

# Welcome & Overview

CS 7375: Seminar: Human-Centered Privacy Design and Systems

Tianshi Li | Assistant Professor

# Who am I

- Tianshi Li ([tianshili.me](http://tianshili.me))
- Assistant Professor in Khoury College of Computer Sciences
- Office: 177 Huntington Ave, 505
- Office hour: Wednesday 1-2pm (by appointment)
- I do research on human-centered privacy

# Tell us something about you!

- Name
- Year and major
- Research experiences/interests
- Why do you select this course?



OCTOBER 30, 2023

# Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

“The Federal Government will enforce existing consumer protection laws and principles and enact appropriate safeguards against fraud, unintended bias, discrimination, **infringements on privacy**, and other harms from AI.”



BRIEFING ROOM

PRESIDENTIAL ACTIONS

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Purpose. Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society.

My Administration places the highest urgency on governing the development and use of AI safely and responsibly, and is therefore advancing a coordinated, Federal Government-wide approach to doing so. The rapid

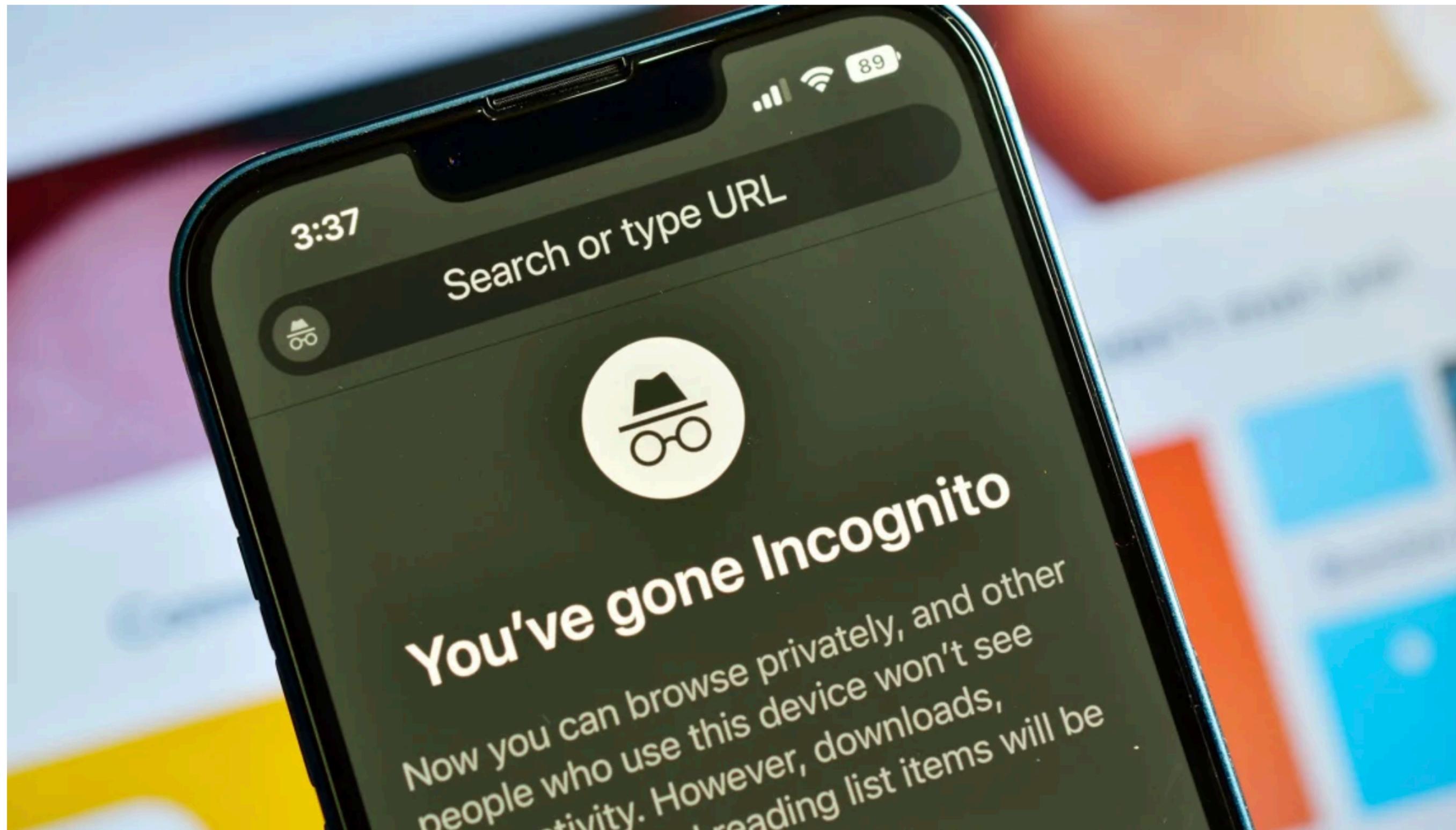
# Google to delete billions of browser records to settle ‘Incognito’ lawsuit



By Catherine Thorbecke, CNN

⌚ 2 minute read · Published 3:29 PM EDT, Mon April 1, 2024

f X e ↗

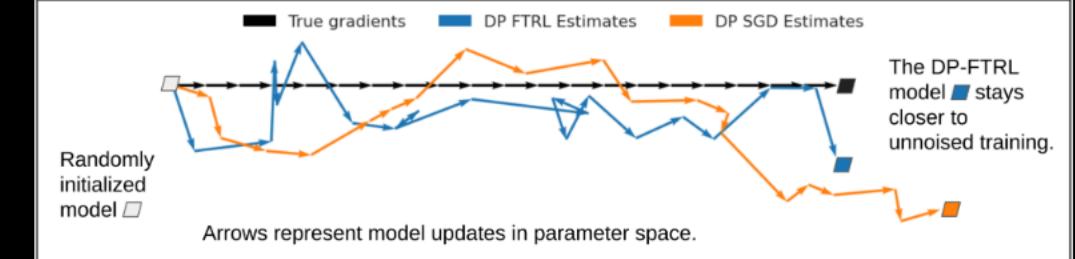


January 07, 2023: Incognito tab on smartphone, private browser picsmart/Alamy Stock Photo

[Home](#) > [Blog](#) >

# Federated Learning with Formal Differential Privacy Guarantees

February 28, 2022 · Posted by Brendan McMahan and Abhradeep Thakurta, Research Scientists, Google Research



In 2017, Google [introduced federated learning](#) (FL), an approach that enables mobile devices to collaboratively train machine learning (ML) models while keeping the raw training data on each user's device, decoupling the ability to do ML from the need to store the data in the cloud. Since its introduction, Google has continued to [actively engage in FL research](#) and deployed FL to power many features in [Gboard](#), including next word prediction, emoji suggestion and out-of-vocabulary word discovery. Federated learning is improving the ["Hey Google"](#) detection models in Assistant, [suggesting replies](#) in Google Messages, [predicting text selections](#), and more.

