How To Think About End-To-End Encryption and AI

Training, Processing, Disclosure, and Consent

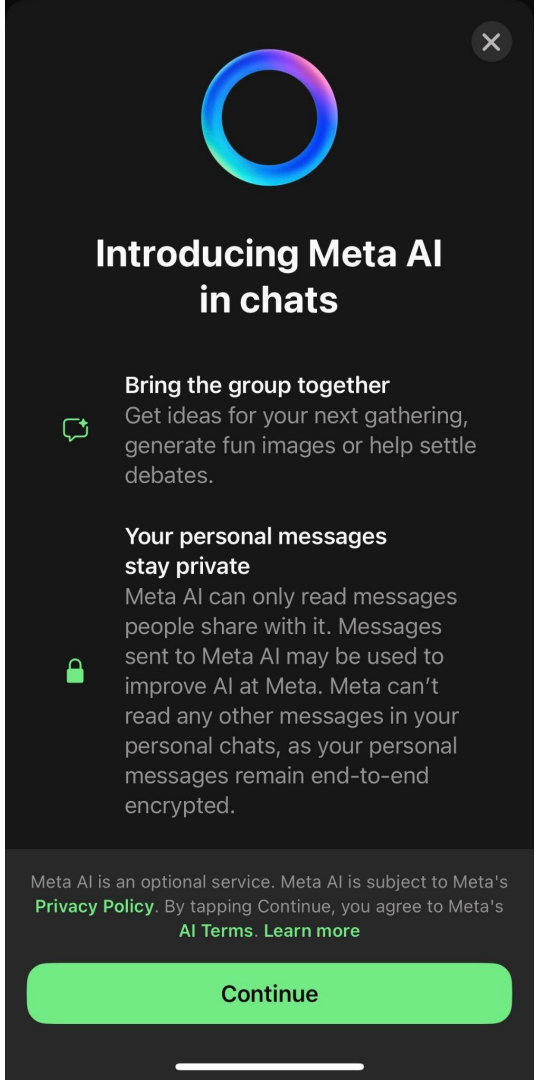
How to Think About E2EE and AI

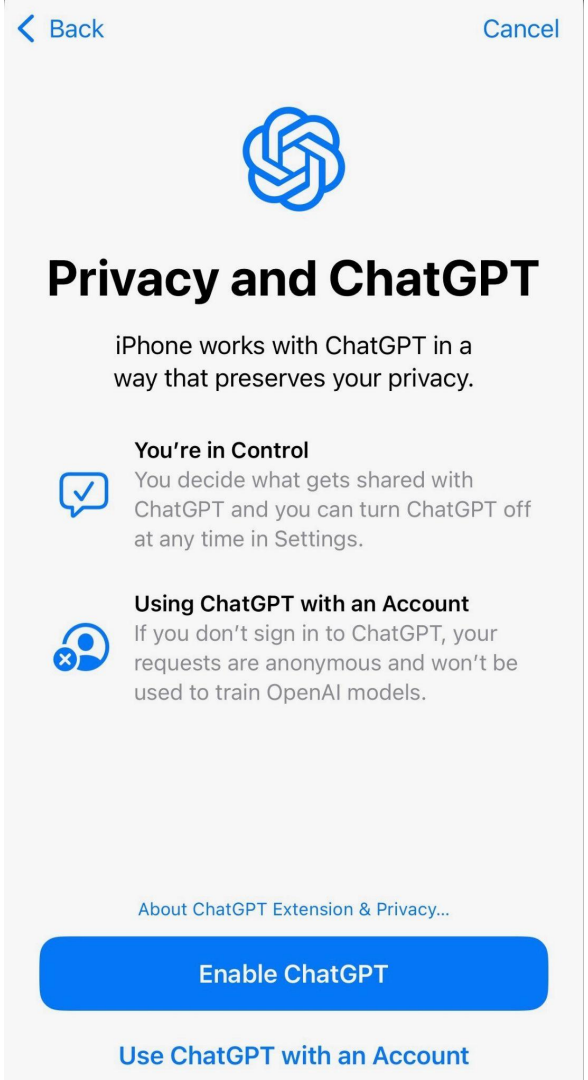
Mallory Knodel, New York University
Andrés Fábrega, Cornell University
Daniella Ferrari, New York University
Jacob Leiken, New York University
Betty Li Hou, New York University
Derek Yen, New York University
Sam de Alfaro, New York University
Kyunghyun Cho, New York University
Sunoo Park, New York University


