# CENG 374E - INTRODUCTION TO COMPUTER SECURITY

**Prof. Dr. Şeref SAĞIROĞLU**
Gazi University
Engineering Faculty
Computer Engineering Department
Maltepe/Ankara

SS@gazi.edu.tr
https://avesis.gazi.edu.tr/ss

# DATA PRIVACY

# Data Privacy and Protection Laws

- Data Privacy and Protection laws refer to legislation that is intended to:
  - protect the right to privacy of individuals
  - ensure that *Personal Data* is used appropriately by organisations that may have
- Personal data is any information that can be used to identify a natural person
  - Name; Phone Number; Email address; etc
- Special Categories of Personal Data require more stringent measures of protection
  - Religion; Ethnicity; Medical information; Criminal Data; Children's Data

# KVKK

Personal Data Protection Board consists of nine members. The selection and appointment process of the Board members was completed at the end of 2016 and the Board started its duty on **January 12, 2017** when the members took the oath in the Court of Cassation Board of First Presidency.

The mission of the Authority is to provide the protection of personal data and develop awareness in this respect in the public eye in line with the fundamental rights related with privacy and freedom stated in the Constitution, as well as to establish an environment to enhance the capability of competition of the public and private organizations in the world of data-driven economy.

# KVKK

Our goal is to be influential in arising the public awareness related with personal data protection and to be a globally accepted authority in this area.

**The Law on the Protection of Personal Data No. 6698** was published in the Official Gazette on **7 April 2016** and 29677 numbered entered into force. **Turkish Data Protection Authority** was established under the same Law in Ankara.

Turkish Data Protection Authority has been established as an independent regulatory authority having organisational and financial autonomy and having a public legal entity in order to fulfill the duties under the Law. The Authority is composed of the Personal Data Protection Board and the Presidency.

02.06.2022

# KVKK

**ARTICLE 2** – (2) The provisions of this Law shall apply to natural persons whose personal data are processed and to natural or legal persons processing such data wholly or partially by automated means or by non-automated means which provided that form part of a data filing system.

**Definitions**

**ARTICLE 3** – (1) For the purposes of this Law:

a. "Explicit consent" means freely given, specific and informed consent,
b. "Anonymization" means rendering personal data impossible to link with an identified or identifiable natural person, even through matching them with other data,
c. "President" means President of the Personal Data Protection Authority,

(ç) "Data subject" (natural person concerned) means the natural person, whose personal data are processed,

d. "Personal data" means any information relating to an identified or identifiable natural person,
e. "Processing of personal data" means any operation which is performed on personal data, wholly or partially by automated means or non-automated means which provided that form part of a data filing system, such as collection, recording, storage, protection, alteration, adaptation, disclosure, transfer, retrieval, making available for collection, categorization, preventing the use thereof,
f. "Board" means the Personal Data Protection Board,
g. "Authority" means the Personal Data Protection Authority,

(ğ) "Data Processor" means the natural or legal person who processes personal data on behalf of the data controller upon its authorization,

h. "Data filing system" means the system where personal data are processed by being structured according to specific criteria,

(ı) "Data Controller" means the natural or legal person who determines the purposes and means of processing personal data and is responsible for the establishment and management of the data filing system.

# Legal Views on Privacy

- Privacy is a fundamental human right that has become one of the most important rights of the modern age
- Each country has a provision for rights of inviolability of the home and secrecy of communications
- Example: In Saudi Law
  - Article 40:
    *"The privacy of telegraphic and postal communications, and telephone and other means of communication, shall be inviolate. There shall be no confiscation, delay, surveillance or eavesdropping, except in cases provided by the Law."*

# Landscape of Privacy Laws

- Two types of privacy laws

1. Comprehensive Laws: General laws that govern the collection, use and dissemination of personal information by public & private sectors
   - Require commissioners or independent enforcement body
   - Difficulty: lack of resources for oversight and enforcement; agencies under government control
2. Sectoral Laws: Avoid general laws, focus on specific sectors instead
   - Advantage: enforcement through a range of mechanisms
   - Disadvantage: each new technology requires new legislation

# Comprehensive Laws In EU

- European Union Council adopted the Privacy Electronic Communications Directive
  - Prohibits secondary uses of data without informed consent
  - No transfer of data to non EU countries unless there is adequate privacy protection

# Sectoral Laws in US

- No explicit right to privacy in the constitution
- A patchwork of federal laws for specific categories of personal information
  - E.g., financial reports, credit reports, video rentals, etc.
- Wide belief that self-regulation is enough and that no new laws are needed (exception: medical records)