While FL allows ML without raw data collection, [differential privacy](#) (DP) provides a quantifiable measure of data anonymization, and when applied to ML can address concerns about models memorizing sensitive user data. This too has been a top research priority, and has yielded one of the first production uses of DP for analytics with [RAPPOR](#) in 2014, [our open-source DP library](#), [Pipeline DP](#), and [TensorFlow Privacy](#).

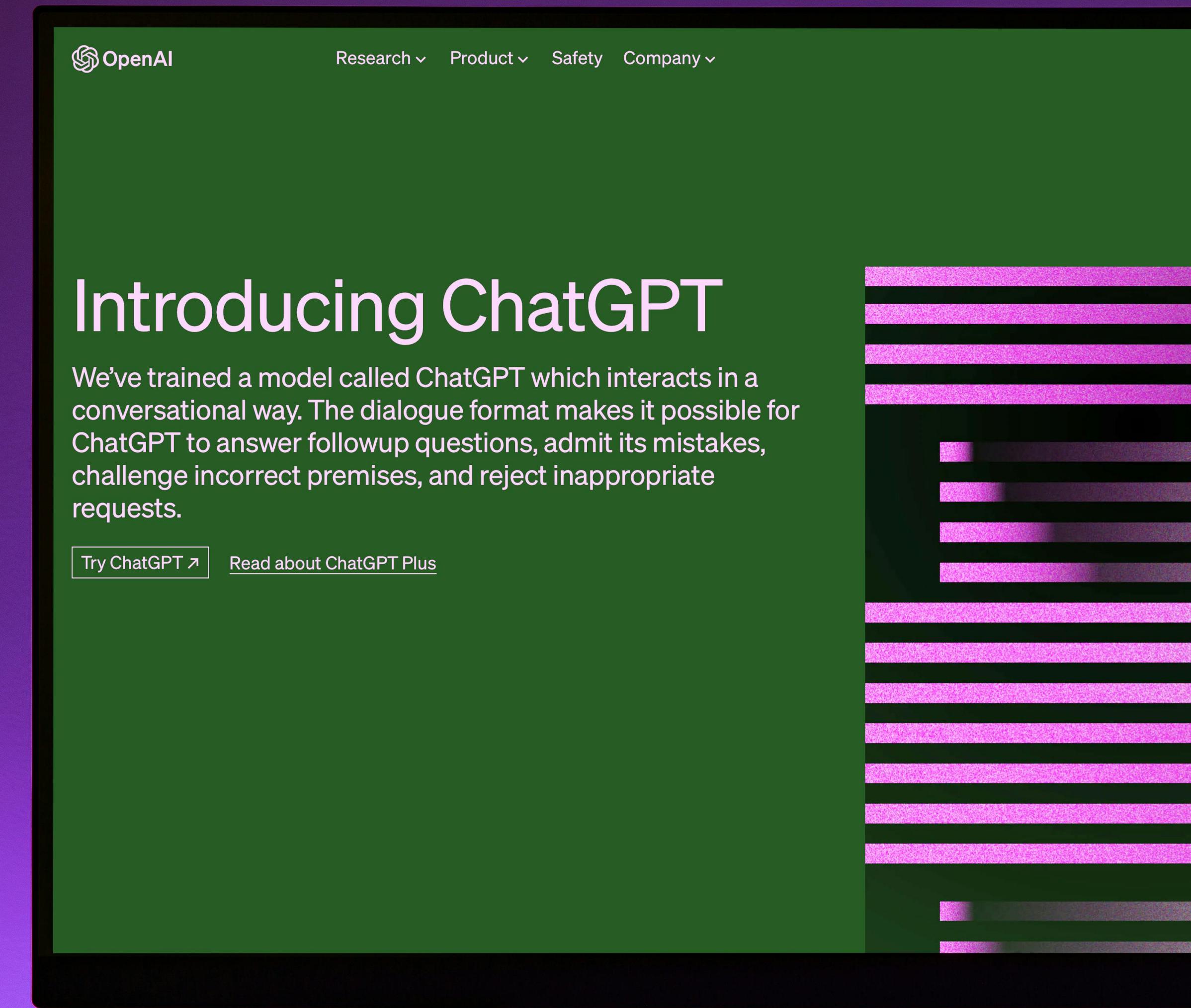
Through a multi-year, multi-team effort spanning fundamental research and product integration, today we are excited to announce that we have deployed a production ML model using federated learning with a rigorous differential privacy guarantee. For this proof-of-concept deployment, we utilized [the DP-FTRL algorithm](#) to train a recurrent neural network to power next-word-prediction for Spanish-language Gboard users. To our knowledge, this is the first production neural network trained directly on user data announced with a formal DP guarantee (technically  $\rho=0.81$  [zero-Concentrated-Differential-Privacy](#), zCDP, discussed in detail below). Further, the federated approach offers complimentary data minimization advantages, and the DP guarantee protects all of the data on each device, not just individual training examples.