Meta AI
Meta AI


Galaxy
AI

Apple Intelligence


Introducing Meta AI in chats
Bring the group together
Get ideas for your next gathering, generate fun images or help settle debates.
Your personal messages stay private
Meta AI can only read messages people share with it. Messages sent to Meta AI may be used to improve AI at Meta. Meta can't read any other messages in your personal chats, as your personal messages remain end-to-end encrypted.
Meta AI is an optional service. Meta AI is subject to Meta's Privacy Policy. By tapping Continue, you agree to Meta's AI Terms. Learn more
Continue


Back
Cancel
Privacy and ChatGPT
iPhone works with ChatGPT in a way that preserves your privacy.
You're in Control
You decide what gets shared with ChatGPT and you can turn ChatGPT off at any time in Settings.
Using ChatGPT with an Account
If you don't sign in to ChatGPT, your requests are anonymous and won't be used to train OpenAI models.
About ChatGPT Extension & Privacy...
Enable ChatGPT
Use ChatGPT with an Account

Is **processing of message content** by integrated AI models compatible with **end-to-end encryption**?

(If so, to what extent and under what circumstances?)

Talk outline: Our recommendations

1. Training
2. Processing
3. Disclosure
4. Consent

Security considerations

What kinds of AI integration are **compatible with E2EE**?

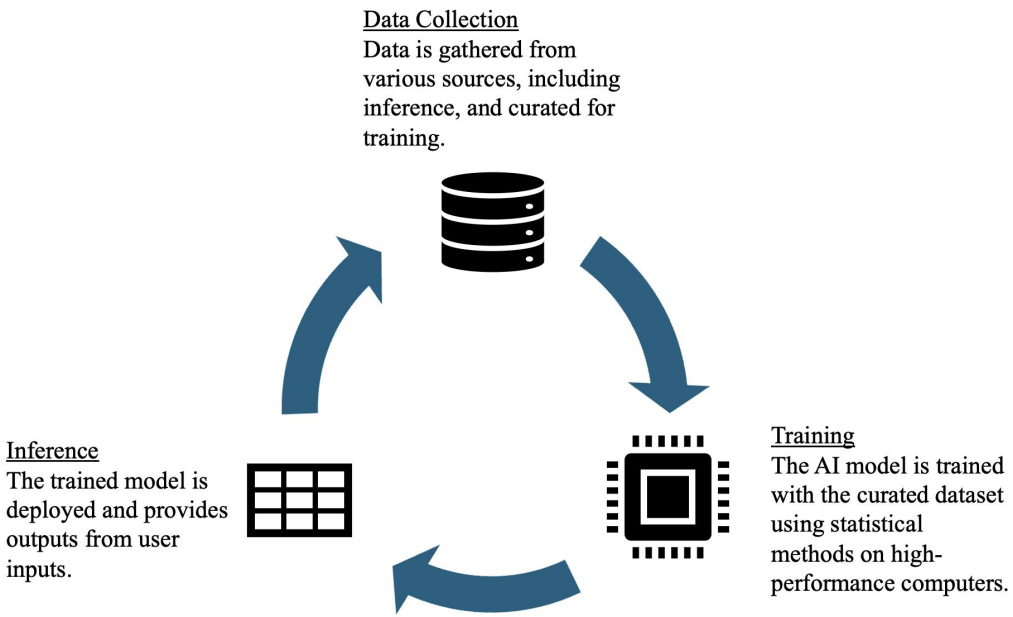
Legal & normative considerations

Given the above, what would **clear disclosure** and **meaningful consent** look like?

