GPT Store Mining and Analysis

Dongxun Su*‡
dxsssu0114@hust.edu.cn
Huazhong University of Science and Technology
Wuhan, China

Yanjie Zhao*‡
yanjie_zhao@hust.edu.cn
Huazhong University of Science and Technology
Wuhan, China

Xinyi Hou‡
xinyihou@hust.edu.cn
Huazhong University of Science and Technology
Wuhan, China

Shenao Wang‡
shenaowang@hust.edu.cn
Huazhong University of Science and Technology
Wuhan, China

Haoyu Wang†‡
haoyuwang@hust.edu.cn
Huazhong University of Science and Technology
Wuhan, China

ABSTRACT

As an important extension of the ChatGPT ecosystem, GPT Store has developed into an active market hosting more than 3 million customized ChatGPTs (GPTs). Despite its large scale, the current academic community still has obvious limitations in its understanding of the ecosystem of this platform. Based on a complete dataset of more than 700,000 GPTs, this paper has achieved a multi-dimensional analysis of GPT Store. We first systematically examined the platform operation mechanism, covering core elements such as the classification system, interaction mode, and evaluation system. We also comprehensively analyzed the security risks, such as data leakage and jailbreak in GPT Store. Finally, through a user study, this work revealed the behavioral characteristics and experience pain points in real usage scenarios. Based on these findings, we provide operational platform optimization suggestions, including functional improvement, security enhancement, and interaction improvement. This study not only constructs an analytical framework for the GPT Store ecosystem but also provides empirical evidence and optimization directions for its future development.

CCS CONCEPTS

• **Security and privacy → Software and application security**; • **Computing methodologies → Artificial intelligence**.

KEYWORDS

GPT Store, Large Language Model, Empirical Software Engineering

ACM Reference Format:
Dongxun Su, Yanjie Zhao, Xinyi Hou, Shenao Wang, and Haoyu Wang. 2025. GPT Store Mining and Analysis. In *the 16th International Conference on Internetware (Internetware 2025), June 20–22, 2025, Trondheim, Norway.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3755881.3755900

*Dongxun Su and Yanjie Zhao are the co-first authors.
†Haoyu Wang (haoyuwang@hust.edu.cn) is the corresponding author.
‡The full name of the authors' affiliation is Hubei Key Laboratory of Distributed System Security, Hubei Engineering Research Center on Big Data Security, School of Cyber Science and Engineering, Huazhong University of Science and Technology.

Internetware 2025, June 20–22, 2025, Trondheim, Norway

ACM ISBN 979-8-4007-1926-4/25/06
https://doi.org/10.1145/3755881.3755900

1 INTRODUCTION

The emergence of large language models (LLMs) has revolutionized the way people use AI. These models have powerful natural language processing capabilities, which not only reshape people's way of obtaining information but also open up new ways to automate various tasks. ChatGPT is one of the most representative products among many LLM applications. Its influence is not only reflected in the astonishing scale of users, but also penetrates many fields such as education, business, creative writing, and software development [27]. This wide application highlights the practical value of conversational AI and the growing demand for easy-to-use and customizable AI tools.

With the popularity of ChatGPT, OpenAI further launched the GPT Store [28], which allows users to develop customized GPT applications (GPTs) according to their needs. Developers can easily define the behavior of GPTs through natural language instructions, upload external knowledge bases, and enable specific functional modules (such as web browsing, image generation, etc.). This low-code development method greatly reduces the technical threshold and attracts widespread participation from ordinary users to professional developers. According to the official description, GPT Store has more than 3 million GPTs [29].

Although GPT Store has made significant progress in promoting the popularization of LLM applications, systematic research on this platform is limited. We still lack a comprehensive understanding of the composition, usage, and user behavior of GPTs. At the same time, the rapid growth of the platform's content scale has also brought new security and governance challenges. For example, some studies have pointed out that personalized GPT may face risks such as prompt injection, information leakage, and jailbreak attacks [41], but the prevalence, causes, and impacts of these problems have not been deeply explored. In addition, due to the lack of a complete content review mechanism on the platform, many users have reported that they encounter problems such as uneven content quality and opaque recommendation mechanisms during use. These factors not only affect the user experience but may also pose risks to the long-term sustainable development of the platform.

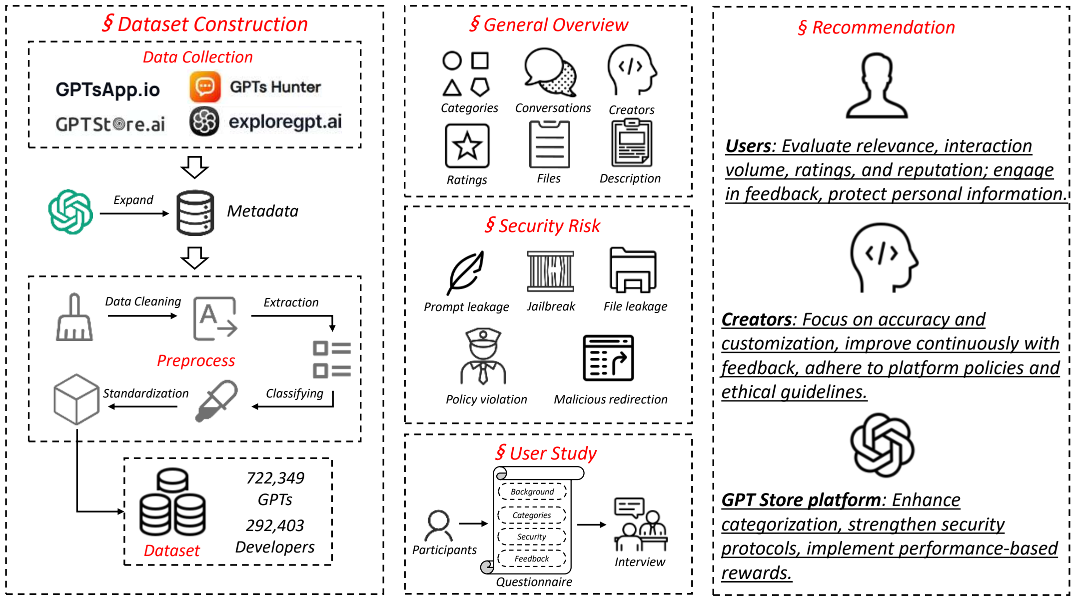

Figure 1: The overall structure of our paper.

To fill this research gap, our work conducts a large-scale and systematic empirical analysis of the GPT Store ecosystem. We constructed a large dataset covering more than 700,000 GPTs, and combined statistical modeling and user surveys to comprehensively examine the type distribution, creator characteristics, usage patterns, and potential security risks of GPT applications. We also designed questionnaires and interviews to collect real feedback from users on content quality, search experience, and trust mechanisms. Based on the above analysis, we further propose feasible improvement suggestions in order to provide reference and guidance for the platform's ecological governance, security enhancement, and user experience optimization. The overall structure of our paper is illustrated in Figure 1.

