



# Public perception of generative AI on Twitter: an empirical study based on occupation and usage

Kunihiro Miyazaki<sup>1\*</sup> , Taichi Murayama<sup>2</sup>, Takayuki Uchiba<sup>3</sup>, Jisun An<sup>1</sup> and Haewoon Kwak<sup>1</sup>

\*Correspondence:

[kunihirom@acm.org](mailto:kunihirom@acm.org)

<sup>1</sup>Indiana University Bloomington,  
Bloomington, USA

Full list of author information is  
available at the end of the article

## Abstract

The emergence of generative AI has sparked substantial discussions, with the potential to have profound impacts on society in all aspects. As emerging technologies continue to advance, it is imperative to facilitate their proper integration into society, managing expectations and fear. This paper investigates users' perceptions of generative AI using 3M posts on Twitter from January 2019 to March 2023, especially focusing on their occupation and usage. We find that people across various occupations, not just IT-related ones, show a strong interest in generative AI. The sentiment toward generative AI is generally positive, and remarkably, their sentiments are positively correlated with their exposure to AI. Among occupations, illustrators show exceptionally negative sentiment mainly due to concerns about the unethical usage of artworks in constructing AI. People use ChatGPT in diverse ways, and notably the casual usage in which they "play with" ChatGPT tends to be associated with positive sentiments. These findings would offer valuable lessons for policymaking on the emergence of new technology and also empirical insights for the considerations of future human-AI symbiosis.

**Keywords:** Twitter; Generative AI; ChatGPT; Public perception; Occupation

## 1 Introduction

Generative AI has garnered significant attention in recent years, demonstrating innovative advancements in various fields, including text, image, and software coding [1]. Especially, ChatGPT, a specialized conversational application of the Generative Pre-trained Transformer (GPT) series [2–4], has gained widespread popularity for its intelligence and seamless conversation capabilities [5–7]. The global debut of ChatGPT has led to substantial societal reactions [8], such as the official ban in public schools [9], a global call for a moratorium on developing GPTs [10], and the national debates about banning access to ChatGPT [11, 12].

The successful incorporation of emerging technologies into the societal fabric, without too much expectation and fear, is posited as an imperative responsibility for both policymakers and corporations [13–16]. Excessive expectations and unwarranted fear are said to harm society potentially [17]: the former can result in subsequent disappointment and

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

hinder research progress [18], while the latter may impede the adoption of potentially beneficial systems with stifling overregulation [19]. These potential negative consequences have been substantiated in the literature, establishing the importance of early understanding of public perception towards new technologies [20]. In fact, some AI researchers criticized ChatGPT, arguing that people's expectations are driven by hype [21], whereas the mainstream media has occasionally perpetuated a narrative suggesting that people are afraid of generative AI taking their jobs [22]. Despite the debate and importance of the topic, there is a lack of quantitative analysis examining people's perceptions of generative AI.

In this work, we investigate the public perception of generative AI on Twitter,<sup>1</sup> especially focusing on the user's occupation and usage of this technology. Public perceptions of emerging science and technology are known to exhibit substantial variability from group to group [24]. In the case of generative AI, the impact on occupations represents a paramount discussion surrounding emerging technologies, especially in the context of job displacement [25–28]. Notably, generative AI has the potential to influence not just IT specialists but also individuals in non-IT sectors due to its significant usability [29]. Also, Twitter serves as an excellent environment for gauging public reactions to generative AI, with a diverse range of individuals showcasing their interactions with these technologies on a daily basis [30]. Therefore, understanding their perception and reaction toward generative AI on social media would be particularly insightful for practitioners and policymakers considering the future human-AI symbiosis.

We set the research questions (RQs) below for analysis.

*RQ1: Which professions mention generative AI?* Our focus lies in understanding the attraction of users from diverse occupations towards generative AI. To achieve this, we compare the frequency of tweets discussing generative AI with randomly selected tweets.

*RQ2: What are the sentiments of various professions regarding generative AI?* We delve into the assessment of whether users on social media express positive or negative sentiments concerning generative AI. Furthermore, we investigate the relationship between the sentiments by occupations and their “AI exposure score” from existing research [26, 27, 31].

*RQ3: How do people interact with generative AI?* We get insights into how people use generative AI through the lens of ChatGPT, which is the most popular generative AI tool providing various use cases. We provide a detailed analysis of the prompts that users provide to ChatGPT, which is accomplished by studying the ChatGPT interfaces shared through screenshots.

Our contributions are as follows: (1) To the best of our knowledge, this is the first study to analyze social media perceptions of generative AIs with respect to occupation and use, which would provide policymakers with valuable insights about human-AI relationships. (2) We propose methodologies for extracting occupations from Twitter user profiles and analyzing texts extracted from ChatGPT images. (3) We create a comprehensive Twitter dataset focusing on generative AI, which will be publicly available upon publication.

---

<sup>1</sup> As of July 2023, Twitter changed its name to X [23], but we use the name Twitter because all of this research was conducted before the name change.

## 2 Background and related works

### 2.1 Public perceptions toward AI

Existing studies have examined people's perceptions of AI. [32] conducted extensive interviews with Americans on AI in 2018, finding that more people support AI development than oppose it (41% vs. 22%). However, more people believe AI will eliminate more jobs than it will create (49% vs. 26%), and 82% say AI should be managed carefully. In a 2022 Pew Research Center survey on AI in the U.S., more Americans expressed anxiety about the growing use of AI in their daily lives than those who felt the opposite about these prospects (37% vs. 18%), citing employment and privacy concerns as reasons for this [33]. [17] conducted a nationally representative survey of the British public, revealing that awareness of AI is not thorough, with 25% of the British public believing AI and robots are synonymous and the majority feeling uneasy about future scenarios regarding the impact of AI. The Pew Research Center also conducted an early interview survey on generative AI [34]. They found that, as of the end of December 2022, Americans who had heard about AI programs for writing news articles were overall unsure whether the development was a major advance for the news media. Apart from these survey-based studies, in [19], over 30 years of data from the New York Times showed that the debate about AI has been consistently optimistic, while concerns about the negative impact of AI on jobs have grown in recent years.

In this study, we conduct social media analysis focusing on users' perceptions of generative AI, which remains relatively unexplored in the literature. A primary advantage of this approach is the immediacy with which users respond to evolving trends, facilitating real-time insights [35]. This characteristic renders the platform apt for surveying perceptions of emerging technologies, which offers a convenient avenue to evaluate the potential societal impact of these novel technological advancements. Also, the availability of texts in posts enables an in-depth exploration of the underlying rationales for these sentiments [36]. Moreover, people often showcase how they use generative AI on Twitter, which provides a unique opportunity to study the usage of those technologies off the shelf. This study leverages those posts to analyze people's perceptions of the actual use of this technology [30].

### 2.2 Social media perception of emerging technologies

Social media has been used to understand public perceptions of emerging technologies, such as IoT [37], self-driving car [38], and solar power technology [39]. These studies found that a majority of social media users were generally positive or neutral toward these emerging technologies. On the other hand, they identified some problematic issues by text analysis, such as privacy and security issues on IoT devices.

While social media has been used to gauge perceptions of emerging technologies, there are relatively fewer studies focusing on the perception of AI or generative AI. [40] collected tweets about AI and found that sentiments expressed in AI discourse were generally more positive than the typical Twitter sentiment, and the public's perception of AI was more positive compared to expert opinions expressed in tweets. [41] analyzed the sentiment of 10,732 tweets from early ChatGPT users and discovered that a majority of them expressed positive emotional sentiments. [42] analyzed over 300,000 tweets and concluded that "#ChatGPT" is generally regarded as high-quality, with positive and joyful sentiments dominating social media. However, they also observed a slight decline in its recognition, with a decrease in joy and an increase in (negative) surprise.

While these studies offer insights into public sentiment on AI or generative AI on Twitter, our research stands out because we focus on analyzing perceptions by occupation, and more specifically, we examine sentiment related to the usage of ChatGPT. This novel approach allows us to provide a more detailed understanding of the various ways different occupations perceive and interact with generative AI technologies like ChatGPT.

### 3 Dataset building

#### 3.1 Selection of generative AI tools

To generalize our research, we consider various types of generative AI. We select them considering functionality and serviceability on social media. As for functionality, we select generative AI that is conversational (Chat), generates images given prompts (Image), complements codes in programming tasks (Code), and serves as a base model for various application tools (Model). As a result, we select the generative AI listed in Table 1. Since they include both models and services of generative AI, we use “generative AI tools” to refer to them.

