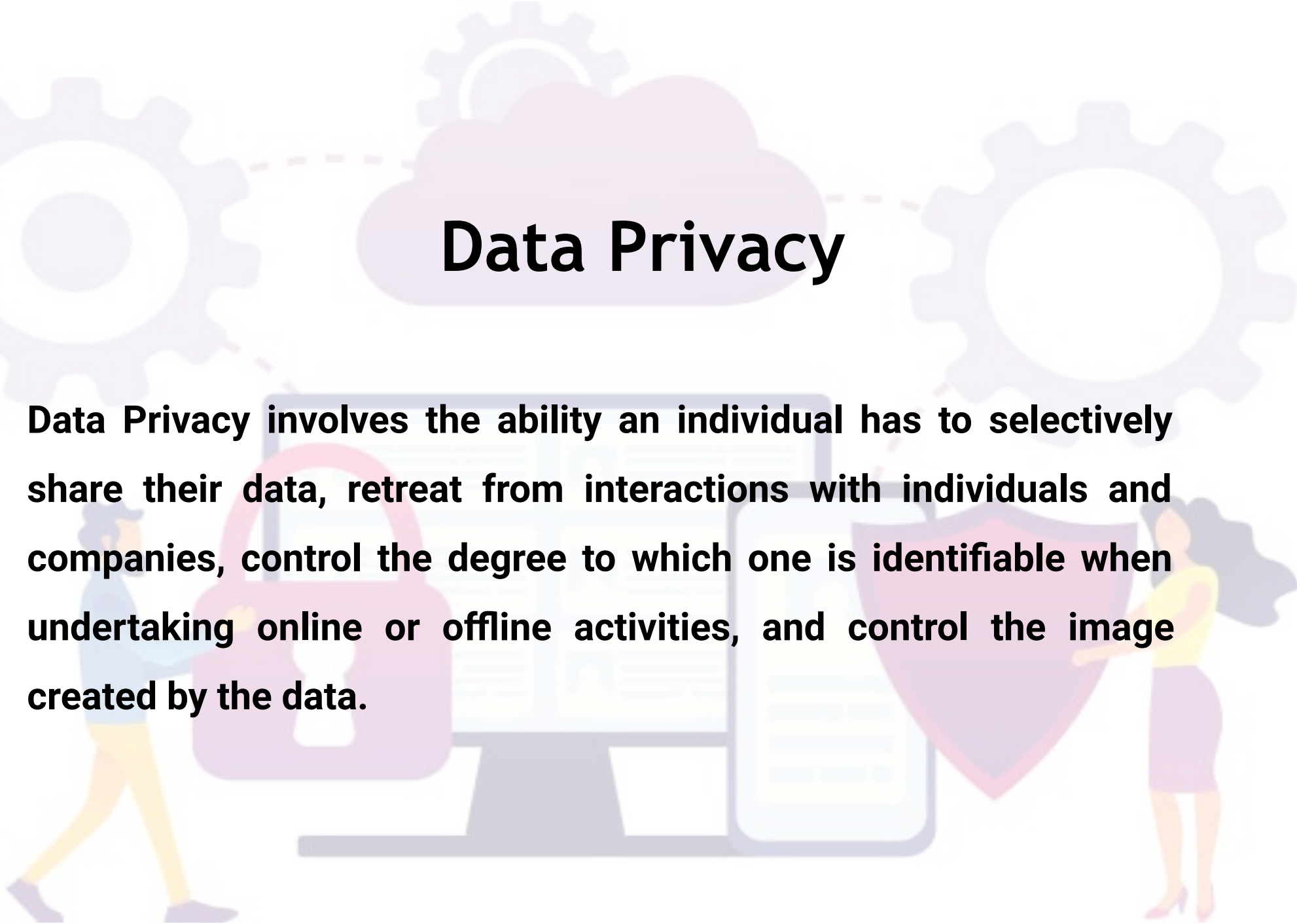
Legal and Ethical Privacy Concepts in Data-Driven Technology

