stringent requirements and is thus considered as a source of valuable real- world perspectives especially on technological, scientific, and engineering topics (Alfayez, 2023). Users must acquire a certain reputation rating before they are allowed to upvote or downvote a certain question or answer, and poorly rated comments and questions are often removed by the community managers, ensuring that the questions and answers posted within the community are of good quality. Among all the subsites, Stack Overflow is considered as the flagship subsite because Stack Exchange originates from Stack Overflow, a website for programmers to ask coding questions and communicate about programming solutions. Stack Overflow has 11.7 million registered users and 257.7 million monthly visits[3].

Our data was collected from all the Stack Exchange subsites split into 4 subcategories: AI-related subsites, technical subsites, non-technical subsites, and Stack Overflow. Technical subsites included communities such as askUbuntu, Android Enthusiasts, Physics, Mathematics, etc., whereas non-technical subsites are about topics such as Coffee, Music Fans, Pets, etc. We used Python language and set 'ChatGPT' as the search keyword to retrieve posts through the website API, with an end date of June 1st, 2023. Most posts were contributed between August 2022 and June 2023, although the earliest post can be traced back to April 2019. A total of 17,957 posts were retrieved. Upon the completion of data collection, all the posts were grouped according to the aforementioned 4 subcategories, resulting in 868 posts from AI-related subsites, 10,049 posts from technical subsites, 2,888 posts from non-technical subsites, and 4,152 posts from Stack Overflow.

After data collection, we performed data preprocessing with Python packages such as the Natural language Toolkit (NLTK) through removing stop words and punctuation, lowercase transformation, stemming, lemmatizing, tokenization, and n-gram procedures (Nadkarni et al., 2011; Nagwani, 2015; Wallach, 2006).

To capture major discussion topics emerged from the online conversations, we primarily applied Latent Dirichlet Allocation (LDA, Blei et al., 2003), a probabilistic topic modeling technique that takes into

[3] https://stackexchange.com/about

account content clustering probability and extracts cohesive topic clusters from a collection of documents. LDA assumes that each text corpus is composed of several topics that can be configured as a probability distribution over words. To analyze our data, we applied Scikit-Learn's LDA model (Raschka et al., 2022; Raschka & Mirjalili, 2019), which is a free machine learning Python library identifying keywords under each emerging topic.

3.2. Topic Modeling

To complement and validate our primary method, we applied several other widely used topic modeling libraries and techniques. First, we tested our data using another LDA algorithm that allows the topics to surface by identifying phrases rather than words, in order to see if there are significant differences between topics with one versus more than one words. Second, BERTopic's topic reduction function was tested on our data, which uses HDBSCAN, a hierarchical clustering algorithm that extracts a flat clustering based on the stability of clusters (Abuzayed & Al-Khalifa, 2021; Alhaj et al., 2022; Egger & Yu, 2022). Latent Semantic Analysis (LSA) model (Guzman et a., 2021) from Scikit-Learn was then applied to our data as the third method. This algorithm can reduce the dimensionality of a text document by singular value decomposition (SVD) and can group topics together by their cosine similarity scores (Hassani et al., 2021). Lastly, Scikit-Learn Non-negative Matrix Factorization (NMF) model was tested using our data, which works by decomposing a text document into non-negative matrices to produce clustered data (Akhter et al., 2019; Flores et al., 2022; Zheng & Xu, 2021).

3.3. Sentiment Analysis

To extract the emotional polarity imbued in the comments within the Stack Exchange community, we utilized the National Research Council of Canada Emotion Lexicon (NRCLex) sentiment analysis library (Ainapure et al., 2023; Bashar, 2021; Smirnov & Hsieh, 2022). Sentiment analysis is a subfield of natural language processing used to determine the polarity of emotions in a given text. Lexicon-based strategies rely on a sentiment library containing an established list of words and phrases where each word is labeled an orientation as positive, negative, or neutral. Utilizing lexicon-based sentiment analysis strategies is simple and efficient, as they are not required for the training of a model. Using NRCLex, we extracted eight different sentiments (i.e., fear, anger, anticipation, trust, surprise, sadness, disgust, and joy) and two polarities (i.e., positive and negative) from our data. To conduct our sentiment analysis, the semantic orientation of the words and phrases was calculated. The sentiment values of the words were determined by matching words from the NRCLex dictionary, and the percent distribution of positive, neutral, and negative comments for each type of subsite was harvested.

