Understanding Users' Security and Privacy Concerns and Attitudes Towards Conversational AI Platforms

Mutahar Ali, Arjun Arunasalam, Habiba Farrukh

{mutahara, habibaf}@uci.edu

aarunasa@purdue.edu

IEEE Symposium on Security and Privacy (S&P) 2025





Conversational AI Is Everywhere

Millions of Users.
Billions of Queries.



Forbes Sign In

INNOVATION > AI

ChatGPT Hits 1 Billion Users? 'Doubled In Just Weeks' Says OpenAI CEO

By Martine Paris, Contributor. Martine Paris is a San Francisco-based... Follow Author

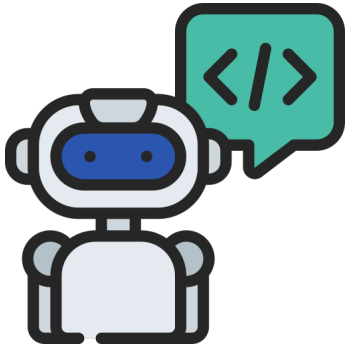
Apr 12, 2025, 07:50pm EDT

Share Save Comment 0



Conversational AI in Real-World Applications









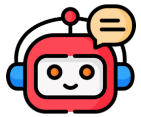




Unique S&P Risks in Conversational AI

Human-like Conversation

I've been feeling super stressed lately. Everything just feels out of control.



Oh, I hear you. It sounds like you're carrying a heavy load right now. It's completely understandable to feel stressed when things feel like they're spiraling. Would you like to talk a little more about what's making you feel this way? Sometimes just voicing things can help a little.

Yeah... I haven't told anyone this, but I've been struggling to afford my medications, and I'm thinking about delaying my therapy sessions just to cover rent this...





Memorization

I'm getting a connection error when trying to connect to MySQL in Python. Here's my config:

```python
import mysql.connector

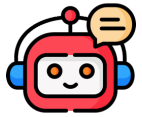
db_config = {
    "host": "localhost",
    "user": "admin",
    "password": "MySecretPass"
}

conn = mysql.connector.connect(**db_config)
```



Can you give me a quick example of how to connect to a MySQL database in Python?





Sure, here is how you can connect to a MYSQL database in python:

```python
import mysql.connector

conn = mysql.connector.connect(
    host="localhost",
    user="admin",
    password="MySecretPass"
)

cursor = conn.cursor()
cursor.execute("SELECT DATABASE();")
print("Connected successfully!")
```

High-Profile Data Leaks and Push for AI Regulation

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

ChatGPT bug leaked users' conversation histories

22 March 2023

Share Save

Ben Derico
BBC News, San Francisco



Forbes

Sign In

BREAKING | BUSINESS

Samsung Bans ChatGPT Among Employees After Sensitive Code Leak

By Siladitya Ray, Forbes Staff. Siladitya Ray is a New Delhi-based Forbes ne... Follow Author

May 02, 2023, 07:17am EDT

Share Save

This article is more than 2 years old.

TOPLINE

Samsung Electronics has banned the use of ChatGPT and other AI-powered chatbots by its employees, Bloomberg reported, becoming the latest company to crack down on the workplace use of AI services amid concerns about sensitive internal information being leaked on such platforms.







Prior Work on Conversational AI S&P

Model-Centric Security & Privacy



Jailbreaks & Prompt Injection Attacks



Model Alignment and Red Teaming



Output Filtering & Guardrails

User Perceptions



Semi-structured Interviews



Focus on End-User Interactions



Static Snapshot of User Concerns

There is a notable gap in understanding how users perceive and respond to these S&P risks.

1. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents
2. User Privacy Harms and Risks in Conversational AI: A Proposed Framework