# Privacy Impact Assessments (PIA)

- An evaluation conducted to assess how the adoption of new information policies, the procurement of new computer systems, or the initiation of new data collection programs will affect individual privacy

- The premise: Considering privacy issues at the early stages of a project cycle will reduce potential adverse impacts on privacy after it has been implemented

# Privacy Laws Framework

- Most data laws were developed alongside three major concepts that implicate our privacy
  - Media
  - Surveillance
  - Personal data
- The laws revolve around privacy "torts"
  - Intrusion upon seclusion
  - Public disclosure of private facts
  - Misappropriation of name or likeness
  - Placing someone in a false light
  - Negligent handling of people's personal information

# Fair Information Practice Principles (1)

- FIPPS are a set of internationally recognized principles that inform information privacy policies both within government and the private sector
- The principles are
  - Collection Limitation
  - Data quality principle
  - Purpose specification
  - Use limitation principle
  - Security safeguards principle
  - Openness principle
  - Individual participation principle
  - Accountability principle

# General Data Protection Regulations (GDPR)



- The GDPR is new EU legislation that comes into effect on May 25th 2018.
- It very clearly sets out the ways in which the privacy rights of every EU citizen must be protected and the ways in which a person's 'Personal Data' can and can't be used.
- It carries significant penalties for non-compliance
  - €20 Millions, or 4% of the entire global revenue
  - Whichever is higher!

# GDPR Entities

- Three entities are defined in GDPR
  1. **A data subject:** the person whose data is collected
  2. **A data controller:** the entity that collects and uses personal data
  3. **A data processor:** the entity that processes data on behalf of the data controller
- Laws and regulations impose different obligations on the controllers and processors
- For example,
  - Data controller: a company has a website that collects data on the pages their visitors visit
  - Data processor: Google Analytics