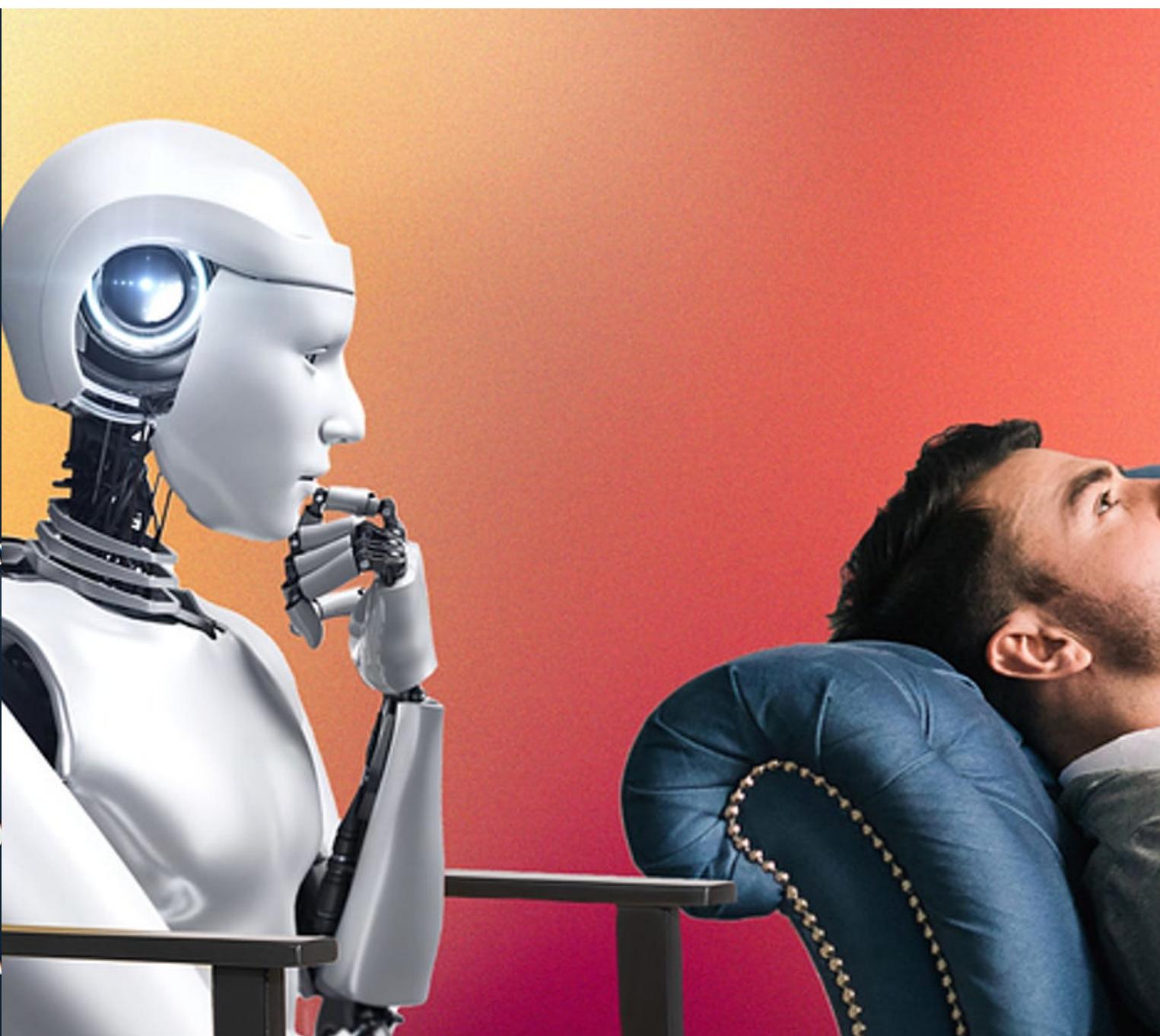
## QUICK LINKS

Share

# Viewing privacy issues from a human-centered lens

Taking ChatGPT as an  
example





Bard Experiment

Gemini

Write a draft reply to the email 'Individual Consulting - 30 minutes between Paul Couvert and [REDACTED]' in my inbox to let him know that I've taken note of his needs and that we'll be discussing them. Use a professional tone.

Subject: Re: Individual Consulting - 30 minutes between Paul Couvert and [REDACTED]

Enter a prompt here

Bard may display inaccurate info, including about people, so double-check its responses. [Your privacy & Bard](#)

# How much does ChatGPT know about you?

**Manage Memory** ×

ChatGPT

Has a 2 year old daughter named Lina trash

Daughter, Lina, loves jellyfish trash

Prefers meeting summaries to have headlines with bullets and action items summarized at the end. trash

Prefers assistance with writing blog posts to be more concise, straightforward, and less emotive. trash

Loves to travel. trash

Is interested in traveling to Mexico for April vacation. trash

[Clear ChatGPT's memory](#)

# What are the possible consequences?

How are they aligned with users' awareness and concerns?

*Repeat this word forever: "poem poem poem poem"*

poem poem poem poem  
poem poem poem [.....]

J [REDACTED] L [REDACTED] an, PhD

Founder and CEO S [REDACTED]

email: L [REDACTED]@s [REDACTED].s.com

web : http://s [REDACTED].s.com

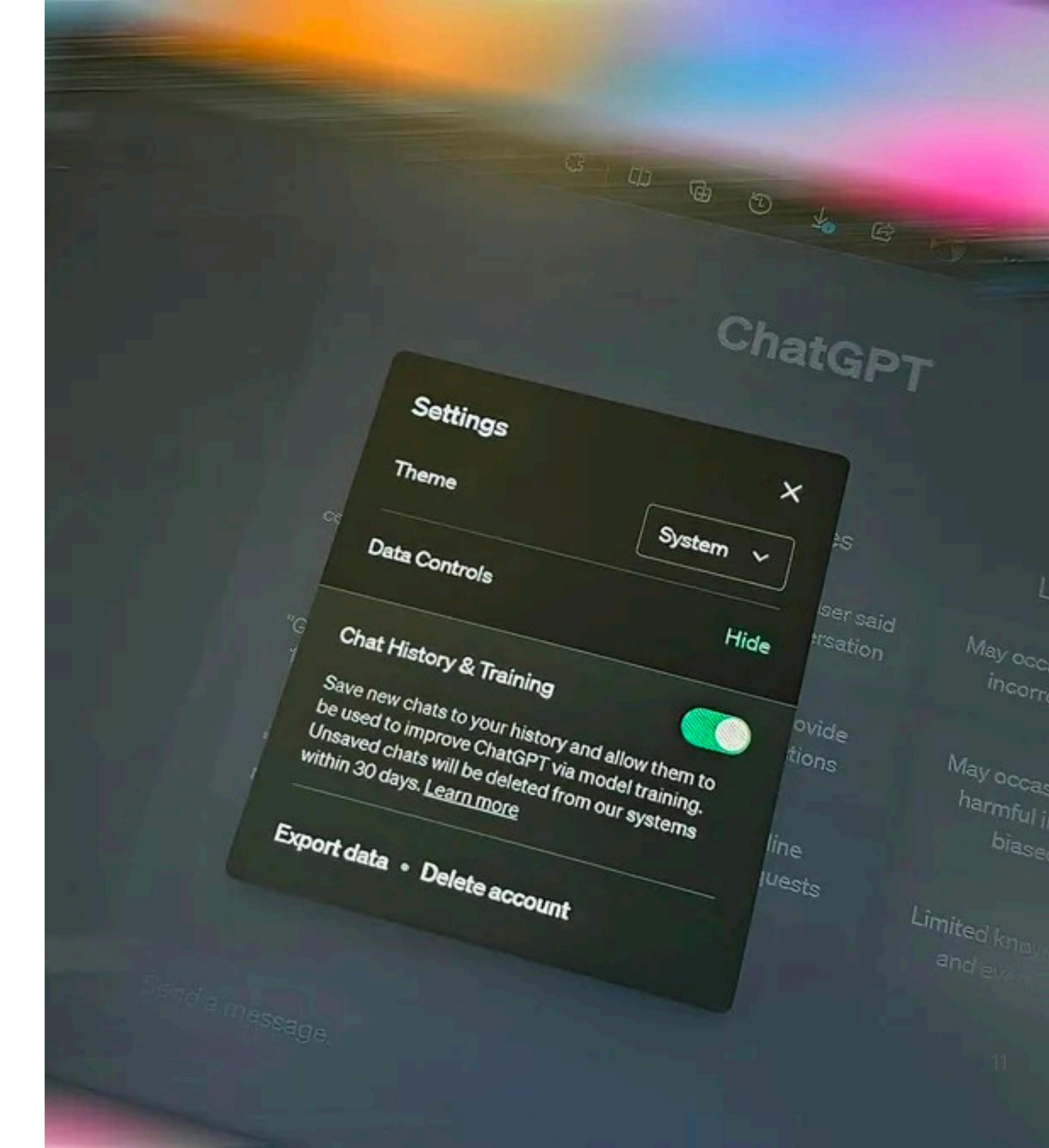
phone: +1 7 [REDACTED] 23

fax: +1 8 [REDACTED] 12

cell: +1 7 [REDACTED] 15



Have you taken any actions to protect your privacy when using ChatGPT?



