

AI WILL ALWAYS NEED YOU:

Privacy and Security



Artificial intelligence (AI) has been around since at least the mid-1950s, but in the last few years (and even weeks) the technology has developed at a rapid rate. It seems like the entire world has its eyes on AI. What can it do, who can it help, who can it hurt, and about one million additional questions, concerns, and insights have folks in both the business and policy worlds activated.

Enter our newest blog series: AI Will Always Need You – an AI-drafted, human-supervised blog series highlighting the pros, cons, and everything in between about the world of artificial intelligence. This means we'll be drafting the questions, and ChatGPT is answering – unedited. We'll discuss how it can be applied across industries and shared responsibility to create ethical, helpful technology. Our last installment covered copyright and IP, but this time around we're diving into how this AI language model addresses privacy and security.

How do artificial intelligence tools like ChatGPT handle user privacy and data security?

AI tools like ChatGPT handle user privacy and data security by employing several strategies. Personal data used for training these models is usually anonymized, which means that personally identifiable information is either removed or altered to prevent linkage to specific individuals.

While OpenAI and other companies might have clear policies, the companies that use those tools may not. Say a business is using a ChatGPT plug-in, its just as essential to ensure the business using the AI is gathering, storing, and processing all of that data in a safe, secure way.

The companies behind these tools, such as OpenAI, have clear policies regarding how long they retain user data and what they use it for. For example, as of my knowledge cut-off in 2021, OpenAI didn't use the data sent via the model for model improvement.

The dramatic irony in this narrative is that governments around the world are working on proposals that would effectively prohibit the use of end-to-end encryption or strong device encryption. ChatGPT's somewhat blithe reference to being able to rely on strong encryption is a reminder of how seriously problematic those proposals are given the rapid development of generative AI.

These organizations also use robust security measures like encryption and access controls to protect data from unauthorized access. They're also required to comply with privacy laws in the jurisdictions they operate, like the European General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA).

While someone may have consented to have their data used in a certain way, there is little to no measure for us to determine if their consent extends to training a large language model (LLM) or AI more generally, making it is nearly impossible to appropriately address the privacy of all potential user data.

Transparent and clear communication about their privacy practices is a crucial aspect of their operation, often provided in the form of comprehensive privacy policies.

At least in the U.S. the term "usually" is a little misleading - only 12 states have some kind of consumer privacy law. And as a rule, the majority of Americans only have federally guaranteed data privacy rights in highly nuanced sector-specific areas like financial privacy, health privacy, and children's online privacy.

Moreover, users usually have certain rights over their data, including the ability to access their data, correct inaccuracies, and sometimes, request for their data to be deleted.

While these methods provide a robust privacy and security framework, it's important to note that practices may continue to evolve with the technology and as such, users are always encouraged to review the most current policies of the AI tools they use.

While these AI models may occasionally generate text or images that resemble existing works, these are coincidences resulting from the vast amount of data they were trained on, not a result of the AI intentionally copying or referencing specific works.

Folks at the FTC are looking into whether or not OpenAI has clear policies as to how it protects data used to train its models is highly contested. For example, the FTC sent a civil investigative demand (CID) last week asking a long list of probing questions as to virtually all of OpenAI's practices associated with ChatGPT (including its privacy and security practices).

What mechanisms does AI employ to protect sensitive data?

AI technologies can protect sensitive data through several mechanisms:

1. **Data Anonymization and Pseudonymization:** AI can remove or alter personally identifiable information (PII) to ensure privacy, preventing data from being linked to specific individuals. Pseudonymization replaces identifiers with pseudonyms.
2. **Data Encryption:** Sensitive data can be encrypted both at rest and in transit, meaning that it's converted into a format that's unreadable without a decryption key.
3. **Access Controls:** AI can enforce strict rules on who can access certain data, which helps protect sensitive information. This can include role-based access control (RBAC) and attribute-based access control (ABAC).
4. **Intrusion Detection Systems:** AI can help monitor and detect any unusual activities or behaviors that might suggest a data breach or other security threat.
5. **Secure Multi-party Computation (SMPC):** This is a cryptographic protocol for distributing a computation process across multiple parties where no single party has access to all the data.
6. **Differential Privacy:** This mathematical technique adds noise to data queries to ensure that statistical analyses can't identify individual entries in the dataset.
7. **Federated Learning:** This is a machine learning approach where an AI model is trained across multiple devices or servers holding local data samples and no data is exchanged or centralized, which protects the privacy of individual data points.

What ChatGPT doesn't say is that LLMs often store data and information indefinitely. Meaning a user could inadvertently communicate sensitive data to the AI tool, years down the road that tool is breached, and that sensitive data is exposed.

These techniques are designed to balance the utility of AI systems with the need to maintain privacy and security, and they are used in various combinations depending on the specific use case.

What are the risks to privacy and security posed by the use of artificial intelligence?

The use of artificial intelligence poses several risks to privacy and security. A significant concern is data breaches, as AI models often rely on vast datasets, including sensitive information. If not properly secured, these datasets can become targets for malicious activity. Furthermore, AI systems can unintentionally reveal sensitive information if they are trained on such data, known as overfitting.

On the bright side, this can help augment existing efforts to identify and prioritize existing vulnerabilities if used by good actors. Experts have pointed out that AI will greatly augment our cybersecurity defenses and countermeasures, providing necessary tools to combat AI that falls into the wrong hands.

AI also has the potential to enable more sophisticated cyberattacks. For instance, AI could be used to automate and scale phishing attacks or to develop malware that can adapt to its environment.