Lecture 10

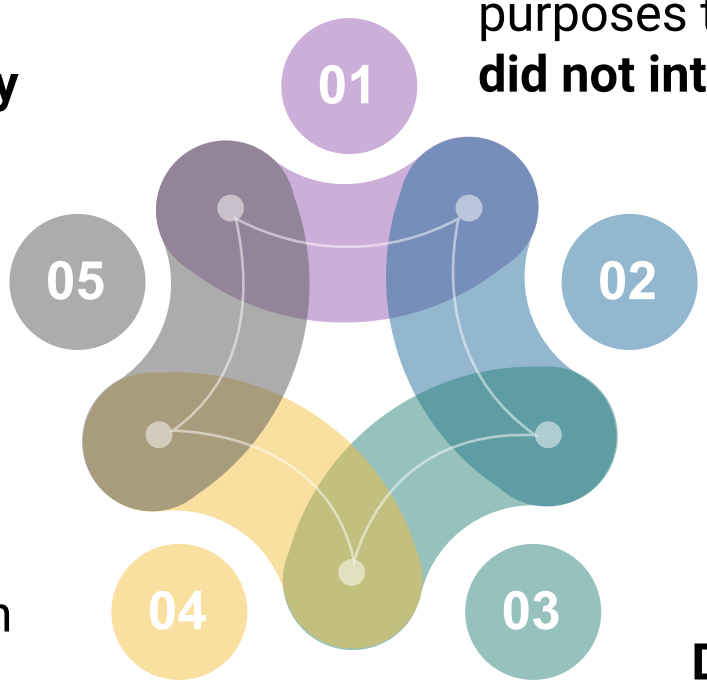
Data Privacy

Data Privacy involves the ability an individual has to selectively share their data, retreat from interactions with individuals and companies, control the degree to which one is identifiable when undertaking online or offline activities, and control the image created by the data.



Privacy Risks Associated with Data Collection by Companies

Data may be used for purposes that the individual **did not intend or approve**.



To build **detailed profiles of individuals**, for targeted advertising or other uncomfortable purposes.

Data breaches and hacks can put personal information at risk

Sell or share data with third-party partners.

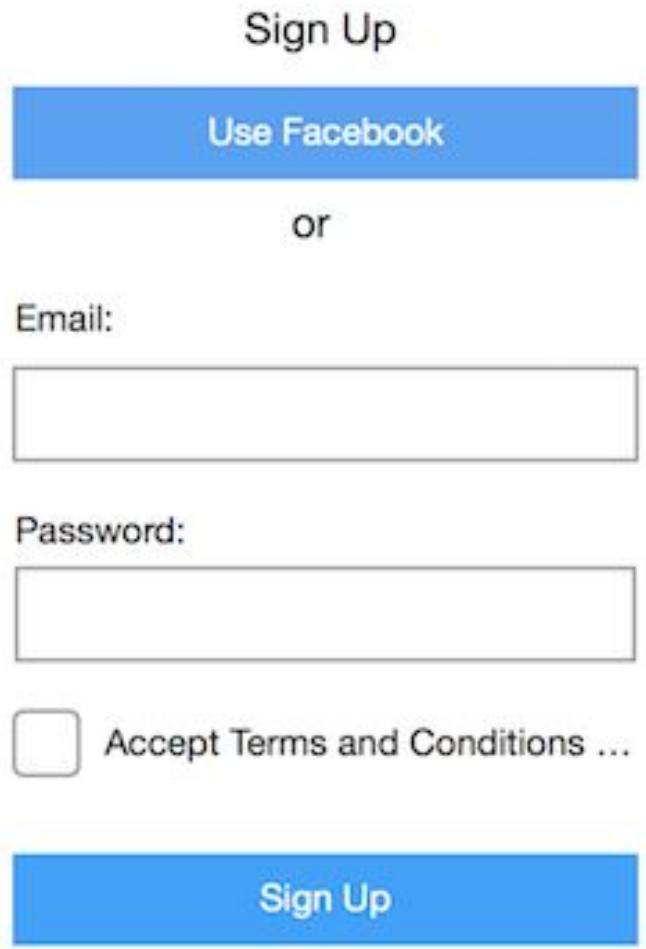
Could be used for **discriminatory purposes**, such as making decisions about employment or credit based on personal characteristics.

The Privacy Paradox

The term "privacy paradox" describes the **contradiction between privacy-related attitudes and behavior.**

We know our privacy is being exploited, and we feel like we're losing control of our data, yet statistics show we're not changing our online behavior to claim it back. Instead, we keep feeding data-collection machines that undermine our privacy.

