Online Social Networks and Media 55 (2026) 100355

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

Machines in the crowd? Measuring the footprint of machine-generated text on Reddit

Lucio La Cava [a], Luca Maria Aiello [b,c], Andrea Tagarelli [a,*]

[a] Department of Computer Engineering, Modeling, Electronics, and Systems Engineering (DIMES), University of Calabria, Rende, 87036, Italy
[b] IT University of Copenhagen, Copenhagen, Denmark
[c] Pioneer Centre for AI, Denmark

ARTICLE INFO

Keywords:
Machine-Generated Text
AI detection
Large language models
Natural Language Processing
Reddit

ABSTRACT

Generative Artificial Intelligence is reshaping online communication by enabling large-scale production of Machine-Generated Text (MGT) at low cost. While its presence is rapidly growing across the Web, little is known about how MGT integrates into social media environments. In this paper, we present the first large-scale characterization of MGT on Reddit. Using a state-of-the-art statistical method for detection of MGT, we analyze over three years of activity (2022–2024) across 51 subreddits representative of Reddit's main community types such as information seeking, social support, and discussion. We study the concentration of MGT across communities and over time, and compare MGT to human-authored text in terms of social signals it expresses and engagement it receives. Our very conservative estimate of MGT prevalence indicates that synthetic text is marginally present on Reddit, but it can reach peaks of up to 9% in some communities in some months. MGT is unevenly distributed across communities, more prevalent in subreddits focused on technical knowledge and social support, and often concentrated in the activity of a small fraction of users. MGT also conveys distinct social signals of warmth and status giving typical of language of AI assistants. Despite these stylistic differences, MGT achieves engagement levels comparable than human-authored content and in a few cases even higher, suggesting that AI-generated text is becoming an organic component of online social discourse. This work offers the first perspective on the MGT footprint on Reddit, paving the way for new investigations involving platform governance, detection strategies, and community dynamics.

1. Introduction

Generative Artificial Intelligence (GenAI) is transforming the Web and reshaping everyday practices related to search [1], news consumption [2], entertainment [3], and online social interactions [4]. In particular, the dynamics of the participatory Web are being revolutionized by the commercial release of Large Language Models (LLMs) that dramatically lowered the barriers to producing Machine-Generated Text (MGT). The broad diffusion of these new tools is paving the way for large volumes of crowd-generated synthetic content to be uploaded to social media and other participatory online platforms.

This socio-technical shift carries profound societal implications. The Web, and especially social media, has become a central arena for public discourse and information sharing, shaping the spread of ideas and influencing democratic participation. The diffusion of MGT may significantly affect these collective processes [5] in ways that scholars are still actively debating. Optimistic perspectives emphasize the potential of GenAI to enhance contributors' creativity [6], increase engagement, and promote inclusivity in online discussions [7]. In contrast, critics raise concerns about misinformation, declining authenticity in user interactions [7,8], deteriorating content quality [8,9], and emergent biases [10]. When produced with strategic or malicious intent, MGT can be deliberately crafted to be persuasive [11] and highly engaging [12], while embedding inaccurate, biased, or harmful information [8]. This could exacerbate systemic risks identified by the European Digital Services Act, including discrimination, declining mental well-being, and the erosion of civic and electoral processes [13]. Developing a quantitative understanding of AI's impact on people's online experiences is therefore both essential and urgent to inform effective policies and design safeguards against potential harms.

Although distinguishing Machine-Generated Text (**MGT**) from Human-Generated Text (**HGT**) text is inherently challenging, promising tools have been developed with high accuracy and inference speed [14]. These tools have enabled initial estimates of the proliferation of synthetic content online [15] and of the growing presence of

* Corresponding author.
E-mail addresses: lucio.lacava@dimes.unical.it (L. La Cava), luai@itu.dk (L.M. Aiello), andrea.tagarelli@unical.it (A. Tagarelli).

https://doi.org/10.1016/j.osnem.2026.100355
Received 5 February 2026; Received in revised form 30 April 2026; Accepted 3 July 2026
2468-6964/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

AI agents posing as human users [12]. However, beyond these initial estimates, the use of MGT in real-world online social environments remains mostly unexplored. To shed light on how online social dynamics might change as result of MGT, it is not only important to estimate its incidence, but also to analyze the nature of the content produced, the context in which it is published, and the response that the public has to it.

To help address this gap, we present the first large-scale characterization of MGT on Reddit. Our choice of Reddit is motivated by its societal significance. Reddit is one of the world's leading and fastest-growing social media platforms, and it has established itself as an active forum of political discussion that had often important societal effects [16], such as the birth of the global climate movement [17]. It is also a very popular alternative source of information to traditional Web Search, especially in the health domain [18].

Reddit offers also a good case-study for the analysis of MGT. It is organized around topic-specific communities (i.e., *subreddits*), making it a uniquely structured environment for exploring whether and to what extent MGT emerges and spreads across diverse types of online discourse. Moreover, the organic and user-driven growth of such communities makes MGT both easier to deploy (users are anonymous and identified by nicknames) and more consequential (with concrete effects on community-level information quality). In this work, we focus on 51 popular subreddits representative of Reddit's main functional community types, analyzing comments and submissions from 2022 to 2024, a period that includes milestones in the release of GenAI tools. We used a state-of-the-art method to detect messages with a high likelihood of being MGT and address four main research questions:

RQ1 — How is MGT adopted across subreddit communities and over time, and how does this reflect the distinct conversational norms of different community categories?
RQ2 — What temporal patterns characterize MGT prevalence, and how might these relate to exogenous events?
RQ3 — How does the distribution of MGT users evolve over time, and are there any significant trends in its adoption?
RQ4 — Do MGT and HGT differ in the type or intensity of social signals and engagement patterns they convey across subreddit categories?

Contributions. By answering these questions, our main contributions in this work are as follows:

- We present the first large-scale, longitudinal characterization of MGT adoption on Reddit, covering 51 subreddits, more than 9M comments and 2M submissions over a 3-year period encompassing the rise of LLMs;
- We show that MGT prevalence, while globally marginal, is unevenly distributed: it peaks up to 9% in specific subreddits and temporal windows, is concentrated in a small fraction of users, is affected by new milestones in GenAI, and varies systematically across community types;
- We unveil that MGT conveys distinct social signals compared to HGT, e.g., consistently higher in status-giving and supporting language;
- We show that, despite stylistic distinctiveness, MGT achieves engagement levels compared to, or even higher than, HGT in various cases, suggesting organic acceptance by communities.

2. Methodology

We present the data (Section 2.1), the technique for estimating the likelihood of text being MGT (Section 2.2), and the methods for assessing the nature of the content (Section 2.3) and engagement (Section 2.4) in MGT and HGT.

2.1. Data

To characterize the emergence and diffusion of AI-generated texts on Reddit, we start by considering the top 1000 subreddits by number of subscribers.[1] The choice of analyzing large and active communities is motivated by (i) the broad visibility and potential impact that the content posted in those communities have on the public, and (ii) the substantial volume of data providing sufficient support for a reliable estimation of the prevalence of MGT.

We parsed the list of subreddits in decreasing order of popularity and selected a representative subset of 51 subreddits, by excluding those that (i) primarily contained images/videos, with minimal textual content (that would not be suitable for our analysis), (ii) are dedicated to aggregating links rather than discussions, (ii) are too narrow or niche to represent a specific "social category". Communities passing these criteria were then mapped into five main functional subreddits categories that have been informed by a taxonomy from prior work [19]:

- **Information Seeking**, including, e.g., r/worldnews, r/askscience, r/explainlikeimfive, r/health;
- **Social Support**, including, e.g., r/GetMotivated, r/mentalhealth, r/AITAH, r/LifeProTips;
- **Discussion**, including, e.g., r/changemyview, r/unpopularopinion, r/SeriousConversation, r/politics;
- **Identity**, including, e.g., r/teenagers, r/asktransgender, r/AskWomen, r/BlackPeopleTwitter;
- **Chit Chat**, including, e.g., r/funny, r/entertainment, r/books, r/Showerthoughts.

We collected the complete data on submissions and comments posted in these subreddits from January 2022 to December 2024 (see Appendix A.1 for the complete list of subreddits) using data dumps obtained from the *PushShift API* [20]. This covers the period during which LLMs and GenAI have become widely accessible to the public (cf. Table A.1 in Appendix A.2 for additional insights), allowing us to study their adoption and diffusion in online discourse through Reddit. After filtering out empty posts (e.g., containing only images or URLs), we collected **38,074,021 comments** and **4,073,586 submissions**. A detailed overview of the number of analyzed comments for each month is reported in Fig. A.1, Appendix A.5.

2.2. Detecting machine-generated text

Problem setting. We are given a collection of text messages (e.g., Reddit comments or submissions) $\mathcal{X} = \{x_1, \ldots, x_n\}$ where each message x_i can be authored by either a human (i.e., HGT) or any machine-generator tool like large language models (i.e., MGT), but the type of author is unknown. We frame the detection of MGT content in online discussions as a *binary text classification task*: given a classifier model f, the goal is to assign a label $y_i \in \{0, 1\}$ to each $x_i \in \mathcal{X}$ such that:

$$y_i = \begin{cases} 1, & \text{if } P(MGT \mid x_i) \geq \tau \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $P(MGT \mid x_i)$ denotes the probability, estimated by f, that x_i is an MGT, and $\tau \in [0, 1]$ is a decision threshold, ideally selected to reflect the intended trade-off between false positives and false negatives.

Approach. To detect signals of generative AI usage within Reddit communities, we considered two major classes of MGT detection techniques, namely metric-based and model-based approaches.

The former typically rely on statistical properties of texts [21], such as the log-likelihood [22], log-probabilities associated with tokens [23], log-rank information [24], or conditional probability curvatures [14].

[1] https://www.reddit.com/best/communities/1/.

Conversely, model-based approaches leverage deep-learning classifiers specifically trained for distinguishing between HGT and MGT [22], including machine learning and Transformer-based models [25] or contrastive-learning approaches [26,27].

Although both perspectives have been widely adopted in the literature and demonstrated strong performance [28], we opted for a metric-based solution because of two main practical constraints. First, model-based approaches are typically trained on datasets from domains like news articles or essays, leading to degraded performance on conversational data such as Reddit comments. Moreover, fine-tuning such approaches on Reddit data would be costly in practice, as it would require a large training set of labeled Reddit-specific data—currently unavailable, to the best of our knowledge. Second, model-based approaches require relatively slow (neural) inference for each input text, which becomes computationally prohibitive at a scale of millions of comments and submissions.

Based on these considerations, we selected the zero-shot metric-based *Fast-DetectGPT* [14], as our primary detection tool. Fast-DetectGPT strikes a good balance between detection capabilities and computational efficiency, ranking among the most effective detectors in recent benchmarks [29] while remaining significantly faster than model-based detectors and commercial alternatives (e.g., *GPTZero*). *Fast-DetectGPT* operates by measuring the divergence between the log-probabilities (extracted through a small, efficient language model) of a text and the distribution of perturbed variants of the same text, based on the hypothesis that a significant shift would follow typical statistical patterns of machine-generated texts [14].

