

Putting AI to Work

14

Privacy and Data

Learning Objectives

- Investigate how personal information can be included in AI training datasets and assess the ethical concerns around consent
- Analyze the risks of AI-enabled impersonation and describe how voice, image, or personal data can be exploited
- Explain how individuals can exercise their rights to opt out of data use and evaluate the effectiveness of consent mechanisms
- Apply strategies to safeguard personal and sensitive data when interacting with AI tools

Module 14.1: Personal and Private Training Data

- AI models are trained on massive datasets that are scraped from internet sources without individual consent.
- Public accessibility does not equal permission for the information to be used for the training of AI.
- Data enters training sets via web scraping, public repositories, archived content, and indirect disclosure.
- Personal information can resurface in AI outputs; for example, names in stories, contact info in responses, and replicated artist styles.
- Ethical questions:
 - Who gave consent?
 - What rights do individuals have over their data?

Module 14.1: Ethics in Action

- Using personal data without consent raises serious ethical concerns, even if doing so is technically legal.
- Developers must consider whether they *should* use data, not just whether they *can*.
- Training datasets should be filtered to remove identifying or sensitive information.
- Respecting privacy means prioritizing user trust over technical performance.

Module 14.1: Techie Dive

- LLMs trained on Common Crawl, Reddit, and similar sources rarely distinguish personal from informational content.
- Filtering methods (for example, automated classifiers, regex, and manual review) are imperfect.
- Even anonymized data can be reidentified when combined with other context.
- Emerging privacy-focused architectures may train without storing raw personal data.

Module 14.1: Business Lens

- Companies are liable if the AI tool they use outputs private or copyrighted information.
- Users must vet AI tools, understand data sources, and monitor outputs.
- Privacy violations can damage customer trust and lead to legal consequences.
- GDPR and HIPAA impose strict requirements for the handling of personal data.

Module 14.2: Identity Theft and Misuse

- AI can generate convincing images, voices, and text from limited information.
- AI impersonation is harder to detect than traditional identity theft.
- Risks:
 - Deepfake videos
 - Voice cloning scams
 - Fake job applications
 - Social engineering
- Public photos and videos can be scraped for model training without consent.
- Fake AI content can trick people into trusting fraudulent messages.

Module 14.2: Ethics in Action

- Impersonating someone using AI is unethical and often illegal.
- Developers are implementing consent checks, watermarking, and detection systems.
- Users are responsible for not imitating real people without permission.
- There are questions of trust, authenticity, and digital identity in the modern world.

Module 14.2: Techie Dive

- Voice and image cloning models map visual or audio patterns from samples.
- They can recreate a face or voice with surprising accuracy from short samples.
- Detection tools (for example, deepfake scanners and voice verification) are constantly evolving.
- Watermarking helps identify synthetic media, but its adoption is not universal.

Module 14.2: Business Lens

- Companies must protect brands, employees, and customers from impersonation.
- Users must monitor for unauthorized content, provide public awareness, and verify identities.
- Failing to act leads to fraud, legal consequences, and a loss of trust.
- Implement multifactor authentication and train employees on recognition.

Module 14.3: Opting Out and Consent

- Online content is often used for AI training without clear consent.
- Types of consent:
 - Implied consent (public posting)
 - Informed consent (active agreement)
 - Opt-out systems
- Examples of regional protections are the GDPR (EU) and the CCPA (California), but this protection is not universal.
- Challenges:
 - Vague terms
 - Hidden opt-out mechanisms
 - Late awareness of data use
- "Click to accept" agreements rarely explain AI training implications.

Module 14.3: Ethics in Action

- Ethical AI respects individual rights to privacy and autonomy.
- Developers should provide clear, accessible opt-out mechanisms.
- Consent must be informed and understandable, not buried in legal jargon.
- Companies should avoid using content never meant for training purposes.

Module 14.3: Techie Dive

- Opting out adds content to "do-not-train" lists for future collection.
- It's difficult or impossible to remove data from already-trained models.
- Patterns are embedded in model weights, not stored as discrete data.
- Tools like Google "Remove Personal Info" are steps in the right direction.

Module 14.3: Business Lens

- Companies must navigate consent laws carefully to avoid fines and lawsuits.
- Transparent policies build trust and competitive advantage.
- Always ask: "Do we have the right to use this data?"
- GDPR and CCPA compliance requires proper documentation and processes.

Module 14.4: Protecting Personal Information

- Risks:
 - Accidental disclosure
 - Data retention
 - Phishing
 - Prompt history
 - Third-party apps
- Personal information includes names, addresses, credentials, and identifying details.
- Combining just a few facts (for example, school name + zip code) can identify individuals.
- Protection strategies:
 - Remove identifiers
 - Use placeholders
 - Avoid sensitive documents
- Review outputs before sharing, use privacy settings, and separate work and personal use.

Module 14.4: Ethics in Action

- Protecting personal information is a part of using AI ethically.
- It's the user's responsibility to avoid exposing private details even without warnings.
- Never use AI to share or investigate others' information without consent.
- Always consider the privacy implications for everyone whose data might appear in prompts.

Module 14.4: Techie Dive

- Some tools use zero data retention modes, while others keep logs for safety.
- Check the AI tool's privacy-documentation and data-retention policies before use.
- Encrypted local models are the safest for sensitive use but require technical setup.
- Understanding AI architecture helps a user make informed decisions about tool trust.

Module 14.4: Business Lens

- Establish clear company guidelines about what data is allowed in AI prompts.
- Careless requests can leak client information or violate privacy laws.
- Train employees to think before they prompt as part of onboarding.
- Implement policies that specify approved tools and prohibited data types.

Key Takeaways

- Personal data included in AI training without knowledge or consent raises ethical and legal concerns.
- AI-enabled impersonation through deepfakes and voice cloning poses significant fraud risks.
- Consent mechanisms are often confusing or ineffective with regional disparities in protection.
- Protecting information requires removing identifiers, using placeholders, and understanding retention policies.
- Businesses face legal and reputational risks without proper AI vetting and data policies.
- The line between public and private data is increasingly blurred in the digital age.
- Ethical AI use balances tool benefits with respect for individual privacy and autonomy.
- Both users and developers share responsibility for protecting personal information.