#### 3.2 Generative AI tweets

*Tweet retrieval* We use Twitter Academic API to retrieve tweets of generative AI tools from January 1, 2019, to March 26, 2023, covering all their release dates. Our search keywords (case-insensitive for search on Twitter) are as exhaustive as possible and include variations (see Table 1). We exclude retweets and non-English tweets. In total, we retrieve 5,118,476 tweets from 1,475,262 users. Table 1 summarizes our dataset.

**Table 1** Summary of tweets about generative AI tools. Num (Orig.) indicates the initial tweet count, and Num (Fin.) indicates the noise-reduced tweet count. The keywords and numbers with parentheses are omitted from the dataset due to their difficulty in distinguishing from homonyms or the scarcity of their volume. Note that the release date can vary depending on definitions, but we basically chose the beta release date found on the Web

Category	Name	Release	Search keywords	Num (Orig.)	Num (Fin.)
Chat	ChatGPT	11/30/2022	chatgpt, chat gpt	2,106,256	1,751,065
	Bing Chat	02/22/2023	bing chat, bingchat	20,867	12,285
	Perplexity AI	12/07/2022	perplexity ai, perplexityai, perplexity.ai, perplexityask, perplexity ask	2395	(2112)
Image	DALL-E 2	07/20/2022	dall-e-2, dall-e 2, dall-e2, dall-e 2, dall-e2, dall-e-2, dalle-2, dalle2, (dalle 2)	196,714	105,270
	DALL-E	01/05/2021	dall-e, dall-e, (dalle)	303,470	139,103
	Stable Diffusion	08/22/2022	stable diffusion, stablediffusion	453,894	230,264
	Midjourney	03/21/2022	midjourney, (mid journey)	517,680	350,419
	IMAGEN	11/02/2022	imagen	173,620	(2805)
	Craiyon	04/21/2022	craiyon, dall-e-mini, dall-e mini, dall-emin, dall-e mini, dall-emin, dall-e-mini, dalle-mini, dallemini, dalle mini	66,873	34,465
	DreamStudio	08/20/2022	dreamstudio	10,713	(3475)
Code	GitHub Copilot	10/29/2021	copilot, co-pilot	325,105	112,315
Model	GPT-4	03/14/2023	gpt-4, gpt4, gpt 4	304,719	245,174
	GPT-3.5	11/30/2022	gpt-3.5, gpt3.5, gpt 3.5	11,226	(5749)
	GPT-3	06/11/2020	gpt-3, gpt3, gpt 3	317,852	237,190
	GPT-2	02/14/2019	gpt-2, gpt2, (gpt 2)	39,417	30,397

**Noise removal** To ensure the robustness of our analysis, we perform five types of noise removal on the acquired tweets. We remove (1) the tweets in which the name of authors or mentioned users include the name of generative AIs, but the text itself does not; (2) the tweets discussing a generative AI tool before its release date so that we gain a clearer understanding of the perception toward it; (3) the keywords that result in low tweet volumes (e.g., Perplexity AI, DreamStudio, and GPT-3.5) that have less than 10,000 tweets (0.3% of the total volume); (4) the tweets containing homonyms, which are often considered difficulties in analyzing social media posts on certain keywords [43]; and (5) the tweets generated by bot-like accounts.

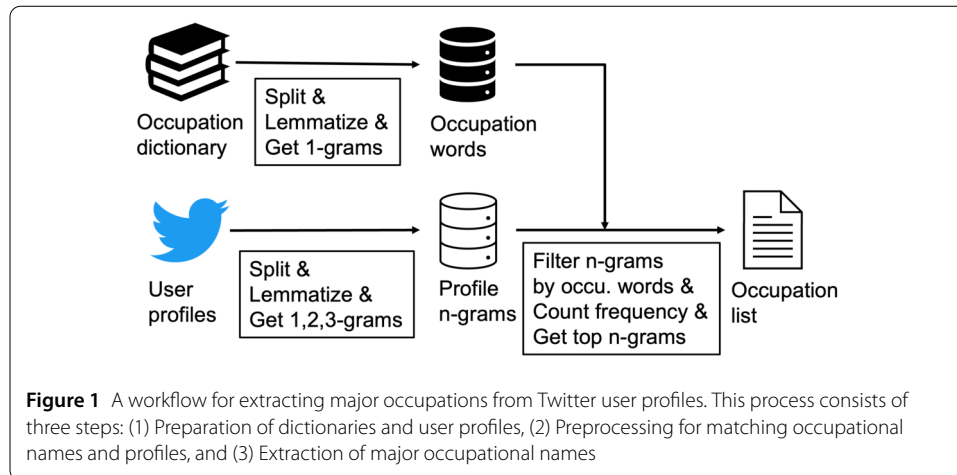
As the first three types of noise removals are straightforward, we provide a detailed explanation for the fourth and fifth types in the following. To identify and remove homonyms, we randomly sample 100 tweets from each keyword and manually annotate them to see if they refer to generative AI tools. We find that most of the keywords refer to the corresponding generative AI tools with almost 100% accuracy. However, several keywords are highly likely homonyms; thus, we remove them, which are annotated with parentheses in Table 1 (see Appendix for the accuracies and the detailed reasons for the homonyms). Among the keywords with homonyms, we retain GitHub Copilot (i.e., copilot and co-pilot) in our dataset as they do not have the alternative in their keywords, unlike other generative AI tools (note that its homonym indicates the second pilot). Furthermore, GitHub Copilot is a generative AI tool mainly for engineers/researchers, which enables us to analyze the potential difference in responses from engineers/researchers and lay people. To ensure GitHub Copilot's inclusion, we build a machine-learning model to classify whether a tweet is truly about GitHub Copilot or not. We first annotate additional tweets containing the two keywords (400 tweets in total) and fine-tune a RoBERTa-based model [44] with these annotations. The model achieves an F1 score of 0.94 in 5-fold cross-validation, allowing us to predict and remove noise from GitHub Copilot tweets. We also remove the tweets with "Microsoft 365 Copilot" because this is a different tool from GitHub Copilot and is not publicly available during this study. Additionally, we find the mixed use of DALL·E and DALL·E 2 in many tweets; thus, we combine them in our analysis and use the notation of "DALL·E (2)".

We then remove bot-like accounts. We use Botometer [45], a widely used tool for bot detection that assigns the bot score (0 to 1 scale) to Twitter accounts. Using Botometer for all the 1.4M users is difficult and costly due to the API limit. Instead, we extract users with more than five tweets in our dataset (127,547 users) to mitigate the impacts of the active bots. Using a score of 0.43 as a threshold following [46], a relatively conservative setting among previous literature [47], we regard 13,315 (10.4%) users as bots and remove their 872,893 tweets.

In total, we obtain 3,065,972 tweets and 1,082,092 users. We denote these generative AI-related tweets as  $T_{genAI}$  in the rest of the paper. The final number of tweets for each generative AI tool is displayed in Table 1. Note that when we count the tweets for each generative AI tool, we avoid overlap between substring models (e.g., when counting GPT-3, tweets with GPT-3.5 are not included).

### 3.3 Occupation extraction

To conduct an occupational analysis of the reaction to the emergence of generative AIs, we infer the major occupations of Twitter users from their profiles, inspired by [48]. AI-



though user profiles on Twitter can present issues with self-disclosure and falsehood, we acknowledge these limitations in our study, as also discussed in [49].