We point out that alternative detectors can be used as reference models. The interested reader can refer to Appendix A.3 for a comparison of the results discussed in the following sections with those obtained by considering another state-of-the-art zero-shot detector, i.e., Binoculars [30].

2.3. Content analysis

Syntactic and Readability Characteristics. We measure main features of HGT and MGT messages, capturing text length, redundancy, and readability, namely (i) number of words and sentences, (ii) compression ratio (i.e., original size divided by gzip compressed size), and (iii) Flesch Reading Ease score indicating the degree of difficulty in comprehending a text passage as a function of the number of words, syllables, and sentences it contains; the score ranges within $-\infty$ (low readability) and 121.22 (high readability).[2]

Social Dimensions. We compare MGT and HGT in terms of the strength of the *social signal* they convey. Social signals are instantiations of different social functions of language [31]. These functions are used to achieve pragmatic communicative acts such as transferring information, creating conflict, or offering moral support. Rather than capturing the themes of conversation, as topic modeling does, the study of social signals indicates the *types* of interaction occurring in a social system. We test whether the introduction of MGT shifts the norms in terms of the typical social signals conveyed in online forums.

Social theorists have identified universal social signals that emerge commonly in everyday communications [32]. These so-called *social dimensions* generalize across topics and provide a *functional* characterization of the text, which we expect to vary across community types and between MGT and HGT. Prior work introduced a method for automatically identifying these social dimensions in conversational text [33].[3] This tool consists of a set of classifiers (one per dimension) trained and previously validated on Reddit comments (average area under the ROC curve of 0.84 across dimensions), making them directly applicable to our study.

Table 1
Summary characteristics of Social Dimensions, as originally reported in [33].

Dimension	Description
Knowledge	Exchange of ideas or information; learning, teaching
Status	Conferring status, appreciation, gratitude, or admiration upon another
Support	Giving emotional or practical aid and companionship
Fun	Experiencing leisure, laughter, and joy
Conflict	Contrast or diverging views
Similarity	Shared interests, motivations or outlooks

Table 2
Dataset statistics across subreddit categories after filtering. Count indicates the number of subreddits within each category.

Category	Count	# Submissions	# Comments
Information seeking	13	246,106	1,635,872
Social support	18	1,716,198	4,541,287
Discussion	5	42,371	1,453,938
Identity	10	99,855	615,020
ChitChat	5	26,029	785,886
Total	51	2,130,559	9,032,003

We use these classifiers according prior work's best practices [34]. Given a message x and a social dimension d, the classifier outputs a score $\phi_d(sen) \in [0,1]$ for each sentence $sen \in x$, representing the likelihood that sen conveys dimension d. In our setting, we associate each message x with the maximum score across its sentences, i.e., $\phi_d(x) = \max_{sen \in x} \phi_d(sen)$. Using the maximum rather than the average avoids dilution effects and aligns with the theoretical interpretation of social dimensions being effectively conveyed even if expressed when expressed only briefly in a message [32].

To identify messages that convey a given dimension with high probability, we discretize the classifier outputs into binary indicators. For each message x and dimension d, we consider x to convey d if $\phi_d(x)$ exceeds a dimension-specific threshold θ_d. Favoring precision over recall, we set θ_d to the 85th percentile of observed scores for each classifier. Because longer messages have a higher chance of containing at least one sentence above the threshold, we mitigate the length bias with a discounting factor. Specifically, we standardize the length of each message x, denoted $\ell_d(x)$, with respect to the distribution of message lengths for dimension d: $\hat{\ell}_d(x) = \frac{\ell_d(x) - \mu_d}{\sigma_d}$, where μ_d and σ_d represent the mean and standard deviation of lengths across all messages conveying d. Finally, to give greater weight to shorter messages that strongly express a given dimension, while down-weighting cases where the dimension appears only due to excessive length, we define the *social dimension score* of x with respect to d, denoted $\varphi_d(x)$, as follows:

$$\varphi_d(x) = \begin{cases} \dfrac{1}{1+\hat{\ell}_d(m)} & \text{if } \phi_d(x) \geq \theta_d \ \wedge \ \hat{\ell}_d(m) \geq 0 \\ 2 - \dfrac{1}{1-\hat{\ell}_d(m)} & \text{if } \phi_d(x) \geq \theta_d \ \wedge \ \hat{\ell}_d(m) < 0 \\ 0 & \text{if } \phi_d(x) < \theta_d. \end{cases} \quad (2)$$

The score ranges in $[0,2]$, and is equal to 1 when the text length is as the average for that dimension, approaches 0 as the text becomes substantially longer, and 2 as it becomes shorter.

For our study, and given the categories identified in Table 2, we considered the dimensions that are most relevant to the Reddit discourse: *knowledge*, *status*, *support*, *fun*, *conflict*, and *similarity*.

A short description of the selected social dimensions, as defined in [33], is reported in Table 1.

[2] https://pypi.org/project/textstat/.

[3] Available as a Python package at https://github.com/lajello/tendimensions.

2.4. Engagement analysis

The public engagement generated by messages in online communities is reflected in the number of *upvotes* and *downvotes* they receive. For any message x_i, we use the difference between upvotes and downvotes as a direct measure of its positive engagement, which we refer to as the *engagement score* (e_i).

Our goal is to compare the engagement levels of MGT with those of HGT. To ensure a sufficiently fine-grained resolution, we organized the data on timespans a *monthly* basis and at the *subreddit level*. Hereinafter, we denote with $I = \{(s,t)\}$ the set of (subreddit, timespan) instance pairs under examination. We assume that messages posted within a subreddit during a given timespan may include some fraction of MGT, although we expect this fraction to be substantially smaller than that of HGT. To compare engagement within a fixed context, we restrict our analysis to each subreddit–timespan pair $(s,t) \in I$ individually. For each pair, we first identify the MGT messages, and then draw $N = 1000$ bootstrap samples (with replacement) from the pool of HGT. Each of the N samples has the same size as the set of MGT messages detected for that subreddit and timespan. We denote by $B^{(M)}_{(s,t)}$ the set of MGT messages detected for (s,t), and by $B^{(j)}_{(s,t)}$ the jth bootstrap sample of HGT messages, with $j = 1 \dots N$.

Given a bootstrap sample of HGT messages ($B^{(j)}_{(s,t)}$) and the corresponding set of MGT messages ($B^{(M)}_{(s,t)}$), we compared their engagement score distributions to assess the significance and magnitude of their difference. To do so, we employed the nonparametric *Mann–Whitney U test* for comparing numerical distributions. Its null hypothesis is that values from the two groups are drawn from the same distribution, and the test evaluates this by comparing the ranks of the observations. If the resulting p-value falls below the significance threshold of 0.05, we reject the null hypothesis, indicating a statistically significant difference in engagement between MGT and HGT.

For each subreddit–timespan pair (s,t), we conducted N two-sided Mann–Whitney tests (one per bootstrap sample). Focusing only on the samples $B^{(j*)}_{(s,t)}$ for which the null hypothesis is rejected, we estimated the effect size relative to $B^{(M)}_{(s,t)}$ using *Cliff's delta* (δ), defined as:

$$\delta_{B^{(j*)}_{(s,t)}} = \frac{1}{k^2} \sum_{x_i \in B^{(j*)}_{(s,t)}} \left(\mathbb{1}[e_i^{\text{MGT}} > e_i^{\text{HGT}}] - \mathbb{1}[e_i^{\text{HGT}} > e_i^{\text{MGT}}] \right) \tag{3}$$

where k is the number of MGT messages detected for (s,t); e_i^{MGT} and e_i^{HGT} denote the engagement scores of the ith MGT and human-written message, respectively; and $\mathbb{1}$ is the indicator function. The term inside the sum reflects the difference between the number of pairwise comparisons in which MGT comments receive higher, versus lower, engagement scores than human-written comments. Cliff's δ ranges in $[-1,1]$, where $\delta > 0$ indicates that MGT comments receive higher engagement, $\delta < 0$ indicates the opposite, and δ close to 0 suggests negligible differences in engagement.

3. Results

Experimental Setting. To ensure that our findings are robust and that the detected signals are not driven by noise, we adopt a conservative detection policy. Specifically, we set the detection threshold τ in Eq. (1) to 0.99, and restrict our analysis to texts (both submissions and comments) containing at least 250 tokens; it should be noted that the choice of this token length threshold is supported by evidence provided in [14].

Table 2 provides an overview of the resulting dataset under these criteria. Although this choice narrows our focus to **9,032,003 comments** (with an average length of 387.39 ± 180.16 tokens) and **2,130,559 submissions** (with an average length of 542.79 ± 331.87 tokens), it is motivated by two key considerations. First, using a high threshold τ in the reference detection method substantially reduces the risk of false positives, ensuring that texts labeled as machine-generated represent highly confident detections. Second, prior work has shown that detection reliability improves with text length [35]. Taken together, these constraints provide conservative yet trustworthy evidence of the presence of MGT on Reddit.

Table 3
Peak prevalence of MGT across subreddit categories. For each category, we report the fraction of MGT comments over the total detected, the maximum observed share of MGT, the subreddit in which it occurred, and the corresponding date (yy/mm).

Community	MGT %	Peak %	Top subreddit	Date
Inf. Seek.	29.83	6.33	`r/askscience`	23/07
Soc. Supp.	26.73	7.69	`r/malefashionadvice`	23/07
Discussion	20.83	1.28	`r/politics`	23/09
Identity	17.81	8.46	`r/teenagers`	23/02
ChitChat	4.80	3.13	`r/funny`	23/05

We investigated the prevalence of MGT in *both* submissions and comments on Reddit. Nonetheless, due to space limitations, we focus our presentation on comments for three main reasons: they provide clear evidence of interaction flow, they are more likely to reflect genuine communication, and we empirically observed that comments exhibit lower sparsity over time, making them better suited for fine-grained temporal analysis.

Therefore, in the following, we address the previously stated RQs across subreddit comments, i.e., *where* and *when* MGT is used, *who* adopts it, and *what* types of content are most likely to be identified as MGT. We replicate the same experimental analysis on submissions; the corresponding figures and tables are collected in Appendix B, while the discussion is integrated within each relevant subsection below.

3.1. Community distribution

The adoption of MGT on Reddit varies considerably across subreddits, reflecting differences in conversational norms determined by the different community categories (cf. Section 2.1). We summarize results in Table 3 and Fig. 1, and elaborate on them next. Information-Seeking communities exhibit among the highest and most consistent adoptions of MGT, with subreddits like `askscience` showing clear spikes in MGT contributions, peaking at 6.33% in July 2023. This is supported by the observation that such communities are structured around question-answering dynamics, where GenAI is expected to produce helpful responses with minimal effort, thus lowering the interaction barrier.