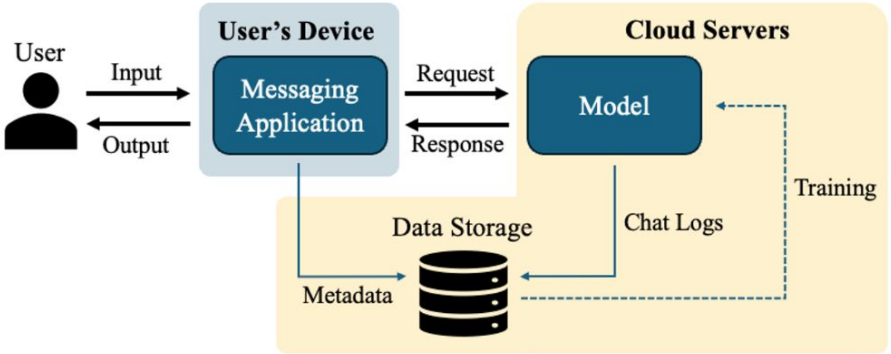
Training

Recommendation #1

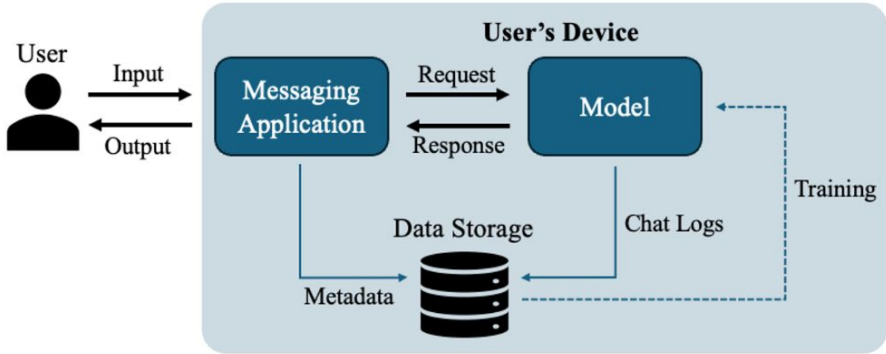
Background: AI



Background: AI



Cloud-Based Architecture



Local Architecture

Recommendation #1: Training

Using end-to-end encrypted content to train shared AI models is not compatible with E2EE.

"Shared AI model" *means a model that may be be used for inference by multiple users, and/or may be trained on multiple users' queries.*

Processing

Recommendation #2

Background: E2EE vs. TEEs

- **Confidentiality** of communication between users
- **Specification security** relies on mathematical hardness
- **Implementation security** relies on software and hardware supply chain

E2EE goals

- **Confidentiality** of compute, machine state between users and cloud
- **Specification security** relies on hardware design
- **Implementation security** relies on (different) hardware supply chain

Trusted Hardware goals

Recommendation #2: Processing

Processing E2EE content for AI features (such as inference or training) may be compatible with end-to-end encryption only if the following recommendations are upheld:

(a) Prioritize endpoint-local processing whenever possible.

(b) If processing E2EE content for non-endpoint-local models,

(i) No third party can see or use any E2EE **content** without **breaking encryption**, &

(ii) A user's E2EE content is exclusively used to fulfill that user's requests

"Content" *includes any derivative of E2EE content (i.e., any non-trivial function of content).*

"Cannot see/use without breaking encryption" *may be achieved with some privacy-preserving technical configurations like FHE.*