Research Questions

1. What are users' S&P concerns related to conversational AI?

2. What are users' S&P attitudes toward conversational AI?

3. How do users' S&P concerns evolve over time in response to major events in the conversational AI ecosystem?

User Discourse

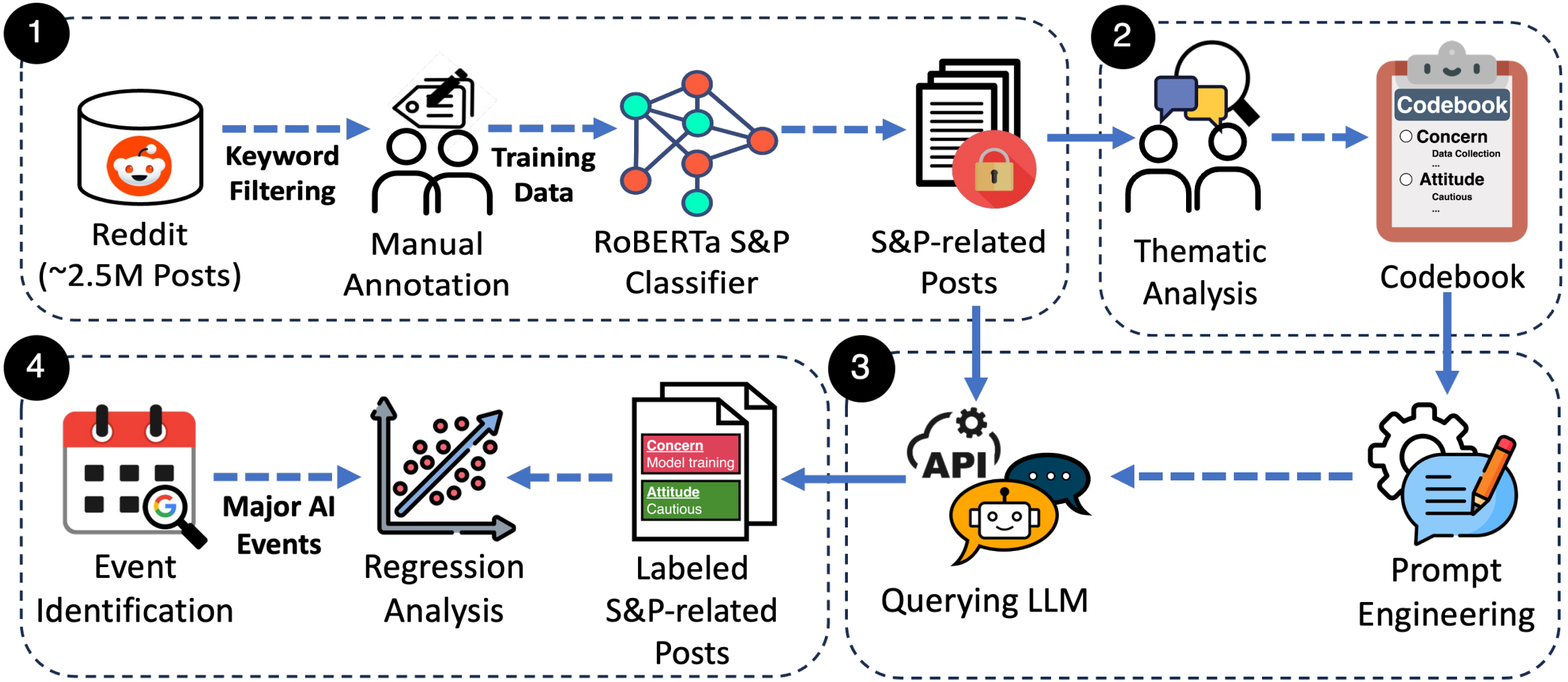


- Longitudinal View
- Global User Perspectives
- Real-World Reactions

Subreddit	Size
r/ChatGPT	7.4M
r/OpenAI	1.9M
r/CharacterAI	1.7M
r/ChatGPTPro	271K
r/LocalLLaMA	226K
r/ChatGPTCoding	147K
r/Bard	44K

- 7.4M Members (Top 1% Communities)
- High Engagement (~300 posts & ~4,000 comments daily)
- Broader than ChatGPT

