FIPPs and Privacy Foundations

Warren and Brandeis: right to be let alone. FIPPs (HEW 1973):

- No secret record systems.
- Notice of what is collected and how used.
- Prevent secondary use without consent.
- Access and correction.
- Security and reliability (for anyone storing identifiable data).

Modern critique: FIPPs assume static databases where risk comes from direct access to stored records. Modern systems rely on **inference** from machine learning and repeated queries. Each query reveals new information and creates more risk than the raw data itself. Risk also depends on how data was generated since high resolution and behavioral data expose patterns. Inference is disclosive because systems can infer sensitive traits without storing identifiers.

PII and Re-Identification

PII (Personally Identifiable Information) is not a fixed list. Any data that distinguishes an individual is identifying. 87 percent re-identification: ZIP + gender + DOB uniquely identify about 87 percent of US population. Classic attacks: AOL logs. Netflix Prize. MA hospital data (Sweeney). **Scrubbing** PII is not enough. De-identification often fails.

US Privacy Law Landscape

Sectoral patchwork.

- **HIPAA**: only applies to providers, plans, clearinghouses. HIPAA gap: wearables, trackers, apps not covered.
- FERPA. FCRA. COPPA (under 13 parental consent). GLBA. GINA.
- VPPA. DPPA. Privacy Act.

Breach notification: CA first in 2003. All states now. Strong driver of corporate security. (Equifax 2017 breach)

FTC Section 5

Unfair or deceptive practices unlawful.

- Deceptive: statements likely to mislead.
- Unfair: substantial injury. Not avoidable. Not outweighed by benefits.
- FTC is de facto privacy regulator.
- Cases: Uber. BetterHelp. GoodRx. Amazon Prime dark patterns. Facebook 5B fine.

GDPR

Omnibus law. Extraterritorial. Applies to anyone processing data of EU residents or monitoring behavior. Rights: notice. access. correction. erasure. consent. lawful basis. minimization. DPIA required. Cross border transfer limits. Right to be forgotten. Debate: protects privacy or strengthens incumbents. (P: strict rules on data use, clear user rights, major fines, global privacy awareness. I: compliance is expensive, large companies afford lawyers but not small - large dominant)

CCPA and CPRA

CCPA (2020):

- Right to opt out of sale (of personal info).
- Applies to for-profit entities with revenue greater than 25M or selling data of 100k or more consumers or with half revenue from data sales.
- Required link: "Do Not Sell My Personal Information".

CPRA (2023):

- Adds opt out of sharing.
- Prohibits dark patterns.
- Required links: "Do Not Sell or Share My Personal Information" or "Limit the Use of My Sensitive Personal Information" or "My Privacy Choices" with icon.
- GPC (Global Privacy Control): frictionless (Browser setting that sends opt-out signal to websites). If honored, link may not be required.

Spillover and Compliance

Spillover effect: protections apply beyond CA. Companies avoid geofencing due to cost and complexity. Automated compliance:

- Link in HTML but hidden by JS.
- Link visible only after render.
- Need scraping plus rendering.
- Misplaced links. Wrong text. Missing icons.
- OneTrust often misconfigured.

Dark Patterns

Definition: UI designs that manipulate or mislead users for privacy. Three categories: **Obstruction:** friction. extra steps. identity verification loops. CAPTCHAs. asymmetric effort. mail-in forms. account creation. mutually exclusive choices. **Interface Interference:** reduced visibility. hidden links. tiny text. asymmetry (1 click accept vs 3 clicks reject). **Misdirection:** confusing text ("Allow sale" next to "Opt-out of sale"). contradictory toggles. double negatives. confirmshaming. unclear "on" vs "off". About one third of opt-out attempts fail.

Informed Consent

Meaningful consent requires understanding. Dark patterns undermine autonomy. Links to Meeting 1 principles.

AI Privacy Risks

Traditional: breach. unauthorized access. data sale. human reviewer visibility. AI specific:

- Memorization: rare or repeated data stored by model. Extractable via targeted prompts. Weakens transformative claims.
- Prompt injection: adversarial prompts override safeguards or extract data.
- Interdependent privacy: users upload data about others (emails. spreadsheets. client info).
- Anthropomorphization: chatbot feels human. Encourages progressive disclosure.
- Bundled consent: users share data without clear understanding that inputs may train future models.

AI User Mental Models

Magic or Super Searcher: incorrect belief model searches a DB. Stochastic Parrot: model predicts tokens. More accurate. (trained on huge datasets, generates text by predicting the next token (Does not understand search retrieve facts only predicts patterns) Fair game fallacy: users think trading data for access is fair but often unaware of training implications.

AI Dark Patterns

Examples: "Allow me to remember chats to improve answers". Presents benefits. Obscures data retention. Confusing consent flows. Hard to find opt-out.

To disable training, you must disable chat history. To keep chat history, you must stay in training. The "keep history but opt out of training" setting is buried.

Copyright Basics

Protects expression. Not facts. Not ideas. Not systems. Software code is copyrighted. Rights: copy. distribute. derivative works. public display. public performance. Reproduce the work, Distribute the work, Create derivative works, Publicly perform, Publicly display. Unauthorized use = infringement.

Fair Use Test

Four factors:

- Purpose and character: transformative use. New meaning or utility. Commercial use counts against fair use.
- Nature: factual favors fair use. creative disfavored. functional code more favorable.
- Amount: snippet vs whole work. Entire copying counts against but may be justified for interoperability or training.
- Market effect: If the new use replaces the original or harms licensing revenue, it is less likely to be fair use. If it does not substitute the work, it favors fair use.