Disclosure & Consent

Recommendations #3 & #4

Areas of Law

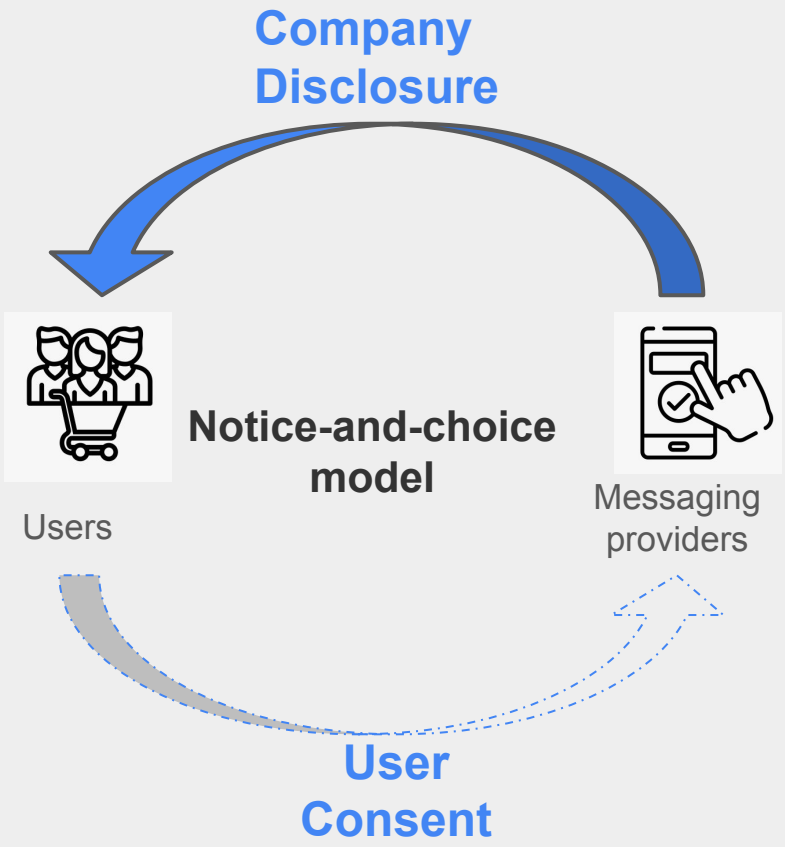
Contract Law

Consumer Protection Law

Data Protection Law

Antitrust Law

Icons designed by FreePik



User Consent

… and problems with consent

- Consumers do not read privacy policies or ToS
- Company data practices and privacy harms are hard to understand
- There are limited alternatives
- User preference may change over time and may vary according to the context
- Individual consent does not capture group settings

… but consent is **legally valid**!

Icons designed by FreePik

US Consumer Protection and the FTC

- The US Federal Trade Commission (FTC) investigates unfair or deceptive trade practices
- Deceptive practices cover privacy/security misrepresentations – when companies' disclosures do not match their practices
  - FTC v. Zoom (2020)
  - If app is marketed as having "E2EE", must:
    - Use E2EE throughout the processing of the relevant data
    - Have E2EE on by default
    - Meet industry standards

EU Data Protection Landscape

General Data Protection Regulation (GDPR) 2016

- User consent is one of six possible legal bases for processing personal data
- Consent requires an affirmative act i.e. opt-in consent
- Other safeguards: data minimization, purpose limitation, DPO appointment, etc.

ePrivacy Directive 2002,

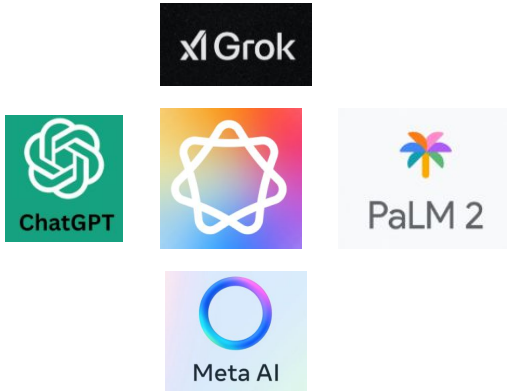
Digital Markets Act 2022,

AI Act 2024, etc.

Ongoing Controversies

Non-consensual AI training in the EU

- EU privacy regulators are questioning AI developers for violating the GDPR when training AI on personal data
  - ChatGPT was initially suspended by the Italian regulator for not collecting user consent, among other things
  - Meta AI, X's Grok, Google's PaLM 2 and Apple Intelligence have delayed/suspended releases in EU because of Irish regulator inquiries
- Training on private chat data will raise more legal concerns



Recommendation #3: Disclosure

Messaging providers should not make **unqualified representations** that they provide E2EE if the **default** for any conversation is that **E2EE content is used** (e.g., for AI inference or training) by any **third party**.

Recommendation #4: Opt-in Consent

AI assistant features, if offered in E2EE systems, should generally be **off by default** and only activated via **opt-in consent**. Obtaining meaningful consent is complex, and requires careful consideration including but not limited to: scope and granularity of **opt-in/out**, **ease and clarity** of opt-in/out, group consent, and management of consent over time.

Talk outline: Our recommendations

1. Training
2. Processing
3. Disclosure
4. Consent

Security considerations

What kinds of AI integration are **compatible with E2EE**?

Legal & normative considerations

Given the above, what would **clear disclosure** and **meaningful consent** look like?

Thank you!

Mallory Knodel,
mallory.knodel@nyu.edu

ia.cr/2024/2086