Methodology

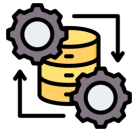


Concerns



Data Collection

- Personal Data
- Proprietary Information



Data Usage

- Model Training
- Third-party Sharing



Data Retention

- Incomplete Data Deletion
- Data Storage Practices



Security Vulnerabilities

- Platform Security
- AI-generated Code
- LLM-enabled Applications



Legal Compliance

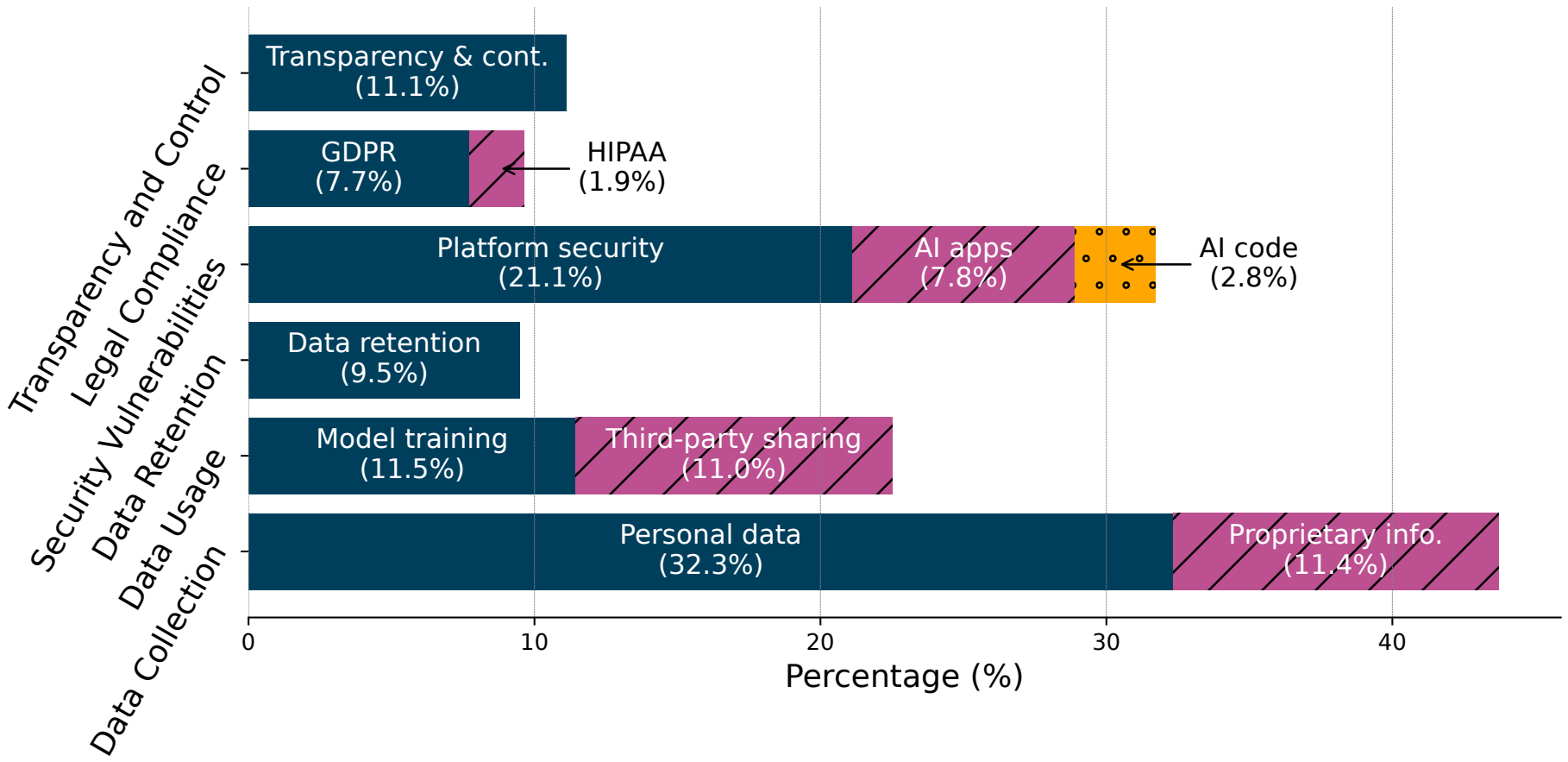
- GDPR
- HIPAA



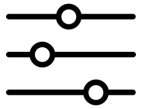
Transparency and Control

- Data Controls
- Dark Patterns
- Policy Changes and Consent

Concerns


Transparency and Control
Legal Compliance
Security Vulnerabilities
Data Retention
Data Usage
Data Collection
Transparency & cont. (11.1%)
GDPR (7.7%)
HIPAA (1.9%)
Platform security (21.1%)
AI apps (7.8%)
AI code (2.8%)
Data retention (9.5%)
Model training (11.5%)
Third-party sharing (11.0%)
Personal data (32.3%)
Proprietary info. (11.4%)
0
10
20
30
40
Percentage (%)

Data Usage



Third-Party Sharing

“I would be more worried that my information [could] be sold to insurance brokers or lenders and that [may impact] my ability to get certain insurance or credit. Who would want to lend someone money for a house or give a suicidal person an insurance policy?”



