NO PREP

AI and Your Personal Data

What Companies Know

Includes: Student reading, case study, differentiated questions, teacher key

Name: Period: Date:

AI and Your Personal Data: What Companies Know

The Data That Feeds the Machine

Every interaction you have with a digital device generates data. When you search for something online, the search terms are recorded. When you watch a video, the duration, the pauses, the rewinds are logged. When you use a navigation app, your location, speed, and route are tracked. When you shop online, your purchases, browsing history, and even your cursor movements may be collected. When you use a social media platform, your posts, likes, comments, follows, and the amount of time you spend looking at different types of content are all data points. When you interact with an AI assistant or chatbot, the content of your conversations

Name: **Period:** **Date:**

AI and Your Personal Data: What Companies Know

The Data That Feeds the Machine

Every interaction you have with a digital device generates data. When you search for something online, the search terms are recorded. When you watch a video, the duration, the pauses, the rewinds are logged. When you use a navigation app, your location, speed, and route are tracked. When you shop online, your purchases, browsing history, and even your cursor movements may be collected. When you use a social media platform, your posts, likes, comments, follows, and the amount of time you spend looking at different types of content are all data points. When you interact with an AI assistant or chatbot, the content of your conversations may be stored and used. This data is extraordinarily valuable: not just for improving services, as companies often explain it, but as a commercial asset. The behavioral data you generate helps companies build detailed profiles of who you are, what you want, how you make decisions, and how you are likely to behave in the future. Those profiles are used to target advertising, but they are also increasingly used to power AI systems that make decisions about you in contexts far beyond advertising: your creditworthiness, your insurance risk, your suitability for a job, your likelihood of defaulting on a loan, whether you should receive extra scrutiny from law enforcement.

What AI Systems Actually Collect

AI assistants and chatbots represent a particularly intimate form of data collection. When someone types a question into an AI system, they often share things they might not tell a friend, a doctor, or a family member (health symptoms, relationship problems, financial worries, questions about sexuality, struggles with addiction). The conversational format creates a sense of confidentiality that may not actually exist. OpenAI's terms of service for ChatGPT, as of 2024, state that the company uses

conversation data to train future versions of the model. Users can opt out of some of this data use, but the option is buried in settings and not the default. AI-powered apps in health, mental health, fertility, and personal finance collect even more sensitive data. Period-tracking apps collect menstrual cycle data. Mental health apps collect therapy session notes. Personal finance apps have access to bank accounts. Research has found that many of these apps share data with third parties, including advertisers and data brokers, even when their privacy policies claim otherwise. A 2021 investigation by researchers at Mozilla Foundation found that many popular mental health apps had privacy practices that fell far short of their stated policies.

The Data Broker Ecosystem

Much of the data that AI systems use to build profiles of individuals does not come directly from users' interactions with those AI systems. It comes from data brokers, companies whose entire business model is collecting, aggregating, and selling personal information. Data brokers buy data from apps, retail loyalty programs, public records, social media platforms, and other sources. They combine information from multiple sources to build comprehensive profiles: your name, address, age, estimated income, political affiliation, religion, health conditions, purchase history, and location patterns. All of it is compiled into a single dossier and sold to whoever is willing to pay. Data brokers are largely unregulated in the United States. The Federal Trade Commission (FTC) has called for stronger regulation, but as of 2026, no comprehensive federal data broker law exists. In 2023, a data broker sold location data from a dating app that was used to identify and out the location of a Catholic priest, who subsequently resigned. In 2021, Motherboard reported that data brokers were selling location data of people who visited Planned Parenthood clinics. These cases illustrate that data originally collected for one purpose (matching people with dates, tracking foot traffic at retail locations) can be repurposed in ways that harm individuals, and that the subjects of data collection have little ability to prevent it.

How AI Makes Data More Powerful and More Dangerous

The combination of abundant personal data and AI inference represents something qualitatively new. AI systems can make accurate inferences about characteristics that users have never disclosed. Researchers have demonstrated that large language models and machine learning systems can infer sexual orientation from purchasing

patterns, political beliefs from social media activity, health conditions from search histories, and emotional states from typing patterns and response times. These inferences can be more accurate than self-reports and can be made from data that the individual considers innocuous. This creates what privacy researchers call a "mosaic effect": each individual piece of data seems harmless, but when assembled by AI into a comprehensive profile, the result reveals far more than any individual piece would suggest. A person's location data at various times, combined with their social connections, search history, and financial transactions, can reveal their religion, immigration status, health conditions, and intimate relationships — information they would never voluntarily disclose and that in some contexts could be used to discriminate against or harm them.