To verify and corroborate the results of the primary sentiment analysis, a validation was performed using an alternative method. In particular, Scikit-Learn's sentiment analysis model NLTK VADER (Elbagir & Yang, 2018; Gupta et al., 2017; Hutto & Gilbert, 2014) was used to calculate an aggregate sentiment score between -1 and 1 for every comment. After testing various thresholds, it was found that a sentiment value below -0.2 for negative comments and a sentiment value above 0.5 for positive comments are most reliable as they allowed the inclusion and examination of the largest corpus of text. Each comment was classified as positive, neutral, or negative. Finally, we randomly selected 0.5% of the entire data for manual labeling and then compared human-labeled results against machine-coded results to verify the accuracy of the algorithm.

4. Results

4.1. Findings from Topic Modeling

Our primary topic modeling approach produced a total of six topics, with 15 keywords identified for each topic. The results are presented in Table 1. The topics and corresponding keywords are shown in descending relevance, meaning that the first topic appears more frequently than the sixth topic, and the first keyword in each topic appears more frequently than the fifteenth keyword. The words displayed in each topic are found to be cohesive with one another in terms of their meaning. Specifically, Topic 1 is about how people, especially students, use ChatGPT in their study and work, with keywords identified as "problem", "make", "work", "people", "students", and "chatgpt". Topic 2 is about using ChatGPT to assist with coding, as it includes keywords such as "API", "code", "function", etc. Topic 3 seems to be about asking questions to ChatGPT on Stack Overflow, as it identified keywords such as "ask", "stack", "answers". Topic 4 is also concerned with asking ChatGPT diverse context-specific questions related to "policy", "generated", "human" and so on. Topics 5 and Topic 6 seem to be about programming and ChatGPT in general, configured by keywords such as "error", "server", "language", "models", "text", etc.

Table 1. Results from topic modeling using primary LDA (all subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	point	problem	make	work	people	use	students	left	way	number	frac	like	right	time	chatgpt
Topic #2	file	help	api	work	need	function	want	data	like	use	error	tried	using	chatgpt	code
Topic #3	overflow	correct	problem	users	good	ask	like	code	stack	post	questions	answers	chatgpt	comment	Stack overflow
Topic #4	actually	policy	good	make	generated	question	human	use	people	know	think	like	content	chatgpt	ai
Topic #5	files	version	like	tried	project	run	js	com	error	chatgpt	app	use	server	using	file
Topic #6	using	like	com	models	language	answers	text	use	openai	generated	data	gpt	model	ai	chatgpt

To compare the discussions on different subsites, we show the topic modeling results on non-technical and technical subsites in Table 2 and Tables 3, respectively. Based on the topics generated, it can be found that the technical subsites are more focused on the functionality and technical details of ChatGPT as compared to the non-technical subsites. We observed in the non- technical subsites words such as "know", "information", "question", "english", "language", "human", "evidence", "answer", and "think." In the technical subsites we found words such as "string", "data", "code", "api", "function", "app", "js", "error", and "file" which are more coding related. The non-technical subsites do have "ai" and "chatgpt" included but it is much less prevalent as compared to the technical subsites; they seem to be more focused on the information that ChatGPT outputs.

Table 2. Results from topic modeling using primary LDA (non-technical subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	different	answer	really	model	word	know	information	way	question	english	ai	chatgpt	language	human	like
Topic #2	example	https	windows	chatgpt	like	looking	hydrogen	make	work	time	symmetry	using	question	com	use
Topic #3	license	judges	said	ve	answers	content	se	people	judge	know	rules	question	chatgpt	evidence	answer
Topic #4	questions	better	question	think	work	use	answer	stack	make	new	time	people	like	chatgpt	ai
Topic #5	post	policy	se	site	use	users	questions	stack	generated	content	question	answer	ai	answers	chatgpt
Topic #6	right	software	use	work	app	question	person	information	say	know	need	people	think	chatgpt	like

Table 3. Results from topic modeling using primary LDA (technical subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	example	solution	string	left	like	data	time	using	function	right	code	api	need	use	frac
Topic #2	time	users	good	stack	like	post	content	people	generated	questions	ai	answers	question	chatgpt	answer
Topic #3	time	line	point	understanding	know	wrong	answer	word	like	think	chatgpt	internal	gradle	org	java
Topic #4	training	prompt	ai	using	like	use	code	models	language	openai	data	text	gpt	model	chatgpt
Topic #5	way	values	type	want	need	work	using	tried	value	chatgpt	function	use	data	like	code
Topic #6	like	work	script	app	js	want	help	api	use	chatgpt	tried	using	error	file	code

Topic modeling on AI-related subsites and on Stack Overflow also resulted in six topics with 15 keywords emerged in each topic (Table 4 and Table 5). Expectedly, conversations on AI-related subsites mainly discussed ChatGPT and its functionality, as shown by the fact that every topic includes the word "ai" or "chagpt", as well as other machine learning related words such as "model", "data", "learning", "training", and "prompt." Conversations on Stack Overflow also mainly discussed coding and machine learning-related subjects.