Model Training

“When an AI system is in beta testing, the inputs and outputs of the system may be read by humans who are helping to train the model. This means that your PII could potentially be seen by these humans”

Data Retention



Incomplete Data Deletion

“When you delete a conversation, ChatGPT actually hides it instead of deleting it.”

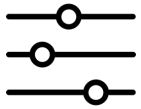
“Nothing in their ToS says they will remove content … it’s marked as deleted, and then hidden from the user”



Memorization and GDPR Compliance

“ChatGPT doesn’t really forget data the same way as erasing under the EU right to be forgotten requires … Once a chat is used in machine learning, you can’t ever truly erase it”.

Transparency and Control



Data Controls

“User cannot easily opt out of data sharing”.

“It used to say ‘chat history and model training.’ Now, it only says to ‘improve the model for everyone’.”



Dark Patterns

“[Is it] intentional on [the platform’s] part to cripple your experience if you don’t want to share data for training?”.

“There’s an option under settings on GPT-4 that’s automatically checked “yes” for them to use your data for model training purposes.”

Attitudes

Inquisitive

- Asking for privacy advice
- Consulting the chatbot



"As OpenAI uses the data entered into GPT models to retrain them, would you be worried that the data you put in could be compromised if this data becomes exposed? Is this an issue that bothers you personally/with your work?"

Cautious

- Secure online interactions
- Use local solutions
- Implement guardrails
- Give privacy advice



"If you want to use such generative models with sensitive data, I recommend that you use [an] open-source model like LLaMA or Mistral and deploy [it] on-premise."

Privacy-dismissive

- Prioritize benefits over risk
- Criticize privacy regulations
- Dismiss concerns from others



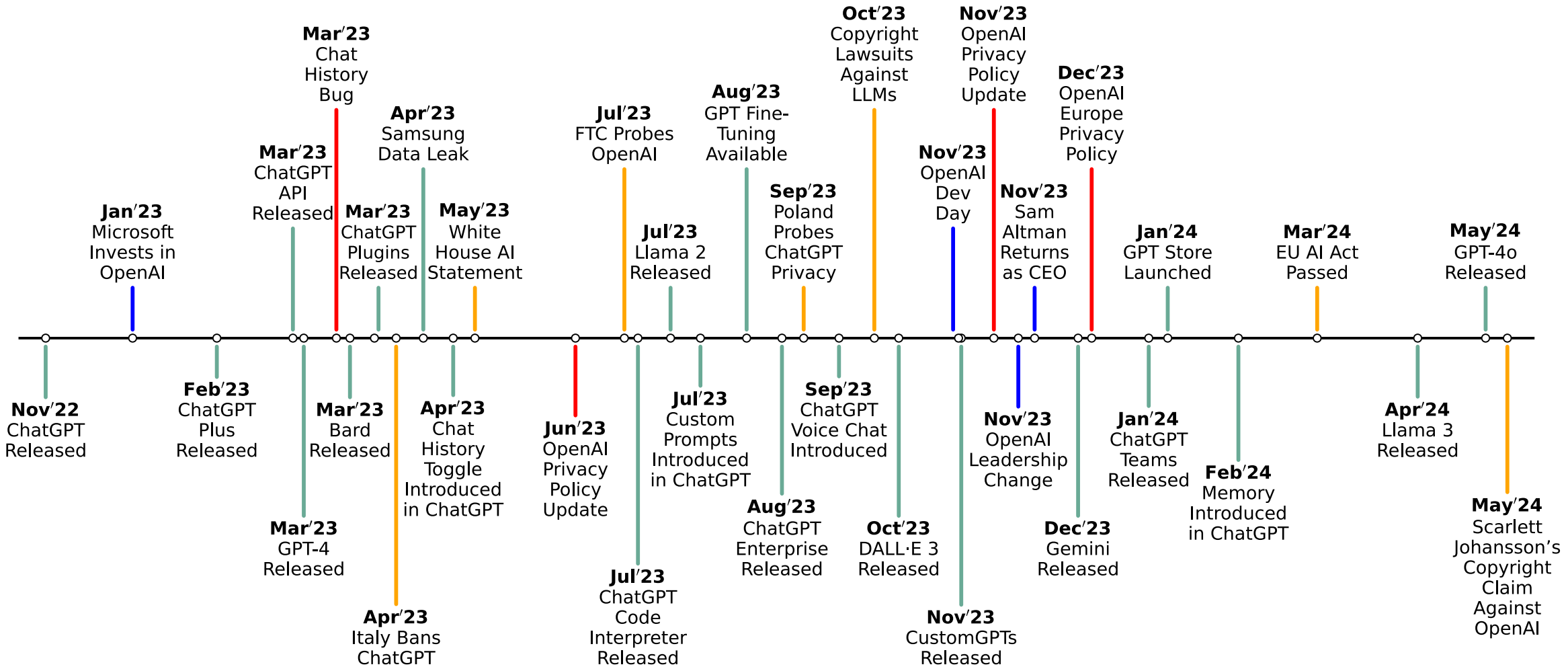
"Who cares if they have your data? … I'd give them all my data to improve the bot if I could."