Social-Support communities show moderate yet unevenly distributed adoption across subreddits, with some more consistent usage in spaces like `GetMotivated` or `lifehacks`, and rare yet significant peaks in others like `malefashionadvice` (up to 7.69% of MGT in July 2023). Interestingly, subreddits addressing more sensitive topics (e.g., `SuicideWatch` or `mentalhealth`) exhibit minimal MGT activity. This hints at more spontaneous interactions, likely driven by community norms.

Notably, discussion-oriented communities consistently exhibit the lowest adoption of MGT, with only sporadic and slight signals of usage in communities like `politics` and `SeriousConversation`, with the former peaking at no more than 1.28% of MGT in September 2023. We ascribe this phenomenon to the nature of these spaces, where users typically want to express their own views and arguments, rather than resorting to GenAI to generate them.

The identity-focused communities display the most noteworthy outlier pattern: the `teenagers` subreddit exhibits consistently high adoption of MGT (up to 8.46%) across all analyzed periods, rendering them as early-adopters of GenAI tools. This phenomenon is likely supported by (i) higher curiosity and willingness to experiment with emerging technologies, (ii) faster peer-to-peer diffusion via social networks, and

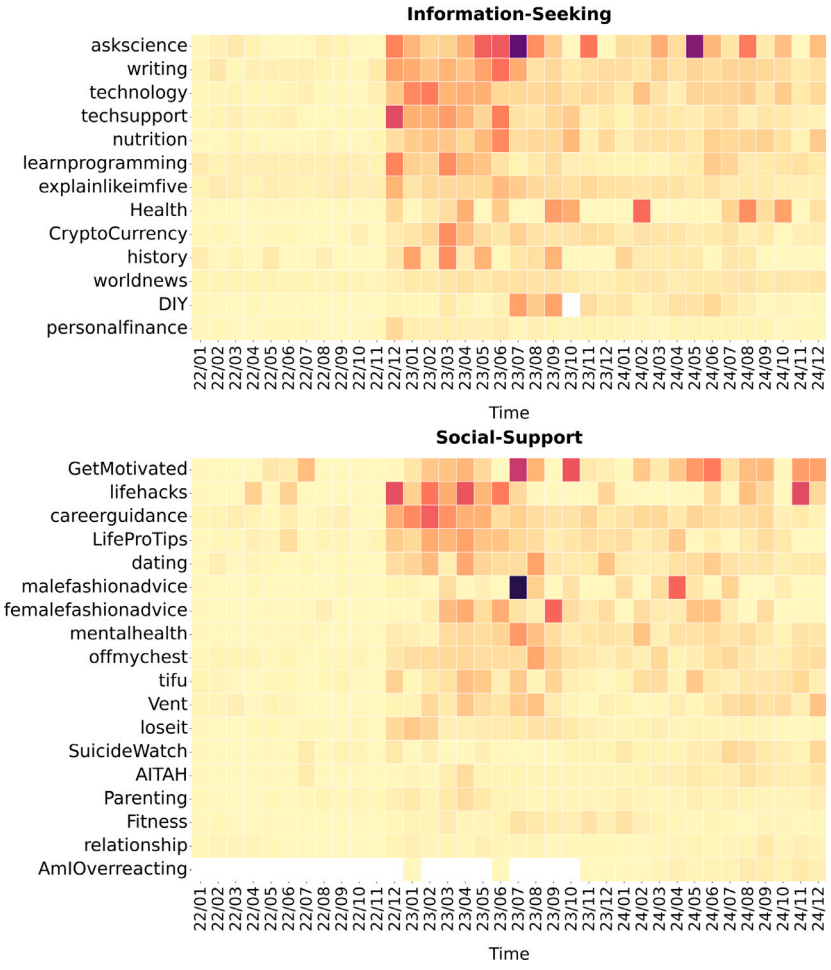


Fig. 1. MGT comment usage across subreddits. Rows are sorted by the sum of monthly MGT adoption. Each cell shows the share of MGT posts for a subreddit-month pair. Darker shades indicate higher usage. Empty cells mean no comments matching our filtering criteria.

(iii) lower barriers to adoption due to higher familiarity with digital tools. Moreover, other identity-based subreddits, like politically driven ones, exhibit observable adoption of MGT.

Finally, for ChitChat communities, we observe lower adoption of MGT compared to other spaces, with some slight peaks at 3.13% for `funny` in May 2023. We ascribe this more contained prevalence to the low-effort and more spontaneous nature of the corresponding subreddits.

Results on submissions. We applied the same analytical framework to Reddit submissions; figures and tables for this analysis are reported in Appendix B, where data are aggregated by quarters rather than monthly to counter sparsity. The community-level distribution of MGT prevalence in submissions closely mirrors that of comments overall, yet three notable differences emerge. First, the subreddits displaying the highest MGT prevalence differ from those observed for comments: `r/CryptoCurrency`, `r/tifu`, `r/changemyview`, and `r/socialism` show elevated MGT levels in submissions but only modest signals in comments. Second, all Discussion communities exhibit more MGT in submissions than in their corresponding comment sections (cf. Table B.1 in Appendix B), suggesting that AI-assisted argumentation finds a natural outlet in debate-oriented submission posts. Third, peak prevalence for submissions reaches approximately 7%, slightly below the 9% peaks observed for comments, confirming that MGT is non-negligible in submissions as well, though its distribution across subreddits follows a distinct pattern from comments.

3.2. Temporal patterns

Fig. A.1 shows that MGT adoption follows clear temporal patterns. We link these fluctuations to exogenous events, particularly major GenAI milestones and the release of new tools during the observation period of our study.

The first measurable increase in machine-generated text appears in November 2022, coinciding with the release of *ChatGPT*. From this point onward, most subreddits in Fig. A.1 display increasing prevalence of MGT, as GenAI tools transitioned from software packages requiring programming expertise to accessible chatbot interfaces, thus lowering barriers to adoption. A notable exception is the `teenagers` subreddit, where modest MGT signals are detected even before the widespread diffusion of ChatGPT. Manual inspection suggests that these early MGT traces are of very low quality (e.g., lacking punctuation, containing repetitive patterns), and likely stem from preliminary statistical tools predating widely usable models such as GPT.

The months immediately following ChatGPT's release mark a phase of rapid adoption across nearly all subreddits, producing the strongest concentration of MGT in the entire observation period. This adoption "wall" is followed by several peaks that may correspond to later LLM releases. For instance, the higher prevalence observed in Q1–Q2 2023 in communities such as `askscience` and `teenagers` coincides with the release of GPT-4 (March 2023). Thereafter, adoption patterns diverge across communities, with some consolidating a persistent prevalence of MGT, while others show a decline in signals. A renewed increase in adoption appears in Q1–Q2 2024, coinciding with major releases such as *Gemini*, *Claude*, *Grok*, and *Llama*.

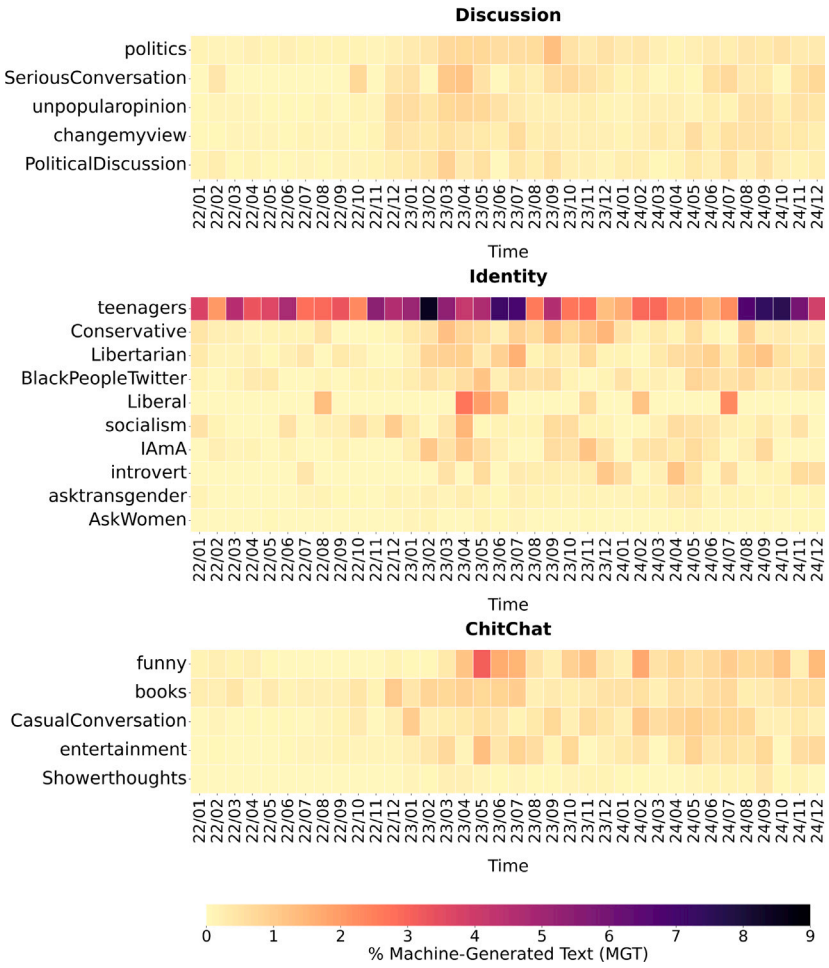


Fig. 1. (*continued*).

Results on submissions. Temporal dynamics for submissions are consistent with those observed for comments. MGT prevalence in submissions also remains stable since late 2022, with peaks reaching approximately 7%, and exhibits similar responsiveness to major GenAI milestones (cf. Fig. A.2 in Appendix A). Due to the lower volume of submissions relative to comments, the temporal analysis is aggregated at the quarterly level to ensure statistical robustness and limit the effect of sparsity.

3.3. User distribution

To understand the evolution in adoption of MGT among Reddit users, we analyzed the average fraction of users adopting MGT across the observation period, as well as the percentage of their comments that are detected as being machine-generated.

Overall, Fig. 2 suggests that the adoption of MGT by Reddit users remained relatively low over time, with mean values peaking at 2% and maximum values reaching 3% in certain periods. Notably, temporal dynamics reveal distinct adoption phases. In particular, after a slight signal of usage traceable to early adopters, a marked increase is detected in correspondence with the widespread release of major GenAI tools, suggesting a likely relationship between tools' accessibility and adoption. Interestingly, this peak led to the highest rate of adoption across our observation period. After this, adoption stabilized at slightly elevated levels compared to the early adoption period, suggesting the establishment of a small yet persistent cohort of MGT-assisted users. The fraction of MGT content that these users produce compared to their total comment production is highly heterogeneous, ranging between 10% and 40% of posts. Notably, the fraction of MGT posts remains substantial (i.e., around 20%) even after the initial phase adoption. This suggests that, once incorporating GenAI tools into their workflow, such users tend to rely on them for a significant share of their produced content.