Table 4. Results from topic modeling using primary LDA (AI-related subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	context	think	language	text	data	questions	conversation	question	gpt	ai	prompt	answers	model	like	chatgpt
Topic #2	model	answers	questions	ai	chatgpt	know	gpt	able	correct	provide	energy	question	number	information	answer
Topic #3	instructgpt	like	model	dentist	com	openai	cat	learning	fine	chatgpt	ai	gpt	data	answer	question
Topic #4	know	like	100	math	devices	training	language	minutes	machines	ai	answer	models	make	model	chatgpt
Topic #5	answer	input	data	language	chinese	understand	models	human	training	output	text	chatgpt	like	understanding	model
Topic #6	output	example	responses	answer	arxiv	google	know	prompt	model	question	human	gpt	ai	openai	chatgpt

Table 5. Results from topic modeling using primary LDA (Stack Overflow)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	work	run	help	problem	use	server	python	like	js	using	chatgpt	error	tried	file	code
Topic #2	time	client	method	page	type	tried	need	like	chatgpt	using	error	code	function	use	data
Topic #3	build	command	problem	run	stackoverflow	execute	project	help	internal	question	com	gradle	code	org	java
Topic #4	like	text	chat	model	request	_x000d_	error	code	using	response	use	gpt	openai	chatgpt	api
Topic #5	different	hidden	data	database	want	create	import	new	using	chatgpt	tried	line	like	user	use
Topic #6	problem	answer	use	trying	error	know	help	like	want	work	tried	need	using	chatgpt	code

To show the validity of our primary LDA, we present our validation results from phrase- based LDA, BERTopic, LSA, and NMF, respectively, across all subsites (Tables 6, Table 7, Table 8, and Table 9). With phrase-based LDA, we see very few 2 or 3-word phrases emerging from the discussion. The topics gathered are similar to the results from our primary LDA, as Topic 1, Topic 3, Topic 4, and Topic 5 are about the use of ChatGPT, Topic 2 is about the information and answers that ChatGPT is able to provide, and Topic 6 is about questions and discussions about the usage of ChatGPT.

Table 6. Results from topic modeling using phrase-based LDA (all subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10	Word #11	Word #12	Word #13	Word #14	Word #15
Topic #1	make	new	using	want	work	model	time	code	example	ve	use	data	chatgpt	need	like
Topic #2	stack	people	question	know	knowledge	new	make	_x000d_	community	answer	like	human	answers	chatgpt	ai
Topic #3	using	know	code	human	language	answer	text	gpt	question	data	use	model	ai	like	chatgpt
Topic #4	value	want	use	work	function	internal	right	make	org gradle	using	need	chatgpt	gradle	java	org
Topic #5	data	work	problem	help	api	like	want	com	file	tried	use	error	using	chatgpt	code
Topic #6	stack overflow	overflow	people	users	like	post	questions	stack	generated	content	ai	question	answer	answers	chatgpt

BERTopic's topics gathered correlate with the results of the primary LDA for the first two topics but are much less cohesive in terms of the remaining eight topics. In other words, only the first two topics held the highest level of relevance, with the remaining topics having comparably lower level of relevance. BERTopic also tended to generate topics with potential bias. For example, Topic 4 is exclusively about firebase and cloud services, Topic 5 is exclusively about downvoting, and Topic 6 is only about trying and thanking.

Table 7. Results from topic modeling using BERTopic (all subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10
Topic #1	chatgpt	answer	question	would	ai	like	code	answers	use	using
Topic #2	chatgpt	would	answer	use	like	question	code	ai	one	using
Topic #3	page	code	json	file	xdebug	javascript	pdf	html	using	port
Topic #4	firebase	files	permissions	folder	firestore	git	directory	cloud	user	file
Topic #5	downvote	downvotes	rep	penalty	downvoting	people	downvoted	meta	voting	answers
Topic #6	thanks	thank	try	ok	much	ill	worked	help	works	tried
Topic #7	docker	command	container	pipe	pipes	image	cpp	run	port	host
Topic #8	related	help	question	look	somewhat	much	stack	questions641 40091	questions620 72056	gotube
Topic #9	temperature	api	web	deterministic	via	gpt4	gpt	openai	chatgpt	default
Topic #10	nan	value	300	float	null	missing	cheer	nans	salary	comparisons

Results from LSA seemed to be less informative compared to the results from our primary LDA models as keywords such as “chatgpt”, “answer”, question”, “ai” repeated several times across all the topics. Similar to our primary results, ChatGPT-related questions and answers appeared most frequently in the topics emerged from using LSA.