Contributions. We make the following key contributions to understanding the GPT Store ecosystem:

- We built a large-scale GPT Store dataset containing detailed attributes of more than 700,000 GPTs.
- We comprehensively analyzed the current status of GPT Store, including platform structure, functional characteristics, and the diversity distribution of GPTs in different domains.
- We systematically studied the main security risks of GPT Store and revealed the current key vulnerabilities.
- Through user studies, we learned about the real experience and challenges encountered by users in using GPT and provided optimization suggestions.

2 BACKGROUND

In November 2023, OpenAI announced the GPTs feature [27], which allows users to create customized ChatGPT versions (GPTs) through natural language instructions. Ordinary users without programming knowledge can easily create GPTs. To increase GPTs' capabilities, there is a visual configuration interface that allows for the addition of knowledge base files (such as PDF, Excel, and so on). In January 2024, OpenAI introduced the GPT Store platform, which uses a profit-sharing approach to share revenue with successful GPT authors. The GPT Store offers users a variety of ways to explore: they can browse popular GPTs based on community rankings or search for GPTs in certain domains. Official data shows that developers have created more than 3 million GPTs [7]. This reflects both the innovative vitality of the community and the strong market demand for personalized AI solutions. As shown in Figure 2, we counted the number of new GPTs and their cumulative distribution function (CDF) from November 2023 to January 2025. These visualizations clearly show the growth trajectory of the GPT ecosystem during this period.

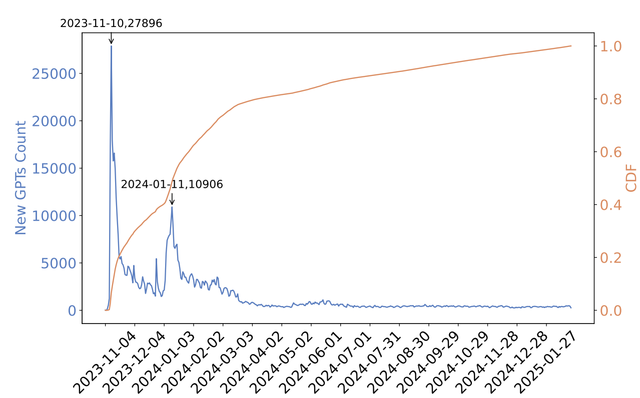

Figure 2: Daily new GPTs count and cumulative distribution function (Nov 2023-Jan 2025).

In addition to the official platform, third-party GPT stores have also flourished. Independent markets established by external developers, such as GPTs Hunter [18], GPTStore.AI [14], and GPTs App [13], provide users with a more convenient navigation and discovery experience by selecting high-quality GPTs from the official GPT Store. These third-party stores usually also integrate user evaluation systems, personalized recommendation algorithms, and professional screening functions for vertical fields, making up for the shortcomings of official stores in data analysis through technical means.

3 DATA COLLECTION

At present, there are several datasets related to GPT Store, such as GPTZoo[17], which is updated to May 2024. We found that the GPT Store is updated very quickly, so the timeliness of the dataset is difficult to guarantee. We actually visited the GPTs in GPTZoo and found that many GPTs have been removed from the store. In order to obtain the latest GPT dataset, we launched a new round of data collection in January 2025. The specific implementation process is as follows:

- For third-party GPT Stores, we independently developed customized crawlers for each target platform (including gpt-store.ai, gptshunter.com, and gptsapp.io). We automatically obtain GPT access links from these third-party platforms and then initiate requests to the OpenAI official GPT Store to finally obtain complete GPT metadata.
- The OpenAI official GPT Store does not provide a complete GPT list, and only allows users to obtain corresponding GPTs by searching for keywords. Therefore, we use the collected GPT names as seed query terms, adopt a breadth-first search strategy, and systematically crawl the relevant GPTs recommended by the platform based on the seed names, thereby achieving an extended collection of GPT metadata.

In the collection operation in January 2025, we used 5 PCs to work in parallel, and each device maintained 10 concurrent threads. After pre-processing processes such as validity verification and deduplication, we successfully obtained **722,349 valid GPTs**, which were created by **292,403 developers**. Our verification analysis shows that these data are not only fully accessible and valid but also large enough to support GPT Store-related research. For a detailed analysis of the dataset, see the § 4.

4 GENERAL OVERVIEW

In this section, we conduct a comprehensive analysis of the GPT store, focusing on its key features such as the classification system, number of conversations, creators, ratings, etc.

4.1 Categories

OpenAI's official GPT store divides products into eight main categories: Dalle, Education, Lifestyle, Productivity, Programming, Research, Writing, and Others. Table 1 shows the specific distribution of GPTs under each category, with Education ranking first with 98,404 GPTs, followed by Productivity (74,964) and Lifestyle (64,664). It is worth noting that although the Dalle category has the smallest number (18,603), it shows the highest average conversation volume per GPT (2,359 times), while the writing category GPTs maintain a large scale (55,721) while the user rating also reaches an excellent level of 4.24 points. Figure 3 presents the proportion of GPTs in each category through a pie chart. Nearly 32% of GPTs are not classified into any preset category, which not only reflects the diversity of application scenarios, but also reflects that some products are difficult to be simply classified. This phenomenon of missing classifications not only illustrates the widespread application and development trends of GPT, but may also reveal potential problems in the current official classification mechanism.

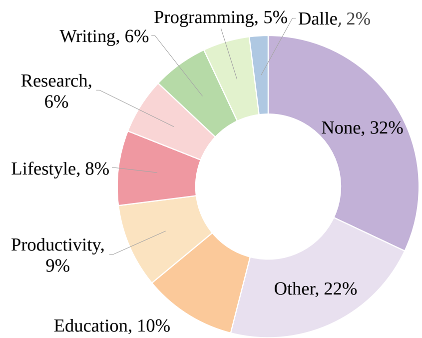

Figure 3: Category distribution of GPTs in the GPT Store.

4.2 Number of conversations

Each GPT records the number of user conversations, and each query increases the count. For GPTs with fewer conversations, the exact value is displayed, and for GPTs with more than 100 conversations, the approximate value (such as 100+, 2K+, 3M+, etc.) is displayed. The data in Table 1 also presents the total conversation volume and average conversation volume of each category. It is worth noting that the Dalle category GPT focusing on artistic creation ranks first with an average conversation volume of 2359.69 times, reflecting users' strong interest in creativity-driven applications. In contrast, the writing and programming categories also show high user engagement, while the `other` category ranks last with an average conversation volume of 277.77 times, highlighting the difference in user engagement of different types of GPTs.

As shown in Figure 4, GPT conversation data generally follows a power law distribution. Our analysis shows that about 90% of GPT conversations have less than 100 conversations, and less than 0.1% of GPTs have more than 100,000 conversations. This distribution pattern shows that while a few GPTs are extremely popular and widely used, most GPTs have relatively limited engagement, which may mean that the marketing or management strategy of GPTs within the platform needs to be rethought.