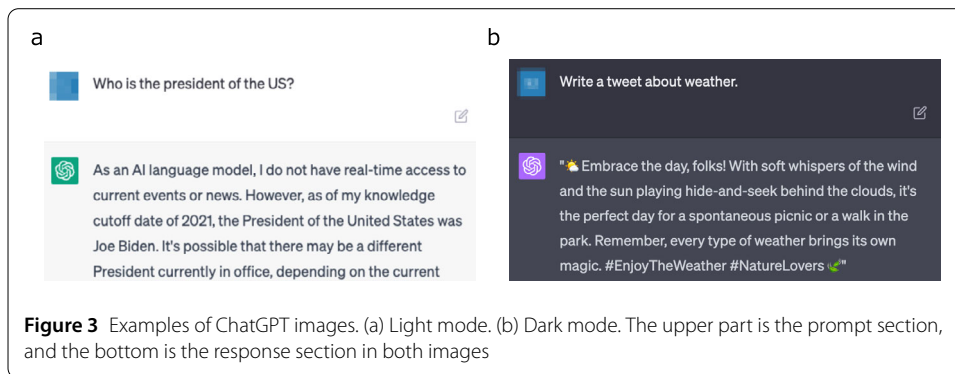
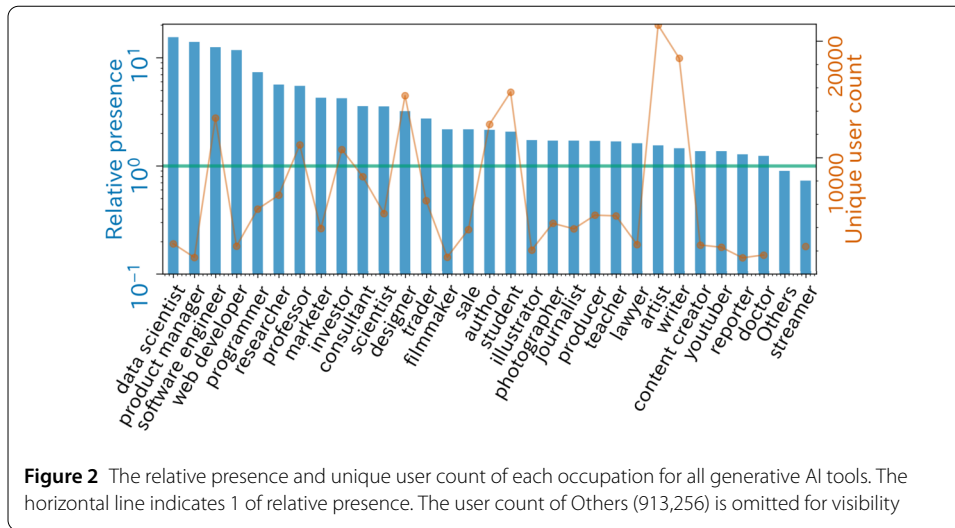
We employ dictionary-based methods to extract occupations. We find the accuracy of machine-learning approaches [50, 51] is unsatisfactory (about 50% of accuracy in nine-class classifications). In contrast, dictionary-based methods offer interpretability and are more commonly used in Twitter research [49, 52]. Figure 1 shows our workflow for inferring the occupations of Twitter users.

*1. Preparation of dictionaries and user profiles.* We start with the occupation names from all classes in the Bureau of Labor Statistics (BLS) Occupation Dictionary [53], which is frequently employed for data-driven occupational analysis with O\*NET data [54]. However, it has formal occupational names, failing to capture emerging occupations (e.g., YouTubers). To address this limitation, we further use Indeed’s occupation list [55], a more up-to-date resource. For user data to use for matching, we retrieve the profile texts of all users from  $T_{genAI}$ .

*2. Preprocessing for matching occupational names and profiles.* As the words in the occupation dictionary may not align with the text in the user profiles, and some occupation names consist of multiple words, we conduct preprocessing to mitigate their impacts. First, we split all occupation names (from the dictionaries) and the profile texts into 1-grams. Next, we lemmatize all these 1-grams after removing stop words. Finally, we extract unique 1-grams from the dictionary and 1, 2, and 3-grams from the profile texts, designating them as “Occupation words” and “Profile n-grams,” respectively.

*3. Extraction of major occupational names.* To extract the major occupation names in Twitter users’ profiles [48], we select the most frequent n-grams related to occupation. Initially, we retain only Profile n-grams that contain Occupation words. We then count the frequency of the Profile n-grams across all profile texts. For cases where multiple occupations are mentioned in the same user’s profile, we count each occupation separately (i.e., instead of selecting only one occupation). Then, two authors of this study review the Profile n-grams in descending order of frequency, discussing from the top whether they indicate occupation, and extract the top 30 occupation names, which can be found in Fig. 2.





### 3.4 Randomly sampled tweets

In addition to the tweets collected by Twitter Academic API, we utilize the dataset of 10% randomly sampled tweets collected by Twitter Decahose API, which spans a three-year period from May 2020 to April 2023, and we detect users' occupations in the same manner as in  $T_{genAI}$ . From this dataset, we extract the subset that is one-tenth of this data (i.e., 1% sample of Twitter) of all English tweets, which we denote  $T_{rand}$  in the rest of the paper. We use  $T_{rand}$  as a reference to the usual tweets of occupational users to mitigate the inherent bias of Twitter in our analysis. After removing accounts with a Botometer score of 0.43 or higher, the volume of  $T_{rand}$  is 3,601,890, excluding retweets.

### 3.5 Extraction of images and prompts

In RQ3, we aim to understand how people interact with generative AI tools, with ChatGPT as our chosen illustrative example. To this aim, we analyze screenshots of interactions with ChatGPT, which many users voluntarily share on Twitter [30]. These images typically feature a single-color background separating the prompt and response sections, with two primary variations: light mode and dark mode, as in Fig. 3. This design allows us to analyze the images by simple rule-based methods.

First, we identify whether or not these images are screenshots of the ChatGPT interface. In essence, we consider the image a ChatGPT screenshot if two of the top three RGB colors of the image correspond to the range of RGBs of the prompt and response sections

(the detail of RGBs are omitted for the reason of space). We use the Python package Pillow [56] for RGB extraction. This simple approach achieves an F1 score of 0.93 when tested on 300 randomly sampled and manually annotated images. The primary cause of misclassifications arises when ChatGPT generates programming code—the edges of the code section displayed by ChatGPT share the same RGB values as the prompt section, which sometimes confuses our rule-based classification process. We address this problem on a rule basis in the text extraction step—since the text on the edge of codes is the only name of the language and “Copy code,” we can remove this section by a simple rule.

Next, from the identified screenshots of ChatGPT, we extract prompts. We isolate the prompt section and apply Optical Character Recognition (OCR) using the Python package pytesseract [57]. Since pytesseract frequently produces errors when processing dark-mode images, we address this issue by inverting the RGB values for dark-mode images. As a result, we get 93,623 prompts extracted from ChatGPT screenshots.

#### 4 RQ1: which occupation mention generative AI?

In addressing RQ1, we begin with counting the users of each occupation who mention each generative AI tool. We employ user counts instead of tweet counts to alleviate the potential impact of particularly vocal users. As these counts are influenced by the inherent distribution of occupations on Twitter, we normalize this by employing the distribution obtained from  $T_{rand}$  (Sect. 3.4). This enables us to compute the *relative presence* of each occupation that reflects the interest in generative AI more accurately. We define the relative presence as follows:

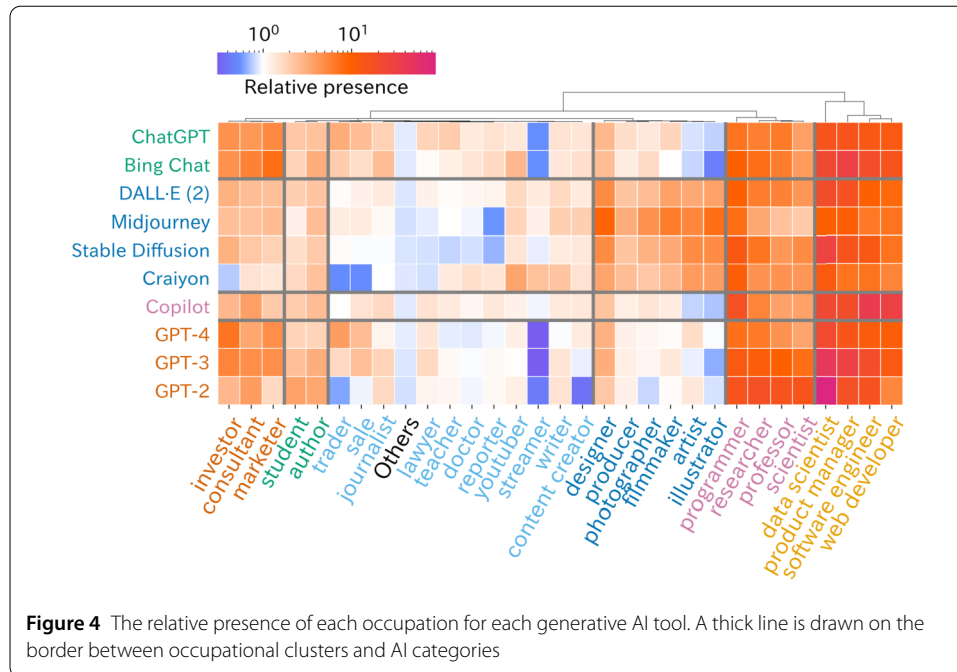
$$RelativePresence_i = \frac{|U(T_{genAI}^i)|/|U(T_{genAI})|}{|U(T_{rand}^i)|/|U(T_{rand})|}, \quad (1)$$

where  $|U(T_{genAI})|$  and  $|U(T_{rand})|$  are the numbers of users in  $T_{genAI}$  and  $T_{rand}$ ,  $i$  is a particular occupation, and  $|U(T_{genAI}^i)|$  and  $|U(T_{rand}^i)|$  are the numbers of users of occupation  $i$  in  $T_{genAI}$  and  $T_{rand}$ , respectively. This measure reveals how much each occupation's presence in discussing generative AI differs from its usual engagement on Twitter. If the value is greater than 1, the occupation exhibits a higher presence in tweets about generative AI than usual.