Results on submissions. The concentration of MGT among users mirrors the pattern observed for comments. The fraction of users generating MGT in submissions remains similarly low and stable over the observation period, and MGT in submissions is likewise concentrated among a small number of prolific contributors (cf. Fig. A.4, Appendix B). This consistency across content types reinforces the interpretation that a small, persistent cohort of users is responsible for the majority of AI-assisted content on Reddit, regardless of whether they contribute via comments or original posts.

3.4. Content features

Text Statistics. Table 4 summarizes aggregated textual statistics, providing a quantitative comparison of the structural properties of MGT and HGT. Across all subreddit categories, MGT is consistently longer than HGT. Moreover, MGT is generally less readable, with this difference particularly pronounced in Identity-oriented communities. We attribute this to the very low-quality content observed in `r/teenagers` (e.g., lack of punctuation, highly repetitive patterns), which impacts the aggregated readability scores for the category. Additionally, MGT is more compressible than human comments, reflecting a higher degree of internal redundancy or repetitive phrasing. Overall, MGT comments are typically more verbose than HGT and slightly more difficult to read.

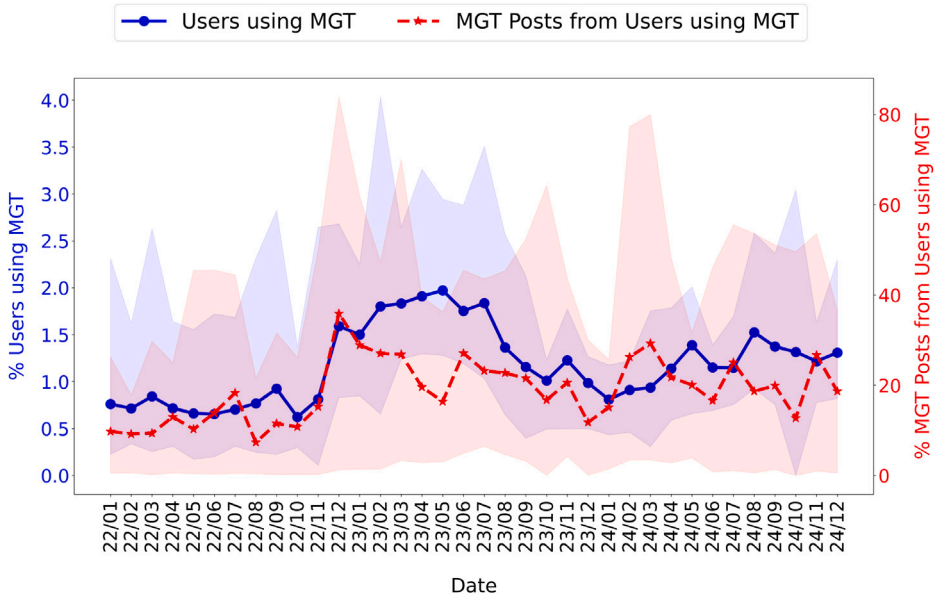


Fig. 2. Average number of users adopting MGT (in blue, left y-axis) and the corresponding percentage of their comments detected to be MGT (in red, right y-axis) across all categories over time. Shaded areas denote minimum and maximum observed values. Users with a single post have been filtered out to mitigate noise due to one-off activities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Aggregated values of text statistics from the MGT and HGT comments across subreddit categories. All category–statistics combinations exhibit statistically significant distributional differences ($p < 0.05$) according to the Mann–Whitney U test.

Category	Text	# Words	# Sentences	Flesch Read.	Compression
Inf. Seek.	HGT	263.94 ± 167.29	15.14 ± 9.67	52.27 ± 29.06	2.94 ± 0.31
	MGT	334.77 ± 187.36	18.59 ± 11.95	52.46 ± 58.32	3.45 ± 4.45
Soc. Supp.	HGT	264.82 ± 124.07	16.93 ± 8.71	64.43 ± 21.96	2.9 ± 0.27
	MGT	339.12 ± 154.21	19.75 ± 10.54	63.58 ± 33.88	3.25 ± 1.23
Discuss.	HGT	253.39 ± 138.77	14.59 ± 7.58	50.63 ± 25.72	2.92 ± 0.31
	MGT	418.88 ± 232.35	22.32 ± 14.4	48.52 ± 37.66	3.45 ± 2.4
Identity	HGT	317.39 ± 226.31	16.84 ± 20.05	38.71 ± 146.23	4.73 ± 11.99
	MGT	459.34 ± 289.56	22.55 ± 24.32	−32.96 ± 280.96	8.39 ± 22.5
ChitChat	HGT	221.42 ± 131.9	13.61 ± 6.72	52.03 ± 42.12	2.97 ± 2.03
	MGT	348.67 ± 182.04	19.11 ± 12.74	50.31 ± 94.24	3.6 ± 5.4

Results on submissions. Machine-generated submissions exhibit textual properties closely aligned with those observed for MGT comments: they are consistently longer and less readable than human-authored submissions across all subreddit categories (cf. Table B.2, Appendix B). Most category–statistics combinations yield statistically significant differences according to the Mann–Whitney U test, with the exception of word count and compression ratio in Social Support communities. Notably, submissions are already longer than comments on average, and the additional verbosity of MGT is proportionally consistent with the trend observed for comments.

Social Dimensions. By applying the social dimension classifier described in Section 2, we performed a qualitative analysis to assess whether and to what extent MGT and HGT differ in the type or strength of social signals they convey across subreddit categories. Fig. 3 illustrates the results, grouped for each combination of subreddit category and dimension, and annotated with the corresponding significance and effect size. Next, we will discuss the statistically significant differences according to a Mann–Whitney U test, and provide interpretations on a per-category basis.

Information Seeking communities. Human-authored content consistently conveys stronger signals of *knowledge* (i.e., fact-based information) and *similarity* (i.e., drawing parallels between the matter of discussion and personal traits or experiences). This might suggest that humans tend to ground their comments on experiential knowledge, which differs from the "pre-training" one achieved by GenAI tools, and with a language that typically aligns with the corresponding community. Conversely, MGT tends to express higher levels of *status* (i.e., expressions of admiration and appreciation) and *support* (i.e., empathy and emotional warmth), likely due to their "authoritative" answers and assistive tone that is typical of instruction-tuned chatbots.

Social Support communities. Human-written messages score substantially higher in *knowledge* and *similarity*. We ascribe the first finding to the hypothesis that supportive knowledge often depends on empathy or experience, which are both unique to humans. Related considerations hold for similarity, where the social and peer support occurring in real life may help in instilling more prominent similarity signals in HGT. On the other hand, MGT tends to contain abundant *support* and *status* expressions, reflecting the optimization (e.g., human preference alignment) of modern GenAI tools in offering instructive and "comforting" answers.

Discussion communities. MGT conveys more *knowledge* and *status*. This is likely due to a particularly fertile ground for GenAI, such as producing knowledge-rich content in discussion, and their authoritative styles. HGT, however, exhibits significantly better signals of *support*, *conflict*, and *similarity*. These findings perfectly align with the social dynamics of online debates: users tend to support like-minded peers, engage in confrontational or conflictual interactions, and emphasize shared identities. Notably, the lower conflict scores exhibited by MGT reflect

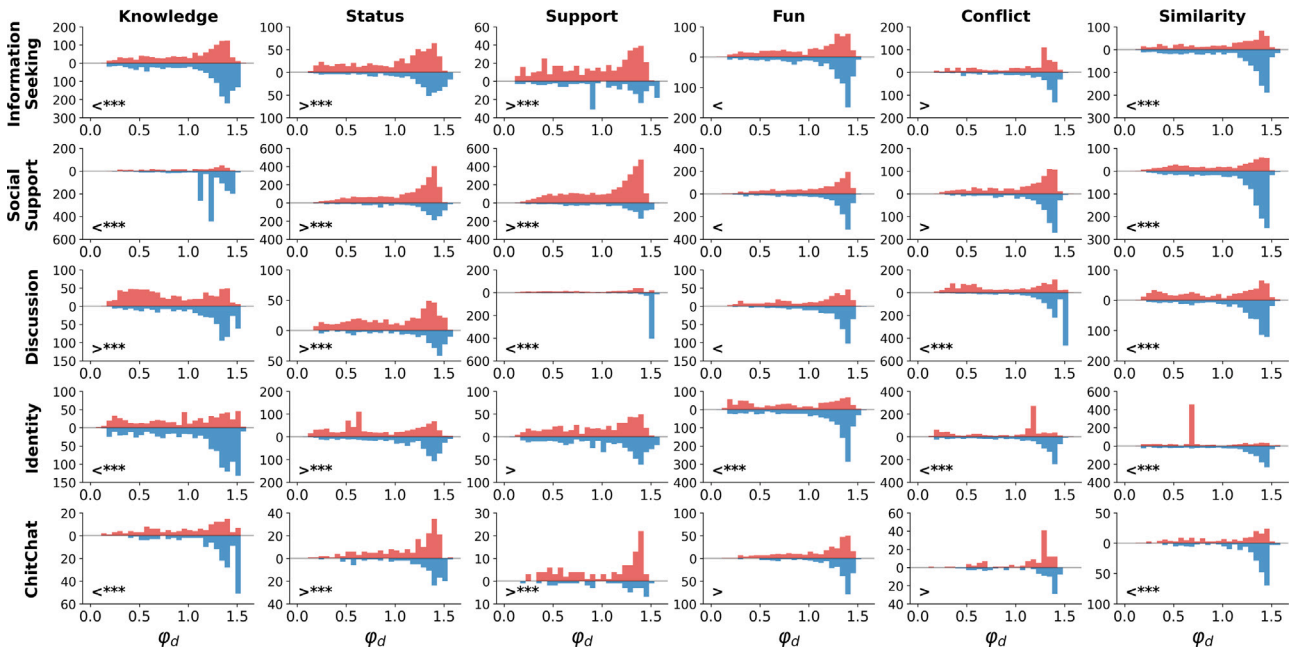


Fig. 3. Comparison of social dimension expressions between MGT (in red) and HGT (in blue) comments across different subreddit categories. Each subplot shows the distribution of social dimension scores $\phi_d(x)$ (excluding instances with $\phi_d(x) = 0$) for all combinations of subreddit and social dimension. Each subplot is labeled with the corresponding effect size, where > indicates that MGT scores higher than HGT, and < indicates the opposite. Marker *** denotes that the difference in social dimension expression scores between MGT and HGT is statistically significant according to the Mann–Whitney U test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the efforts of preventing such tools from moving towards conflictual behaviors with their users.

Identity communities. MGT is found to convey better signals than HGT only for *status*. This suggests that while GenAI tools might adopt a declarative tone, they struggle in expressing the emotionally and experientially grounded communication typical of identity discourse. Indeed, we found human-authored content to better convey *knowledge*, *fun*, *conflict*, and *similarity*—drawing from personal identity and experience.

ChitChat communities. We found HGT to outperform MGT in conveying *knowledge* and *similarity*, suggesting a higher capacity to express shared tastes. Again, MGT confirms its ability to convey better *status* and *support*, reflecting its consistent behavior in using explanatory and supportive language even in informal contexts, where these signals are less socially expected.