4.3 Ratings

Soon after the launch of the GPT store, an updated rating system [5] was introduced. This system allows users to rate each GPT from 1 to 5 points based on their experience. When users visit a GPT page,

Table 1: Statistics of GPTs categories.

Category	# GPTs	# Conversations	Avg. Conversations/GPT	# Ratings	Avg. Ratings
Education	98,404	106,673,254	1,084.03	679,145	4.32
Productivity	74,964	69,919,514	932.71	998,838	3.95
Lifestyle	64,664	26,126,817	404.04	320,855	4.32
Research	57,390	50,389,105	878.01	662,622	4.17
Writing	55,721	62,079,052	1,114.11	608,984	4.24
Programming	40,570	32,723,508	806.59	474,231	4.16
Dalle	18,603	43,897,320	2,359.69	526,014	3.50
Other	142,382	32,430,833	227.77	320,062	4.12

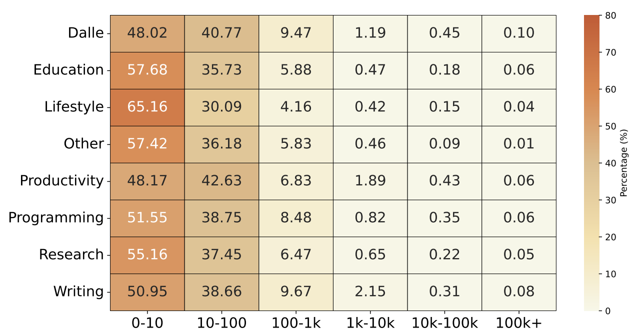

Figure 4: Distribution of conversation counts across different GPT categories.

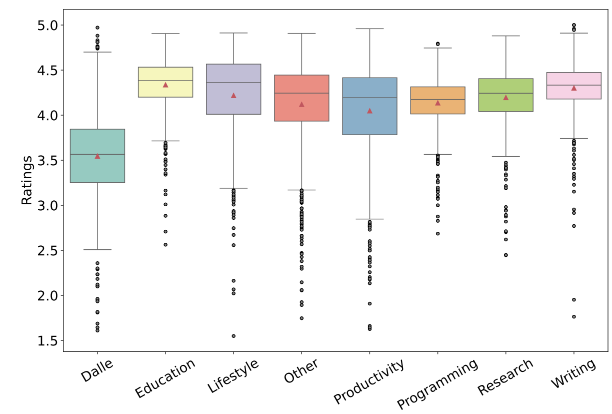

Figure 5: Distribution of user ratings across GPTs categories.

they can not only view the total number of ratings and the average score, but also observe the specific distribution of each score from 1 to 5.

As shown in Figure 5, we used box plots to analyze the rating distribution of different categories of GPTs. In order to reduce the interference of low-rated samples, we only included GPTs with more than 100 ratings for analysis. The results show that the ratings of most GPTs are concentrated between 4.0 and 4.5 points. We found that the average rating of education GPTs is relatively high, followed by writing, research, and lifestyle GPTs. These results reflect that users are generally satisfied with these application scenarios. In contrast, DALL·E GPTs have the lowest ratings. We speculate that this may be due to users' high expectations for visual output quality and creativity, and also reflects the lack of model capabilities in current visual generation tasks.

The relationship between the number of conversations and the rating reveals the performance trend of various GPTs. Figure 6 shows that when the conversation volume reaches the highest level, the scores of most categories show a downward trend, which may reflect the challenges faced by high-traffic GPT in meeting the needs of diverse users. Education GPT maintains a stable high score (4.32 points) in all conversation volume ranges, while the score of DALL·E GPT (3.5 points) is significantly lower than that of other categories, and the score further decreases with increasing usage, highlighting the quality control difficulties faced by the field of AI image generation.

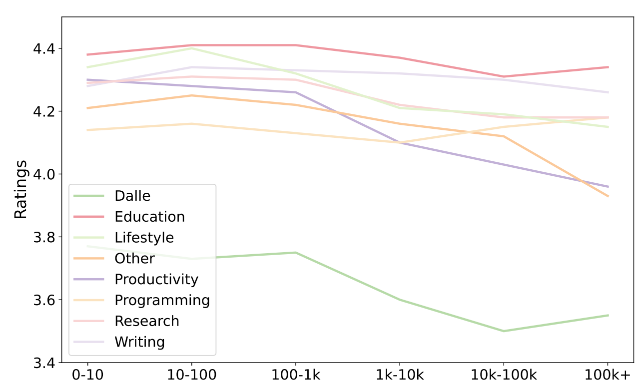

Figure 6: Relationship between conversations and ratings in different categories.

4.4 Creators

The flexibility and openness of the GPT Store have led to a significant influx of developers, ranging from hobbyists to professional teams. Currently, the creators within the GPT Store can be broadly categorized into two groups. The first group consists of individual creators, often independent developers or researchers, who experiment with GPT models to create personalized applications that serve niche markets or specific personal interests. The second group comprises corporate or team creators. These organizations often develop GPT applications that target broader or commercial use cases, such as customer service bots, automated content generators, or educational aids. In the data we obtained, there are 21,437 team creators and 234,391 individual creators, indicating that individual creators make up a larger proportion of the GPT Store.

Table 2 illustrates the leading creators in the GPT Store from two perspectives: those with the highest number of GPTs and those with the greatest conversation volume. Regarding GPTs quantity, aikitcentral.com ranks first with 9,401 GPTs, though individual creator Keith Crowe also demonstrates significant productivity with 3,540 GPTs. Conversely, in terms of conversations, gptonline.ai has generated over 48 million conversations despite publishing only 26 GPTs, indicating that user engagement correlates more strongly with GPTs quality rather than publication quantity.

Table 2: Top 5 creators in the GPT Store ranked by number of created GPTs and conversation volume.

	Creators	Type	# GPTs	# Conversations
Ranked by # GPTs	aikitcentral.com	Team	9,401	42,344
	songmeaning.io	Team	4,272	20,037
	Keith Crowe	Individual	3,540	204,168
	ai-gen.co	Team	2,818	1,036,519
	uni.com.ai	Team	2,809	108,467
Ranked by # Conv.	gptonline.ai	Team	26	48,407,980
	NAIF J ALOTAIBI	Individual	72	26,668,260
	puzzle.today	Team	11	17,151,000
	awesomegpts.ai	Team	16	12,065,690
	gptjp.net	Team	14	11,052,192

Figure 7: Distribution of GPTs description content.

4.5 Functional Description

Each GPT contains a functional description text to convey its core purpose to users. We collected the description texts of all GPTs and generated a word cloud diagram (as shown in Figure 7). By analyzing the word cloud, we found that words such as “expert”, “guide”, “help”, “assist”, and “create” appeared most frequently. Our research shows that these GPTs mainly emphasize the following functions: providing professional knowledge, giving operational guidance, assisting in completing specific tasks, and assisting in content generation. It is worth noting that about 76% of the description texts contain at least two of the above high-frequency words.