Table 8. Results from topic modeling using LSA (all subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7
Topic #1	chatgpt	answer	question	answers	ai	code	like
Topic #2	answer	answers	question	ai	generated	questions	stack
Topic #3	question	answer	code	edit	error	ask	com
Topic #4	code	ai	generated	answers	error	problem	answer
Topic #5	com	question	questions	stackoverflow	code	meta	stackexchange
Topic #6	chatgpt	answers	code	answer	questions	openai	wrong
Topic #7	com	answer	questions	stackoverflow	stackexchange	meta	error
Topic #8	gpt	ai	answer	model	openai	code	api
Topic #9	error	content	file	generated	post	tried	api
Topic #10	answers	data	gpt	error	model	question	api

Similar to BERTopic and LSA, when using NMF, the topics provide less information compared to those generated by our primary LDA, with Topic 1 and Topic 6 discussing AI- and ChatGPT-generated content, and Topic 2 mainly focusing on the online interaction among programmers such as JavaScript, trying, and thanking.

Table 9. Results from topic modeling using NMF (all subsites)

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6	Word #7	Word #8	Word #9	Word #10
Topic #1	site	tools	se	users	ban	human	policy	content	generated	ai
Topic #2	trying	app	thanks	js	try	help	using	tried	file	error
Topic #3	comment	help	op	correct	ask	asked	post	edit	question	answer
Topic #4	works	better	provide	working	python	reproduce	underst and	write	problem	code
Topic #5	network	policy	community	site	sites	meta	post	exchange	overflow	stack
Topic #6	content	asking	text	ban	generated	using	ask	banned	use	chatgpt
Topic #7	way	make	people	time	need	use	know	data	think	like
Topic #8	continue	text	discussion	models	turbo	chat	model	api	openai	gpt
Topic #9	help	related	site	topic	ask	stackexchange	meta	stackoverflow	com	questions
Topic #10	answer	low	post	wrong	good	quality	people	correct	users	answers

4.2. Findings from Sentiment Analysis

One of the goals of this study was to also understand the prevalent sentiment towards ChatGPT in Stack Exchange, as it is crucial in revealing users' diverse attitudes and opinions regarding ChatGPT and the development of AI. Using NRCLex, our results demonstrate the prevalence of the following sentiments: positive, negative, neutral, trust, anticipation, joy, fear, sadness, anger, surprise, and disgust. In Figures 1 – 4 we present a summary of the percentages of different sentiments on technical subsites, non-technical subsites, AI-related subsites, and Stack Overflow, respectively.

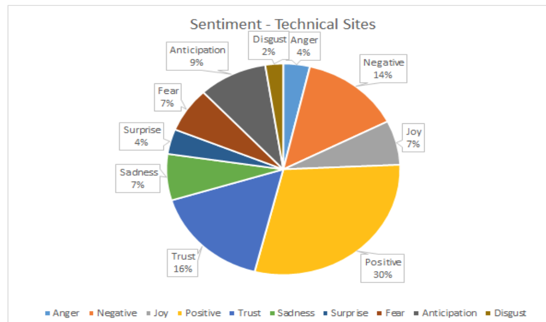

Figure 1. Sentiments on ChatGPT from technical subsites

Overall, the general sentiment towards ChatGPT is more leaning toward positive and neutral. This echoes prior research findings that in general, people's opinion towards ChatGPT is positive (e.g., Shoufan, 2023). Among all the sentiment categories, positive, negative, and trust were the most dominant sentiments, followed by joy and anticipation. AI-related subsites demonstrate the biggest percentage of positive views on ChatGPT compared to other subsites in Stack Exchange.