Results on submissions. The social dimension analysis for submissions (cf. Fig. A.5, Appendix B) is broadly consistent with the patterns observed for comments. MGT submissions convey more signals of *social support* but fewer of *knowledge*, *conflict*, and *similarity*, echoing the characteristic linguistic style of instruction-tuned language models. One notable departure is that MGT submissions tend to express *less status*-related language compared to their comment counterparts, a difference especially pronounced in Discussion communities. We interpret this as a consequence of the absence of conversational dynamics in submissions: without a direct interlocutor to respond to, the opportunities for expressing admiration or appreciation—signals typically triggered by social interaction—are naturally reduced.

3.5. Engagement

After characterizing MGT both quantitatively and qualitatively, we examined whether it exhibits systematically different engagement patterns compared to HGT. In particular, we compared the distributions of their *engagement scores* for each *subreddit–month* pair. To ensure meaningful comparisons, we restricted the analysis to cases where both MGT and human comment distributions contained at least 50 samples, resulting in a total of 102 tests.

Table 5
Summary of subreddit–month pairs where engagement distributions between MGT and HGT differ significantly. Positive, resp. negative, values of Cliff's δ indicate higher, resp. lower, engagement for MGT. For each case, %p reports the fraction of significant tests.

Category	Subreddit	# Months	Avg. δ	% p
Information seeking	r/technology	3	0.27	0.72
	r/explainlikeimfive	1	0.17	0.55
	r/worldnews	1	−0.28	0.53
Social support	r/relationship	10	0.28	0.91
Discussion	r/politics	2	0.30	0.94
Identity	r/teenagers	7	0.20	0.66

Following the procedure described in Section 2.4, we applied a bootstrap resampling approach ($N = 1000$) to balance sample sizes between MGT and HGT comments, performing a Mann–Whitney U test on each resample. For robustness, we considered a difference statistically meaningful only if a majority of bootstrap runs (i.e., more than 50%) yielded significant results ($p < 0.05$). This filtering step isolates contexts in which divergences between distributions are consistent rather than incidental, yielding 26 significant cases out of 102 evaluated.

For these statistically significant cases, we complemented the significance testing with an effect size estimation using Cliff's δ. Remarkably, in all but one case, the effect size was positive ($\delta > 0$), indicating that when significant differences in engagement scores arise, MGT consistently attracts higher engagement than human-authored comments. The only exception was observed in `r/worldnews` (June 2023), where MGT exhibited significantly lower engagement.

Interestingly, the 26 identified subreddit-month pairs in which engagement scores between MGT and HGT differ significantly are concentrated in specific conversational domains (i.e., subreddit categories), as summarized below and in Table 5.

- **Information-seeking communities**—We found **five** significant cases in explanatory (**r/explainlikeimfive**) or technical subreddits (**r/technology**), with a modest effect size ($\delta \approx 0.20$). Notably, this category yields an interesting outlier, **r/worldnews** in June 2023, where MGT engagement was significantly lower than for HGT ($\delta = -0.28$), suggesting that users in factual information environments might be more skeptical towards MGT.
- **Social-support discussions**—With **r/relationship_advice** accounting for **ten** significant instances between December 2022 and November 2024. These exhibit some of the strongest observed effect sizes ($\delta \approx 0.28$), suggesting that MGT might receive higher engagement in supportive and advice-oriented contexts.
- **Discussion communities**— **r/politics** shows **two** significative instances occurring within August and October 2024, both associated with a relatively large effect ($\delta \approx 0.36$). This suggests that MGT content might attract more engagement during periods of heightened political activity (e.g., 2024 US elections).
- **Identity-oriented spaces**—With **seven** significant instances (i.e., subreddit-month pairs) primarily in **r/teenagers** between April 2022 and May 2023, this category accounts for a substantial share of cases. Despite a relatively small effect size ($\delta \approx 0.19$), these are consistently positive, suggesting that MGT tends to receive slightly more engagement than the human counterpart among young people.

Results on submissions. The engagement analysis for submissions (cf. Table B.3, Appendix B) reveals that, when significant differences between the engagement distributions of MGT and HGT emerge, they tend to disfavor MGT. Specifically, subreddits such as `r/CryptoCurrency` and `r/writing` show negative effect sizes ($\delta \approx -0.29$ and -0.25, respectively), suggesting that users in these communities are more likely to penalize machine-generated posts. We attribute this asymmetry to the greater visibility and scrutiny that submission posts attract compared to comments: being the primary entry point of a discussion thread, submissions may be evaluated more carefully by community members, leading to lower tolerance for content perceived as inauthentic or of lower quality. Interestingly, three subreddits in Social Support communities (`r/AITAH`, `r/offmychest`, `r/tifu`) still show mixed results, with `r/AITAH` and `r/offmychest` exhibiting higher engagement for MGT, while `r/tifu` leans negative—consistent with the high narrative expectations of that community.

4. Discussion

With this first large-scale study of Machine-Generated Text (MGT) on Reddit, we contribute to advancing the understanding of how Generative AI is reshaping discussions and information sharing in online social media. We moved beyond estimating the overall prevalence of MGT and instead examined its *concentration* and *nature* within an online ecosystem.

Our results show that different communities are affected to varying degrees, with MGT more concentrated in subreddits oriented toward technical knowledge exchange and social support (**RQ1**). After the public release of ChatGPT, the use of MGT has been steady over time, with peaks of prevalence corresponding to major releases of new LLMs (**RQ2**). Most of the MGT is concentrated in the activity of a few users, with 2% of the active users being responsible for the entire MGT production (**RQ3**). We also found MGT often conveys social signals that differ from those of HGT. Most notably, across many subreddits, MGT exhibits strong patterns of social support and status-giving, reflecting the characteristic style of modern AI assistants. Finally, MGT comments are often indistinguishable from HGT in terms of positive engagement, and in some cases even outperform them, whereas more differences arise in the engagement distributions of MGT and HGT submissions (**RQ4**).

Impact. These findings have several implications. Practically, they highlight the need for platforms to adapt moderation and transparency policies, particularly in communities where trust and authenticity are central, such as knowledge-sharing and support fora. Theoretically, they suggest that GenAI is not merely amplifying existing discourse but actively shaping new communicative norms, challenging established notions of authenticity and prompting to consider how human and AI voices co-evolve in online ecosystems. Ethically, the stronger engagement elicited by MGT opens questions about the risks of manipulation and representational imbalance.

Limitations. Our contribution has limitations that future work could address. First, we study a single platform, making the findings not generalizable to other social media. Reddit is well-suited for this type of analysis due to its relatively unconstrained comment length, which facilitates MGT detection. Extending the analysis to platforms where textual content is typically much shorter (e.g., 𝕏, Bluesky) would require the development of specialized MGT detectors for short-form text—a research challenge in itself.

Second, while our selection of subreddits includes popular, active, and diverse communities, it still represents a small portion of the activity on Reddit. These communities are therefore not necessarily representative of the full spectrum of GenAI use or MGT diffusion across the platform. Nevertheless, by demonstrating that the spread of MGT on Reddit is non-negligible, our study highlights the need for broader, more systematic investigations.

Third, although informative temporal patterns emerge from our analysis, the public adoption of GenAI tools is still too recent to draw any conclusions about long-term trends of MGT use. Within our observation window, MGT prevalence appears relatively stable after the initial surge of adoption between November 2022 and September 2023. However, only a systematic monitoring of the platform over longer periods can shed light on the persistence of this phenomenon. Furthermore, prior work suggests that the diffusion of GenAI may influence communication norms in human-authored text as well [36], raising new questions about the co-evolution of human and AI-generated writing styles that more detailed temporal analyses could help uncover.

Fourth, our comparison of levels of engagement between MGT and HGT is based on a weak form of matching based on a combination of subreddit and month of publication. Stronger forms of matching that account for other factors including topic and the author's popularity would yield a more accurate comparison.

Finally, our adopted detection methodology does not allow us to make claims about the exact prevalence of MGT on Reddit. Our estimates are based on a highly conservative classification threshold ($\tau = 0.99$), chosen to demonstrate that even under strict criteria, MGT prevalence is non-negligible in many communities. The actual use of GenAI to produce text on Reddit is likely broader than our estimates suggest. For instance, lowering the classification threshold from 0.99 to 0.95 already doubles the estimated prevalence.

5. Related work

The growing presence of MGT across online platforms has motivated a growing body of research aimed at estimating its prevalence and understanding its potential impact on digital communication. Studies using BERT-based detectors on 15M news articles revealed that by mid-2023, since the release of ChatGPT the proportion of AI-generated news increased by over 57% on mainstream outlets and by an astonishing 474% on misinformation and propaganda sites [37]. Similarly, supervised detectors estimated that AI-generated articles account for roughly 40% of news contributions on Medium and Quora [15]. Analyses of Wikipedia entries based on the GPTZero detector suggest that over 5% of newly created English articles may be AI-generated [38]. Even the scientific review process appears affected: recent estimates indicate that between 6% and 17% of peer reviews for leading Machine Learning conferences may be produced by LLMs [39].

In contrast, fewer studies have quantified the prevalence of AI-generated text on social media. Earlier research primarily detected

machine-generated posts by identifying bot-like authors exhibiting coordinated or automated behaviors [40]. Text-based approaches relying on stylometry [41] offered limited accuracy and generalizability. More recent studies have leveraged crowdsourced signals (such as 𝕏's Community Notes) to examine the diffusion and nature of AI-generated discourse [42].

On Reddit, community responses to AI-generated content have been characterized by skepticism and concern [43], leading moderators and administrators to introduce new governance rules soon after the public release of ChatGPT [44]. Crowdsourced data indicate that AI-generated visual content remained relatively rare through late 2023, representing fewer than 0.5% of image-based posts [45]. In this context, the most closely related study to ours is the recent work by Sun et al. [15], who applied a supervised detection method to 982K Reddit posts published between January 2022 and July 2024, estimating that around 2.5% were AI-generated. However, it should be noted that the latter work presents some gaps that we aim to fill in this paper, as outlined next. First, Sun et al. consider Reddit within a multi-platform framework, limiting the analysis to temporal prevalence trends and top AI-generated words. Our work overcomes these limitations by considering the effects of MGT on engagement, social dimensions, and user-level adoption. Second, Sun et al. focus on fewer than 1M comments and no submissions, whereas we analyze more than 9M comments and 2M submissions. Third, our temporal window is larger than Sun et al. (July 2024 vs. December 2024). Finally, Sun et al. do not specify how many subreddits were analyzed, whereas we consider a principled sample of 51 subreddits, enabling a deeper community-level perspective.

CRediT authorship contribution statement

Lucio La Cava: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Luca Maria Aiello:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Andrea Tagarelli:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

AT, resp. LLC, is supported by project "Future Artificial Intelligence Research (FAIR)" spoke 9 (H23C22000860006), resp. project SERICS (PE00000014), both under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU. LMA acknowledges the support from the Carlsberg Foundation through the COCOONS project (CF21-0432).