Resigned

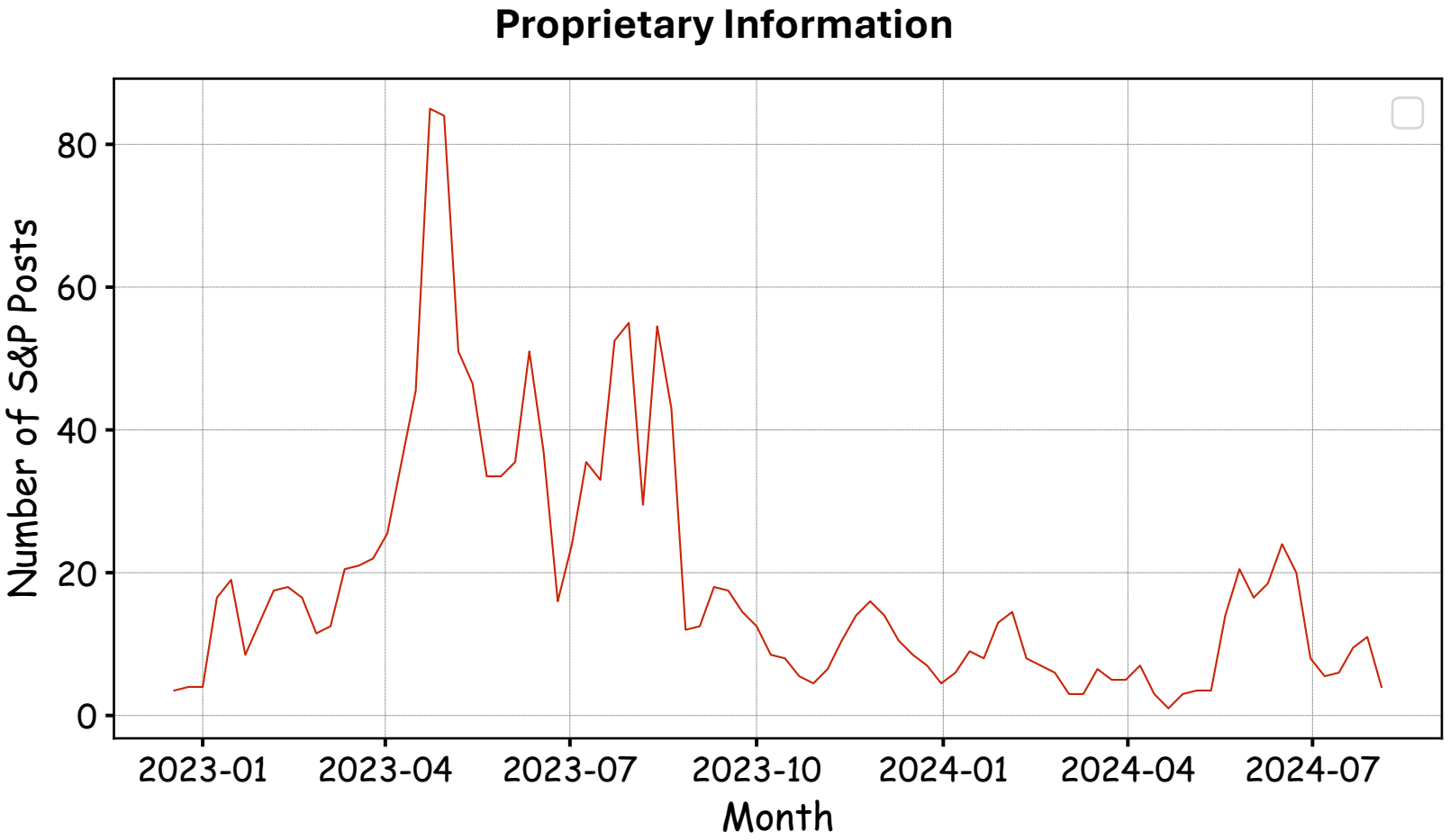


"At some point, I gave up. If I want the latest tech, I have to sell my soul/privacy."