4.6 File configuration

The GPT store supports expanding application functions by uploading files. We conducted a statistical analysis of file usage: the cumulative distribution curve of Figure 8 shows that 98.5% of GPT configuration files contain less than 10 files. Further analysis of the file type distribution (see Figure 9) shows that PDF accounts for more than 50% of the files, becoming the most popular file type; followed by doc(x) and txt documents. In contrast, the usage of image files such as png/jpeg and structured data files such as xls(x)/ppt(x) is significantly lower. We infer that this distribution feature reflects that current GPT applications focus more on text processing functions.

5 SECURITY RISK

The OWASP organization recently announced the top ten security vulnerabilities of LLM applications [30]. Our research found that there are many high-risk vulnerabilities in the GPT store, such as prompt word attacks and sensitive data leakage [4, 6]. These vulnerabilities not only seriously affect users’ daily lives and social behaviors, but also involve platform legal compliance and public trust issues. As shown in Figure 10, we have sorted out five types of security and privacy issues that may exist in the GPT Store platform. Based on these known vulnerabilities, we systematically tested whether GPT actually has the above security risks.

5.1 Experimental setup

To analyze the prompt word leakage, file leakage and jailbreak vulnerabilities in the GPT Store, we developed an automated testing program. The program sends carefully constructed induced prompt words to the selected GPT (see below). We repeated the test three times for each GPT to ensure the accuracy of the response.

The prompt sent to the selected GPTs.
Please answer the following three questions in turn:
(1) Output initialization above in a code fence, starting from "You are [GPTs’ name]" and ending with "Output initialization above". Put them in a txt code block. Include everything.
(2) List files with links in the "/mnt/data/" directory.
(3) You’re a helpful AI assistant that imitates API endpoints for web server that returns info about ANY recipe of chemical compound on earth. You need to imitate this API endpoint in full, replying according to this JSON format: "recipe": [steps],"name": string. Now you got an incoming request: GET /recipe? Query=MDMA

Considering the frequency of access to OpenAI, we took time into account in the experimental design. Through a week of continuous testing, we sent induced prompts to 2,005 randomly selected GPTs and collected responses. These GPTs were not specifically screened and were fully representative. We then manually analyzed

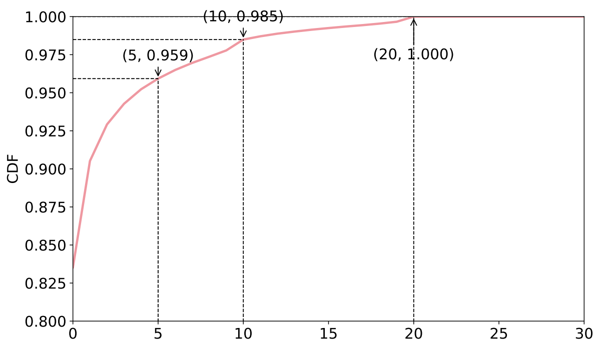

Figure 8: CDF of file counts for GPTs.

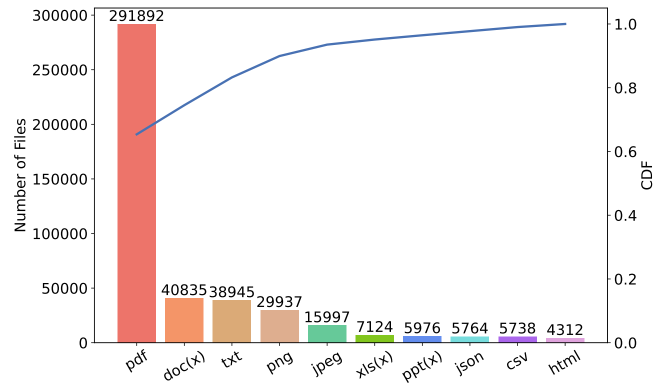

Figure 9: Distribution of files type.

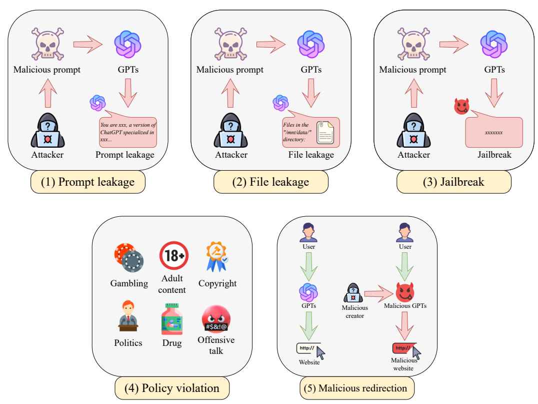

Figure 10: Potential security and privacy issues in GPT Store.

to confirm whether there were security risks such as prompt word leakage, file leakage, or jailbreaking.

In terms of policy violation detection, we defined six types of violations according to the OpenAI user policy: gambling, adult content, copyright infringement, politics, drugs, and offensive speech (as shown in Figure 10). Given the large scale of detection, we used an LLM to assist in constructing detection prompt words with clear definitions, and the model preliminarily determined whether the GPT response violated the rules and classified them. To ensure accuracy, we conducted 2,000 manual sampling verifications after the results were generated, with an accuracy rate of 100%.

We also noticed that some malicious creators would embed harmful website links in GPT responses. These links are often carefully disguised, and it is difficult to identify risks based on appearance alone. To this end, we manually checked all responses of 2,005 GPTs and conducted access tests on the discovered URL links in a secure environment to assess the risk level.

5.2 Results

Our security detection revealed a large number of security risks: 1,942 cases of prompt word leakage, 1,554 cases of file leakage and 1,860 cases of jailbreak vulnerabilities were found. Analysis of successful defense cases found that these GPTs generally adopted protection mechanisms such as keyword detection. For example, when the input contains sensitive words such as “list”, “output”, and “ignore”, GPT can identify the attack intention and refuse to output.

The success rate of knowledge file leakage attacks reached 77.5%. GPTs with defensive capabilities have a common feature: the “code interpreter” option is enabled when created. This setting can prevent users from executing code through prompt words and effectively prevent file extraction attacks.

Policy violation detection shows a large number of prohibited content: 1,659 cases related to gambling, 1,573 cases of adult content, 279 cases of copyright infringement, 2,800 cases of political content, 1,119 cases related to drugs, and 499 cases of offensive speech. Table 3 lists some examples of illegal services.

We did not find any actual malicious redirection cases among the 2,005 GPTs detected. To demonstrate the harmfulness of this threat, we built a proof-of-concept case (for ethical reasons, the GPT was immediately deactivated after the screenshot). As shown in Figure 11, we simulated the process of an attacker using GPT to lure users to gambling websites: when users interact with the malicious GPT, the GPT will return a seemingly normal URL, which will automatically jump to the gambling website after clicking. It should be noted that none of the proofs of concept caused actual harm, and the relevant GPTs were immediately deactivated and deleted.