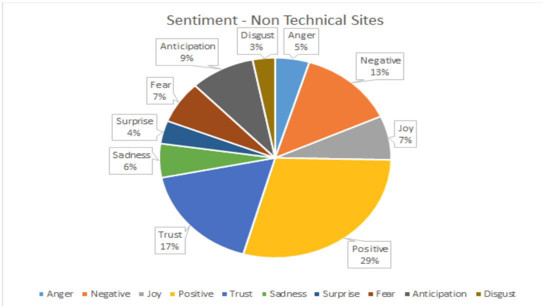

Figure 2. Sentiments on ChatGPT from non-technical subsites

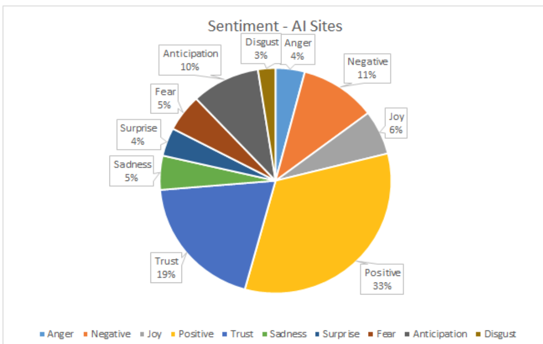

Figure 3. Sentiments on ChatGPT from AI-related subsites

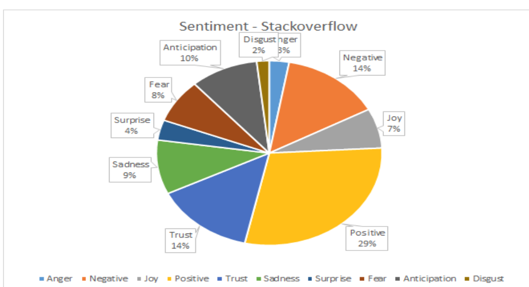

Figure 4. Sentiments on ChatGPT from Stack Overflow

To elucidate our results, we further focused on three sentiments and polarities: positive/neutral, negative, and trust, as they are the most outstanding categories of sentiment emerged from our data. Due to the characteristics of NRCLex (Chiny et al., 2021), positive and neutral sentiments are considered as an integrative aggregate when calculating the percentages.

Positive and/or Neutral. Figure 5 shows the difference across the four categories of subsites in terms of the prevalence of positive and neutral sentiment toward ChatGPT. AI-related subsites predictably demonstrate the highest percentage of positive sentiment, likely due to the fact that ChatGPT represents the most cutting-edge development in the field of artificial intelligence. Contrary to AI-related subsites, non-technical, technical subsites, and Stack Overflow all demonstrated similar percentages. This reveals that the major sentiment towards ChatGPT in the Stack Exchange community is positive and/or neutral.

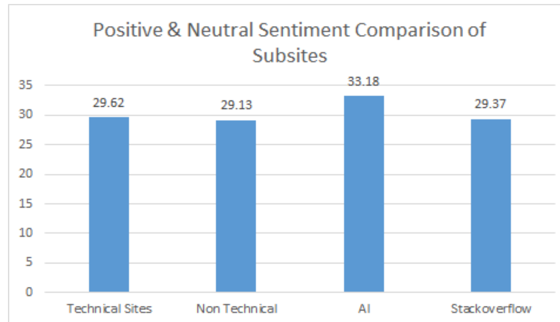

Figure 5. Prevalence of positive/neutral sentiment across different subsites

Negative. Figure 6 compares the levels of negative sentiment on ChatGPT across the four types of subsites. Concurring with the previous sentiments, AI-related subsites demonstrate the lowest percentage of negative sentiment by a large margin, with Stack Overflow displaying the highest percentage of negative sentiment. The prevalence of negative sentiment in the Stack Overflow community may be due to the inconsistency and inaccuracy of ChatGPT when it comes to coding problems. Technical and non-technical subsites demonstrate a similar composition of different sentiments, although they are generally more negative toward ChatGPT compared to AI-related subsites.

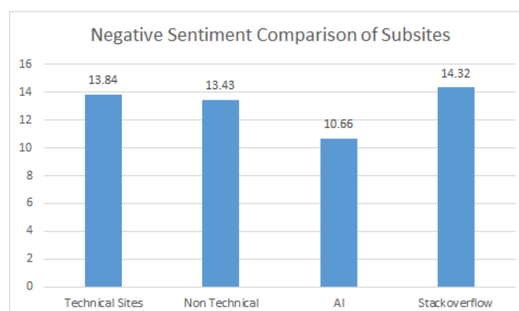

Figure 6. Prevalence of negative sentiment across different subsites

Trust. Figure 7 compares the prevalence of trust towards ChatGPT across the four types of subcommunities. It can be found that among all these subsites, members of AI-related subcommunities generally demonstrate a higher level of trust toward ChatGPT compared to members of other subcommunities. On the contrary, the Stack Overflow community is the most skeptical toward ChatGPT, with more than 5% less trust compared to the AI-related subcommunities. This may be due to the fact that programmers and developers of ChatGPT still occasionally struggle with coding questions that are particularly complex. For example, many programmers and developers who work with ChatGPT often have to post their entire coding projects to the Stack Overflow subsite in order to seek solutions. Another interesting finding to note is that members from technical subcommunities overall tend to demonstrate a lower level of trust compared to members from non-technical ones. This could be because ChatGPT is still in its early stage of development, and thus it lacks the nuances that can be leveraged to solve advanced technical problems.