Figure 2 shows the relative presence concerning all generative AI tools by each occupation, along with the corresponding number of unique users. For reference, we calculate the relative presence for tweets that do not include the top 30 occupations and include them as “Others.” The data reveals that most occupations in this study (i.e., occupations with a large volume of tweets regarding generative AI) exhibit a relative presence greater than 1, indicating their higher presence than usual. Notably, attention to generative AI is evident not only among IT-related occupations but also among those with less connection to IT, including lawyers, sales, traders, and teachers; however, only streamers show a relative presence lower than 1. This suggests that generative AI has captured the interest of a wide range of occupations.

Then we ask, is the attention to generative AI equally distributed to various generative AI tools? Fig. 4 presents the relative presence of each generative AI tool across different occupations. The predominance of warmer hues, which indicate a relative presence greater than 1, shows that most occupations mention the variety of generative AI tools





**Figure 4** The relative presence of each occupation for each generative AI tool. A thick line is drawn on the border between occupational clusters and AI categories

more than usual. In the figure, we sort the occupations by Ward's hierarchical clustering method, thereby creating distinct clusters. IT-related occupations (e.g., data scientists), researchers, and business-related groups (e.g., marketers) express a broad interest in generative AI, although their interest levels vary. Conversely, occupations associated with visual content (e.g., designer and photographer) display a heightened interest in image-generative AI tools. For the remaining occupations, there is largely an increase in interest in various generative AI tools. Overall, the correspondence between clusters of occupations and the generated AI tools seems reasonable.

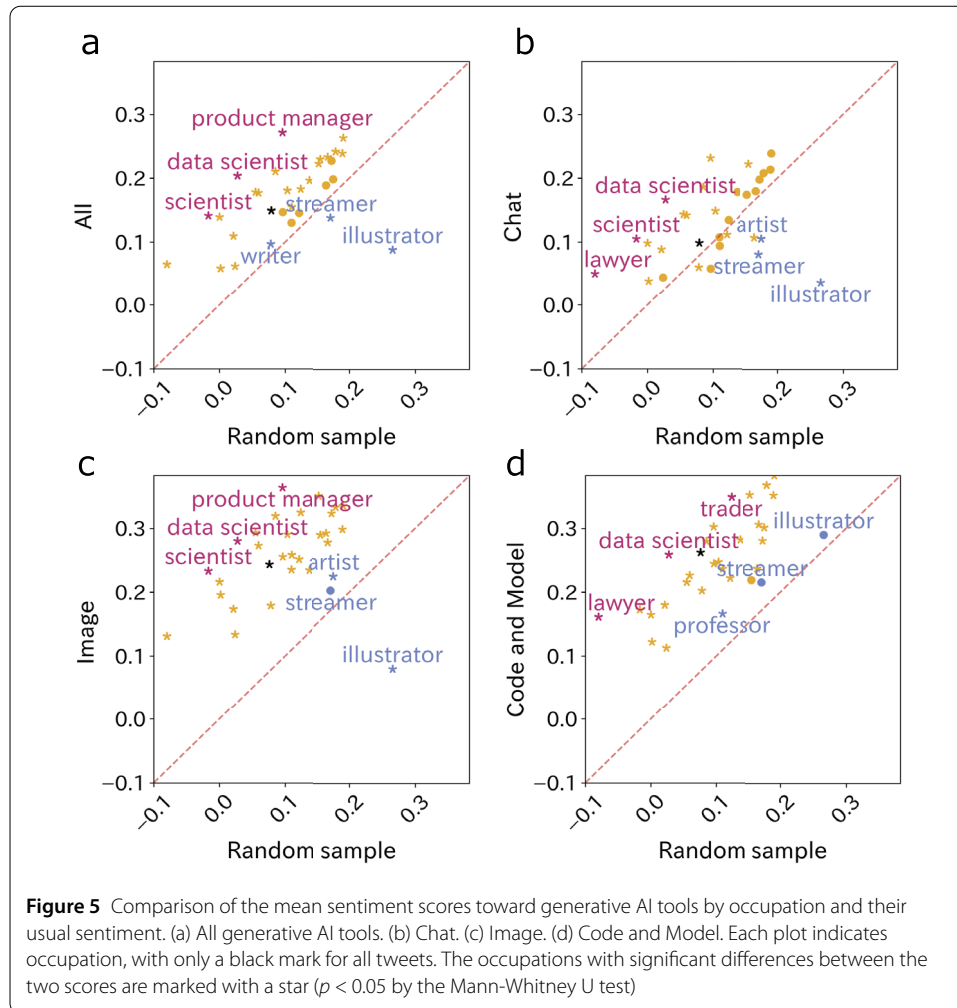
## 5 RQ2: what are the sentiments of different occupations toward generative AI?

### 5.1 Sentiments expressed by each occupation

As a proxy of the perception of each occupation toward generative AI, we analyze sentiment expressed in tweets about generative AI tools and conduct topic modeling to understand the sentiment better.

We employ a RoBERTa-based model pre-trained on Twitter data [44] to classify tweets into three classes: Positive, Neutral, and Negative. We manually validated this model with 99 tweets, 33 tweets for each class, resulting in an F1 score of 0.768, which is comparable to the reported performance of the model [44]. Since this model can output the probabilities of each model, in order to facilitate comparison between tweets, we aggregate the probabilities into a single value per tweet, namely *sentiment score*, by subtracting the Negative probability from the Positive probability (eventually, the score is scaled from  $-1$  to  $1$ ). This value approaches  $0$  if the Positive and Negative values have similar values or if the Neutral probability gets greater.

We compute this sentiment score for every tweet in  $T_{genAI}$  and then aggregate them regarding each generative AI tool by occupation. To reduce the influence of vocal users, we average the scores over the same users and then average them for each occupation.



To account for the inherent bias in sentiment by occupation, we also compute their usual sentiments from the randomly sampled tweets,  $T_{rand}$ .

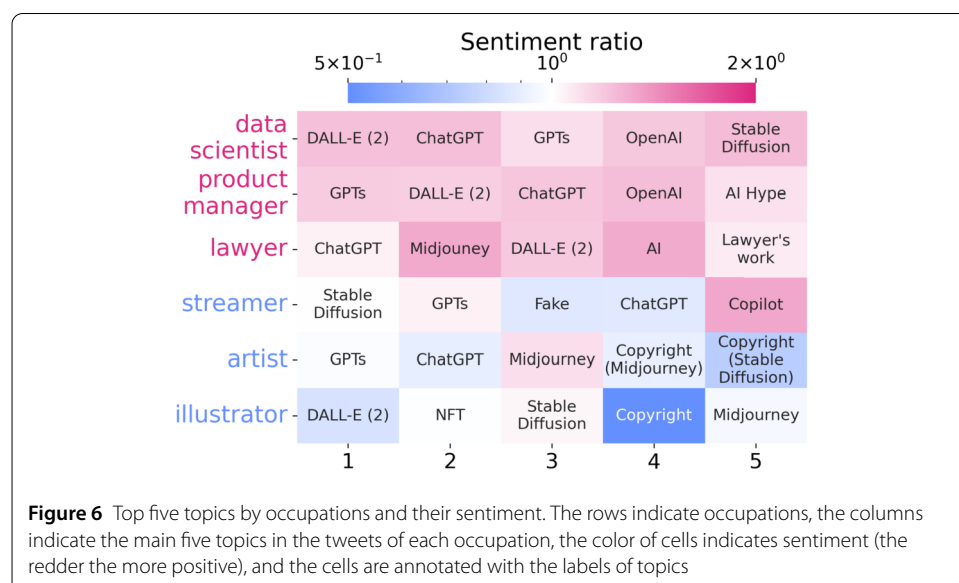
Figure 5 displays the comparison of the sentiment of each occupation with their usual sentiments for all generative AI tools as well as for the Chat, Image, and Code/Model categories (Sect. 3.1). The  $y$ -axis represents the sentiment in  $T_{genAI}$ , while the  $x$ -axis shows the sentiment of tweets in  $T_{rand}$ . For clarity, we label a few occupations that have the most pronounced differences in sentiment between  $T_{genAI}$  and  $T_{rand}$ . Specifically, after standardizing each value to the  $[0, 1]$  scale (since the sentiment score originally ranges  $[-1, 1]$  scale), we divide the sentiment score in  $T_{genAI}$  by that in  $T_{rand}$  (called the *sentiment ratio*) and highlight the top and bottom three values each. The red dotted line in the figure indicates the  $y = x$  line, which means no difference in sentiments when an occupation mentions generative AI compared to usual. We also plot the black marks representing all tweets that include all occupations for reference.

