outrageously
AMBITIOUS

Module 2:
Data Privacy and AI

Duke
PRATT SCHOOL of ENGINEERING

Module 2 Objectives:

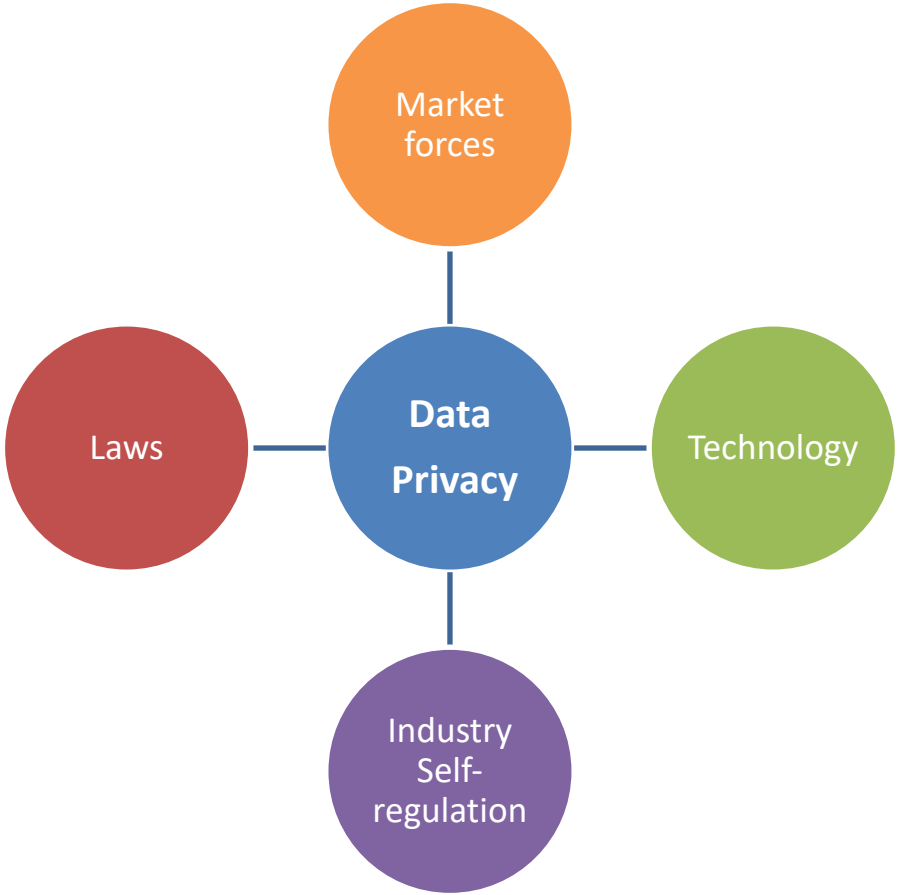
At the conclusion of this module, you should be able to:

1) Implement privacy policies and practices which comply with the Fair Information Practices (FIPS)
2) Identify the key laws related to data privacy in the U.S. & Europe
3) Explain methods to protect user privacy in AI systems


outrageously
AMBITIOUS
Introduction to
Data Privacy
Duke
PRATT SCHOOL of
ENGINEERING

Data privacy

Right of users to have control over how their information is collected, used and shared



What is protected?

- Laws generally cover **Personally Identifiable Information** (PII)
  1. Non-public information
  2. Can be tied back to an "**identified**" or "**identifiable**" person
- A sub-set of personal information is **sensitive information**, which may have additional stricter privacy rules
  - Social security numbers, financial information, medical records

Personally identifiable

- May be directly identifiable or indirectly identifiable:
  - **Directly identifiable:** name, address
  - **Indirectly identifiable:** collection of attributes that can allow re-identification
- Generally excludes:
  - Fully anonymized / de-identified data
  - Aggregated data

Which laws apply

If your organization:

- **Offers online services** (even free) to users in a country
- Or **analyzes or processes data** of users who live in that country