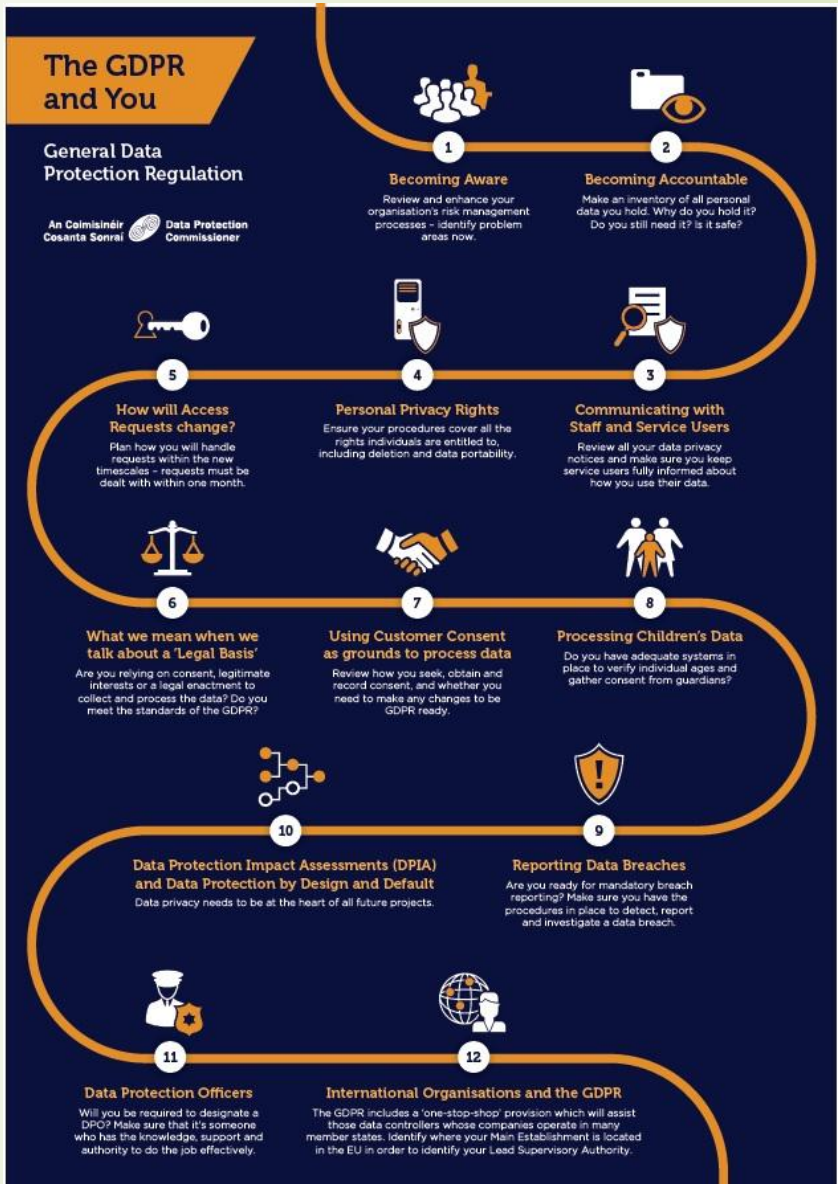
# Seven Principles of Data Protection

1. Lawfulness, Fairness, Transparency
2. Purpose Limitation
   - Use only for one or more specified purposes
3. Data Minimisation
   - Collect only the amount of data required for the specified purpose(s)
4. Accuracy
   - Ensure data is kept up to date, accurate and complete
5. Storage Limitation
   - Kept for no longer than necessary for the specified purpose(s)
6. Integrity and Confidentiality
   - Processed ensuring appropriate security of data
7. Accountability
   - Essential not only to be compliant, but to be able to demonstrate compliance

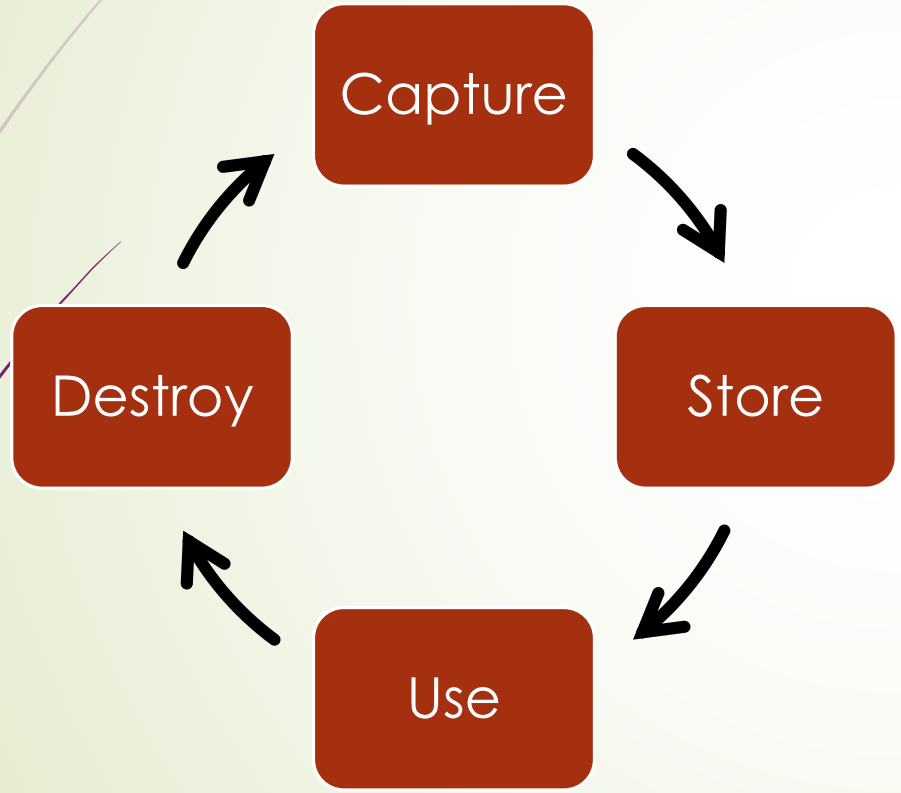
# How to Comply with GDPR?

**The Data Protection Commissioner has issued a guide to compliance, consisting of 12 steps.**

1. Becoming Aware
2. Becoming Accountable
3. Communication with members
4. Personal Privacy Rights
5. Subject Access Requests
6. Legal Basis
7. Consent
8. Children's Data
9. Reporting Breaches
10. Impact Assessments
11. Data Protection Officers
12. International Organisations