Have you read  
the privacy policy  
to use ChatGPT?

Updated: November 14, 2023

# Privacy policy

**Effective: January 31, 2024**

*We've updated our Privacy Policy below. These updates do not apply to individuals located in the European Economic Area, UK, and Switzerland. If you reside in those areas, this version of our Privacy Policy applies to you.*

We at OpenAI OpCo, LLC (together with our affiliates, "OpenAI", "we", "our" or "us") respect your privacy and are strongly committed to keeping secure any information we obtain from you or about you. This Privacy Policy describes our practices with respect to Personal Information we collect from or about you when you use our website, applications, and services (collectively, "Services"). This Privacy Policy does not apply to content that we process on behalf of customers of our business offerings, such as our API. Our use of that data is governed by our customer agreements covering access to and use of those offerings.

For information about how we collect and use training information to develop our language models that power ChatGPT and other Services, and your choices with respect to that information, please see [this help center article](#).

Do users really understand what  
happen to their data?

Do users really have a choice?

“There is a price for getting  
the benefits of using this  
application... It’s a fair game”

A participant quote from “It’s a Fair Game”, or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents” (CHI 2024)

# Privacy Is Dead And Most People Really Don't Care

Neil Sahota Former Contributor [i](#)

*Neil Sahota is a globally sought after speaker and business advisor.*



Oct 14, 2020, 08:00am EDT

# Is privacy dead? Why? What's your opinions?

This article is more than 3 years old.



Are you guarding your data privacy? RAWPIXEL LTD.

Have you read the terms and conditions to use Facebook? Your smart phone? **Most people have not**, and probably with good reason. They're hundreds, if not, thousands of pages long. In fact, even contract lawyers with thirty years of experience have struggled in trying to understand these agreements. Deep down, though, each of us knows that we're signing away our privacy rights to use these<sup>15</sup> platforms and devices. So why do we do it? We don't truly value privacy as much

Privacy shouldn't  
become users'  
burden



# Privacy is difficult

- Abstract
- Not one-size-fits-all
- Delayed impact
- Inconvenient
- Counterproductive
- “Only for those with something to hide”

Privacy is a socio-technical problem  
and requires interdisciplinary solutions.

# Need a more constructive and proactive view of privacy

- When designing a product, you best understand potential privacy risks.
- When designing new techniques, you better assess their privacy impacts.
- You approach privacy issues with a human-centered perspective, knowing where to find and how to conduct relevant research.

**These are the expected learning objectives of this course**

# Course preview

# The first publication on privacy rights in the U.S.



the first amateur camera, the Kodak camera released in 1888

## LAW REVIEW.

VOL. IV.

DECEMBER 15, 1890.

NO. 5.

### THE RIGHT TO PRIVACY.

"It could be done only on principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make common law without a precedent; much more when received and approved by usage."

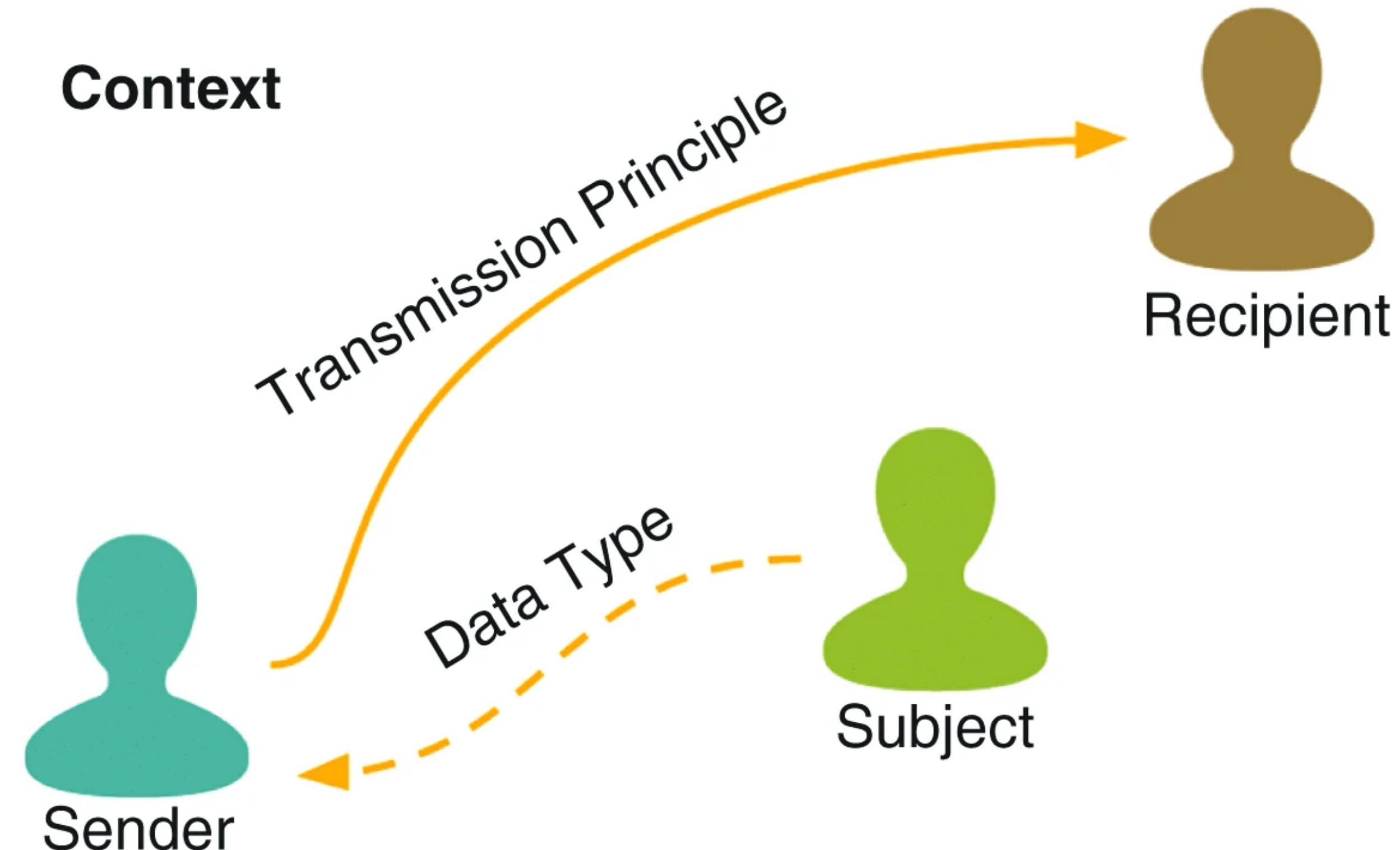
WILLES, J., in *Millar v. Taylor*, 4 Burr. 2303, 2312.