Appendix A. Additional details on methodology

A.1. Categorization of Reddit communities

Below, we provide the full list of our 51 selected subreddits across the five community types:

Information Seeking. r/worldnews, r/DIY, r/askscience, k r/personalfinance, r/technology, r/history, r/CryptoCurrency, r/nutrition, r/learnprogramming, r/explainlikeimfive, r/health, r/techsupport, r/writing.

Social Support. r/LifeProTips, r/GetMotivated, r/relationship_advice, r/lifehacks, r/malefashionadvice, r/dating, r/femalefashionadvice, r/careerguidance, r/loseit, t/tifu, r/Parenting, r/fitness, r/AITAH, r/AmIOverreacting, r/offmychest, r/SuicideWatch, r/vent, r/mentalhealth.

Discussion. r/politics, r/unpopularopinion, r/changemyview, r/SeriousConversation, r/PoliticalDiscussion.

Identity. r/BlackPeopleTwitter, r/AskWomen, r/IAmA, r/teenagers, r/introvert, r/asktransgender, r/Conservative, r/Liberal, r/Libertarian, r/Socialism.

ChitChat. r/funny, r/Showerthoughts, r/entertainment, r/books, r/CasualConversation.

Table A.1
Key milestones in the development of GenAI during the time period covered by our data.

Date	Event
Nov 2022	Release of ChatGPT by OpenAI, marking the mainstream breakthrough of conversational AI
Feb 2023	Release of Llama by Meta, fostering access to open-weight large language models
Mar 2023	Release of GPT-4 by OpenAI, introducing multimodal capabilities and improved reasoning
Mar 2023	Release of Claude by Anthropic, focusing on AI safety and alignment
Jul 2023	Release of Llama 2 by Meta, expanding access to open-weight large language models
Dec 2023	Launch of Gemini by Google, a multimodal AI model family competing with GPT-4
Jan–Apr 2024	Rapid wave of model releases: Grok (xAI), Claude 3 (Anthropic), Llama 3 (Meta), and GPT-4o (OpenAI), emphasizing multimodality, speed, and real-time interaction

A.2. Key GenAI releases

Table A.1 summarizes major milestones in the evolution of generative AI over the period covered by our dataset.

A.3. Usage of alternative detectors

While our study was conducted by using Fast-DetectGPT as a reference detector following the efficiency and effectiveness criteria discussed in the main text, we have also performed a comparison with Binoculars [30], another state-of-the-art zero-shot detector which we found to be 19× slower than our reference model.

In particular, by using the latter in a "low-FPR mode" on a stratified sample of 46,310 Reddit comments, we observed a substantial agreement with Fast-DetectGPT (Cohen's $\kappa = 0.676$). Furthermore, we found all major findings from our main text (i.e., observed by using Fast-DetectGPT as a reference detector) to replicate even when considering an alternative detector.

First, the subreddit-rankings by considering MGT prevalence observed in Fig. 1 are nearly identical (Spearman = 0.987). Second, the user-adoption trends shown in Fig. 2 correlate strongly across detectors (Pearson = 0.8). Furthermore, we note that 19 out of 20 text-statistics comparisons (cf. Table 3) remain significant, aligned, and directionally consistent, as well as 26 (subreddit-type,social-dimension)-pair differences (cf. Fig. 3) remain significant and with similar effect-size direction. Finally, we point out that also engagement effect-sizes remain consistent across Fast-DetectGPT (mean = 0.194, median = 0.238) and Binoculars (mean = 0.205, median = 0.247).

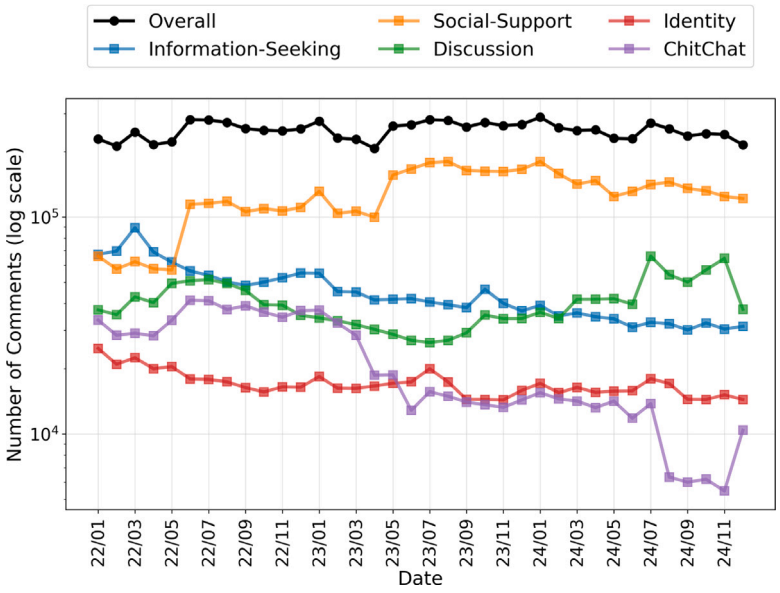


Fig. A.1. Monthly volume of comments for each subreddit category.

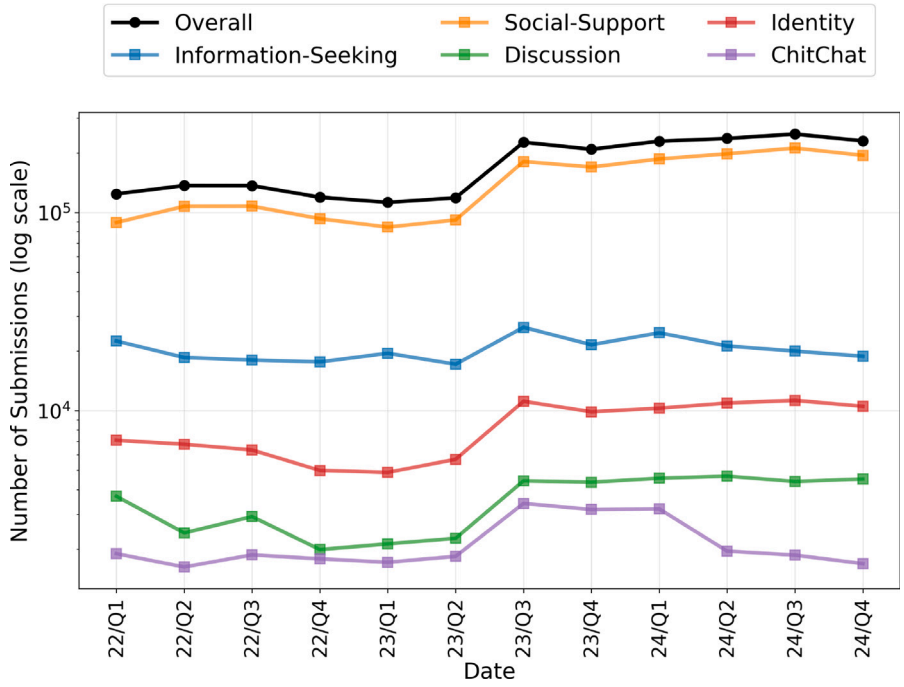


Fig. A.2. Quarterly volume of submissions for each subreddit category.

A.4. Details on experimental setup

All experiments have been performed on an 8x NVIDIA A30 GPU 24 GB NVRAM server with 24 GB of RAM each, 764 GB of system RAM, a Double Intel Xeon Gold 6248R with a total of 96 cores, and Ubuntu Linux 20.04.6 LTS as OS.

To ease reproducibility, we adopted Fast-DetectGPT via its default parameter settings, as in their official GitHub repository.[4]

[4] https://github.com/baoguangsheng/fast-detect-gpt.

A.5. Temporal distribution of analyzed data

Figs. A.1–A.2 report the number of comments and submissions over time, across the different community types. Submission data is aggregated by quarters instead of monthly to counter sparsity.

Appendix B. Detailed results on the analysis of machine-generated text in submissions

We applied to Reddit submissions the same analytical framework used for comments. Fig. A.3 illustrate the prevalence of MGT across

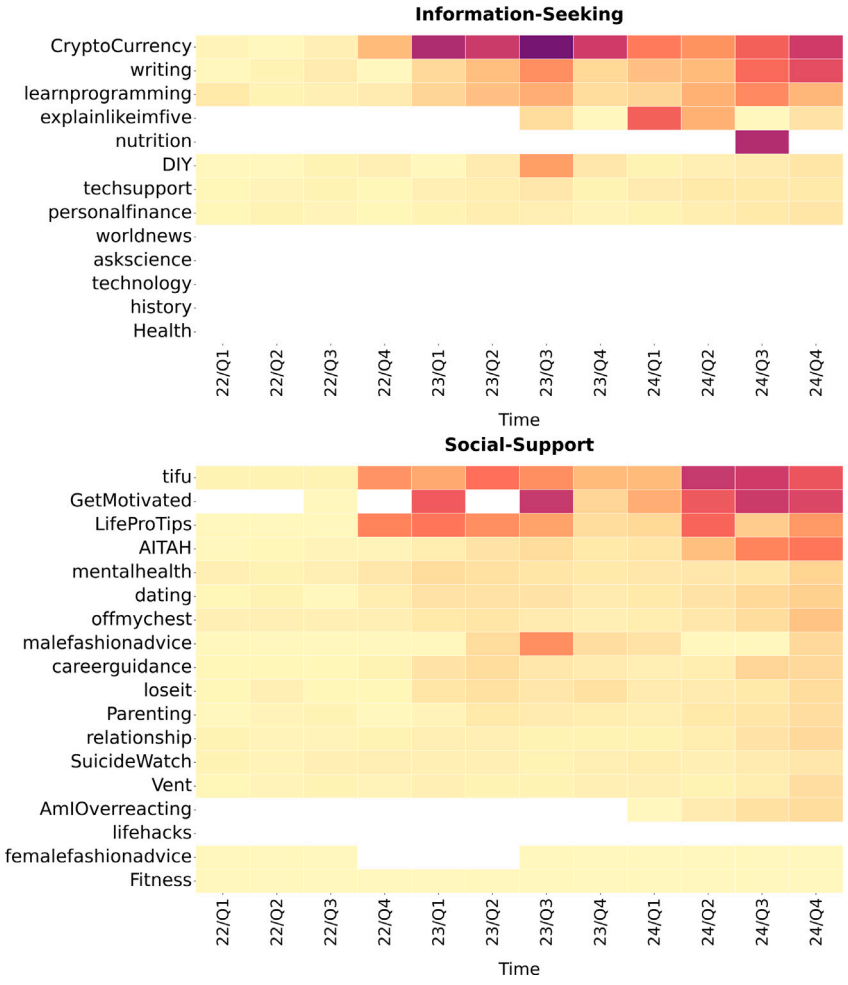


Fig. A.3. MGT submission usage across subreddit categories. Rows are sorted by total quarterly adoption. Each cell shows the share of MGT posts for a subreddit–quarter pair. Darker shades indicate higher usage. Only subreddits with ≥1 post/day on avg. are shown; empty cells mean no matching submissions.