In the figures, most occupations are located above the red line, suggesting that they express more positive sentiments about generative AI than their random tweets on Twitter. By category, Image and Code/Model exhibit highly positive sentiment compared to Chat. Some occupations are skewed toward negative sentiments for Image and Chat, while all occupations are more positive than usual for Code/Model. Overall, the occupations ex-

pressing particularly positive sentiment include product managers, data scientists, scientists, traders, and lawyers. Lawyers, interestingly, have a more positive sentiment despite having a relatively lower sentiment usually. Indeed, their sentiment scores are 0.07, 0.13, and 0.16 for Chat, Image, and Code and Model related tweets, respectively, while it is  $-0.08$  for the random tweets. On the other hand, occupations with comparatively negative sentiments include illustrators, streamers, artists, and writers. In particular, illustrators are the most negative by a significant margin.

We conduct topic modeling to better understand sentiments expressed by each occupation. We use transformer-based methods, such as BERTopic [58], a combination of transformer-based text embedding and density-based clustering HDBSCAN [59], which can identify cohesive groups of texts in the embedding space. It is well known that BERTopic is effective for extracting more detailed topics [60] compared to other topic modeling methods such as Latent Dirichlet Allocation [61]. Since BERTopic yields various numbers of clusters depending on the datasets and the parameters (we set it as default), we focus on the largest five topics.

Figure 6 shows the top five topics and their sentiment by occupation. The authors assign labels to the topics through the manual examination. We select the most pronounced occupations in the difference in sentiment ratio. Occupations with positive sentiments generally praise generative AI tools or admire the progress of AI, e.g., “Seeing all these use cases of ChatGPT gets me super excited. I can’t wait to see what it’s going to be doing in a couple of years,” and we assign the names of generative AI tools to topics like them. As a more specific topic, lawyers discuss the use of generative AI in the lawyer’s work [62], e.g., “This is amazing. You’re wonderful for sharing this. I am getting to hear about ChatGPT for the first time, but I think I have to leverage on its opportunities especially in my field as a lawyer.” Regarding negative sentiments, illustrators and artists highlight concerns related to copyright issues, expressing notably negative sentiments, e.g., “Stable Diffusion uses datasets based on art theft. Don’t pretend you’re doing the right thing here, CSP. You know full well this is a shitty, unethical move.”



## 5.2 Comparison to AI exposure score

The results so far indicate that people in occupations with high activity on Twitter have largely positive sentiments about generative AI tools. We then ask what factors may tie to these impressions. In particular, we investigate how much the exposure of AI to people's jobs, which is frequently discussed in many contexts [25, 26, 28], shapes their views on AI.

AI's impact on occupation has been a crucial topic of the relationship between humans and AI. The fear of emergent technology displacing occupations has been a historical concern [17]. As exemplified by the Luddite movement [63], people may express fear and hostility toward automation technologies that threaten their jobs, which is perhaps related to the negative sentiments of each occupation.

Here, we compare AI exposure scores, which are the proxy of the potential impact of AI on occupations, with the sentiment scores expressed by occupations on Twitter. We utilize two AI exposure scores that were developed independently: one by [26], who considered the overlap between job task descriptions and patent texts, namely *text-based* AI exposure score, denoted by  $E_t$ , and another by [27], who conducted a survey on the relationship between AI and occupational tasks, namely *survey-based* AI exposure score, denoted by  $E_s$ . A higher score for an occupation means that the occupation is likely to be more exposed to AI.

The purpose of these two scores is the same: to quantify how much exposure each occupation has to AI, but we recognize two major differences in how the scores are calculated. First, although both commonly break down each occupation into its factors to calculate their relevance to AI, the nature of the factors is different. In  $E_t$ , each occupation is broken down into task levels, which use the task descriptions for each occupation in the O\*NET database.  $E_s$  breaks down each occupation into ability levels, which use categories of AI applications (e.g., image recognition, speech recognition, etc.) based on definitions by the Electronic Frontier Foundation (EFF).<sup>2</sup> Second, the relevance of each task/ability to AI is quantified in different ways.  $E_t$  employs a text processing-based approach, counting how many verb-noun pairs that appear in AI-related patents appear in each occupation's task description, and integrating them to compute the AI exposure score.  $E_s$  asked crowd workers (using the crowdsourcing service Amazon Mechanical Turk<sup>3</sup>) to answer the values of how relevant the AI applications by EFF are to each occupation, and integrated them to compute the AI exposure score.

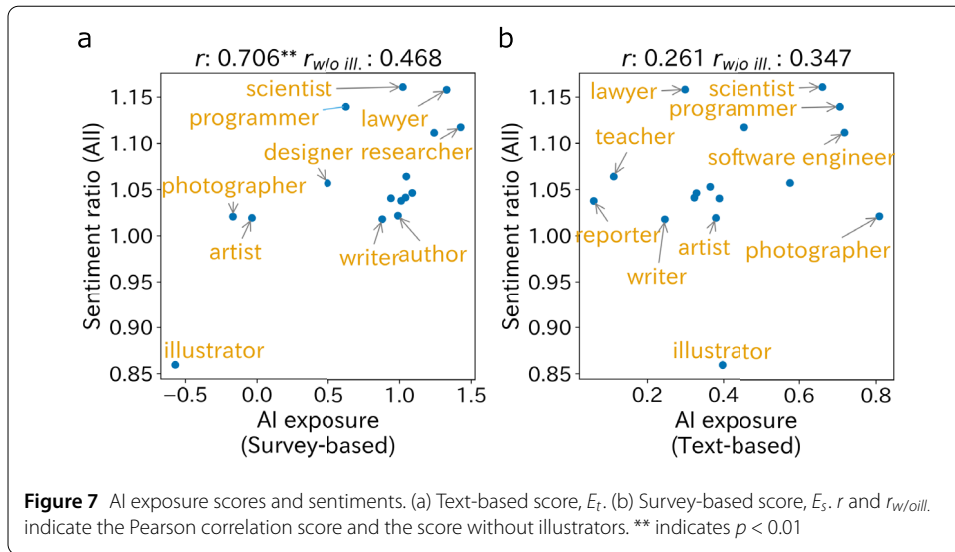
According to the original papers,  $E_t$  is an indicator of "exposure to automation", and  $E_s$  is "agnostic as to whether AI substitutes for." Both scores are linked to the occupation list from the BLS, making them applicable to our research. We compute the correlation coefficient between  $E_t$  and  $E_s$  and find the coefficient is 0.027 with  $p = 0.486$ , meaning that both scores capture different perspectives. Following [25], this study does "not distinguish between labor-augmenting or labor-displacing effects," and uses them as indicators of the strength of the relationship between AI and occupation.

We match our top 30 occupations with their  $E_t$  and  $E_s$ . As our occupation list incorporates the Indeed dictionary in addition to the BLS dictionary, 14 occupations are not matched among 30 occupations. Also, if more than one occupation in BLS is matched, we use their mean scores (e.g., "designer" matches both "Fashion Designers" and "Graphic

---

<sup>2</sup><https://www.eff.org/>.

<sup>3</sup><https://www.mturk.com/>.

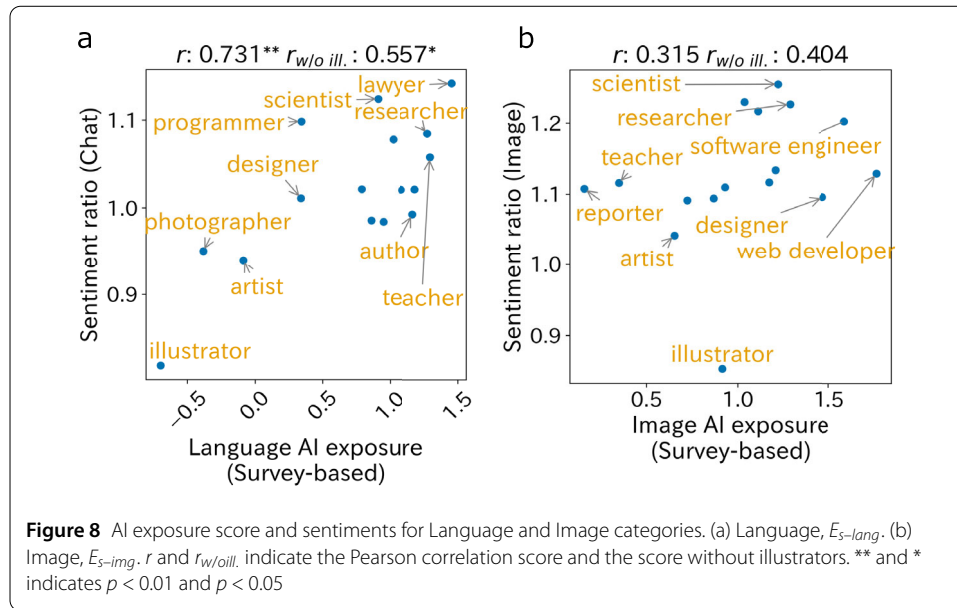


Designers”). For the corresponding sentiment, we use the sentiment ratio, which is the sentiment score for  $T_{genAI}$  adjusted by the one for  $T_{rand}$  (Sect. 5.1).