Google v Oracle

Copied 11500 lines of Java API declaring code. Supreme Court said fair use. Transformative since used to create mobile platform. Different market. Nature functional (ideas and systems). Amount reasonable for compatibility.

AI Training and Fair Use Arguments

For fair use:

- Training learns patterns. Not copying expression.
- Outputs not substitutes for original books or images.
- Analogy to human learning.
- Promotes progress.

Against fair use:

- Training uses full works at scale.
- Memorization shows reproduction.
- Outputs can substitute. Summaries. images.
- Copying from pirate datasets.
- Market harm. Licensing markets exist.

AI Lawsuits

NYT v OpenAI. Authors Guild cases. Getty v Stability AI. Music publishers v Anthropic. Claims: reproduction. substitution. memorization. lost licensing market.

Content Moderation

Section 230:

- Platforms not liable for user content.
- Good Samaritan: protection if moderating in good faith.

Moderation challenges:

- Scale: billions of posts. human review impossible.
- Context: AI fails at detecting satire. counter speech. educational use.
- Legal vs norms moderation: copyright takedowns required by law (DMCA). hate speech and misinformation based on community standards.

Censorship Methods

Filtering: blocking keywords. IPs. domains. **Friction**: slowdown. login. paywalls. captchas. hiding search results. throttling. burying content **Flooding**: propaganda. spam. noise. **Fear**: legal pressure. surveillance. Porous censorship often intentional.

AI Accountability and Liability

Responsibility ambiguous. Developer. deployer. user. fine tuner. **Strict liability:** strong incentives for safety. kills innovation and open source. **Negligence:** liability when ignoring standard of care. Hard to define. **Risk based model:** EU AI Act. High risk requires documentation. testing. oversight. **Open vs closed models:** Open = security through visibility. Auditable. Risks misuse. Closed = security through control. Harder to audit. Concentrates power.

Algorithmic Fairness

Definitions: **statistical parity. equalized odds**: FP = FN. predictive parity (Positive predictions must be equally accurate across groups). calibration (For any score output by the model, it must mean the same thing for each group.). Tradeoffs inevitable. Cannot satisfy all fairness definitions simultaneously.

Assignment 3: DNS Tracking

Unencrypted DNS: ISP sees queries. DNS resolver sees queries. Network intermediaries see queries. Beyond DNS: CDNs. ad networks. analytics. embedded third parties. Many domains belong to same company. Encrypted DNS: ISP cannot see DNS queries. DNS provider sees everything. Some companies still see visits via direct IP connections. Shifts trust. Less visibility for networks. More visibility for DNS provider. Concerns: profiling. targeted ads. sensitive browsing histories. corporate surveillance.

Always Tested

FIPPs. PII pitfalls. spillover. CPRA text. GPC. dark pattern taxonomy. FTC Section 5. fair use arguments both sides. memorization. prompt injection. moderation errors. censorship taxonomy.

High-Yield Exam Questions

Privacy Law and CPRA

Explain spillover effects: why companies outside CA still implement CPRA links. Hard to geofence. Cheaper to apply uniformly. Benefits all users. Explain GPC: browser signal for frictionless opt-out. CPRA requires honoring if implemented. Reduces need for links. Explain exact CPRA compliance: correct link text, correct icon, and GPC acceptance.

FIPPs and PII

Explain why PII is hard to define. Distinguishing info can re-identify. ZIP + DOB + gender re-identifies 87 percent of US population. Explain why de-identification fails. Linkage. inference. auxiliary data.

Dark Patterns

Identify category given an interface. Obstruction = friction. Interface interference = hidden or low-visibility controls. Misdirection = confusing toggles or contradictory text. Explain why a flow violates CPRA dark pattern rules. CPRA requires minimal steps. clear choices. no manipulative design. symmetry for opt-in and opt-out.

Content Moderation

Explain why automated systems struggle: lack of context. sarcasm. nuance. dialects. high false positives and false negatives. Explain Section 230 scope: platforms not liable for user content. Good Samaritan protection for moderation in good faith. Exceptions: federal criminal law. intellectual property.

Censorship

Explain filtering. friction. flooding. fear. Give examples. Explain how filtering can increase access. Misconfigured filters accidentally unblock content. Circumvention spreads. Mirror sites multiply. Known as "porous censorship". Identify which examples are friction vs flooding vs filtering.

Copyright and Fair Use

Give fair use argument (for or against). For: transformative use. new purpose. not substitutive. functional code. human-learning analogy. Against: market harm. full copying. memorization. substitution. Explain Google v Oracle reasoning: transformative purpose. functional API. interoperability. Explain reverse engineering legality for interoperability (Sega v Accolade).

AI Privacy

Explain memorization risk. Extracting training data. Rare sequences stored. Explain interdependent privacy: users share data of others (emails. spreadsheets). Explain why LLM interfaces use dark patterns. Example: "remember chats" framed as quality improvement.

DNS and Tracking (Assignment 3)

Explain entities who see unencrypted DNS: ISP. resolver. intermediaries. Explain who sees encrypted DNS: DNS provider only. traffic endpoints still know visits. Explain privacy tradeoff: encrypted DNS hides from ISP but shifts trust to provider. Some third parties still see traffic via direct IP.

Net Neutrality (older exams)

Explain bright line rules: no blocking. no throttling. no paid prioritization. Explain why congestion-throttling of streaming video is or is not a violation: depends on application-specific discrimination vs reasonable network management. Explain paid peering: improves performance. not a net neutrality violation because it is interconnection, not differential treatment on last-mile network.

Bug Bounty and Disclosure Explain benefits: more eyes. faster discovery. coordinated disclosure improves safety. Explain drawbacks: duplicates. false reports. incentive abuse. Explain why coordinated disclosure period exists.