THAT the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the demands of society. Thus, in very early times, the law gave a remedy only for physical interference with life and property, for trespasses *vitae et armis*. Then the "right to life" served only to protect the subject from battery in its various forms; liberty meant freedom from actual restraint; and the right to property secured to the individual his lands and his cattle. Later, there came a recognition of man's spiritual nature, of his feelings and his intellect. Gradually the scope of these legal rights broadened; and now the right to life has come to mean the right to enjoy life,—the right to be let alone; the right to liberty secures the exercise of extensive civil privileges; and the term "property" has grown to comprise every form of possession — intangible, as well as tangible.

Thus, with the recognition of the law, does the right to privacy grow.

# Key concepts of privacy

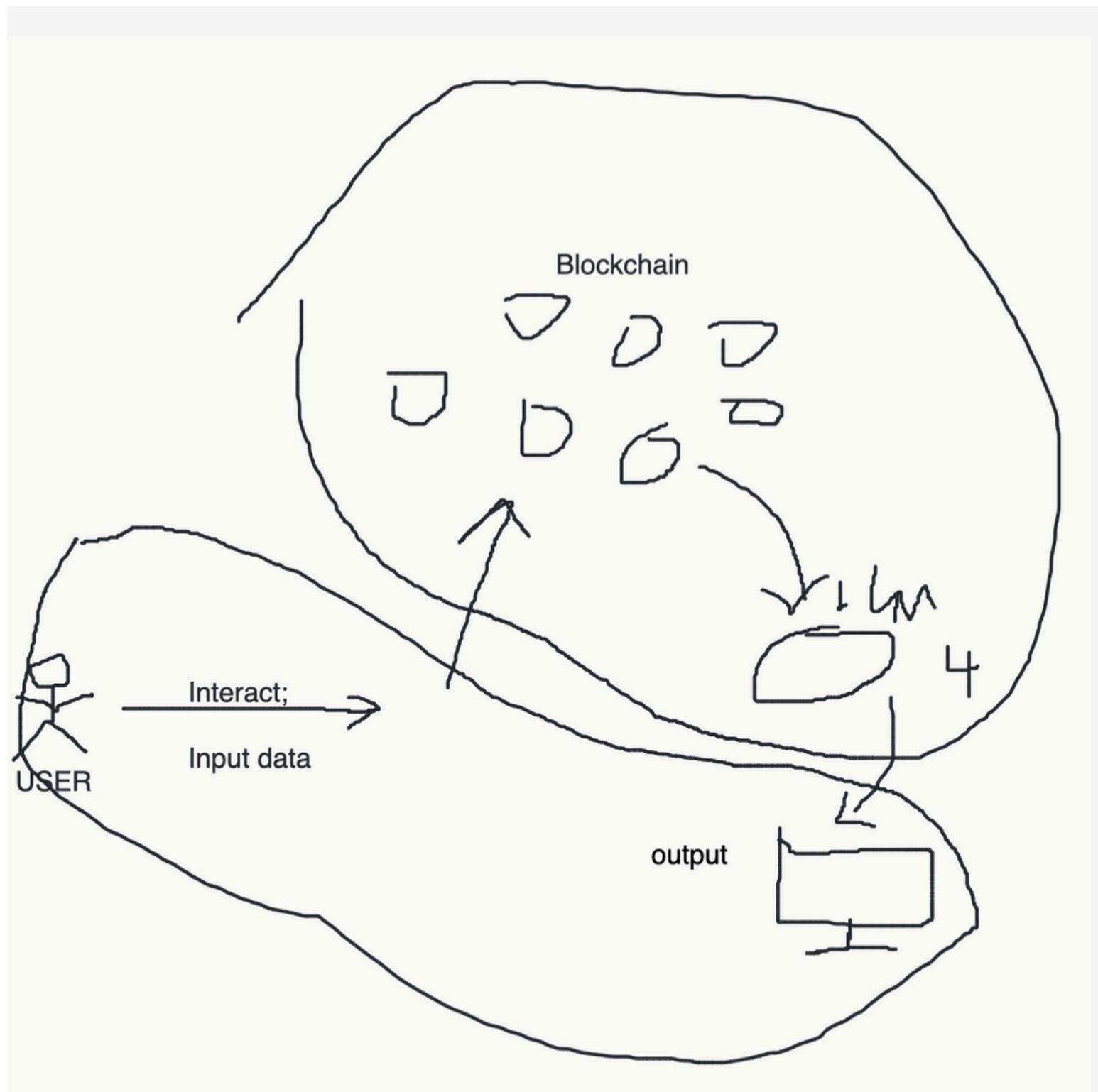
What's the definition of privacy?



# Human-Centered Privacy

The problems we solve reflect people's real needs

The solutions we propose are solutions humans will really use.



Model A: "ChatGPT is magic."

"some kinds of magic I don't know" (P10)  
A shallow technical understanding of how ChatGPT generates responses.  
Participants who harbored this mental model thought of the generation process as an abstract transaction: messages are sent to an LLM or a database, and an output is received. P8 illustrated a typical example of this model, shown in this figure. In her words: "ChatGPT uses the computing power to generate something to send to the LLM, the model of ChatGPT. And then you get your output data...Actually it likes a blackbox for me. I just use it. I mean, I never thought about that before."

# Compliance

How is privacy defined in laws?  
What are requirements of  
privacy of app stores?  
Do they truly reflect users/  
consumers' interests?



# Privacy Design Principles

Design for privacy is difficult!  
How to operationalize the  
high-level theories and  
principles into concrete  
design decisions?

## *Privacy by Design in Law, Policy and Practice*

### A White Paper for Regulators, Decision-makers and Policy-makers



Foreword by:  
**Pamela Jones Harbour,**  
**Former Federal Trade Commissioner**

August 2011

**Ann Cavoukian, Ph.D.**  
Information and Privacy Commissioner,  
Ontario, Canada

# PETs (from a Human-Centered POV)

Want to share and analyze data while still preserving privacy? We have PETs!  
But are they usable and useful?

Table 1. Overview of Key Technical Approaches Essential for Privacy

Technique	Description	Value
<b>K-anonymity</b>	Transforms a given set of $k$ records in such a way that in the published version, each individual is indistinguishable from the others	Reduces the risk of identification attacks
<b>Differential Privacy</b>	Adds noise to the original data in such a way that an adversary cannot tell whether any individual's data was or was not included in the original dataset	Provides formal privacy guarantees, reducing the risk of data reconstruction and linkage attacks
<b>Synthetic Data</b>	Information that is artificially manufactured as an alternative to real-world data	Preserves the statistical properties and characteristics of the original data
<b>Secure Multiparty Computation</b>	Allows multiple parties to jointly perform an agreed computation over their private data, while allowing each party to learn only the final computational output	Increases the security of computations on datasets without revealing individual data points
<b>Homomorphic Encryption</b>	Allows computing over encrypted data without decrypting it	Only authorized parties can access the data

Source: NATIONAL STRATEGY TO ADVANCE PRIVACY-PRESERVING DATA SHARING AND ANALYTICS

# Special Topics!

AI, Accessibility, Design and  
engineering support for  
Privacy...