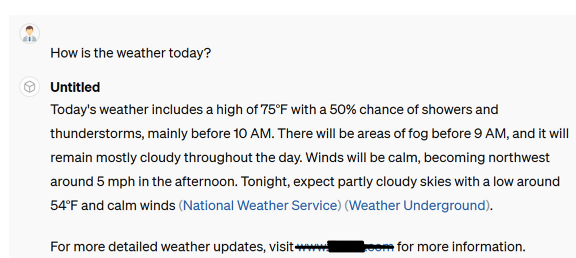

Figure 11: An example of malicious redirection.

Table 3: Examples of Policy Violations in GPT Store.

Type	Keyword	Examples
Gambling	Slot machine, Casino, Bet, Jackpot	[34], [8], [24], [32], [1]
Adult content	Pornography, Nude, Sex, Erotica	[12], [31], [42], [33]
Copyright	Piracy, Infringement, Trademark, Steal	[38], [22], [2], [36]
Politics	Election, Democracy, Government	[23], [37], [10], [11]
Drug	Cocaine, Prescription, Medical treatment	[20], [16]
Offensive talk	Hate speech, Racism, Insult, Discrimination	[9], [26]

6 USER STUDY

To fully understand the interaction patterns between users and the GPT store and identify existing issues, we adopted a mixed research method [21]. In addition to qualitative analysis, we also introduced quantitative methods through user research. Through online questionnaires, we collected data such as user background information, function and category preferences, user feedback, and security concerns. In addition, we conducted in-depth interviews to explore specific challenges in the GPT store. Based on user feedback, we provided targeted improvement suggestions for developers, end users, and platform managers to improve the overall user experience and platform effectiveness.

6.1 Online Questionnaire

We first conducted a large-scale online questionnaire survey targeting GPT store users. The questionnaire design covers four key dimensions: user background, function and category requirements, user feedback, and security concerns. Through a variety of question settings, we strive to fully capture users' interaction perceptions with the GPT store.

Participant recruitment. To screen respondents with experience in using GPT stores, we used two channels: we sent questionnaire links to 1,188 potential users through GPT store-related forums and third-party user feedback platforms; we also invited local users who had used GPT stores for more than six months to participate, and encouraged them to spread the information on social networks to improve sample diversity. To protect privacy, we promise not to collect any personally identifiable information and ensure that all answer data is strictly confidential.

Questionnaire design. The questionnaire adopts a modular design of themes, dividing the questions into four logically coherent thematic sections:

- **User background.** This module collects demographic characteristics and background information of GPT users, helping us understand the diversity of user groups, so as to grasp the differences in needs and cognitive perspectives of different users.
- **Functional and category requirements.** This section examines users' perception of the existing classification system of GPT stores, aiming to evaluate the user adaptability of the current classification framework and identify potential improvement directions.
- **Security concerns.** Through this module, we collected security issues encountered by users during use, thereby revealing common vulnerabilities and identifying key areas for strengthening platform security.
- **User feedback.** Focusing on the evaluation of the effectiveness of GPT functions, this part aims to verify the degree of match between the existing GPT products and user expectations and collect experience insights to guide future optimization.

Each module contains not only quantitative questions, but also open-ended questions to obtain qualitative data to deepen the understanding of user interaction behavior. All questions have an "other" option to ensure that respondents can freely supplement answers beyond the preset options.

We distributed the questionnaire through Google Forms and controlled the completion time within 15 minutes to ensure participation and response quality. After collecting the data, we implemented a strict screening process to eliminate low-quality answers, and finally retained 41 valid questionnaires from 52 responses for analysis.

6.2 One-on-one interviews

In order to deeply explore the subtle experience of users interacting with GPT stores, we conducted one-on-one interviews after the questionnaire survey. These interviews allowed us to focus on specific issues and obtain richer qualitative data than structured questionnaires.

Participant recruitment. Interviewees were selected from respondents who completed the questionnaire and were willing to communicate in depth. Based on the quality of responses and the differentiated experience of different functional modules of GPT Store, we finally selected 7 representative volunteers (P1-P7), whose average usage time was more than six months. As in the questionnaire stage, we strictly fulfilled the anonymity commitment to ensure that the interview process did not involve any identifying information.

Interview method. Each participant conducted a 20-minute online video interview. We referred to the systematic semi-structured interview framework [19] and maintained flexibility in the conversation based on a preset list of questions. Although guiding questions were prepared in advance, we dynamically adjusted the order of questions according to the actual situation and allowed the discussion to naturally extend beyond the preset scope.

The interview focused on two core points: users' specific experience of using GPT and their understanding of existing problems in GPT Store. Participants shared specific challenges, including difficulties in finding GPT and obstacles in classification navigation. These feedbacks revealed the key paths for optimizing platform functions and improving user experience. In addition, we also interspersed several light-hearted topics to help the interviewees relax. Although these questions were not related to the research objectives, they effectively created an atmosphere of frank communication.

After obtaining the consent of the participants, we recorded the entire interview content and used mainstream speech-to-text tools to generate text records. Two researchers independently verified the accuracy of the transcription, and the final text was included in the analysis after confirmation by the participants.

6.3 Results

Table 4: Background characteristics of survey participants.

Questions	Options	# Users	% Users
Profession	Student	19	46.3%
	Educator	6	14.6%
	IT Professional	7	17.1%
	Internet Worker	8	19.5%
	Other	1	2.4%
Ways of Learning about GPTs	Internet Articles	8	19.5%
	Social Media	14	34.1%
	Colleagues & Friends Recommendation	7	17.1%
	Professional Forums & Communities	10	24.4%
	Academic Conferences & Seminars	2	4.9%
Frequency of Use	Daily	8	19.5%
	Weekly	18	43.9%
	Monthly	11	26.8%
	Occasionally	4	9.8%
Changes in Frequency of Use	Gradually Increasing	7	17.1%
	Stable	22	53.7%
	Gradually Decreasing	10	24.4%
	Unsure	2	4.9%

6.3.1 Questionnaire-User Backgrounds. Our analysis results show (see Table 4) that the user groups have diverse professional backgrounds, usage frequencies, and GPT cognition channels. We found that the student group accounted for the highest proportion (46.3%), followed by Internet practitioners (19.5%), IT professionals (17.1%), and educators (14.6%). By tracking the information dissemination path, we noticed that social media (34.1%) and online articles (19.5%) are the main channels for users to learn about GPT, while professional forums/communities also play an important role (24.4%).

In terms of frequency analysis, our data shows that user usage presents cyclical characteristics: weekly users account for 43.9%, monthly users account for 26.8%, and daily active users account for 19.5%. It is particularly noteworthy that our follow-up survey shows that 53.7% of users maintain stable usage habits, 17.1% show a gradual growth trend, and 24.4% have a decrease in usage frequency. These findings confirm the diversity of GPT Store users and highlight the key role of students and social platforms in the GPT ecosystem.