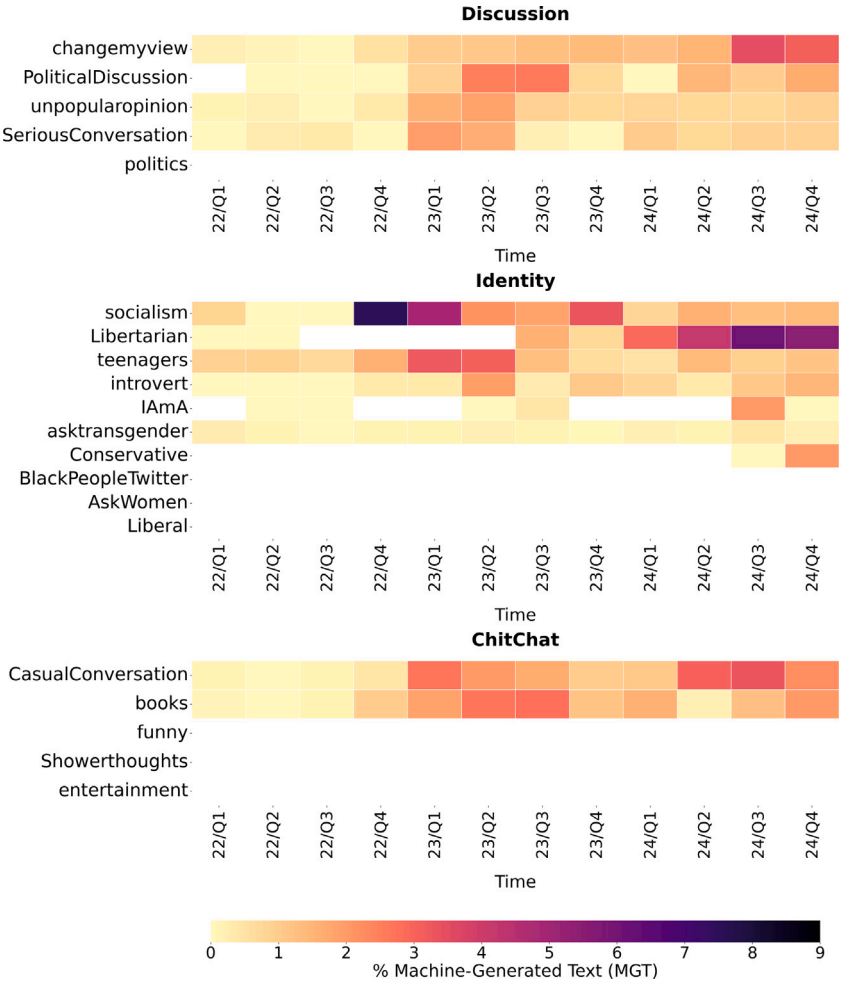


Fig. A.3. (*continued*).

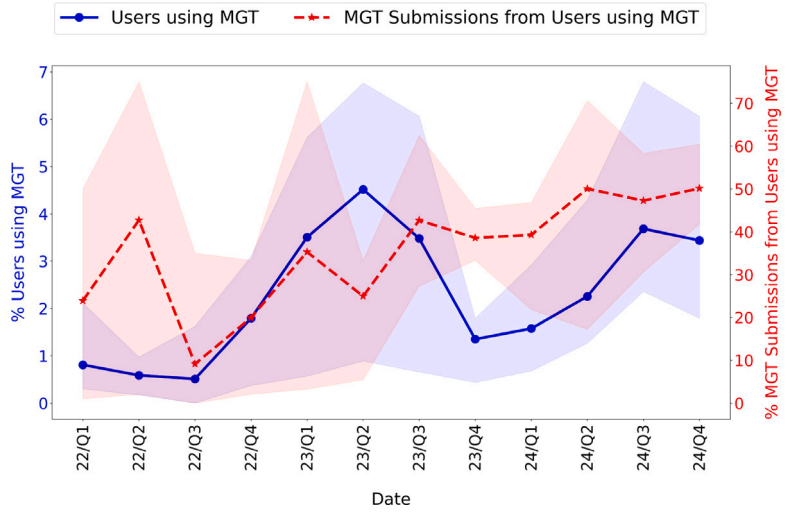


Fig. A.4. Average number of users adopting MGT (in blue, left y-axis) and the corresponding percentage of their submissions detected to be MGT (in red, right y-axis) across all categories over time. Shaded areas denote minimum and maximum observed values. Users with a single post have been filtered out to mitigate noise due to one-off activities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

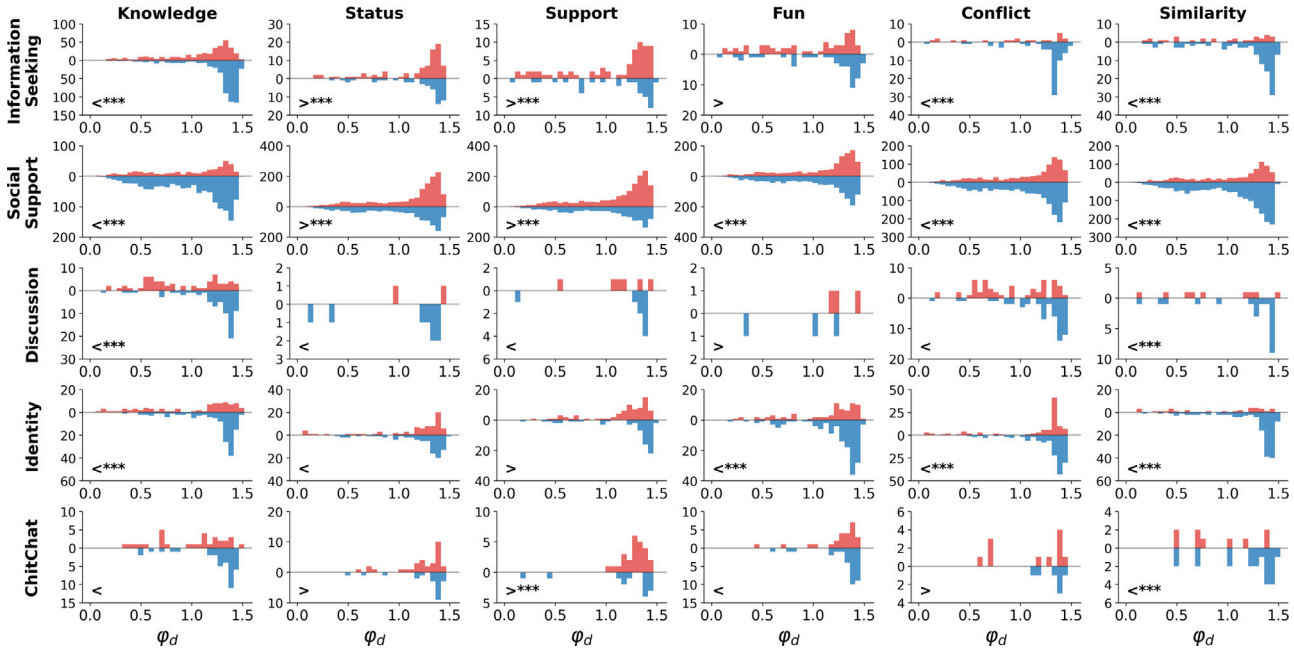


Fig. A.5. Comparison of social dimension expressions between MGT (in red) and HGT (in blue) submissions across different subreddit categories. Each subplot is labeled with the corresponding effect size, where > indicates that MGT scores higher than HGT, and < indicates the opposite. Marker *** denotes that the difference in social dimension expression scores between MGT and HGT is statistically significant according to the Mann–Whitney U test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table B.1
Peak prevalence of MGT across subreddit categories. For each category, we report the fraction of MGT submissions over the total detected, the maximum observed share of MGT, the subreddit in which it occurred, and the corresponding date (yy/quarter).

Community	MGT %	Peak %	Top Subreddit	Date
Inf. Seek.	14.23	6.00	`r/CryptoCurrency`	23/Q3
Soc. Supp.	71.42	4.16	`r/tifu`	24/Q2
Discussion	3.72	3.45	`r/changemyview`	24/Q3
Identity	7.26	7.53	`r/socialism`	22/Q4
ChitChat	3.37	3.30	`r/CasualConversation`	24/Q3

Table B.2
Aggregated values of text statistics from the machine-generated and human-authored submissions across subreddit categories. All category–statistics combinations (but # Words and Compression for Social Support) exhibit statistically significant distributional differences ($p < 0.05$) according to the Mann–Whitney U test.

Category	Text	# Words	# Sentences	Flesch Read.	Compression
Inf.Seek.	HGT	352.8 ± 338.95	20.79 ± 22.12	58.3 ± 28.88	3.07 ± 0.72
	MGT	527.16 ± 488.73	31.49 ± 34.13	51.59 ± 21.63	3.48 ± 0.57
Soc.Supp.	HGT	493.05 ± 357.3	24.78 ± 18.87	71.96 ± 51.61	3.17 ± 0.42
	MGT	471.2 ± 337.14	28.52 ± 20.82	74.86 ± 18.06	3.2 ± 1.27
Discussion	HGT	404.86 ± 400.81	20.31 ± 22.08	56.92 ± 26.25	3.11 ± 0.31
	MGT	650.27 ± 443.12	34.45 ± 22.75	43.06 ± 28.13	3.51 ± 0.42
Identity	HGT	356.66 ± 234.28	17.39 ± 15.16	59.95 ± 80.08	3.53 ± 8.47
	MGT	552.99 ± 802.74	33.97 ± 86.57	−10.09 ± 484.72	10.03 ± 48.86
ChitChat	HGT	345.11 ± 266.0	20.08 ± 19.77	74.36 ± 12.13	2.96 ± 0.29
	MGT	430.22 ± 381.33	23.84 ± 18.42	61.45 ± 29.28	3.19 ± 0.46

subreddits and over time, while Table 3 summarizes statistics on peak prevalence across subreddit types. Table B.2 reports measures of text length, readability, and compressibility for submissions, and Fig. A.5 shows the distribution of social dimensions for MGT and HGT. Engagement statistics are presented in Table B.3. All measurements are aggregated by quarters instead of monthly to counter sparsity.

Data availability

Data will be made available on request.

Table B.3
Summary of subreddit–quarter pairs where engagement distributions between MGT and HGT differ significantly. Positive, resp. negative, values of Cliff's δ indicate higher, resp. lower, engagement for MGT. For each case, %p reports the fraction of significant tests.

Category	Subreddit	# Quarters	Avg. δ	% p
Information seeking	r/CryptoCurrency	4	−0.29	0.97
	r/writing	2	−0.25	0.86
Social support	r/AITAH	4	0.15	0.88
	r/offmychest	2	0.24	0.61
	r/tifu	1	−0.26	0.89
Discussion	r/changemyview	1	−0.21	0.65

References

[1] R.W. White, Advancing the search frontier with ai agents, Commun. ACM 67 (9) (2024) 54–65.