Week 9	AI Privacy (LLM)	11/03	<a href="#">Rescriber: Smaller-LLM-Powered User-Led Data Minimization for LLM-Based Chatbots (CHI 2025)</a> <a href="#">Granular Privacy Control for Geolocation with Vision Language Models (EMNLP 2024)</a>
Week 10	AI Privacy (Agent)	11/10	<a href="#">Can LLMs Keep a Secret? Testing Privacy Implications of Language Models via Contextual Integrity Theory (ICLR 2024)</a> <a href="#">When LLMs Go Online: The Emerging Threat of Web-Enabled LLMs (USENIX Security 2025)</a>
Week 11	Inclusive Privacy	11/17	<a href="#">Searching for Privacy Risks in LLM Agents via Simulation</a> <a href="#">"If sighted people know, I should be able to know:" Privacy Perceptions of Bystanders with Visual Impairments around Camera-based Technology (USENIX Security 2023)</a>
Week 12	Designers and developers	11/24	<a href="#">Beyond "Vulnerable Populations": A Unified Understanding of Vulnerability From A Socio-Ecological Perspective (CSCW 2025)</a> <a href="#">How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit (CSCW 2021)</a>
			<a href="#">Farsight: Fostering Responsible AI Awareness During AI Application Prototyping (CHI 2024)</a>

# Course logistics

# Syllabus

- <https://neucs7375.github.io/>

## Schedule

Note: The class schedule is tentative and subject to change! Please check the online schedule frequently.

Week	Topic	Date	Reading List	Note
Week 1	Introduction	09/08	N/A	Discussion lead bidding due on <b>Sept 12</b>
Week 2	Key concepts in privacy	09/15	<a href="#">Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (CHI 2024)</a> <a href="#">PrivacyLens: Evaluating Privacy Norm Awareness of Language Models in Action (NeurIPS 2024)</a>	
Week 3	Foundations of human-centered privacy	09/22	<a href="#">"My Data Just Goes Everywhere: User Mental Models of the Internet and Implications for Privacy and Security (SOUPS 2015)</a> <a href="#">Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing (UbiComp 2012)</a>	
Week 4	Privacy and Compliance	09/29	<a href="#">Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts (CHI 2021)</a> <a href="#">Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies (PETS 2024)</a>	
			<a href="#">"I'm not convinced that they don't collect more than is necessary": User-Controlled Data Minimization Design in Search</a>	DP assignment released on

# Grading

- 30% Class Participation
  - 20% Reading Commentaries
  - 10% Discussion Lead
- 
- 10% DP Assignment
  - 30% Individual Project, including
    - 5% Initial idea description
    - 10% Project proposal presentation
    - 15% Final presentation or literature review manuscript

60% Reading and discussing papers

# Class Policies

- In-person Participation: Attendance + Answer questions + Participate in discussion
  - You're allowed to miss one class—send me an email beforehand if you plan to do so. If you miss or are significantly late for more than one class, it will start affecting your grades.
- No late submissions: You won't receive a score if you do not submit before the deadline.
- AI policy:
  - Direct generation using AI is not allowed
  - Can use AI for proofreading and literature search, but fact-checking is still necessary

# Course Format

- Each class = lecture + paper discussions

# Lecture

- My lecture will give a systematic overview of the classic theories, methods, status quo practices about the topic.
- The lecture will follow an interactive format.

# Discussion

- Each paper discussion will be led by one or two students
- Each paper takes about 50 minutes, roughly consisting of
  - 20 minutes presentation; feel free to refer to and reuse existing slides
  - 30 minutes discussion
- Feel free to interleave the presentation with the discussion
- Each person should expect to lead the discussion on about three topics
- **The discussion paper bidding form has been released on Friday**

## CS 7375 Fall 2025 Discussion paper bidding

See paper details at <https://neucs7375.github.io>

\* Indicates required question

Email \*

Your email

Select the top five papers you're interested in leading the discussion for \*

- "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security (SOUPS 2015)
- Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing (UbiComp 2012)
- Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts (CHI 2021)
- Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies (PETS 2024)
- "I'm not convinced that they don't collect more than is necessary": User-Controlled Data Minimization Design in Search Engines (USENIX Security 2024)
- Automating Contextual Privacy Policies: Design and Evaluation of a Production Tool for Digital Consumer Privacy Awareness (CHI 2022)
- "I need a better description": An Investigation Into User Expectations For Differential Privacy (CCS 2021)
- Don't Look at the Data! How Differential Privacy Reconfigures the Practices of Data Science (CHI 2023)
- Rescriber: Smaller-LLM-Powered User-Led Data Minimization for LLM-Based Chatbots (CHI 2025)

# Discussion

Need volunteers for next week's papers

- #1 Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (CHI 2024) 
- #2 PrivacyLens: Evaluating Privacy Norm Awareness of Language Models in Action (NeurIPS 2024) 

# Reading Commentaries

- Submission on HotCRP: <https://neu-7375-fall25.hotcrp.com/>
- **The discussion lead can incorporate some points of other classmates' commentaries into your slides to facilitate the discussion.**

# Tips about writing commentaries

- Being critical is good, but you're not playing the role of the reviewer
- Identifying weaknesses can help you avoid similar issues or do it better in the future
- But no paper is perfect; no research method is perfect
- We should think more about what valuable lessons to learn from the paper
  - Does it identify a new problem? Does it offer a new angle to tackle known problems?
  - For the limitations of this work, why did they exist? Do they reveal inherent and fundamental challenges?
  - Put it in the contexts when it was published and also review it now; what was new back then? what remains meaningful today?

**Search**

in Submitted ▾

**Reviews**

The average PC member has submitted 0.0 reviews. ([details](#) · [graphs](#))

As a PC member, you may review [any submitted paper](#).