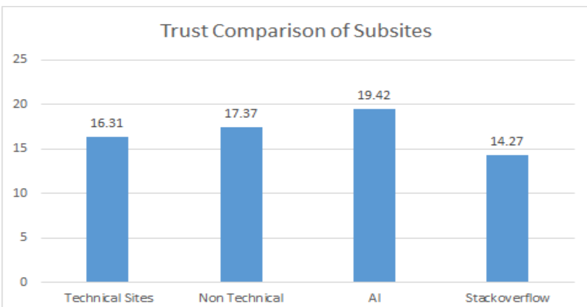

Figure 7. Prevalence of trust sentiment across different subsites

Finally, validation results from the application of the human labeling (Table 10) demonstrate a very low (< 0.05%) rate of false classification performed by the computational sentiment analysis, confirming that the algorithms that we applied were generally accurate.

Table 10. Comparison between manually and computationally classified sentiments

Total Number of Posts: 82	NLTK Classifications				
Manual Classifications		negative	neutral	positive	totals:
	negative	13	0	3	16 (20%)
	neutral	1	34	25	60 (73%)
	positive	0	3	3	6 (7%)
	totals:	14 (17%)	37 (45%)	31 (37%)	82

5. Conclusion and Discussion

5.1. Implications and Discussion

In this paper, we investigated the prevalent topics and sentiments regarding ChatGPT in an online Q&A community, Stack Exchange, and demonstrated how human intelligence tend to perceive artificial intelligence in collective knowledge sharing. We applied and cross-validated several computational tools to provide explorative answers to two specific research questions about 1) the major topics and sentiments regarding ChatGPT generated in the online Q&A community, as well as 2) how these topics and sentiments vary across different sub-communities. In sum, popular themes mainly include questions and remedies about ChatGPT-related technical aspects, and those who are familiar with AI technology tend to generate more trust and positive sentiment toward ChatGPT. To perform topic modeling, we applied and compared several commonly used algorithms: word-based and phrase-based LDA, BERTopic, LSA, as well as NMF. Our results highlight the importance of preprocessing steps and feature extraction in improving the accuracy of the topic models. With regard to sentiment analysis, we compared commonly used approaches such as lexicon-based NRCLex and Scikit-Learn's sentiment analysis model NLTK VADER, and a validation was also performed by incorporating the results from human labeling. Findings from sentiment analysis showed the critical role of threshold selection in accurately capturing the patterns of prevalent sentiments.

Our study is among the first group of research examining the social impact of ChatGPT, a most recent exemplary development of generative AI. Through the lenses of collective intelligence and knowledge sharing, we demonstrated ChatGPT-related public opinion from both professional programmers and lay users without interference, and presented the differences across AI-focused, technical, non-technical, and Stack Overflow subcommunities in terms of the prevalent discussion themes and sentiments toward ChatGPT. The various sentiments demonstrated in our study suggest that the public is generally enthusiastic about newly emerged AI technology, which can help make technology adoption easy and smooth. On the contrary, programmers and technology practitioners may be able to see the not-so-good aspect and may generate critiques which may seem discouraging in the short term but will be beneficial for technology development in the long run. Theoretically, our study contributes to the advanced understanding of the dynamic relationship between human intelligence and artificial intelligence, suggesting that the theoretical framework of AI-human research can benefit from adding the layer of human perceptions and emotions toward AI technology. Practically, our research offers preliminary insights for technology design and technology-related policy making by directing practitioners' attention to human-centered thinking and human-empowered technology management.

5.2. Limitations and Future Directions

There are several limitations of our study, setting the stage for future advanced research. First of all, our study focuses on a knowledge sharing community that may require a certain level of knowledge background in order for the members to engage in the discussion; and this can be a barrier when future researchers seek to generalize our findings. Therefore, we recommend that future researchers pay attention to the characteristics of their target online community and critically analyze the difference between the target online community and knowledge sharing community that our research studied. suggesting that improvements could be made by considering hashtags, emoticons, and other linguistic cues to obtain a more comprehensive understanding of ChatGPT and other AI algorithms. Further analysis also needs to be conducted on the factors that influence the distribution and evolution of ChatGPT- and AI-related themes and sentiments, such as user demographics and contextual nuances, to gain a deeper understanding of the reasons behind various public perceptions of artificial intelligence. From the perspective of computational methods, our study examined several existing language processing methods for topic modeling and sentiment analysis and suggests a few future directions that may worth exploring. First, advanced pre-processing methods that can help extract non-traditional linguistic components need to be developed to improve the efficiency of the language processing models. Moreover, researchers should be attentive to the balance between word stemming and information retaining when selecting and determining the language processing algorithms to be used. Last but not least, sufficient human-labeled dataset should be constructed in future research so that fine-tuning pretrained large language models in downstream tasks can be employed for future sentiment analysis. Finally, our explorative study calls for more in-depth research regarding the interaction between human intelligence and artificial intelligence, and we suggest that future experimental studies should be done in order to examine the behavioral impact of artificial intelligence as well as how human intelligence is influenced by artificial intelligence in both positive and negative ways.