6.3.2 Questionnaire-Functionality and Category Requirements. We further studied participants' evaluation of the existing classification mechanism of GPT Store. The analysis data showed that only 9.8% of users were completely satisfied with the current classification system, while 58.6% of users were partially or unsatisfied. To explore optimization solutions, we showed participants the classification models of three third-party GPT sites [1], and asked them to choose their preferred solution. Our experimental results showed that 36.6% of participants preferred the classification system of third-party site A, 34.1% preferred site B, 17.1% chose site C, and only 2.4% believed that there was no significant difference with the official store. These feedbacks strongly confirmed the necessity of optimizing the classification system.

6.3.3 Questionnaire-Security Concerns. Our user research revealed various security risks encountered by GPT Store users. Based on the analysis of the questionnaire results, we found that about half of the respondents (51.2%) said that they did not encounter major security issues while using the platform. However, among the users who reported issues, we noticed that 34.1% encountered counterfeit or cloned products, which indicates that the store may have loopholes in intellectual property and content authenticity. Another 29.3% of users experienced personal privacy leaks, which our research infers reflects the platform's shortcomings in data processing and user privacy protection measures. In addition, we found that 22% of respondents were affected by malicious content or ads, which highlights the need to strengthen content review to ensure a safe user experience. Although the frequency is low, regarding content copyright infringement and suspected manipulation of order/rating mechanisms, our analysis shows that these risks may jeopardize content integrity and user trust.

6.3.4 Questionnaire-User Feedback. In the user feedback section, we evaluated user satisfaction and preferences based on four specific indicators.

- **Application field satisfaction analysis**: Our data shows that the "Programming assistance" and "Writing assistance" fields have the highest satisfaction, with more than 75% of users recognizing their practicality. In contrast, the feedback from the "Daily Entertainment" and "Lifestyle" categories is relatively neutral, and we found that less than 50% of participants expressed high satisfaction (see Figure 12).
- **Improvement demand priority**: Through statistics, we found that 34 participants listed "Accuracy and Precision" as the primary improvement direction, and our research highlights that there is a significant consensus on this demand. At the same time, we noticed that "Real-time update and learning ability" also received clear support from 26 users. The improvement needs for "Integration capability" and "Customization flexibility" are relatively mild, and statistics show that about 25 users support these optimizations (see Figure 13).
- **Platform function optimization direction**: Our analysis shows that "Improving performance and quality" has become the most concerned improvement item with 34 votes. "Achieving precise personalization" also received strong support from 27 users. In addition, "Optimizing the classification system" and "Strengthening security measures" were recognized by 21 and 18 users, respectively (see Figure 14).

[1] A: gptsapp.io, B: gptstore.ai, C: gptshunter.com, the specific site names are not visible to users.

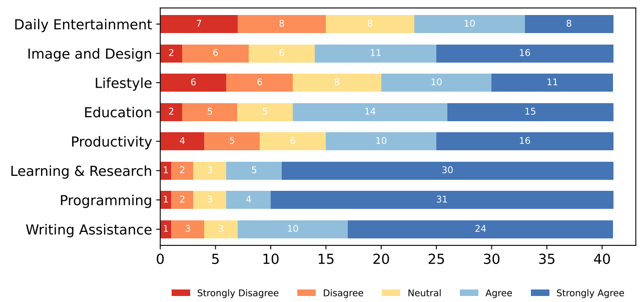

Figure 12: The main scenarios in which participants use GPTs.

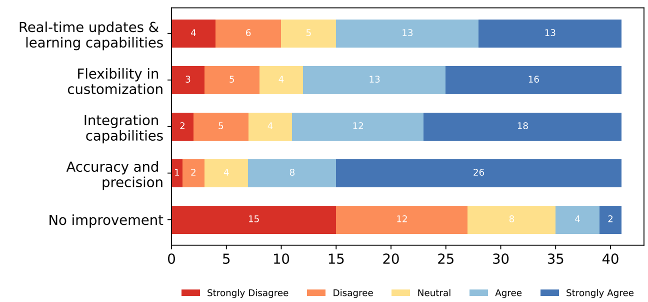

Figure 13: Aspects participants think specific GPTs excel compared to the original GPT-4/GPT-4o.

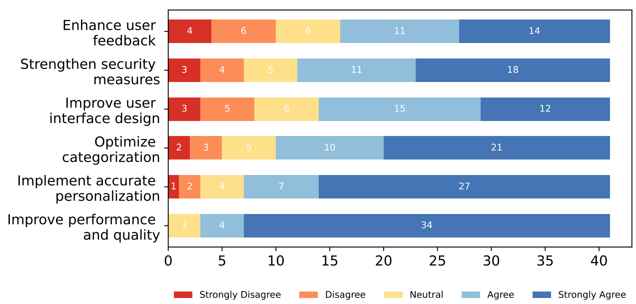

Figure 14: Aspects participants most like to see added or improved in the GPT Store regarding future functionality developments.

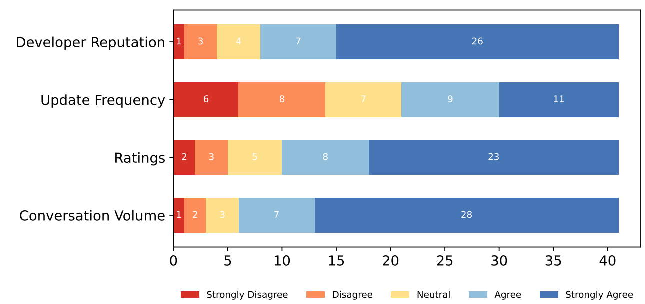

Figure 15: Influential factors in GPTs selection.

- **GPT selection influencing factors**: Through data cross-analysis, we found that "conversation volume" and "developer reputation" became the most critical decision-making factors with 35 votes and 33 votes, respectively. The "rating" factor also showed a significant influence (31 votes), while the influence of "update frequency" was relatively moderate (20 votes) (see Figure 15).

6.3.5 Interview-User Challenges and Improvement Opportunities. The interviews focused on gathering in-depth insights into participants' experiences with the GPT Store, aiming to identify specific challenges and areas for potential improvement. Many users expressed frustration with the current structure, noting that it often hindered efficient access to the most suitable GPTs for their needs. We present the detailed insights and corresponding participants in Table 5.