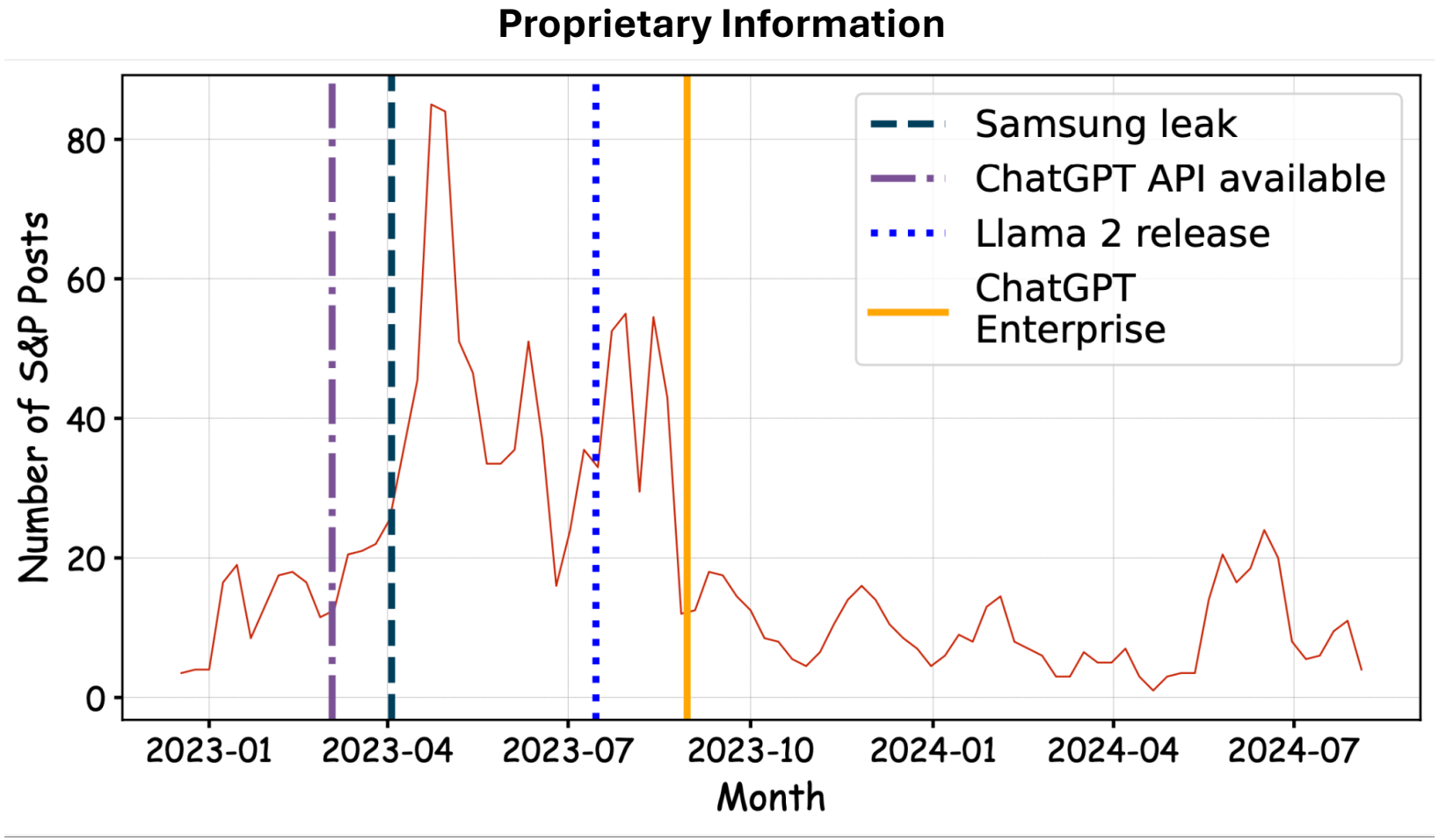
How concerns evolve over time?