Another issue is the risk of AI being used for surveillance and profiling. With the capability to analyze and interpret large amounts of data, AI can be used to track individuals, analyze their behavior, and potentially infringe on their privacy.

Finally, decisions made by AI systems could unintentionally incorporate and perpetuate bias, leading to unfair outcomes. Although this is more of a fairness issue than a privacy and security risk, it does relate to the broader theme of ethical concerns in AI.

An AI application often employed in the app economy are companies of all sizes using ChatGPT or other LLMs to write code for their websites, apps, etc. Knowing what we know about LLMs sourcing their info, the code created by AI could very well include security flaws. This could open both businesses that use LLMs and their customers and clients to security flaws.

It's important to note that while these risks are real, many strategies can be used to mitigate them, and they are an active area of research in the AI community.

How is consent managed when AI systems collect and process personal data?

Consent management when AI systems collect and process personal data largely depends on the regulations of the jurisdiction, like the European General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA). Essentially, it's vital that users are clearly informed about what data will be collected, how it will be used, and who it will be shared with before the collection begins. This is often accomplished through a detailed and understandable privacy policy.

How often are these policies actually understandable to the user? "Detailed and understandable" are often mutually exclusive when it comes to privacy policies. Good consumer privacy is about communication and context.

Consent should be explicitly given by the user, typically in the form of an opt-in rather than an opt-out system. The user should have the right to withdraw their consent at any time and have clear instructions on how to do so. Additionally, consent must be freely given, meaning the user shouldn't feel pressured or be subjected to negative consequences for not providing consent.

Several state laws also include a right to object to inclusion in “automated decision-making” (AI algorithms). Honoring those requests could prove quite difficult if the data has already been used to train an algorithm by the time the request is made, or if the data is no longer practically identifiable but fails to meet a state law’s definition of “anonymized” or “de-identified.”

For children or people under a certain age, additional protections are often in place, requiring parental or guardian consent in most cases. Companies need to implement mechanisms for data rectification, where users can correct their information, and data erasure, also known as the “right to be forgotten,” where users can request their data to be deleted.

These principles ensure that personal data collection and processing are done ethically and transparently while providing users control over their data.

In what ways can AI potentially infringe upon personal privacy if not properly managed?

Artificial intelligence, without proper management, can potentially infringe upon personal privacy. For one, AI systems often need large amounts of data to function effectively. If sensitive personal information is gathered without appropriate consent or protective measures, it can lead to privacy concerns.

There are countless debates about what it means to properly anonymize data, sometimes making it easy for cybercriminals to put the pieces together.

Secondly, even when data is anonymized, AI systems can sometimes re-identify individuals from anonymous data, especially when it is combined with other data sources. This process, known as de-anonymization, can result in privacy breaches.

Something to consider: as we see AI implemented into more use cases, like refining targeting advertising, this will be a critical aspect of privacy, equity, and safety discussions around the use of AI in those (and similar) contexts.

Another risk comes from AI's predictive capabilities. AI can analyze seemingly benign data to infer sensitive information about individuals, such as their health status, sexual orientation, or political beliefs, potentially leading to unwanted disclosure of personal information.

Additionally, the use of AI in surveillance technologies, such as facial recognition systems, can lead to unwarranted intrusions into personal privacy.

Furthermore, poorly managed AI systems might retain data longer than necessary, or they might not provide users with an opportunity to review, correct, or delete their personal data.

This proves the point we made in the second comment: many folks in the U.S. do not actually have the ability to review/delete their personal data. Until we have a national privacy law, that will continue to be the case.

How can AI technologies be better designed to respect and uphold individual privacy?

Designing AI technologies to respect and uphold individual privacy requires a holistic and thoughtful approach. Here are some strategies:

1. Privacy by Design: Incorporate privacy considerations from the very beginning of the AI system development process. Privacy shouldn't be an afterthought; it should be integral to the system design.

2. Data Minimization: Collect and process only the minimum amount of data necessary for a particular function. This reduces the risk of sensitive data being compromised.

This one seems unrealistic in the context of AI given that the whole purpose is to use rather large or robust data caches for the purpose of training the model...and many of the LLMs are general rather than specific to one subject, we're curious to see how this could be implemented.

3. Use of Privacy-Enhancing Technologies: Implement technologies such as differential privacy, homomorphic encryption, or federated learning which can help protect user data while still enabling effective machine learning.

There are definitely diminishing returns to transparency when it comes to AI algorithms. Too much disclosure would be inherently inexplicable (by design, we often don't know exactly how algorithms work) and add little value for consumers.

4. Transparency and Explainability: AI systems should be transparent in their operations, and users should be able to understand how their data is being used and for what purpose. This includes providing clear and accessible privacy policies.

5. User Control over Data: Users should have control over their data, including the ability to access, modify, and delete their data. Consent mechanisms should be clear and easy to use.

6. Regular Auditing and Updating of AI Systems: Regular audits can help ensure that AI systems are functioning as intended and respecting privacy. AI models and systems should also be updated regularly to incorporate advancements in privacy-protection measures.

7. Bias Mitigation: Care should be taken to ensure that AI systems do not reinforce or perpetuate harmful biases, which can infringe on the privacy and dignity of certain groups.

8. Legal Compliance: AI systems should be designed to comply with all applicable privacy laws and regulations, which can vary depending on the region.

In general, protecting privacy in AI is a complex challenge that requires ongoing effort and a multi-faceted approach, considering both technical and ethical aspects.

Stay tuned for part two where we dive into more privacy and AI connections and concerns.