Figure 7 shows the results. Note that not all occupations are annotated due to clarity. Surprisingly, we observe no negative correlation; rather, the result leans toward a positive correlation. For  $E_t$ , the Pearson correlation is 0.262 ( $p = 0.326$ ), and without the illustrator, it is 0.344 ( $p = 0.209$ ). For  $E_s$ , the Pearson correlation is 0.706 ( $p = 0.002$ ), and without the illustrator, it is 0.468 ( $p = 0.079$ ). In other words, negative discourse about generative AI is not more likely to be observed even if one’s job has a strong degree of exposure to AI; if anything, positive arguments tend to be observed. Given that both scores indicate the potential impact of AI on occupations, it is counterintuitive that occupations more likely to be exposed to AI express more positive sentiments on Twitter. Several remarks of commendation on generative AI highlighted during the experiment related to Fig. 5 could offer insight into this observation. Paradoxically, people more likely to be exposed to AI might be better equipped to recognize the potential benefits and practicality of integrating AI into their work, thereby appreciating its value more.

Furthermore, the research team that developed  $E_s$  recently created an additional survey-based AI exposure score specifically assessing the impact of generative AI for image and language, denoted by  $E_{s-img}$  and  $E_{s-lang}$ , in [31]. Here, we investigate the connections between these scores and the sentiment ratios associated with the generative AI tools for both images and language. The results again show a noticeable positive correlation between the two, depicted in Fig. 8. For image-generative AI, the Pearson correlation is 0.315 ( $p = 0.235$ ), and without the illustrator, it is 0.404 ( $p = 0.135$ ). For language-generative AI, the Pearson correlation is 0.731 ( $p = 0.001$ ), and without the illustrator, it is 0.557 ( $p = 0.031$ ). We observe a notably high correlation for language-generative AI, potentially due to people easily associating it with a labor-augmenting feature.

In summary, an overview of the sentiment expressed by Twitter users highlights a prevailing positive outlook towards generative AI, with a few exceptions, notably among certain professions like illustrators. Contrary to our initial assumption that individuals with higher AI exposure might exhibit more pessimistic views about job stability, we do not find any concrete evidence to support the notion that a greater AI impact on a profession cor-



responds to a more negative discourse. Rather, the result indicates those who have higher exposure to AI tend to have more positive sentiments. This was also the case when the analysis is focused on generation AIs for image and language.

## 6 RQ3: how do people interact with generative AI?

Going beyond how people perceive generative AI, in our RQ3, we aim to get insights into how people “use” generative AI through the lens of ChatGPT, which has brought significant global interest in generative AI [5]. In particular, a trend of posting screenshots of their usage of ChatGPT on social media [30] provides an invaluable resource for our investigation. In contrast to image-generative AI tools that have a very focused usage, ChatGPT can be used for many different scenarios. Thus, we focus on ChatGPT as a starting point for understanding the use of generative AI. Exploring upcoming generative AI tools designed for different tasks presents an intriguing avenue for future research.

We classify ChatGPT prompts (Sect. 3.5) into topics using BERTopic. The top 10 topics are presented in Table 2. Examples include (1) work assistance such as translation and tweet writing, (2) casual and entertaining applications like rap, games, poetry, and ASCII art, and (3) more serious topics such as COVID-19 and AI models.

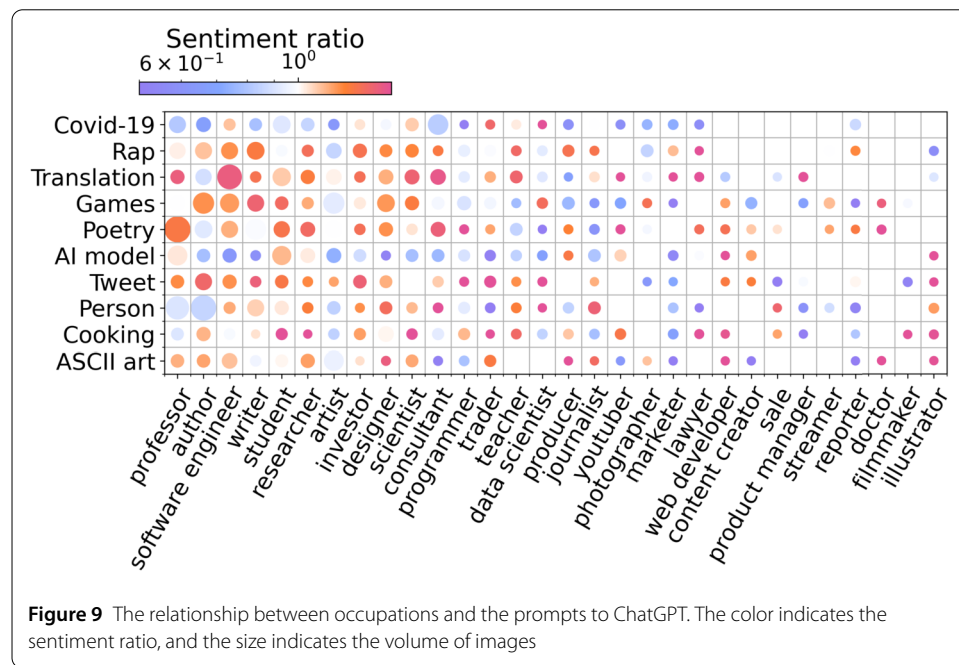
Figure 9 shows the different usages of ChatGPT and sentiment expressed in tweets when people share screenshots of ChatGPT’s results based on occupations. In this plot, the size of the circle represents how many times an occupation shares the corresponding prompt (minimum 1, maximum 25). An empty cell without a circle indicates that the occupation has not shared the screenshots of the prompt with the corresponding topic. The color shows the sentiment ratio of the tweet texts. Occupations are listed in the order of the volume of their ChatGPT images, with professors posting the most.

Prompts querying serious topics such as COVID-19 and AI models have many circles in cooler colors, indicating a negative sentiment (the mean sentiment ratios are the lowest for these two topics, 0.871 and 0.900, respectively). Conversely, it is evident that many circles are of a warmer color, and most of the occupations have positive sentiments with work assistance (e.g., translation and tweet writing) and casual interaction with ChatGPT



**Table 2** Top 10 topics of prompts to ChatGPT. We hide personal names with (personal name). Topic words are determined by cluster-wise TF-IDF (c-TF-IDF) and representative text indicates the core points of each cluster as output by HDBSCAN

Topic	Count	Topic words	Representative text
Covid-19	650	vaccine, covid, covid19, virus, pandemic	Are the covid vaccines safe?
Rap music composition	637	rap, lyric, song, verse, eminem	Write me a rap song in the style of eminem
Multilingual translation	583	translate, spanish, english, japanese, translation	Translate that into English, please.
Role-playing games	547	game, adventure, rpg, play, zork	do you want to play a game with me
Rhymed poetry	445	rhyme, poem, rhyming, poetry, twofer	write a poem about AI
AI language model	408	large, language, model, trained, assistant	do you or any other large language model understand what you say?
Tweet writing	363	tweet, chatgpt, twitter, post, thread	Write a Tweet about ChatGPT
Personal identification	359	chris, who, matt, giuseppe, sean	But who is exactly this (personal name)?
Cooking and recipes	352	recipe, ingredient, dish, meal, garlic	using only these ingredients find me an easy recipe for a meal. beans and bread.
ASCII art creation	347	ascii, art, draw, drawing, face	do you know ascii art?



(e.g., poetry, rap, and ASCII art creation). This suggests that ChatGPT's answers to those prompts may satisfy users' expectations, and users expressed their contentment with the results, fostering a positive sentiment within these interactions.

## 7 Discussion and conclusion

### 7.1 Main findings

Our investigation finds that many occupations, even those unrelated to IT, posted more tweets about generative AI compared to their usual tweeting frequency (RQ1). Although AI remained relatively unfamiliar to the general public until recent years [17, 33], our re-

sults show that the emergence of generative AI may have played a significant role in bridging the gap between AI and the general public.