6. References

Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic topic modeling: An experimental study on BERTopic technique. *Procedia Computer Science*, 189, 191-194.

Alfayez, R., Ding, Y., Winn, R., Alfayez, G., Harman, C., & Boehm, B. (2023). What is asked about technical debt (TD) on Stack Exchange question-and-answer (Q&A) websites? An observational study. *Empirical Software Engineering*, 28(2), 35.

Ainapure, B. S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S., & Bizon, N. (2023). Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches. *Sustainability*, 15(3), 2573.

Akhter, N., Vangara, R., Chennupati, G., Alexandrov, B. S., Djidjev, H., & Shehu, A. (2019). Non-negative matrix factorization for selection of near-native protein tertiary structures. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 70-73). IEEE.

Alhaj, F., Al-Haj, A., Sharieh, A., & Jabri, R. (2022). Improving Arabic cognitive distortion classification in Twitter using BERTopic. *International Journal of Advanced Computer Science and Applications*, 13(1), 854-860.

Ali, M.J & Djalilian, A. (2023) Readership awareness series–paper 4: chatbots and ChatGPT-ethical considerations in scientific publications. *Seminars in Ophthalmology*, March, pp. 1–2.

Almansoury, F., Kpodjedo, S., & El Boussaidi, G. (2022). Game development topics: A tag-based investigation on game development stack exchange. *Applied Sciences*, 12(21), 10750.

Ardichvili, A., Page, V., & Wentling, T. (2003). Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, 7(1), 64-77.

Baclic, O., Tunis, M., Young, K., Doan, C., Swerdfeger, H., & Schonfeld, J. (2020). Artificial intelligence in public health: Challenges and opportunities for public health made possible by advances in natural language processing. *Canada Communicable Disease Report*, *46*(6), 161.

Bashar, M. K. (2021). Exploration of Public Emotion Dynamics in Japan from Twitter Data during COVID-19. *6th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (Vol. 6, pp. 310-315). IEEE.

Biswas, S. S. (2023). Potential use of chat gpt in global warming. *Annals of Biomedical Engineering*, 51(6), 1126-1127.

Chiny, M., Chihab, M., Bencharef, O., & Chihab, Y. (2021). Analysis of sentiments conveyed through Twitter concerning COVID-19. *SHS Web of Conferences* (Vol. 119, p. 07003). EDP Sciences.

Choudrie, J., Patil, S., Kotecha, K., Matta, N., & Pappas, I. (2021). Applying and understanding an advanced, novel deep learning approach: A Covid 19, text based, emotions analysis study. *Information Systems Frontiers*, *23*, 1431-1465.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in Sociology*, 7.

El aoun, M. R., Li, H., Khomh, F., & Openja, M. (2021) Understanding Quantum Software Engineering Challenges An Empirical Study on Stack Exchange Forums and GitHub Issues. *IEEE International Conference on Software Maintenance and Evolution (ICSME)*, Luxembourg, pp. 343-354, doi: 10.1109/ICSME52107.2021.00037.

Elbagir, S., & Yang, J. (2018). Sentiment analysis of twitter data using machine learning techniques and scikit-learn. *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 1-5.

Flores, E., Ouyang, J., Lapointe, F., & Finnie, P. (2022). Nanotube abundance from non-negative matrix factorization of Raman spectra as an example of chemical purity from open source machine learning. *Scientific Reports*, 12(1), 11666.

Gazan, R. (2011). Social q&a. *Journal of the American Society for Information Science and Technology*, 62(12), 2301-2312.

George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1), 9-23.

Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., Badhani, P., & Tech, B. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. *International Journal of Computer Applications*, 165(9), 29-34.

Guzman, E. M., Konduru, R., Yassin, M., Konduru, S., & Jalali-Najafabadi, F. (2021). Application of Latent Semantic Analysis and Supervised Learning Methods to Automate Triage of Referral Letters for Spinal Surgery. *International Conference on e-Health and Bioengineering (EHB)*, pp. 1-4. IEEE.

Hassani, A., Iranmanesh, A., & Mansouri, N. (2021). Text mining using nonnegative matrix factorization and latent semantic analysis. *Neural Computing and Applications*, 33, 13745-13766.

Hern, A. (2023) AI Bot ChatGPT Stuns Academics with Essay-Writing Skills and Usability. Retrieved from https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Kaplan, C. A. (2001). Collective intelligence: A new approach to stock price forecasting. *IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace* (Cat. No. 01CH37236) (Vol. 5, pp. 2893-2898). IEEE.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Korzynski, P., Mazurek, G., Altmann, A., Ejdys, J., Kazlauskaite, R., Paliszkiewicz, J., Wach, K. and Ziemba, E. (2023), Generative artificial intelligence as a new context for management theories: analysis of ChatGPT, *Central European Management Journal*, 31(1), pp. 3-13. https://doi.org/10.1108/CEMJ-02-2023-0091

Krause, J., Ruxton, G. D., & Krause, S. (2010). Swarm intelligence in animals and humans. *Trends in Ecology & Evolution*, 25(1), 28-34.

Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1927-1936.

Leimeister, J. M. (2010). Collective intelligence. *Business & Information Systems Engineering*, 2, 245-248.

Liu, Z., Zhao, Y. C., Chen, S., Song, S., Hansen, P., & Zhu, Q. (2021). Exploring askers' switching from free to paid social Q&A services: A perspective on the push-pull-mooring framework. *Information Processing & Management*, 58(1), 102396.

Naumova, E.N. (2023) A mistake-find exercise: a teacher's tool to engage with information innovations, ChatGPT, and their analogs, *Journal of Public Health Policy*, 1–6.

Ng'ambi, D. (2013). Effective and ineffective uses of emerging technologies: Towards a transformative pedagogical model. *British Journal of Educational Technology*, 44(4), 652-661.

Novielli, N., Calefato, F. & Lanubile, F. (2015), The challenges of sentiment detection in the social programmer ecosystem. *Proceedings of the 7th International Workshop on Social Software Engineering*, ACM, pp. 33-40.

Panch, T., Pearson-Stuttard, J., Greaves, F., & Atun, R. (2019). Artificial intelligence: opportunities and risks for public health. *The Lancet Digital Health*, *1*(1), e13-e14.

Papoutsoglou, M., Kapitsaki, G. M., German, D., & Angelis, L. (2022). An analysis of open source software licensing questions in stack exchange sites. *Journal of Systems and Software*, 183, 111113.

Paul, J., Ueno, A., & Dennis, C. (2023). ChatGPT and consumers: Benefits, pitfalls and future research agenda. *International Journal of Consumer Studies*, 47(4), 1213-1225.

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 78(1), 84-93.

Qi, Y., Zhao, Y., & Zhu, Q. (2018). A Review on Participants' Knowledge Behaviors in Online Q&A Community. *Document, Information & Knowledge*, 3, 103-112.

Raschka, S., Liu, Y. H., Mirjalili, V., & Dzhulgakov, D. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python*. Packt Publishing Ltd.

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.

Rheingold, H. (2000). *The virtual community, revised edition: Homesteading on the electronic frontier*. MIT press.

Roose, K. (2023). The Brilliance and Weirdness of ChatGPT. Retrieved from https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html.

Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences*, 12(3), 148-156.

Rivas, P., & Zhao, L. (2023). Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot technology. *AI*, 4(2), 375-384.

Rudolph, J., Tan, S., & Tan, S. (2023) ChatGPT: bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6 (1).

Santos, T., Walk, S., Kern, R., Strohmaier, M., & Helic, D. (2019, May). Self-and cross-excitation in stack exchange question & answer communities. *World Wide Web Conference Proceedings*, pp. 1634-1645.

Shoufan, A. (2023). *Exploring Students' Perceptions of CHATGPT: Thematic Analysis and Follow-Up Survey*. IEEE Access.

Smirnov, O., & Hsieh, P. H. (2022). COVID-19, climate change, and the finite pool of worry in 2019 to 2021 Twitter discussions. *Proceedings of the National Academy of Sciences*, 119(43), e2210988119.

Tian, Y., Lo, D., & Lawall, J. (2014). Automated construction of a software-specific word similarity database. *Software Evolution Week-IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE),* pp. 44-53. IEEE.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.

Vincent, J. (2022). The scary truth about AI copyright is nobody knows what will happen next. *Verge (November 2022). URL: https://www. theverge. com/23444685/generative-ai-copyright-i nfringement-legal-fair-use-training-data*, *1*(4), 8.

Wedel, I., Palk, M., & Voß, S. (2021). A bilingual comparison of sentiment and topics for a product event on Twitter. *Information Systems Frontiers*, 1-12.

Yarger, L., Cobb Payton, F., & Neupane, B. (2020). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*, 44(2), 383-395.

Ye, D., Xing, Z., Li, J., & Kapre, N. (2016). Software-specific part-of-speech tagging: An experimental study on stack overflow. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pp. 1378-1385.

Zheng, C., & Xu, R. (2021). Molecular subtyping of Alzheimer's disease with consensus non-negative matrix factorization. *Plos One*, 16(5), e0250278.

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang,G., Zhang, K., Ji, C., Yan, Q., He, L., & Peng, H. (2023). A Comprehensive Survey on Pretrained Foundation Models: A History from Bert to Chatgpt. *arXiv preprint arXiv:2302.09419*.