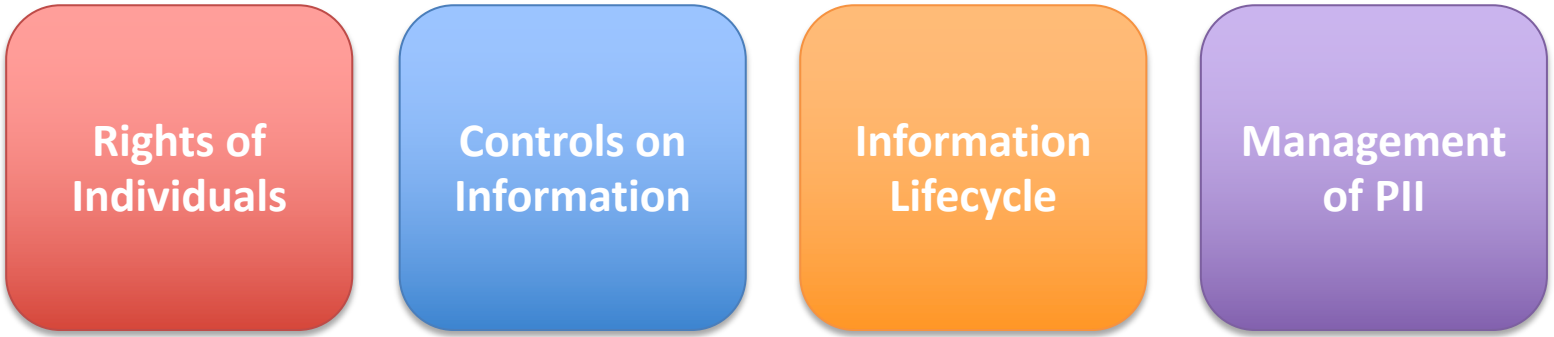
Then regardless of physical location, you must follow the privacy laws of that country

In the US, you must follow both national rules and the rules of individual states


outrageously
AMBITIOUS
Fair Information Practices (FIPs)
Duke
PRATT SCHOOL of ENGINEERING

Fair Information Practices

- Guiding set of principles behind privacy law globally since the 1970s
- The FIPs provide important guidance on **organizational responsibilities** around data privacy
- Organized into four themes:



Rights of Individuals

- **Notice**
  - Provide notice of privacy policies
  - Identify what PII is collected and for what purpose
- **Choice and consent**
  - Describe the choices available to individuals
  - Get explicit or implicit consent
- **Data access**
  - Provide access to their individual PII for review

Controls on information

- **Information security**
  - Use reasonable technical and administrative safeguards to protect PII
- **Information quality**
  - Should maintain accurate and complete PII

Information lifecycle

- **Collection**
  - Should collect PII only as described in the privacy notification
- **Use & retention**
  - Should use PII only for purposes consistent with the privacy notification for which users have provided consent
  - Should retain PII only for as long as necessary to fulfill the stated purpose

Management of PII

- **Accountability**
  - Should document & communicate accountability for privacy procedures
- **Enforcement**
  - Should monitor compliance with privacy policies and have procedures to address complaints from users


outrageously
AMBITIOUS
U.S. Privacy Regulation
Duke
PRATT SCHOOL of
ENGINEERING

U.S. approach to privacy

- Currently no overarching national privacy regulation
- Individual states have passed state laws, with the California Consumer Privacy Act (CCPA) being the strictest
- Regulation on specific types of data exists in several industries:
  - Healthcare
  - Education
  - Financial

Medical data privacy

- **Modern Hippocratic Oath**: “I will respect the privacy of my patients, for their problems are not disclosed to me that the world may know”
- The **Health Insurance Portability and Accountability Act** (HIPAA) of 1996 governs the collection and use of protected health information (PHI) by covered entities and business associates

HIPAA – what it covers

Protected health information (PHI)

- Information that relates to:
  - A past, present or future physical or mental condition of a patient
  - Provision of medical service
  - Payment for healthcare
- Identifies the individual to which it relates
- Created, held or received by a **covered entity** or an **employer**

HIPAA – to whom it applies

Covered Entity

- Healthcare providers – hospital, clinic, etc.
- Health insurance providers
- Healthcare data clearinghouses

Business Associates

- Vendors or contractors who provide services for a covered entity
- Services provided involve the use of protected health information (PHI)

HIPAA requirements

- Requires covered entities to provide **detailed privacy notices**
- Authorizes use of PHI for medical **treatment**, **payment** and **operations**
  - Any other use of PHI require users' opt-in authorization
- Users have the **right to access their PHI**
- Covered entities must implement **safeguards to protect PHI**
  - Must designate a privacy official and ensure monitoring of compliance

Educational data privacy

- The **Family Educational Rights and Privacy Act (FERPA)** of 1974 provides students with control over the disclosure of their educational records
- FERPA applies to all educational institutions that receive federal funding
- Covers all records related to a student maintained by a school or vendor
  - Includes grades, financial information, disciplinary records etc.
  - Does not include directory information

FERPA

- Allows disclosure only under the following conditions:
  - Disclosure is made to the student (if over 18) or parent (if under 18)
  - Consent is provided by the student (if over 18) or parent (if under 18)
  - Information is not personally identifiable
- Requires institutions to provide students the right to access and review their information