Moreover, many occupations, with only a few exceptions, predominantly convey positive sentiments regarding generative AI on Twitter (RQ2). We do not find strong evidence of widespread fear among typical Twitter users, as suggested by certain news reports [22]. The trend of positive sentiment toward AI is consistent with existing studies [19, 40]. Our robust experiment design, comparing each occupation's sentiments about generative AI to their "usual" sentiments, lends further support to this finding.

Surprisingly, occupations with higher exposure to AI exhibited more positive sentiments toward generative AI (RQ2). This result was unexpected because intuitively we anticipated the occupations exposed to AI tend to be scared of the displacement of jobs, but this study indicated the opposite. One possible explanation is that people who have high AI exposure may possess a better understanding of the technology's advantages, which enables them to discern the potential benefits and pragmatism of AI integration within their professional domain, leading to a greater appreciation of its intrinsic value. Indeed, studies have shown that familiarity can be a factor in positive perception toward emerging technology. For example, the previous survey reported that the more people know about self-driving cars or related services, the more likely they are to be supportive of autonomous vehicle [64] and to believe that such technology is a good idea [33].

On the other hand, illustrators generally have negative sentiments toward generative AI, with the primary concern revolving around the use of their artwork for training AI without consent (RQ2) [65]. The concerns raised by illustrators highlight the ethical implications surrounding AI technology [66]. While new technologies generally garner positive sentiments, a few topics, such as this ethical issue, can have exceptionally negative sentiments; thus, it is essential for companies and governments to address these specific concerns without overlooking them to ensure the responsible development of AI technologies.

The use of ChatGPT is often in the topics about work assistance or, interestingly, casual context and is associated with more positive sentiments (RQ3). Indeed, previous studies in human-robot interaction suggest that active engagement in playful activities with robots can foster closer human-robot relationships [67]. The findings in our study may carry implications for the evolution of human-AI relationships.

The issue of hallucinations [68], which has been extensively discussed in the research community, does not emerge in our analysis. Only streamers notably mentioned "Fake" as a negative aspect (RQ2). This suggests that, as of the time of this study, hallucinations may not be a dominant concern in public discussions around generative AI. However, as generative AI becomes more widely used over time, it is plausible that this issue may gain more prominence in public discourse.

## 7.2 Limitations and future work

*Generalization of our findings* It is well known that Twitter demographics are skewed [69]. Thus, we acknowledge that our findings based on Twitter users should not directly apply to the entire population. For example, since people on Twitter are said to be younger than the average [70], they may have a bias with heightened information sensitivity and more forward-thinking toward new technologies, making userbase of our dataset potentially more tech-oriented. Yet, even with this potential skew, our comparisons between occupations and the topic analysis provide meaningful insights, although mitigating these biases remains an essential task for future research.

*Further comparisons of generative AI tools* While our study primarily focuses on occupations, it would indeed be intriguing to delve deeper into the comparison of different generative AI. In particular, we analyzed usage only for ChatGPT with consideration of its huge popularity. As more generative AI chat services, such as Bard by Google, are arriving and more functionality (e.g., handling multi-modality) will be expected in the future, it would be interesting to see how user interaction with those services evolves over time.

*Selection of generative AI* Our analysis included a wide variety of generative AI tools, from image processing to code generation, available as of March 2023. Nevertheless, this choice may limit the generalizability of our findings to other generative AI models and technologies. Future research could expand the analysis to include a wider range of generative AI tools and services to obtain a more comprehensive understanding of the perception and usage of generative AI across occupations.

*Detection, selection, and sampling bias of occupations* The detection of users' occupations using Twitter profiles may not be accurate due to potential self-disclosure and falsehood [49]. Moreover, our analysis focuses solely on the major 30 occupations, not encompassing all professions, which may be affected by inherent bias on Twitter. Future research should strive for a more comprehensive study, possibly leveraging external data sources, that mitigates these biases. Also, the occupations selected for this study were the ones that appeared most frequently in our dataset. For this reason, it is assumed that there is a bias in the selection from the Twitter population. However, we believe that we were able to mitigate these biases by comparing them with the usual tweets in the analysis.

Additionally, the occupations we selected are the ones that tweeted about genAI the most, which may be in favor of knowledge-intensive or creative professions and more common in social media spaces. The conventional or manual forms of occupation (e.g., bricklayers), did not seem to manifest in the data. It is widely known that the distribution of Twitter users does not correspond to the real-world distribution [71], and an analysis that selects occupations aligning with real-world distributions and a focus on traditional occupations would be important in future works.

*Other types of user attributes* In addition to occupations, there may be other relevant factors influencing users' perception and usage of generative AI, such as age, education, number of followers, or cultural background [72], which would provide a more nuanced understanding of how different user segments engage with generative AI in future work.

*Other languages* Our study focuses on English-language tweets; however, conducting an analysis of how responses to ChatGPT varied across different languages would be particularly intriguing, given the multilingual capabilities of the AI model. Future research could extend the analysis to include tweets in other languages and from users in different countries to explore cross-cultural differences in the perception and usage of generative AI.

## Appendix

### A.1 The accuracies and the detailed reasons for the homonyms

The following are keywords related to generative AI tools, but were not adopted because of their many homonyms. The noise probability and reasons noise are also noted:

- “dalle”: 46% (Italian word, human’s name)
- “dalle 2”: 37% (Italian word)
- “imagen”: 94% (Spanish word, typo of “imagine”)
- “mid journey”: 26% (mid trip)
- “copilot”: 34% (second pilot)
- “co-pilot”: 85% (second pilot)
- “gpt 2”: 22% (e.g., GPT 2024)

#### Acknowledgements

We appreciate the comments from Yong-Yeol Ahn, Rachith Aiyappa, and Francesco Pierri on this work. We also appreciate the support concerning Twitter’s randomly collected dataset from the OSoMe engineering team, especially Nick Liu.

#### Funding

This work does not receive any funding.

#### Abbreviations

AI, Artificial Intelligence; API, Application Programming Interface; BERT, Bidirectional Encoder Representations from Transformers; BLS, Bureau of Labor Statistics; COVID-19, Coronavirus Disease 2019; GPT, Generative Pre-trained Transformer; IT, Information Technology; O\*NET, Occupational Information Network; RGB, red, green and blue; RoBERTa, Robustly Optimized BERT Pretraining Approach; TF-IDF, Term Frequency – Inverse Document Frequency.

#### Data availability

The dataset is publicly available here: [https://github.com/Mmichio/public\\_perception\\_of\\_genAI\\_public](https://github.com/Mmichio/public_perception_of_genAI_public).

## Declarations

#### Ethics approval and consent to participate

Not applicable. We gathered publicly available data for the analysis. As for privacy concerns, we did not include personal names or account names in our analysis. We will share only a list of tweet IDs, according to Twitter’s guidelines.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author contributions

KM designed the research, collected and preprocessed the data, analyzed the data, and wrote the main part of the paper. KM, TM, JA, and HK interpreted the results, discussed the direction of the research, and reviewed and revised the manuscripts. KM and TU elaborated on the evaluation measures. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup>Indiana University Bloomington, Bloomington, USA. <sup>2</sup>SANKEN, Osaka University, Osaka, Japan. <sup>3</sup>Sugakubunka, Tokyo, Japan.

Received: 14 September 2023 Accepted: 20 December 2023 Published online: 08 January 2024

## References

1. Brynjolfsson E, Li D, Raymond LR (2023) Generative AI at work. Technical report, National Bureau of Economic Research
2. OpenAI (2022) Introducing ChatGPT. <https://openai.com/blog/chatgpt>. (Accessed on 04/12/2023)
3. Floridi L, Chiriatti M (2020) Gpt-3: its nature, scope, limits, and consequences. *Minds Mach* 30:681–694
4. OpenAI (2023). Gpt-4 technical report. arXiv preprint
5. Ortiz S (2023) What is ChatGPT and why does it matter? Here’s what you need to know. ZDNET. <https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-everything-you-need-to-know/>
6. Buchholz K (2023) Chart: ChatGPT sprints to one million users. Statista. <https://www.statista.com/chart/29174/time-to-one-million-users/>. (Accessed on 04/12/2023)
7. Hu K (2023) ChatGPT sets record for fastest-growing user base – analyst note. Reuters
8. Metz C (2022) The new chatbots could change the world. Can you trust them?. *The New York Times*
9. Rosenblatt K (2023) ChatGPT banned from New York City public schools’ devices and networks. NBC news
10. Paul K (2023) Agencies: letter signed by Elon Musk demanding AI research pause sparks controversy. *The Guardian*
11. Rudolph J, Tan S, Tan S (2023) War of the chatbots: bard, bing chat, chatgpt, ernie and beyond. the new AI gold rush and its impact on higher education. *J Appl Learn Teach* 6(1)
12. McGleenon B (2023) Germany considers following Italy in banning ChatGPT. Yahoo
13. Heßler M, Hitzer B (2019) Introduction: tech-fear. Histories of a multifaceted relationship/einleitung: technikangst. Zur geschichte eines vielgestaltigen verhältnisses. *TG Technikgeschichte* 86(3):185–200
14. Otte M, Hoorn J (2009) Standardization in virtual worlds: prevention of false hope and undue fear. *J Virtual Worlds Res* 2(3):3–15