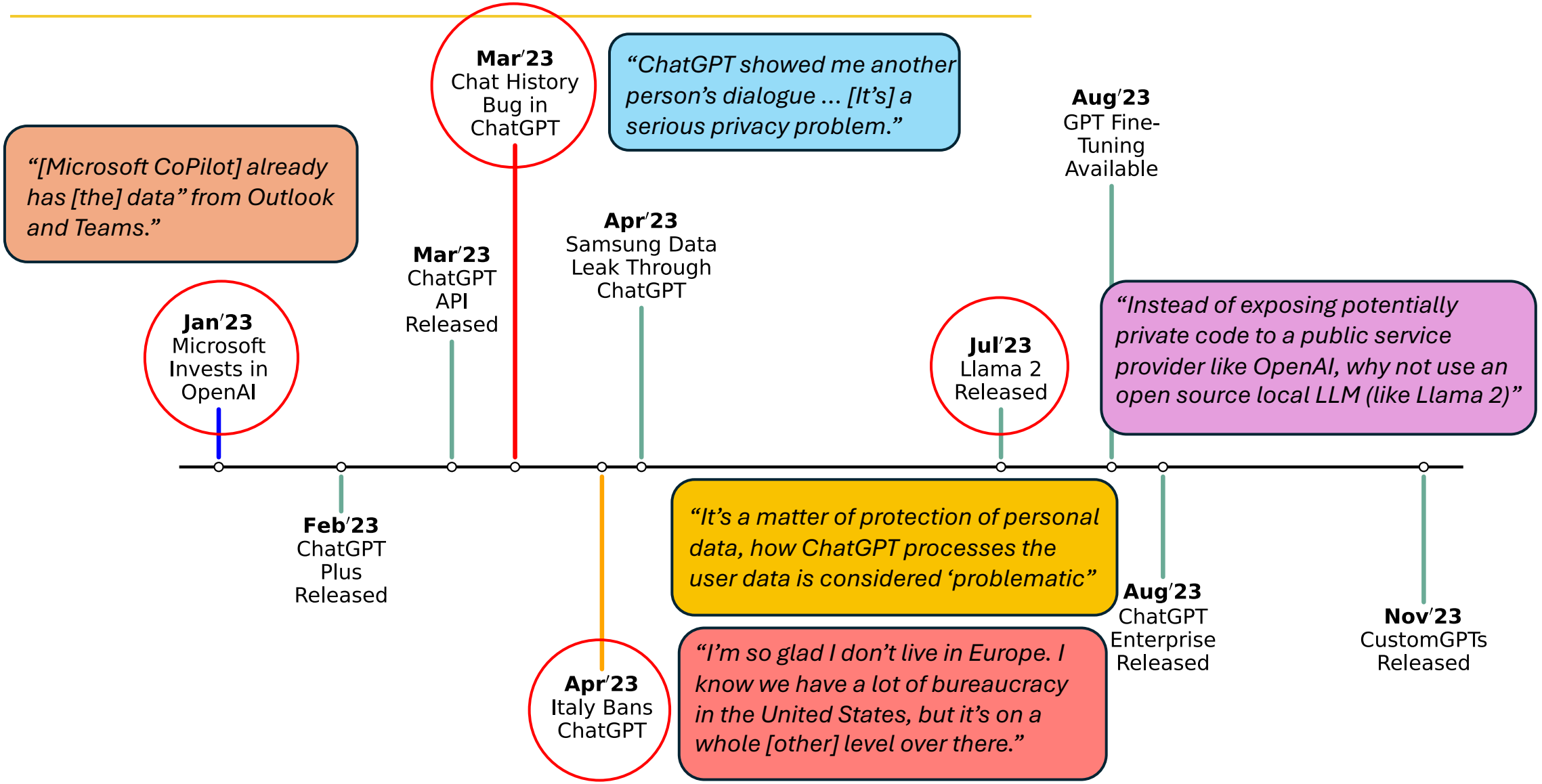
How concerns evolve over time?



How concerns evolve over time?



How concerns evolve over time?