Legal Protections and Their Limits

The United States lacks a comprehensive federal privacy law as of 2026. Privacy protections are fragmented across sector-specific laws: HIPAA protects health data in clinical contexts but not health data collected by wellness apps; FERPA protects student records but not data collected by educational technology products; COPPA protects children under 13 but has significant enforcement gaps. Some states, led by California's CPRA and followed by Colorado, Virginia, Connecticut, and others, have enacted general-purpose data privacy laws that give residents rights to access, correct, and delete their data. The European Union's GDPR provides stronger protections and has levied billions in fines against major technology companies, but its enforcement against AI-specific practices is still developing. Advocates argue that meaningful privacy protection requires more than disclosure requirements — it requires limits on what data can be collected in the first place, restrictions on how it can be used, and real enforcement with consequences that are proportionate to the scale of harm. As AI systems become more capable of inferring sensitive information from innocuous data, the stakes of inadequate data protection rise. The choices people make today about which apps to use, which permissions to grant, and which services to trust are being made in an information environment that does not adequately communicate what those choices really mean. One dimension of personal data that is often overlooked is its permanence. Data collected today does not expire. Information about your location, health, political beliefs, and relationships gathered by an AI system or a data broker may be stored indefinitely,

potentially accessible to future uses that cannot be anticipated now. A political affiliation or health condition that carries no consequences today may be consequential in a different political or social climate years from now. Unlike a spoken conversation that fades from memory, digital data persists. The person who collected it may sell it, be acquired by another company, or be hacked, transferring the data into entirely different hands under conditions entirely different from the ones in which it was first gathered.

Key Information: AI and Personal Data

Detail	Information
Types of data collected	Searches, location, purchases, social media activity, app interactions, AI conversations
AI chatbot data use	OpenAI uses conversation data for training by default; users can opt out via buried settings
Data brokers	Companies that collect, aggregate, and sell personal data profiles with minimal regulation
Mosaic effect	Individually harmless data points combined by AI reveal sensitive information never disclosed
AI inference capability	Can infer orientation, political beliefs, health status, emotions from non-sensitive data
U.S. privacy law gap	No comprehensive federal privacy law as of 2026; protections fragmented across sector laws
GDPR	European Union privacy law providing stronger protections than U.S. equivalents
2021 Mozilla investigation	Found many mental health apps shared data with third parties despite privacy policy claims

Case Study: Carlos's Mental Health App

Carlos is 16 years old and has been using a mental health app for three months. The app lets him log his mood, write journal entries, and chat with an AI support bot. He tells the bot things he has not told anyone else — that he is struggling with anxiety, that he has been having intrusive thoughts, that he is worried about his immigration status. Carlos assumes his conversations are private. He has not read the privacy policy — it is 14 pages long. His school counselor mentioned the app as a helpful resource. A classmate who studies technology tells Carlos that the app's privacy policy allows sharing of anonymized data with research partners and analytics companies. "Anonymized doesn't mean private," his classmate explains. "If they know your age, ZIP code, and what you wrote, they can probably figure out it's you." Carlos feels sick. "I thought I was talking to something that didn't judge me," he says. "I didn't think about who else might be listening."

1. Based on the reading, what specific privacy risks apply to Carlos's situation that go beyond what he understood when he started using the app?

2. Carlos's classmate says "anonymized doesn't mean private." Which concept from the reading best explains why this is true, and how does the reading describe the risk it creates?

Multiple Choice

1. What is the "mosaic effect" as the reading describes it?

a) A method that data brokers use to organize profiles by geographic region and demographic category levels

b) When individually harmless data points are combined by AI to reveal sensitive undisclosed information

c) The visual pattern that AI systems use to identify faces in surveillance footage

d) The fragmented structure of U.S. privacy law across different sectors

2. According to the reading, what did a 2023 case involving a data broker and a dating app demonstrate?

a) Data collected for one purpose can be repurposed to harm individuals in unexpected ways

b) Data brokers can only sell data when users give explicit written consent

c) Dating apps are exempt from data privacy protections under current U.S. law

d) The Federal Trade Commission successfully prosecuted the data broker under existing law

3. What does the reading say is the limitation of U.S. privacy law as of 2026?

a) All U.S. privacy laws expired in 2023 when Congress failed to renew them

b) American privacy law is stronger than the EU's GDPR in most areas regarding individual data protection

c) U.S. protections are fragmented across sector-specific laws with no comprehensive federal privacy law

d) The U.S. has no privacy laws of any kind at either state or federal level

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ Data broker	A) U.S. law protecting health data in clinical contexts but not wellness apps
2. ___ Mosaic effect	B) European Union privacy law levying billions in fines against major technology companies
3. ___ GDPR	C) The ability of AI to determine characteristics a person never disclosed from indirect data signals
4. ___ HIPAA	D) When individually innocuous data points are combined to reveal sensitive undisclosed information
5. ___ Inference	E) A company that collects, aggregates, and sells personal data profiles to paying customers

Fill-in-the-Blank

1. A 2021 investigation by Mozilla Foundation found that many popular _____ health apps shared data with third parties despite claiming otherwise.

2. COPPA protects children under _____ from certain data collection practices but has significant enforcement gaps.

3. AI systems can make accurate inferences about characteristics that users have never disclosed, including from data the individual considers _____ .

Short Answer

Answer each question in 2-4 complete sentences.

1. Explain what makes the combination of AI and personal data qualitatively different from just having large databases of information.

2. What is a data broker, and why does the reading suggest they pose significant privacy risks?

3. The reading says meaningful privacy protection requires more than disclosure requirements. What else does it argue is needed, and why?