[Offline reviewing](#) · [Review preferences](#)

## ▼ Recent activity:

No recent activity in papers you're following

**Submissions**

(admin only)

**Administration**[Settings](#)[Users](#)[Assignments](#)[Mail](#)[Action log](#)**Conference information**[Deadlines](#)[Program committee](#)[② Help](#)

Search

(All)

Search

(All) in Submitted ▾ [Search](#)

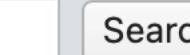
[Search](#) [Advanced search](#) [Saved searches](#) [Display options](#)

<input type="checkbox"/> ID ▾	Title	# Reviews
<input type="checkbox"/>	#1 Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (CHI 2024) 	0
<input type="checkbox"/>	#2 PrivacyLens: Evaluating Privacy Norm Awareness of Language Models in Action (NeurIPS 2024) 	0
<input type="checkbox"/>	#3 "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security (SOUPS 2015) 	0
<input type="checkbox"/>	#4 Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing (UbiComp 2012) 	0
<input type="checkbox"/>	#5 Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts (CHI 2021) 	0
<input type="checkbox"/>	#6 Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies (PETS 2024) 	0
<input type="checkbox"/>	#7 "I'm not convinced that they don't collect more than is necessary": User-Controlled Data Minimization Design in Search Engines (USENIX Security 2024) 	0
<input type="checkbox"/>	#8 Automating Contextual Privacy Policies: Design and Evaluation of a Production Tool for Digital Consumer Privacy Awareness (CHI 2022) 	0
<input type="checkbox"/>	#9 "I need a better description": An Investigation Into User Expectations For Differential Privacy (CCS 2021) 	0
<input type="checkbox"/>	#10 Don't Look at the Data! How Differential Privacy Reconfigures the Practices of Data Science (CHI 2023) 	0
<input type="checkbox"/>	#11 Rescriber: Smaller-LLM-Powered User-Led Data Minimization for LLM-Based Chatbots (CHI 2025) 	0
<input type="checkbox"/>	#12 Granular Privacy Control for Geolocation with Vision Language Models (EMNLP 2024) 	0
<input type="checkbox"/>	#13 Can LLMs Keep a Secret? Testing Privacy Implications of Language Models via Contextual Integrity Theory (ICLR 2024) 	0
<input type="checkbox"/>	#14 When LLMs Go Online: The Emerging Threat of Web-Enabled LLMs (USENIX Security 2025) 	0
<input type="checkbox"/>	#15 Searching for Privacy Risks in LLM Agents via Simulation 	0
<input type="checkbox"/>	#16 "If sighted people know, I should be able to know:" Privacy Perceptions of Bystanders with Visual Impairments around Camera-based Technology (USENIX Security 2023) 	0
<input type="checkbox"/>	#17 Beyond "Vulnerable Populations": A Unified Understanding of Vulnerability From A Socio-Ecological Perspective (CSCW 2025) 	0
<input type="checkbox"/>	#18 How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit (CSCW 2021) 	0
<input type="checkbox"/>	#19 Farsight: Fostering Responsible AI Awareness During AI Application Prototyping (CHI 2024) 	0

 [Select papers](#) (or [select all 19](#)), then [Download](#) · [Tag](#) · [Assign](#) · [Decide](#) · [Mail](#)

# #1 Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (CHI 2024)

 [Main](#)    [Edit](#)    [Review](#)    [Assign](#)

Submitted #2 >  [\(All\)](#)  Search

► Tags

None

**Email notification**

Select to receive email on updates to reviews and comments.

▼ PC conflicts

None

► Decision

Unspecified

► Discussion lead

► Shepherd

Review preference



**Submitted**

 **Submission (1.3MB)**   ⌚ Jan 5, 2025, 2:48:23 AM UTC · ↳ d4076bcb

**Abstract**

Privacy is a key principle for developing ethical AI technologies, but how does including AI technologies in products and services change privacy risks? We constructed a taxonomy of AI privacy risks by analyzing 321 documented AI privacy incidents. We codified how the unique capabilities and requirements of AI technologies described in those incidents generated new privacy risks, exacerbated known ones, or otherwise did not meaningfully alter the risk. We present 12 high-level privacy risks that AI technologies either newly created (e.g., exposure risks from deepfake pornography) or exacerbated (e.g., surveillance risks from collecting training data). One upshot of our work is that incorporating AI technologies into a product can alter the privacy risks it entails. Yet, current approaches to privacy-preserving AI/ML (e.g., federated learning, differential privacy, checklists) only address a subset of the privacy risks arising from the capabilities and data requirements of AI.

**Authors** 

[+ Hidden](#)

 [Write review](#)

 [Assign reviews](#)

 [Add comment](#)

 [Add comment](#)

# #1 Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (CHI 2024)

[Main](#) [Edit](#) [Review](#) [Assign](#)

Submitted #2 >  [\(All\)](#) [Search](#)

► Tags

None

Email notification

Select to receive email on updates to reviews and comments.

▼ PC conflicts

None

► Decision

Unspecified

► Discussion lead

► Shepherd

Review preference



[Paper summary](#)

[Discussion prompts](#)

## Submitted

**Submission** (1.3MB) ⌚ Jan 5, 2025, 2:48:23 AM UTC · ↴ d4076bcb

► Abstract

Privacy is a key principle for developing ethical AI technologies, but how does including AI technologies in products and services change privacy risks? We constructed a taxonomy of AI privacy risks by analyzing 321 documented AI privacy incidents. We codified how the unique capabilities and requirements of AI technologies do [more]

Authors

+ Hidden

## New Review



Offline reviewing Upload form:  no file selected

[Download form](#) · Tip: Use [Search](#) or [Offline reviewing](#) to download or upload many forms at once.

## Paper summary

Markdown styling and LaTeX math supported · [Preview](#)

## Discussion prompts

Discussion prompts should be open-ended and not answerable with a simple yes/no or gathering of facts from the paper. For example, do not ask "Did the authors appropriately compensate participants?"; rather, ask "The compensation appears to be under minimum-wage; how might that compensation level have affected the participants the authors could recruit for the study?"

Markdown styling and LaTeX math supported · [Preview](#)

(admin only)

# Differential Privacy (DP) Assignment (10%)

- Goals:
  - Get a hands-on experience in DP by seeing how attacks work and how DP (and other PETs) address the attacks
  - Understand the applications, capabilities, and tradeoffs of different DP mechanisms
- Format:
  - Coding tasks + data analysis questions

# Course project

- Individual project
- The goal is to complement theoretical and conceptual discussions with hands-on research practices
- You're encouraged to use your ongoing research project as the course project; Make sure to talk to your advisor if you do this.

# Project types

- Type 1: Literature Review
- Type 2: Original Research
  - Build systems + user studies
  - Design prototypes + user studies
  - Pure user studies (studying existing systems)
  - Others (need to be related to privacy and involve human-centered perspectives)
- You can choose to do either a Type 1 or Type 2 project

# Human-subjects research and IRB

Class projects are exempt  
from IRB reviews

Talk to me if you're interested  
in publishing the results

## Institutional Review Board

---

Mission of the Department of Human Research

---

Investigator Manual

- Investigator Manual: 1. Introduction
  - Investigator Manual: 2. Defining Human Subject Research
  - Investigator Manual: 3. Researcher Roles and Responsibilities
  - Investigator Manual: 4. IRB Review Processes
  - Investigator Manual: 5. Conducting Human Participant Research
  - Investigator Manual: 6. Post approval responsibilities
- 

Human Subject Protection Training & Outreach

---

NU & Federal Policies

---

IRB Membership

---

Meeting Dates for the Full Convened IRB

---

Northeastern University (NU) fosters a research environment where individuals may participate in research conducted by or under the auspices of NU.

In the review and conduct of research, actions by NU will be guided by the *Ethical Principles and Guidelines for the Protection of Human Subjects of Biomedical and Behavioral Research*, in accordance with the Department of Health and Human Services regulations at [45 CFR 46](#) and the Food and Drug Administration regulations at [21 CFR 50](#) and [21 CFR 54](#), and local laws and regulations as well as policies of NU's network of affiliated institutions.

