AI Explained · Confidentiality and Privacy, Part 1

Who Can See What You Share?

By Dr Peter Fields

THE 2-MINUTE SUMMARY

The app may keep your questions, files and replies so you can return to them. What it saves and what the model provider keeps are separate questions.[1,2]

Who can see the material depends on your account, the provider's rules and what you share. A reviewer may have limited access. A share link may let someone read and copy the conversation.[3,4,5]

Training can leave traces of the material a model learnt from. Researchers have recovered personal details, but their experiments give no reliable odds for recovering a particular private chat.[6]

The email you come back to

A friend sends you a private email about a dispute at work. You paste it into a chatbot and ask for a tactful reply. The next morning the email and the suggested reply are both still in the chat. The reply on its own could tell someone quite a lot about your friend.

What happens when you press Send

The app supplies your request and the material the model needs to answer. That might include part of an attachment or information from a connected app. For tomorrow's follow-up, the app can send earlier messages again from its own records. The model provider needn't keep a transcript for that to work.[1,2]

The chat and the file can have separate lives. ChatGPT says it schedules deleted chats for permanent deletion within 30 days, with exceptions. If you saved the attachment separately, you may have to delete that copy as well.[1]

Who could read it or take a copy?

What someone can take away depends on how they got access. A screenshot is easy. Exporting the full chat or downloading an attachment may require different permissions.

Who might see it	What they might be able to read or copy
Someone with account access or a share link	It depends on their access. They may be able to read the chat and copy the text. A link can be forwarded.[1,5]
An administrator at work	Some managed accounts let authorised staff access or export conversations or files, depending on the permissions and features in use.[4]
A provider reviewer	Authorised staff or contractors may be able to inspect content for specified purposes.[3]
An authority using legal process	A provider may have to keep or hand over records. In 2025, a US court ordered OpenAI to keep logs that would otherwise have been deleted. OpenAI says this included deleted chats and Temporary Chats. Preserving a copy is a separate step from handing it over.[7]

In 2026, Cybernews researchers found 302.ai chat logs in publicly accessible storage. They read prompts and replies, some with personal details, directly from the public store. How widely anyone else read or copied the logs remains unknown.[8]

Could someone find it by searching or asking AI?

In July 2025, journalists found shared ChatGPT transcripts in search results. One identified a person from details in a chat about rewriting a CV. OpenAI said users had to opt into search visibility when sharing and ended the experiment.[5]

A model can also reproduce material it encountered in training. Nasr and colleagues got an earlier version of ChatGPT to produce an email signature with contact details under experimental conditions. The study does not trace that signature to a private chat. A chatbot with a connection to a mailbox or file store may also retrieve information from there, within the access it has for the person asking. A genuine address in a reply tells you little about which route brought it there.[2,6]

What review and training mean

'Humans may review' tells you who may be allowed to look and why. OpenAI's consumer guidance lists abuse investigations, support, legal matters and model improvement where the user has not opted out. Whether a reviewer opened your chat is a question of fact. In 2019, Belgian broadcaster VRT NWS listened to more than 1,000 Google Assistant recordings. Google confirmed that a language reviewer had leaked confidential audio.[3,9]

Training uses examples to change how a future model responds. Depending on your settings and service, these can include questions, replies and attachments. Turning off training can leave saved chats and safety access in place. OpenAI says it may keep a Temporary Chat for up to 30 days for safety. Google says Gemini can keep chats for around two months with history off.[1,3]

Geek Box · Can a list of numbers give away your words?

Some services turn passages into lists of numbers called embeddings, which help them find text related to your question. Morris and colleagues reconstructed short passages from embeddings under controlled conditions, with very different results across the systems and lengths tested. That tells us why the index deserves protection too. To apply the result to your email, someone would first need access to its numerical representation and a way to reconstruct it.[10]

How to use it better

- **Send only what you need to get an answer.** If you just want a less abrupt reply, try a stand-in email without your friend's details.
- **Check the account and the settings before you share.** Look separately at training, history, saved files, sharing and what goes to connected apps.
- **Share only what the recipient needs.** Copy the reply if that is enough. A link may give the recipient the exchange that led to it.

If the email is from a client

Suppose that private email came from a client. You may never find out whether somebody else opened it. Part 2 asks the question a lawyer has to answer before pressing Send. May I give this service the email on these terms?

Sources and notes

1. OpenAI, Chat and file retention in ChatGPT and Temporary chat in ChatGPT. Google, Gemini Apps Privacy Hub, history settings. Source 1 · Source 2 · Google

2. OpenAI, Conversation state, 'Manually manage conversation state', and Microsoft, Data, Privacy, and Security for Microsoft Copilot, 'How does Microsoft Copilot use your proprietary organizational data?' Examples of replay and connected retrieval. Source 1 · Source 2

3. OpenAI, How OpenAI handles data in consumer services, 'Model improvement for consumer and business accounts' and 'Authorized human access to content'. Source

4. OpenAI, Data access for your managed ChatGPT account, 'Your administrator can access and control data in this account' and 'Workspace policies and role-specific access'. Source

5. Chris Stokel-Walker, Fast Company, 30 July 2025, and Amanda Silberling, TechCrunch, 31 July 2025. Search indexing, a CV example and OpenAI's statement on ending the opt-in experiment. See also OpenAI's guidance on shared links. Source 1 · Source 2 · Source 3

6. Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr and Katherine Lee (2023), Scalable Extraction of Training Data from (Production) Language Models, arXiv:2311.17035, Fig. 5 and §§5.4, 5.6.3. The researchers checked examples against an internet-text corpus and, for a subset, against web search. Source

7. 18 U.S.C. §§ 2703 and 2713. See also the 13 May 2025 preservation order in The New York Times Co v Microsoft Corp, No. 23-cv-11195, ECF 551, and OpenAI's account of its scope and the end of the general hold for new records. Source 1 · Source 2 · Court order · OpenAI

8. Vilius Petkauskas, Cybernews, 22 October 2026, updated 27 October, 'What did the 302.ai data leak expose?' Direct reporting of the research team's discovery, inspection and notification. Source

9. VRT NWS and NOS, 10 July 2019, on the recordings they received. David Monsees, Google, 11 July 2019, on the reviewer's leak. Source 1 · Source 2 · VRT NWS

10. John Morris, Volodymyr Kuleshov, Vitaly Shmatikov and Alexander Rush (2023), Text Embeddings Reveal (Almost) As Much As Text, EMNLP, pp 12448–12460, §5 and Table 1. Source