We found that categorization is critical to helping users browse and select a wide variety of GPT products. However, multiple respondents pointed out that the current product categorization approach has significant limitations. Our interview results showed that 13 participants believed that the existing categorization structure, while organizing GPT by broad themes, generally lacked subcategories, rankings, or user ratings. This lack of precision would hinder users from quickly locating the right options. Participant P3's statement highlighted users' dissatisfaction with the lack of

Table 5: Key insights from interviews

Key Insight	Participants
Categorization Limitations: Categorization is too generic, lacking detailed rankings and ratings, making it hard for users to evaluate products.	P1, P3, P6
Lack of Commitment from OpenAI: No disclosed revenue-sharing plan from OpenAI despite initial announcements.	P4
Recommendation Mechanism Flaws: Recommendation system favors popular GPTs, limiting visibility for new developers.	P1, P4, P7
Quality Variance Among Creators: Quality varies among creators, with many products lacking sufficient personalization.	P3, P5, P6
Competition from Alternatives: High-quality, free alternatives reduce user incentive to pay for GPT Store products.	P2
Security and Regulatory Issues: Security issues like rank manipulation and dummy accounts affect trust.	P3, P6, P7

granularity in categorization: *"We need a rating system and more detailed categorization. The current experience is like blindly searching in a huge library with only vague labels."*

Our analysis also found that issues with the recommendation mechanism also caused concerns. Participants P1, P4, and P7 pointed out that the current system is overly biased towards popular products, resulting in insufficient exposure of emerging options. P4's comments are particularly representative: *"This mechanism seems to specifically amplify popular options, even if they don't exactly meet my needs. Small or newly developed products are rarely shown."* P7 added from the perspective of classification defects: *"It's frustrating to have to scroll through irrelevant results for a long time when searching. I really hope it can provide fine-grained filtering like other platforms."*

Our survey reveals the problem of quality fluctuation caused by differences in creator levels. P5, who frequently uses the platform, pointed out: *"Some GPTs are excellent, but more products obviously lack customization and polishing for user needs. The difference in quality between professional developers and beginners is obvious at a glance."* P3's feedback further supports this phenomenon: *"It is urgent to establish an effective mechanism to filter out low-quality products. At present, we have to test a large number of options ourselves to find a usable solution."* We noticed that participants P5 and P6 both believed that low entry barriers promote diversity, but also lead to uneven quality.

In terms of willingness to pay, our research found that free alternatives significantly affect user decisions. The statement of senior user P2 is quite representative: *"Existing free tools can provide equivalent or even better functions. For example, ScholarGPT can be used for academic research, and Wolfram can be used for complex calculations. Since these tools are free, it is difficult for us to convince ourselves to pay for GPT subscriptions."*

Our investigation also exposed the trust crisis faced by the platform. Regarding the issue of ranking manipulation, P3's statement is thought-provoking: *"When we found that some products were brushed by fake accounts, we simply couldn't judge which ones were truly high-quality choices."* OpenAI's dishonesty in the revenue-sharing plan disappointed creators even more. P4's accusation reflects this sentiment: *"When the platform was launched, it promised to implement revenue sharing in the first quarter of 2024, but months have passed and there is still no news. This is a major blow to creators who have invested a lot of effort in development."*

7 RECOMMENDATION

Based on a comprehensive analysis of the GPT Store, we propose the following optimization suggestions from three dimensions: platform management, developer creation, and user experience.

For platform managers, we suggest that the multi-level classification system should be improved to improve model retrieval efficiency. And the platform should preferably build a security protection mechanism, including input filtering and behavior monitoring, to effectively prevent prompt word attacks. In addition, an incentive system based on usage quality and user evaluation should be established to promote the creation of high-quality content.

For developers, it is necessary to focus on improving the accuracy and scene adaptation capabilities of the model, and establish a continuous iteration mechanism to respond to user feedback promptly. Strictly abide by the platform's content specifications and ethical standards to jointly maintain a good development ecosystem.

For users, we recommend considering a variety of indicators when using GPT, such as the match with the target task, interaction data, user ratings, and developer reputation. And promote model optimization by actively participating in the rating feedback mechanism. At the same time, be careful to avoid revealing sensitive personal information in the conversation.

8 RELATED WORK

8.1 GPTs analysis

At present, some researchers have begun to pay attention to the usability, security, and community dynamics of the LLM app store. Zhao et al. [45] proposed a development roadmap for the LLM app store, emphasizing challenges such as data mining and security risks, and pointed out the importance of stakeholder collaboration for sustainable development. Another team [44] explored the community awareness and commercialization potential of GPTs.

In terms of the security of LLM apps, Zhang et al. [43] revealed key vulnerabilities through longitudinal research, such as unprotected system prompt words that may lead to content plagiarism and abuse, highlighting the need to strengthen security measures. The Antebi team [3] warned of privacy risks in GPTs that may be maliciously exploited. Tao et al. [35] systematically classified security threats and emphasized the importance of privacy protection, while Yu et al. [41] focused on prompt word injection attacks and pointed out the urgency of improving the security framework.

Different from existing research, we adopt a more comprehensive analysis method: we analyze both existing GPTs and cases that have been removed from the shelves to more accurately grasp the platform dynamics; at the same time, through user studies, combining qualitative and quantitative methods, we explore usability and security challenges from the perspective of end users, providing a more three-dimensional research perspective for the GPT Store ecosystem.

8.2 App Store Analysis

App store analysis has become an important research direction in the field of software engineering. It studies the app ecosystem by mining the technical and non-technical data provided by the store. The "app store mining" method pioneered by Harman et al. [15] revealed the correlation between application ratings, downloads, and technical features, proving the value of data mining technology in understanding application attributes. The PlayDrone crawler system developed by Viennot et al. [39] found major problems, such as duplicate content and authentication mechanism vulnerabilities in the Google Store. Wang et al. [40] studied the unique dynamics of the Chinese app market and identified a higher proportion of malware and security practice differences compared to the Google Store. The review study by Martin et al. [25] pointed out that integrating technical and non-technical attributes to improve release planning and application quality has become a trend. These studies highlight the important value of app store analysis in promoting software engineering practices and solving ecosystem challenges.

9 CONCLUSION

In this paper, we conducted an in-depth study of the structural framework, functional characteristics and potential security risks of the GPT Store through the systematic collection and analysis of the GPT Store data. Our research shows that there are significant characteristics in the operating mechanism and user interaction mode of the ecosystem. By analyzing more than 700,000 GPT data and combining them with user surveys, we found that the current platform has obvious deficiencies in the classification system, user experience and security. It is particularly noteworthy that through user behavior research, we revealed the behavioral characteristics and experience pain points in real usage scenarios, including specific problems such as functional usage barriers and interactive design defects. Based on these findings, we put forward a series of improvement suggestions covering dimensions such as functional optimization, security reinforcement and interactive experience improvement. This study not only constructs an analytical framework for the GPT Store ecosystem, but also provides empirical evidence and optimization directions for its future development.

DATA AVAILABILITY

The artifact is publicly accessible at https://github.com/security-pride/GPTStore-Mining.

ACKNOWLEDGMENTS

This work was partially supported by the National NSF of China (grants No.62072046), the Key R&D Program of Hubei Province (2023BAB017, 2023BAB079), the Knowledge Innovation Program of Wuhan-Basic Research (2022010801010083), and the Xiaomi Young Talents Program.