Financial Privacy

- Financial information is protected through multiple laws, e.g.:
  - Fair Credit Reporting Act (FCRA) of 1970
  - Fair & Accurate Credit Transactions Act (FACTA) of 2003
  - Gramm-Leach-Bliley Act (GLBA) of 1999
- The financial industry is also subject to many rules requiring disclosure in certain cases e.g. to prevent money laundering

FCRA & GLBA

Fair Credit Reporting Act (1970)

- Regulates the consumer credit reporting industry
- Applies when consumer reporting data is used for offering credit, insurance, and background checks
- Ensures privacy safeguards:
  - Limiting the use of consumer reports to defined permissible purposes
  - Provides consumers the right to access and correct their information
  - Requiring companies to notify consumers when data is used to make adverse decisions

Gramm-Leach-Bliley Act (1999)

- Passage of GLBA in 1999 led to major changes in financial services industry
- **Privacy Rule:**
  - Standard for privacy notices
  - Opt-out of data being shared externally
- **Safeguards Rule:**
  - Required financial institutions to implement a comprehensive information security program
  - Put in place **administrative**, **technical** and **physical** safeguards to ensure confidentiality of customer data

outrageously
AMBITIOUS

E.U. General Data Protection Regulation (GDPR)

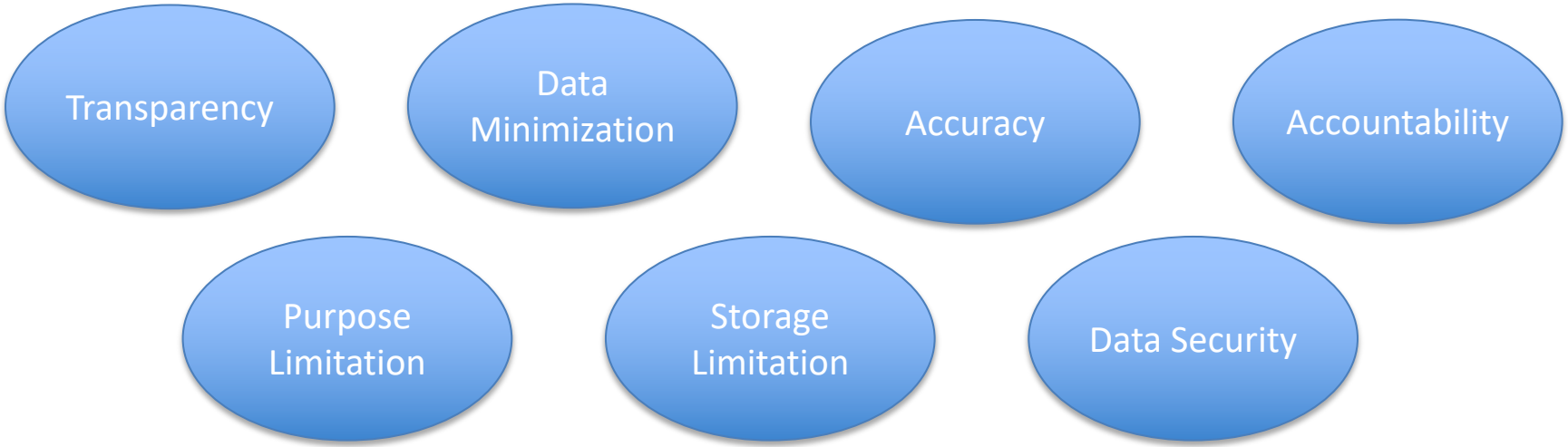
Duke
PRATT SCHOOL of ENGINEERING

GDPR

- EU's General Data Protection Regulation (GDPR) enacted in 2018
- Broad applicability to companies which:
  - Have assets or employees in the EU
  - Sell to users in the EU
  - Store data in the EU
- Applies to data that can be reasonably linked to an identifiable individual
- Fines for violations of GDPR are severe – as much as 4% of **worldwide** revenues

Obligations of data controllers

Seven key principles of GDPR create obligations for organizations who collect and manage PII (“data controllers”)



Rights of individuals

GDPR provides individuals with control over their data through eight rights:

Right to be informed	Right of access	Right to rectify	Right to be forgotten
Right to restrict processing	Right to object	Right to data portability	Right not to be subject to automated decision-making