15. Lucivero F, Lucivero F (2016) Democratic appraisals of future technologies: integrating ethics in technology assessment. In: Ethical assessments of emerging technologies: appraising the moral plausibility of technological visions, pp 3–36
16. Palavicino CA (2016) Mindful anticipation: a practice approach to the study of expectations in emerging technologies
17. Cave S, Coughlan K, Dihal K (2019) “Scary robots” examining public responses to AI. In: AAAI/ACM AIES
18. Bakker S, Budde B (2012) Technological hype and disappointment: lessons from the hydrogen and fuel cell case. *Technol Anal Strateg Manag* 24(6):549–563
19. Fast E, Horvitz E (2017) Long-term trends in the public perception of artificial intelligence. In: AAAI
20. Binder AR et al (2012) Measuring risk/benefit perceptions of emerging technologies and their potential impact on communication of public opinion toward science. *Public Underst Sci* 21(7):830–847
21. Stokel-Walker C, Noorden RV (2023) What chatgpt and generative AI mean for science. *Nature* 614(1)
22. Brower T (2023) People fear being replaced by AI and ChatGPT: 3 ways to lead well amidst anxiety. *Forbes*. <https://www.forbes.com/sites/tracybrower/2023/03/05/people-fear-being-replaced-by-ai-and-chatgpt-3-ways-to-lead-well-amidst-anxiety>
23. CBS News (2023) Twitter is now X. Here’s what that means. <https://www.cbsnews.com/news/twitter-rebrand-x-name-change-elon-musk-what-it-means/>. (Accessed on 08/15/2023)
24. Krinsky S, Golding D (1992) Social theories of risk
25. Eloundou T et al (2023) GPTs are GPTs: an early look at the labor market impact potential of large language models. arXiv preprint
26. Webb M (2019) The impact of artificial intelligence on the labor market. SSRN 3482150. Available at
27. Felten E, Raj M, Seamans R (2021) Occupational, industry, and geographic exposure to artificial intelligence: a novel dataset and its potential uses. *Strateg Manag J* 42(12):2195–2217
28. Frank M, Ahn Y-Y, Moro E (2023) AI exposure predicts unemployment risk. arXiv preprint. [arXiv:2308.02624](https://arxiv.org/abs/2308.02624)
29. Roose K (2023) Why a conversation with bing’s chatbot left me deeply unsettled – The New York Times. <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>
30. Roose K (2022) The brilliance and weirdness of ChatGPT – The New York Times. <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>
31. Felten EW, Raj M, Seamans R (2023) Occupational heterogeneity in exposure to generative AI. Available at SSRN 4414065
32. Zhang B, Dafoe A (2019) Artificial intelligence: American attitudes and trends. Available at SSRN 3312874
33. Rainie L, Funk C, Anderson M, Tyson A (2022) How Americans think about AI. Pew Research Center
34. Funk C, Tyson A, Kennedy B (2023) How Americans view emerging uses of artificial intelligence, including programs to generate text or art. Pew Research Center
35. Gaglio S, Re GL, Morana M (2016) A framework for real-time Twitter data analysis. *Comput Commun* 73:236–242
36. Yang M-C, Rim H-C (2014) Identifying interesting Twitter contents using topical analysis. *Expert Syst Appl* 41(9):4330–4336
37. Bian J et al (2016) Mining Twitter to assess the public perception of the “Internet of things”. *PLoS ONE* 11(7):0158450
38. Kohl C et al (2018) Anticipating acceptance of emerging technologies using Twitter: the case of self-driving cars. *J Bus Econ* 88:617–642
39. Nuortimo K, Härkönen J, Karvonen E (2018) Exploring the global media image of solar power. *Renew Sustain Energy Rev* 81:2806–2811
40. Manikonda L, Kambhampati S (2018) Tweeting AI: perceptions of lay versus expert twitterati. In: ICWSM
41. Haque MU et al (2022) “I think this is the most disruptive technology”: exploring sentiments of chatgpt early adopters using twitter data. arXiv preprint
42. Leiter C, Zhang R, Chen Y, Belouadi J, Larionov D, Fresen V, Eger S (2023) Chatgpt: a meta-analysis after 2.5 months. arXiv preprint
43. Miyazaki K et al (2022) Characterizing spontaneous ideation contest on social media: case study on the name change of Facebook to meta. In: IEEE BigData
44. Barbieri F et al (2020) TweetEval: unified benchmark and comparative evaluation for tweet classification. In: EMNLP. <https://aclanthology.org/2020.findings-emnlp.148>
45. Yang K-C, Ferrara E, Menczer F (2022) Botometer 101: social bot practicum for computational social scientists. *J Comput Soc Sci*, 1–18
46. Keller TR, Klinger U (2019) Social bots in election campaigns: theoretical, empirical, and methodological implications. *Polit Commun* 36(1):171–189
47. Rauchfleisch A, Kaiser J (2020) The false positive problem of automatic bot detection in social science research. *PLoS ONE* 15(10):0241045
48. Zhao Y, Xi H, Zhang C (2021) Exploring occupation differences in reactions to Covid-19 pandemic on Twitter. *Data Inf Manag* 5(1):110–118
49. Sloan L et al (2015) Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS ONE* 10(3):0115545
50. Preoțiuc-Pietro D, Lamos V, Aletras N (2015) An analysis of the user occupational class through Twitter content. In: ACL-IJCNLP
51. Pan J et al (2019) Twitter homophily: network based prediction of user’s occupation. In: ACL
52. Ahamed SHR et al (2022) Doctors vs. nurses: understanding the great divide in vaccine hesitancy among healthcare workers. In: IEEE BigData
53. U.S. BLS (2021) List of SOC occupations. [https://www.bls.gov/oes/current/oes\\_stru.htm](https://www.bls.gov/oes/current/oes_stru.htm). (Accessed on 04/13/2023)
54. Peterson NG et al (2001) Understanding work using the occupational information network (o\* net): implications for practice and research. *Pers Psychol* 54(2)
55. Indeed (2023) Find Jobs. Indeed.com. <https://www.indeed.com/browsejobs/>. (Accessed on 04/13/2023)
56. Clark A et al (2015). Pillow. <https://pillow.readthedocs.io/en/stable/>
57. Hoffstaetter S et al (2021) Pytesseract. <https://pypi.org/project/pytesseract/>

58. Grootendorst M (2022) Bertopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint
59. McInnes L, Healy J, Astels S (2017) hdbscan: hierarchical density based clustering. *J Open Sour Softw* 2(11)
60. Ebeling R et al (2022) Analysis of the influence of political polarization in the vaccination stance: the Brazilian Covid-19 scenario. In: ICWSM
61. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022
62. Reuters (2023) New report on ChatGPT & generative AI in law firms shows opportunities abound, even as concerns persist. Thomson Reuters Institute
63. Hobsbawm EJ (1952) The machine breakers. *Past Present* 1:57–70
64. Horowitz MC, Kahn L (2021) What influences attitudes about artificial intelligence adoption: evidence from us local officials. *PLoS ONE* 16(10):0257732
65. Appel G, Neelbauer J, Schweidel DA (2023) Generative AI has an intellectual property problem. *Harv Bus Rev* 11
66. NHK (2023) Japanese artists call for AI regulation to protect copyright. NHK. [https://www3.nhk.or.jp/nhkworld/en/news/20230509\\_02/](https://www3.nhk.or.jp/nhkworld/en/news/20230509_02/)
67. François D et al (2009) A long-term study of children with autism playing with a robotic pet. *Interact Stud* 10(3):324–373
68. Alkaissi H, McFarlane SI (2023) Artificial hallucinations in chatgpt: implications in scientific writing. *Cureus* 15(2)
69. Pokhriyal N, Valentino BA, Vosoughi S (2023) Quantifying participation biases on social media. *EPJ Data Sci* 12(1):26
70. Wojcik S, Hughes A (2019) How Twitter users compare to the general public. Pew Research Center
71. An J, Weber I (2015) Whom should we sense in “social sensing”-analyzing which users work best for social media now-casting. *EPJ Data Sci* 4:22
72. Brossard D, Scheufele DA, Kim E, Lewenstein BV (2009) Religiosity as a perceptual filter: examining processes of opinion formation about nanotechnology. *Public Underst Sci* 18(5):546–558

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)