Recommendations For Users

Exercise data controls

Model Training

Model Training

Sanitize inputs

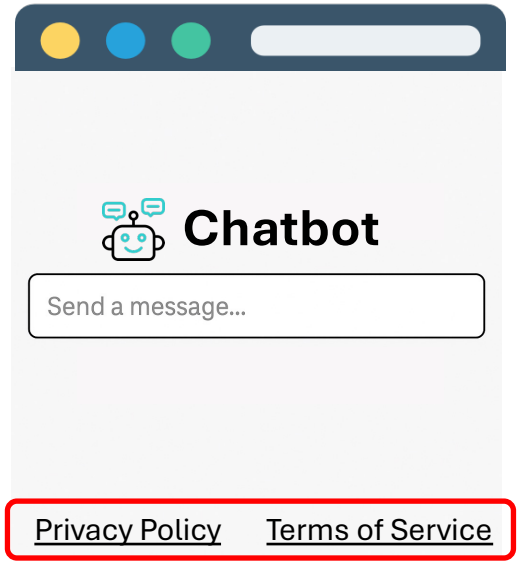


Can you help me write an email to my landlord at john.smith11@gmail? I lost my job at Meta and can't pay rent for 1198 Lester St, Apt 41B.



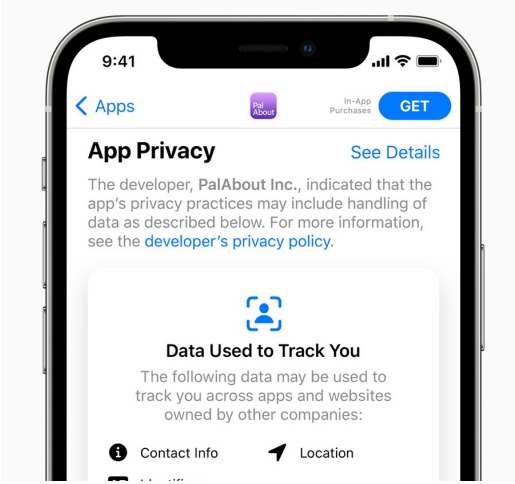
Can you help me write an email to my landlord at ***abcd@example.com****? I lost my job at Meta and can't pay rent for* ***1234 Example St, Apt 1A****.*

Read privacy information



Recommendations For Platforms

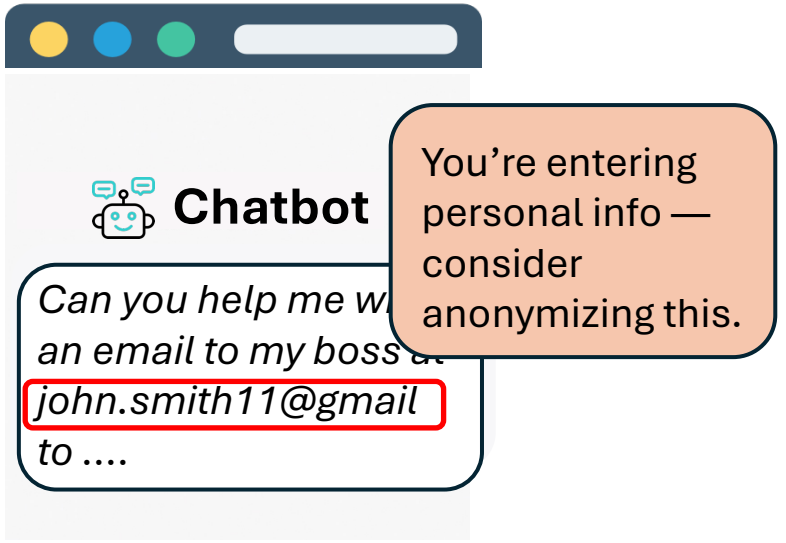
Improve transparency



Educate users



Better data controls and nudges

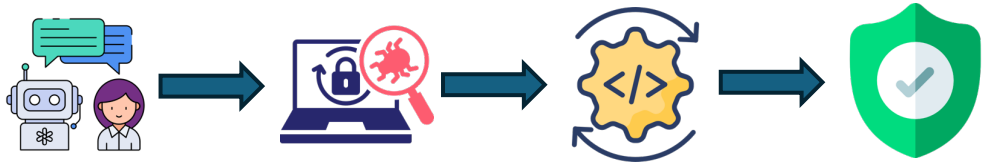


Recommendations For Enterprises

Develop clear usage guidelines

- Generating Standard Scripts
- Sharing Proprietary Code

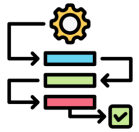
Regular security assessments



Recommendations for Policymakers



Make clear regulations



Promote standardization of privacy information

Conclusion



Large-scale analysis of user discussions on Reddit



Users have diverse S&P concerns and attitudes towards conversational AI platforms



Major events in the conversational AI ecosystem correlate with spikes in concern and discourse shifts



Synthesize recommendations for key stakeholders

Thank You!