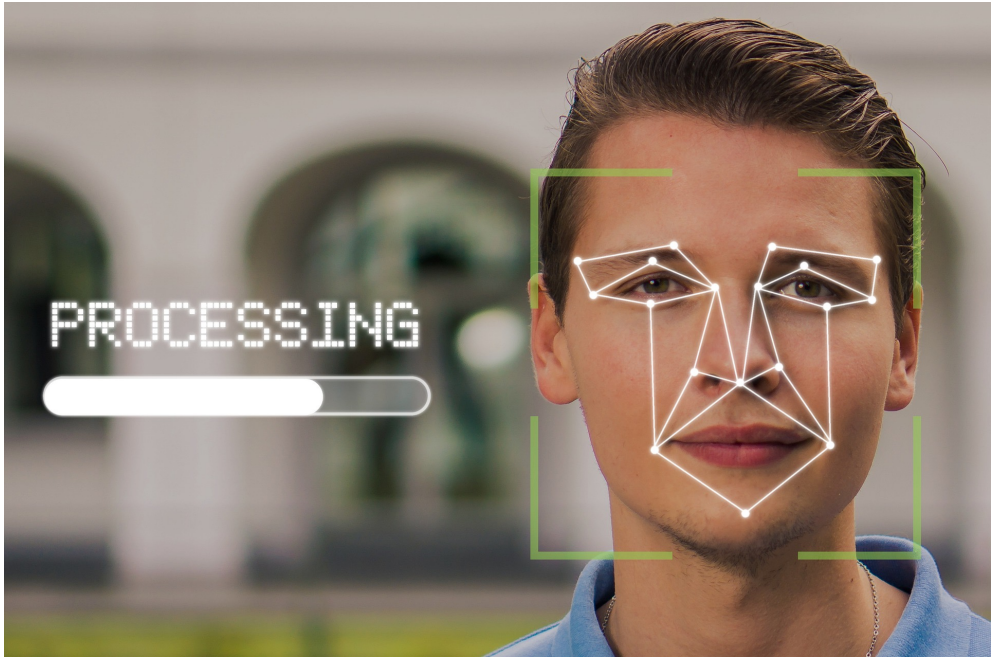

outrageously
AMBITIOUS
Privacy Challenges in AI
Duke
PRATT SCHOOL of
ENGINEERING

AI data needs vs. privacy

- AI data needs often in tension with privacy:
  - AI systems need lots of data
  - Data should also be as feature-rich as possible
  - Difficult to anticipate beforehand which features will be useful
- This can create risks
  - Large-scale data collection without full transparency
  - Ability to infer sensitive data

Police facial recognition

- Over half of all Americans' faces are in a police facial detection database
- Are individuals aware of how their data is being used?
- What laws govern the use of the technology?
- Who audits the systems for accuracy and bias?



Target's pregnancy prediction

"We are very conservative about compliance with all privacy laws. **But even if you're following the law, you can do things where people get queasy**."

- Andrew Pole, Target


outrageously
AMBITIOUS
Protecting Privacy in AI
Duke
PRATT SCHOOL of
ENGINEERING

Why protect user data privacy?

- **Avoid violation of privacy laws**
  - Legal action
  - Significant financial consequences
- **Gain trust of users**
  - Willingness of users to allow you to use their data
- **Maintain reputation**
  - Attract users & employees

How to protect user privacy

There are multiple complimentary approaches to ensuring privacy in AI systems:

- Compliant policy & practices
- Privacy by design
- Technological approaches

Compliant policy

- Privacy policies must comply with all applicable laws
- Policies should clearly state:
  - What is collected
  - For what purpose
  - With whom it is shared
  - Choices available to users
  - Opt-out & contact
- Consent should be **explicit**

Privacy by design

- Best way to mitigate privacy risks is not to create them
- Key principles:
    1. Proactive not reactive
    2. Privacy as the default setting
    3. Privacy embedded into design
    4. End-to-end security
    5. Visibility and transparency
    6. User-centric

Technological approaches

Federated learning

- Allows users/devices to contribute towards improving a shared model without sharing their data
- Device downloads model, updates model using its data, and sends model update back
- Keeps user in control of their data

Differential privacy

- Calculation/modeling approaches where one cannot tell from the output whether any individual's data was included in the input dataset
- Makes it possible to utilize aggregate user data while maintaining privacy of individual users


outrageously
AMBITIOUS
Wrap-up
Duke
PRATT SCHOOL of
ENGINEERING

Wrap-Up

- In building AI systems, data privacy should be considered from the start
- The FIPs provide good guidance for setting privacy policy
- Plan for compliance with the strictest regulation you expect your users to fall under
- Privacy protection involves compliant policies, privacy by design, and technology solutions