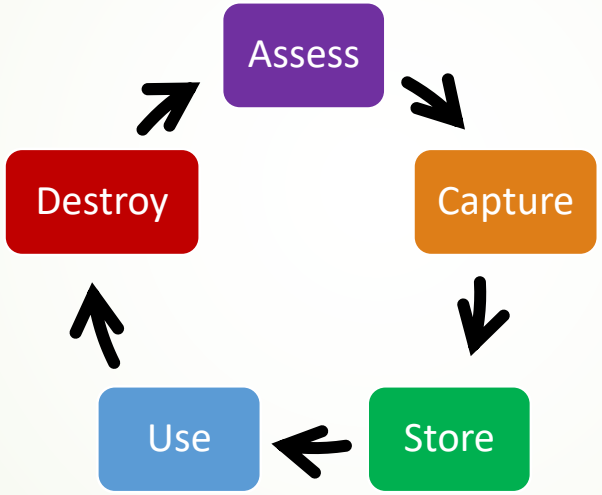
# Information Life Cycle




1. **Capture** Obtain and record information
2. **Store** – Save the information electronically or in paper format
3. **Use** – Use or reuse information
4. **Destroy** – Delete, erase or shred information

# GDPR Information Life Cycle

Data Protection by Design and by Default
Data Protection Impact Assessment (DPIA)
Documentation




Retention Period
Right to erasure
Portability
Third Party copies

Data Minimisation
Privacy Notices
Privacy Rights
Obtain Consent

Appropriate use Consent
Manage Consent
Restricted International Transfers

Safe and Secure
Restricted Access
Data Inventory
Subject Access Requests
Contracts with Data Processors

# The Seven GDPR Sins

- Seven lethal mistakes when designing a new IT system
  1. Storing data forever
     - Data can take long time to be completely deleted
  2. Reusing data indiscriminately
     - E.g. Google used user's data for ad personalization
  3. Walled gardens and black markets
     - Ability to download your personal data instantly
     - Third-party ad companies were blocked from accessing data
  4. Risk-agnostic data processing
     - *"Unless you are breaking stuff, you are not fast enough"*
  5. Hiding data breaches
  6. Making unexplainable decisions
  7. Security as secondary goal

# The Seven GDPR Sins

- Seven lethal mistakes when designing a new IT system

8. Hiding data breaches
   - Prior to GDPR, victims have to check themselves whether they are impacted or not
   - Now, companies must send early notifications to all impacted users
9. Making unexplainable decisions
   - Taking care of privacy when using algorithmic decision making
10. Security as secondary goal
   - Proactive Vs. Reactive security

# Designing GDPR Compliant Systems

- GDPR is intentionally *vague* in terms of technical specifications
- Features for GDPR-Compliant storage systems
  1. Timely deletion
  2. Monitoring and logging
  3. Indexing via metadata
  4. Access control
  5. Encryption
  6. Managing data location

# Examples of Data Laws Breaches

- Marriot International Inc.
  - ~339 million guest records leaked including payment details
  - ~30 million are EU
  - fined **£99,200,396** for the violation
- British Airways
  - ~500K customers information leakes
  - Resulted in a fine of **£183.39 million**.
- Google
  - failing to get valid consent from the users for personalized ads.
  - Google was fined **€50 million**
- Facebook
  - Related to Cambridge
  - Fined **£500,000**
- List of GPDR fines

# Data Privacy Issues in Generative AI

Data privacy in generative AI presents several significant challenges that have become increasingly important as these technologies gain widespread adoption. Here are the key privacy concerns in the current landscape:

**Major Privacy Challenges**

- **1. Training Data Provenance**
- **2. Data Memorization and Extraction**
- **3. User Input Privacy**
- **4. Regulatory Challenges**

# Data Privacy Issues in Generative AI

## 1. Training Data Provenance

- **Unauthorized data usage**: Companies often train models on data scraped from the internet without explicit consent from content creators
- **Copyright implications**: Creative works used in training may be reproduced in similar forms without attribution or compensation
- **Personal information inclusion**: Training datasets may contain personal information that can later be reproduced by the models

## 2. Data Memorization and Extraction

- **Memorization risk**: Large language models can memorize portions of their training data
- **Prompt engineering attacks**: Carefully crafted prompts can sometimes extract verbatim training data
- **PII vulnerabilities**: Names, addresses, and other personally identifiable information may be reproduced from training data

# Data Privacy Issues in Generative AI

## 3. User Input Privacy

- **Conversation logging**: User interactions with AI systems may be stored and used for further training
- **Sensitive information processing**: Users may inadvertently share confidential information with AI systems
- **Cross-contamination**: Information from one user's session might influence responses to other users