[2] J.V. Pavlik, Collaborating with chatgpt: Considering the implications of generative artificial intelligence for journalism and media education, Journal. Mass Commun. Educ. 78 (1) (2023) 84–93.

[3] N. Anantrasirichai, D. Bull, Artificial intelligence in the creative industries: a review, Artif. Intell. Rev. 55 (1) (2022) 589–656.

[4] C.A. Bail, Can generative ai improve social science? Proc. Natl. Acad. Sci. 121 (21) (2024) e2314021121.

[5] Y. Bengio, G. Hinton, A. Yao, D. Song, P. Abbeel, T. Darrell, Y.N. Harari, Y.-Q. Zhang, L. Xue, S. Shalev-Shwartz, et al., Managing extreme ai risks amid rapid progress, Science 384 (6698) (2024) 842–845.

[6] A.R. Doshi, O.P. Hauser, Generative ai enhances individual creativity but reduces the collective diversity of novel content, Sci. Adv. 10 (28) (2024) eadn5290.

[7] M.H. Tessler, M.A. Bakker, D. Jarrett, H. Sheahan, M.J. Chadwick, R. Koster, G. Evans, L. Campbell-Gillingham, T. Collins, D.C. Parkes, et al., Ai can help humans find common ground in democratic deliberation, Science 386 (6719) (2024) eadq2852.

[8] L. Floridi, M. Chiriatti, Gpt-3: Its nature, scope, limits, and consequences, Minds Mach. 30 (4) (2020) 681–694.

[9] A.G. Møller, D.M. Romero, D. Jurgens, L.M. Aiello, The impact of generative ai on social media: An experimental study, 2025, arXiv preprint arXiv:2506.14295.

[10] A.F. Ashery, L.M. Aiello, A. Baronchelli, Emergent social conventions and collective bias in llm populations, Sci. Adv. 11 (20) (2025) eadu9368.

[11] S.M. Breum, D.V. Egdal, V.G. Mortensen, A.G. Møller, L.M. Aiello, The persuasive power of large language models, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 18, 2024, pp. 152–163.

[12] D.T. Schroeder, M. Cha, A. Baronchelli, N. Bostrom, N.A. Christakis, D. Garcia, A. Goldenberg, Y. Kyrychenko, K. Leyton-Brown, N. Lutz, et al., How malicious ai swarms can threaten democracy, 2025, arXiv preprint arXiv:2506.06299.

[13] A. Turillazzi, M. Taddeo, L. Floridi, F. Casolari, The digital services act: an analysis of its ethical, legal, and social implications, Law Innov. Technol. 15 (1) (2023) 83–106.

[14] G. Bao, Y. Zhao, Z. Teng, L. Yang, Y. Zhang, Fast-detectgpt: Efficient zero-shot detection of machine-generated text via conditional probability curvature, in: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7–11, 2024, OpenReview.net, 2024, https://openreview.net/forum?id=Bpcgcr8E8Z.

[15] Z. Sun, Z. Zhang, X. Shen, Z. Zhang, Y. Liu, M. Backes, Y. Zhang, X. He, Are we in the ai-generated text world already? quantifying and monitoring aigt on social media, in: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 22975–23005.

[16] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, M. Zimmer, Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics, Soc. Media+ Soc. 7 (2) (2021) 20563051211019004.

[17] J. Lenti, L.M. Aiello, C. Monti, G.D.F. Morales, Causal modeling of climate activism on reddit, in: Proceedings of the ACM on Web Conference 2025, 2025, pp. 590–600.

[18] X. Zhao, V. Yang, A. Menta, J. Blum, P. Ranasinghe, Exploring the use of social media for medical problem solving by analyzing the subreddit r/medical_advice: quantitative analysis, JMIR Infodemiology 5 (2025) e56116.

[19] G. Weld, A.X. Zhang, T. Althoff, What makes online communities 'better'? measuring values, consensus, and conflict across thousands of subreddits, in: C. Budak, M. Cha, D. Quercia (Eds.), Proceedings of the Sixteenth International AAAI Conference on Web and Social Media, ICWSM 2022, Atlanta, Georgia, USA June 6–9, 2022, AAAI Press, 2022, pp. 1121–1132, https://ojs.aaai.org/index.php/ICWSM/article/view/19363.

[20] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, J. Blackburn, The pushshift reddit dataset, in: M.D. Choudhury, R. Chunara, A. Culotta, B.F. Welles (Eds.), Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM 2020, Held Virtually, Original Venue: Atlanta, Georgia, USA June 8–11, 2020, AAAI Press, 2020, pp. 830–839, https://ojs.aaai.org/index.php/ICWSM/article/view/7347.

[21] S. Gehrmann, H. Strobelt, A. Rush, GLTR: Statistical detection and visualization of generated text, in: M.R. Costa-jussà, E. Alfonseca (Eds.), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Florence, Italy, 2019, pp. 111–116, http://dx.doi.org/10.18653/v1/P19-3019, https://aclanthology.org/P19-3019/.

[22] I. Solaiman, M. Brundage, J. Clark, A. Askell, A. Herbert-Voss, J. Wu, A. Radford, G. Krueger, J.W. Kim, S. Kreps, et al., Release strategies and the social impacts of language models, 2019, arXiv preprint arXiv:1908.09203.

[23] E. Mitchell, Y. Lee, A. Khazatsky, C.D. Manning, C. Finn, DetectGPT: Zero-shot machine-generated text detection using probability curvature, in: Proceedings of the 40th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 202, PMLR, 2023, pp. 24950–24962, https://proceedings.mlr.press/v202/mitchell23a.html.

[24] J. Su, T. Zhuo, D. Wang, P. Nakov, DetectLLM: Leveraging log rank information for zero-shot detection of machine-generated text, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, Singapore, 2023, pp. 12395–12412, http://dx.doi.org/10.18653/v1/2023.findings-emnlp.827, https://aclanthology.org/2023.findings-emnlp.827/.

[25] B. Guo, X. Zhang, Z. Wang, M. Jiang, J. Nie, Y. Ding, J. Yue, Y. Wu, How close is chatgpt to human experts? comparison corpus, evaluation, and detection, 2023, arXiv preprint arXiv:2301.07597.

[26] X. Guo, Y. He, S. Zhang, T. Zhang, W. Feng, H. Huang, C. Ma, Detective: Detecting AI-generated text via multi-level contrastive learning, in: The Thirty-Eighth Annual Conference on Neural Information Processing Systems, 2024.

[27] L.L. Cava, D. Costa, A. Tagarelli, Is contrasting all you need? contrastive learning for the detection and attribution of ai-generated text, in: ECAI 2024–27th European Conference on Artificial Intelligence, in: Frontiers in Artificial Intelligence and Applications, vol. 392, IOS Press, 2024, pp. 3179–3186, http://dx.doi.org/10.3233/FAIA240862.

[28] X. He, X. Shen, Z. Chen, M. Backes, Y. Zhang, Mgtbench: Benchmarking machine-generated text detection, in: Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, CCS'24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 2251–2265, http://dx.doi.org/10.1145/3658644.3670344.

[29] L. La Cava, A. Tagarelli, OpenTuringBench: An open-model-based benchmark and framework for machine-generated text detection and attribution, in: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Suzhou, China, 2025, pp. 26655–26671, http://dx.doi.org/10.18653/v1/2025.emnlp-main.1354, https://aclanthology.org/2025.emnlp-main.1354/.

[30] A. Hans, A. Schwarzschild, V. Cherepanova, H. Kazemi, A. Saha, M. Goldblum, J. Geiping, T. Goldstein, Spotting llms with binoculars: zero-shot detection of machine-generated text, in: Proceedings of the 41st International Conference on Machine Learning, ICML'24, JMLR.org, 2024.

[31] D. Hovy, D. Yang, The importance of modeling social factors of language: Theory and practice, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 588–602.

[32] S. Deri, J. Rappaz, L.M. Aiello, D. Quercia, Coloring in the links: Capturing social ties as they are perceived, Proc. ACM Hum.-Comput. Interact. 2 (CSCW) (2018) http://dx.doi.org/10.1145/3274312.

[33] M. Choi, L.M. Aiello, K.Z. Varga, D. Quercia, Ten social dimensions of conversations and relationships, in: Y. Huang, I. King, T. Liu, M. van Steen (Eds.), WWW'20: The Web Conference 2020, Taipei, Taiwan, April 20–24, 2020, ACM/IW3C2, 2020, pp. 1514–1525, http://dx.doi.org/10.1145/3366423.3380224.

[34] C. Monti, L.M. Aiello, G. De Francisci Morales, F. Bonchi, The language of opinion change on social media under the lens of communicative action, Sci. Rep. 12 (1) (2022) 17920.

[35] S. Chakraborty, A.S. Bedi, S. Zhu, B. An, D. Manocha, F. Huang, Position: On the possibilities of ai-generated text detection, in: Forty-First International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21–27, 2024, OpenReview.net, 2024, https://openreview.net/forum?id=CJbhtpcyGL.

[36] D. Agarwal, M. Naaman, A. Vashistha, Ai suggestions homogenize writing toward western styles and diminish cultural nuances, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–21.

[37] H.W. Hanley, Z. Durumeric, Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 18, 2024, pp. 542–556.

[38] C. Brooks, S. Eggert, D. Peskoff, The rise of ai-generated content in wikipedia, 2024, arXiv preprint arXiv:2410.08044.

[39] W. Liang, Z. Izzo, Y. Zhang, H. Lepp, H. Cao, X. Zhao, L. Chen, H. Ye, S. Liu, Z. Huang, et al., Monitoring ai-modified content at scale: A case study on the impact of chatgpt on ai conference peer reviews, 2024, arXiv preprint arXiv:2403.07183.

[40] F. Cinus, M. Minici, L. Luceri, E. Ferrara, Exposing cross-platform coordinated inauthentic activity in the run-up to the 2024 us election, in: Proceedings of the ACM on Web Conference 2025, 2025, pp. 541–559.
[41] T. Kumarage, J. Garland, A. Bhattacharjee, K. Trapeznikov, S. Ruston, H. Liu, Stylometric detection of ai-generated text in twitter timelines, 2023, arXiv preprint arXiv:2303.03697.
[42] C. Drolsbach, N. Pröllochs, Characterizing ai-generated misinformation on social media, 2025, arXiv preprint arXiv:2505.10266.
[43] T. Lloyd, J. Reagle, M. Naaman, "There has to be a lot that we're missing": Moderating ai-generated content on reddit, 2023, arXiv preprint arXiv:2311.12702.
[44] T. Lloyd, J. Gosciak, T. Nguyen, M. Naaman, Ai rules? characterizing reddit community policies towards ai-generated content, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–19.
[45] H. Matatov, M.A.L. Quéré, O. Amir, M. Naaman, Examining the prevalence and dynamics of ai-generated media in art subreddits, 2024, arXiv preprint arXiv:2410.07302.