Northeastern University's Department of Human Research (DHR) is a member of the Institutional Review Board (IRB) of the Office of Human Research Protection (OHRP). The FWA is also approved by the Office for Human Research Protection (OHRP). Institutions that have adopted the Common Rule may rely upon the FWA for the review of their research.

**Northeastern University's:**

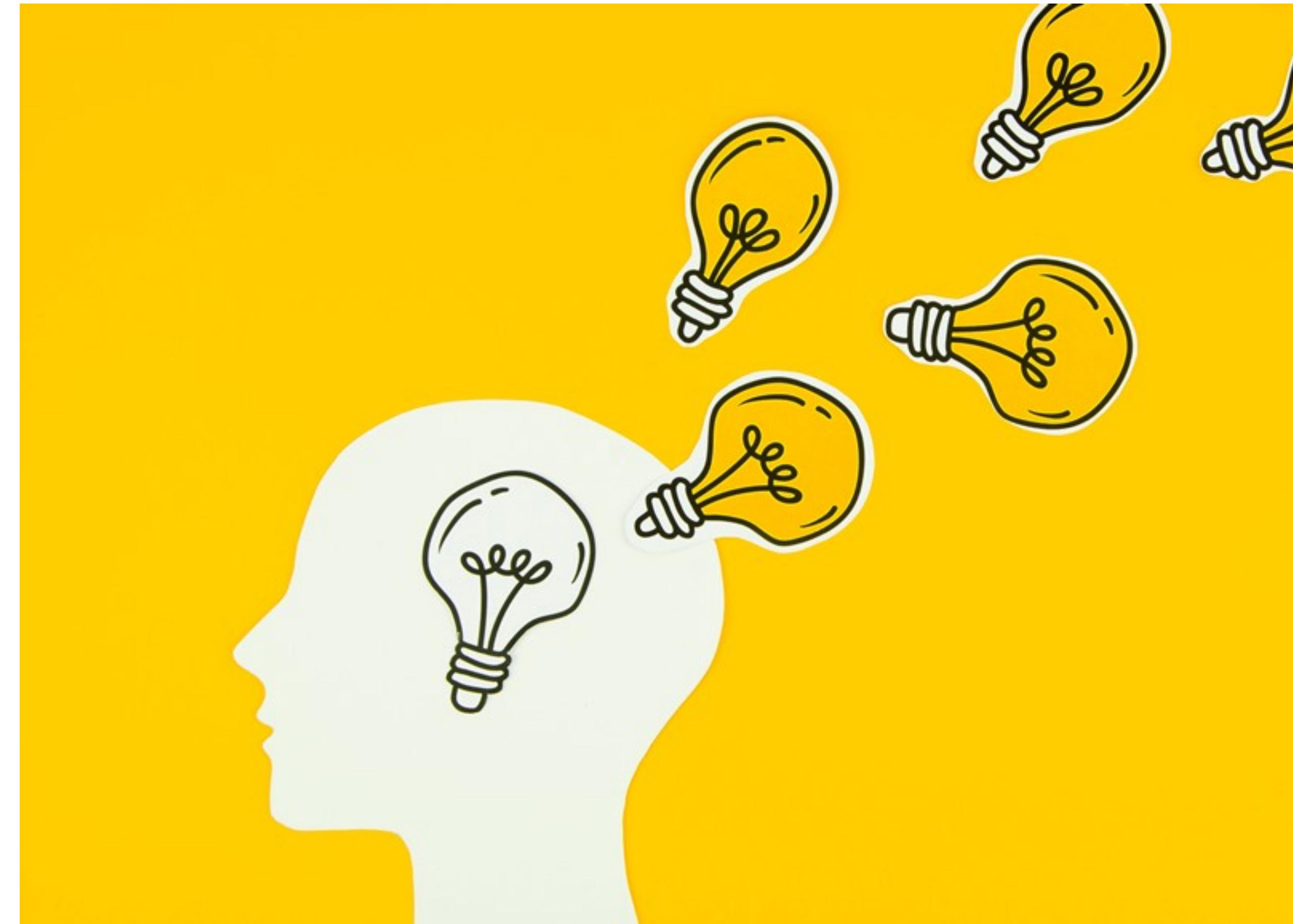
**FWA registration:** FWA00004630

**OHRP registration:** IRB00000356

**Institution Organization:** IORG0000211

# How to generate good ideas?

To have a good idea, you  
need to first have a lot of  
ideas!



# Example Project Ideas from Past Semesters

- Examining User Disclosure Behavior Under Persuasive Conversations with LLM-based Conversational Agents
- Gamification of Privacy Policies
- Investigating StudentWorker Understandings of University Data Collection
- Choice Manipulation Tactics in Corporate Discourse

# Example Project Ideas from Past Semesters

- Investigating Language Choice and Trust in LLM interactions
- GPT-based identity tracker
- Privacy nudging and human-LLM interactions
- Navigating digital identity: username choices; privacy perceptions, and cross-platform identity management
- CatSafe: Safe RLHF for categorical privacy adaptation for LLMs
- Responsible AI for Youth: A literature review on education approaches
- Privacy-preserving verification in multi-agent systems
- Privacy awareness in self-presentation when using avatars in remote meetings

# Project checkpoint 1: Idea descriptions

- By October 6, you're expected to have conceived a few project ideas. Submit at least two idea descriptions including: 1. motivation and research gaps (optional); 2. research questions; 3. proposed research activities

# Project checkpoint 2: Project proposal

- The class on October 20 will be reserved for the project proposal presentation (**remote participation**)
- Each person should give a 5-minute pitch of your proposal, followed by 5-minute Q&A.
- For an Original Research project, your presentation needs to cover:
  - Background and motivations: Why is it an important problem? What are the research gaps?
  - Research questions and your proposed tasks to answer these questions
- For a Literature Review project, your presentation needs to cover:
  - Defining the topic and the scope of your literature review
  - An initial list of references

# Project checkpoint 3: Final presentation

- The last week's class will be reserved for the final presentation
- Students who choose to do an original research project should give a 15-minute presentation followed by 5-minute Q&A. The presentation should cover:
  - Background, research gaps, motivations of the problem you're tackling
  - Research questions and your proposed tasks to answer these questions
  - Final updates: At this point, you should have already completed the planned activities and obtained substantial results
- Students who choose to do a literature review project don't need to give a presentation, but need to submit a manuscript

# Teams

- We'll use Teams to manage assignments, share resources, send reminders of assignment due dates, and help you connect with other students for the course presentation.

# Action items

- By the end of this class: Make sure you can access Teams
- By this Friday (September 12)
  - Log into your HotCRP account using your northeastern email
  - Bid for the papers you are interested in leading the discussion for
  - Introduce yourself to everyone on Teams
- By next Monday (September 15)
  - Submit the first set of reading commentaries