# Data Privacy Issues in Generative AI

## 4. Regulatory Challenges

- **Regulatory fragmentation**: Different regions have varying privacy laws (GDPR, CCPA, etc.)
- **Right to be forgotten**: Difficulty in removing specific data points after models are trained
- **Data sovereignty**: Questions about where data is processed and stored across borders
- **Current Approaches to Mitigate Privacy Issues**
- **Technical Solutions**
- **Differential privacy**: Adding noise to training data to protect individual data points
- **Federated learning**: Training models on local devices without transferring raw data
- **Secure enclaves**: Processing sensitive data in isolated computational environments
- **Synthetic data generation**: Creating artificial training data that preserves patterns without using real personal information
- **Governance Approaches**
- **Opt-out mechanisms**: Allowing content creators to exclude their work from training datasets
- **Data provenance tracking**: Documenting sources of training data for transparency
- **Model cards**: Documenting model limitations and potential biases
- **Privacy-by-design**: Incorporating privacy considerations from the earliest stages of AI development

# The Future of Data Privacy in Generative AI

Privacy considerations are becoming central to their development and deployment.

Organizations are increasingly focusing on:

- Creating clear data governance frameworks specific to AI systems
- Developing technical methods to "unlearn" specific data when requested
- Establishing ethical guidelines for data collection and usage
- Building privacy-preserving AI systems that minimize data collection
- The tension between data hunger (models generally perform better with more data) and privacy protection remains one of the central challenges in advancing generative AI in a responsible manner.
- **Would you like me to explore any particular aspect of generative AI privacy in more depth?**

# Privacy vs. Security in Generative AI

**Privacy in Generative AI:** Privacy focuses on protecting personal data and ensuring appropriate data usage:

- **Key Privacy Concerns:**
- **Data ownership**: Who owns the data used to train models and user interactions
- **Consent management**: Whether proper consent was obtained for data usage
- **Information exposure**: How personal or sensitive information might be revealed
- **User control**: Ability of individuals to control how their data is used
- **Data minimization**: Collecting only necessary information
- **Rights preservation**: Maintaining individuals' rights over their personal information

# Privacy vs. Security in Generative AI

**Privacy in Generative AI:** Privacy focuses on protecting personal data and ensuring appropriate data usage:

- **Privacy Violations Examples:**
- A model reproducing private information from its training data
- User conversations being used for purposes beyond what was disclosed
- Models trained on data without proper consent or attribution
- Inability to have personal data removed from training sets

# Privacy vs. Security in Generative AI

**Security in Generative AI:** Security deals with protecting systems from unauthorized access and malicious use:

- **Key Security Concerns:**
- **System integrity**: Protecting AI systems from tampering or manipulation
- **Attack resistance**: Preventing exploitation through adversarial inputs
- **Access control**: Limiting who can access or modify AI systems
- **Vulnerability management**: Identifying and addressing technical weaknesses
- **Prompt injection**: Preventing attacks that manipulate model behavior
- **Infrastructure protection**: Securing the computational environment

# Privacy vs. Security in Generative AI

**Security in Generative AI:** Security deals with protecting systems from unauthorized access and malicious use:

- **Security Violations Examples:**
- Jailbreaking attacks that bypass content filters
- Data poisoning attacks during model training
- Lateral movement from AI systems to other network resources
- Using AI to create more sophisticated phishing or social engineering attacks

# Privacy vs. Security in Generative AI

## Critical Differences

- **Focus Area**
  - **Privacy**: Centered on individual rights and appropriate data usage
  - **Security**: Centered on system protection and preventing exploitation
- **Regulatory Framework**
  - **Privacy**: Governed by laws like GDPR, CCPA, HIPAA
  - **Security**: Often addressed through technical standards and best practices
- **Implementation Approach**
  - **Privacy**: Often requires policy-based solutions and governance frameworks
  - **Security**: Typically relies on technical controls and monitoring systems

# Privacy vs. Security in Generative AI

## Failure Consequences

- **Privacy**: Violations of individual rights, regulatory penalties, trust damage
- **Security**: System compromise, data breaches, operational disruption

## Overlapping Challenges

Several issues span both privacy and security domains:

- **Prompt injection attacks** can be both a security vulnerability and a privacy risk
- **Model inversion attacks** attempt to reconstruct training data (security breach) which may expose private information
- **Data access controls** protect both system integrity (security) and personal information (privacy)
- **Audit trails** serve both security monitoring and privacy compliance purposes

# Conclusions

- More work to be done to ensure the security of personal information for all individuals in all countries
- Technological solutions to protect privacy are implemented to a limited extent only
- Not enough being done to encourage the implementation of technical solutions for privacy compliance and enforcement

# Q&A