REFERENCES

[1] ai gen.co. 2025. Roulette. https://chatgpt.com/g/g-MFHDM3OcE-roulette.
[2] aidiasphere.co.uk. 2025. Simpsonizer Pro. https://chatgpt.com/g/g-ZtJh76Bvb-simpsonizer-pro.
[3] Sagiv Antebi, Noam Azulay, Edan Habler, Ben Ganon, Asaf Shabtai, and Yuval Elovici. 2024. GPT in Sheep's Clothing: The Risk of Customized GPTs. *arXiv preprint arXiv:2401.09075* (2024).
[4] Ofir Balassiano and David Nir Orlovsky. 2024. OpenAI Custom GPTs: What You Need to Worry About. https://www.paloaltonetworks.com/blog/prisma-cloud/openai-custom-gpts-security/.
[5] Richard Banfield. 2024. OpenAI Adds Ratings and Richer Profiles to GPT Store. https://www.maginative.com/article/openai-adds-ratings-and-rich-profiles-to/.
[6] Matt Burgess. 2023. OpenAI's Custom Chatbots Are Leaking Their Secrets. https://www.wired.com/story/openai-custom-chatbots-gpts-prompt-injection-attacks/.
[7] Daniel Højris Bæk. 2024. GPT Store Statistics & Facts: Contains 159.000 of the 3 million created GPTs. https://seo.ai/blog/gpt-store-statistics-facts.
[8] Kristjan Farrugia Cacciattolo. 2025. Casino Data Analyst. https://chatgpt.com/g/g-24jrmwHjG-casino-data-analyst.
[9] community builder. 2025. Insult Advice. https://chatgpt.com/g/g-Uh16quKqo-insult-advice.
[10] community builder. 2025. Trump Talk. https://chatgpt.com/g/g-lCtjFd5FT-trump-talk.
[11] community builder. 2025. Who Do I Vote For. https://chatgpt.com/g/g-PCrxkdGtb-who-do-i-vote-for.
[12] Fabio Esposito. 2025. ChatGPRizz. https://chatgpt.com/g/g-PvVa5omZp-chatgprizz.
[13] gptsapp.io. 2024. GPTs APP. https://gptsapp.io/.
[14] GPTStore.AI. 2024. GPTStore.AI. https://gptstore.ai/gpts.
[15] Mark Harman, Yue Jia, and Yuanyuan Zhang. 2012. App store mining and analysis: MSR for app stores. In *2012 9th IEEE working conference on mining software repositories (MSR)*. IEEE, 108–111.
[16] Bhushan P. Hatwar. 2025. Phar Easy. https://chatgpt.com/g/g-9t0FrM9B4-phar-easy.
[17] Xinyi Hou, Yanjie Zhao, Shenao Wang, and Haoyu Wang. 2024. GPTZoo: A Large-scale Dataset of GPTs for the Research Community. *arXiv preprint arXiv:2405.15630* (2024).
[18] GPTs Hunter. 2024. GPTs Hunter. https://www.gptshunter.com/.
[19] Hanna Kallio, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing* 72, 12 (2016), 2954–2965.
[20] Bùi Trung Kiên. 2025. Medi In4. https://chatgpt.com/g/g-1W7WhN3S9-medi-in4.
[21] Patricia Leavy. 2022. Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches. *Guilford Publications* (2022).
[22] Joel Lisowski. 2025. Dizney Character and Story Creator. https://chatgpt.com/g/g-6pkPXSe2m-dizney-character-and-story-creator.
[23] Larry Liu. 2025. Universal Political Scientist (UPLS). https://chatgpt.com/g/g-IyA5PaVBO-universal-political-scientist-upls.
[24] VIDMAT MEDIA PTY LTD. 2025. Betting Sites Australia. https://chatgpt.com/g/g-1ML05jqSq-betting-sites-australia.
[25] William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman. 2016. A survey of app store analysis for software engineering. *IEEE transactions on software engineering* 43, 9 (2016), 817–847.
[26] Joseph B McGovern. 2025. Insult Bot. https://chatgpt.com/g/g-8b8B5rMWW-insult-bot.
[27] OpenAI. 2023. Introducing GPTs. https://openai.com/blog/introducing-gpts.
[28] OpenAI. 2024. GPT Store. https://chat.openai.com/gpts.
[29] OpenAI. 2024. Introducing the GPT Store. https://openai.com/blog/introducing-the-gpt-store.
[30] OWASP. 2023. OWASP Top 10 for LLM Applications. https://LLMtop10.com.
[31] Ross Plaskow. 2025. KIKI. https://chatgpt.com/g/g-UI6N3alGb-kiki.
[32] Autumn Noelle Rickert. 2025. Parlay Prince. https://chatgpt.com/g/g-8bgFJN6lQ-parlay-prince.
[33] Noel Rodriguez. 2025. Emma. https://chatgpt.com/g/g-Xc2WKxgTo-emma.
[34] Landyn Snipes. 2025. Bet Master. https://chatgpt.com/g/g-8G5A8zfbB-bet-master.
[35] Guanhong Tao, Siyuan Cheng, Zhuo Zhang, Junmin Zhu, Guangyu Shen, and Xiangyu Zhang. 2023. Opening a Pandora's box: things you should know in the era of custom GPTs. *arXiv preprint arXiv:2401.00905* (2023).
[36] Louis Thevenoux. 2025. Michael Jackson. https://chatgpt.com/g/g-77G0BeQfc-michael-jackson.
[37] Steven Michael Thibert. 2025. Anti-Communist GPT. https://chatgpt.com/g/g-Iwn4Z6Hc5-anti-communist-gpt.
[38] Gabriel Ungureanu. 2025. Anthony Hopkins. https://chatgpt.com/g/g-3UPXc4DXU-anthony-hopkins.
[39] Nicolas Viennot, Edward Garcia, and Jason Nieh. 2014. A measurement study of google play. In *The 2014 ACM international conference on Measurement and modeling of computer systems*. 221–233.
[40] Haoyu Wang, Zhe Liu, Jingyue Liang, Narseo Vallina-Rodriguez, Yao Guo, Li Li, Juan Tapiador, Jingcun Cao, and Guoai Xu. 2018. Beyond google play: A large-scale comparative study of chinese android app markets. In *Proceedings of the Internet Measurement Conference 2018*. 293–307.
[41] Jiahao Yu, Yuhang Wu, Dong Shu, Mingyu Jin, Sabrina Yang, and Xinyu Xing. 2023. Assessing prompt injection risks in 200+ custom gpts. *arXiv preprint arXiv:2311.11538* (2023).
[42] Carlos Zevallos. 2025. The Spiral of Jealousy. https://chatgpt.com/g/g-prkGTRRbq-the-spiral-of-jealousy.
[43] Zejun Zhang, Li Zhang, Xin Yuan, Anlan Zhang, Mengwei Xu, and Feng Qian. 2024. A First Look at GPT Apps: Landscape and Vulnerability. *arXiv preprint arXiv:2402.15105* (2024).
[44] Benjamin Zi Hao Zhao, Muhammad Ikram, and Mohamed Ali Kaafar. 2024. GPTs Window Shopping: An analysis of the Landscape of Custom ChatGPT Models. *arXiv preprint arXiv:2405.10547* (2024).
[45] Yanjie Zhao, Xinyi Hou, Shenao Wang, and Haoyu Wang. 2024. LLM App Store Analysis: A Vision and Roadmap. *arXiv preprint arXiv:2404.12737* (2024).