What is your Password?



Main Contributing Factors to The Privacy Paradox

01 We instinctively perform a **cost-benefit analysis**, weighing the risk of privacy loss against convenience gained.

02 Our privacy cost-benefit analysis tips heavily toward **convenience** when it comes to privacy policies.

03 Privacy policies are, essentially, **written in a style and length that deter people** from reading them.

04 We continue to opt for the **fast and easy option** of ticking and clicking away our privacy.

Is Privacy Paradox a Major Problem?

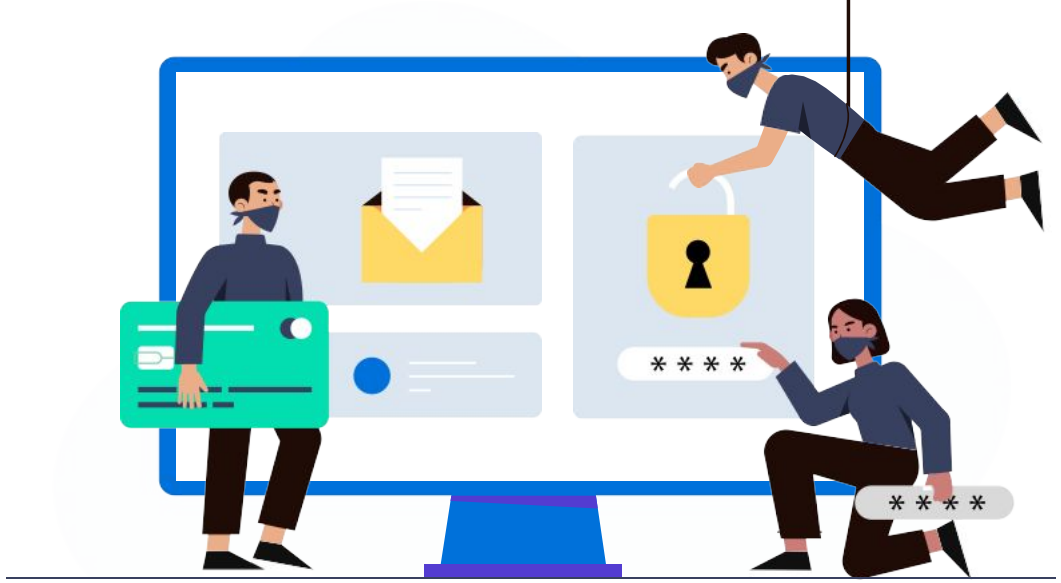
01 Companies will continue to track, gather and sell more and more information.

02 Creates an ideal opportunity for exploitation, creating opportunities for coercion, for our thoughts and behavior to be controlled.

03 They can predict our browsing and buying habits, our political leanings and affiliations.

How to Resolve the Privacy Paradox?

- Society-wide effort – Legislation and government policies, such as GDPR and CCPA, need to be strengthened globally to amplify individuals' rights and regulate the tech industry.
- Tech companies should prioritize user preferences by designing products with privacy in mind, including default privacy settings and clear, understandable policies.
- Terms of service, privacy policies, cookie and other notices must be written in ways that are clear, easy to understand, and allow us to give consent that is truly informed



Further Reading

Privacy Risks in Ambient Intelligence

- Refers to a network of IoT devices that **collect real-time data** to provide personalized services.
- Ambient Intelligence systems capture a **vast amount of information** about individuals within the environment, posing several potential risks to privacy in terms of data collection, transmission, storage, and access.
- Data transmission for most IoT devices is **unencrypted**, which makes it vulnerable to being stolen and used for criminal purposes.
- Improperly stored data **can compromise privacy**, and retention policies should be in place to ensure that data is discarded when it is no longer needed.
- **Example**: Thermostat, Smart TV

Does Facebook sell your data?

Privacy Protection Through Individual Authorization

Informed consent

Consent means voluntarily agreeing to what is occurring.

Right to withdraw

The person must also have the right to withdraw consent at any time as easily as they gave consent in the first place

Clear privacy policies

Often the policy is hard to find, long and full of legal terminology. So even if people can find it, they don't read the entire policy or can't understand the details. One way to counteract this is to have privacy policies modeled after nutrition labels.

Creative commons license

A legal policy written by lawyers, a human-readable layer and a machine-readable version

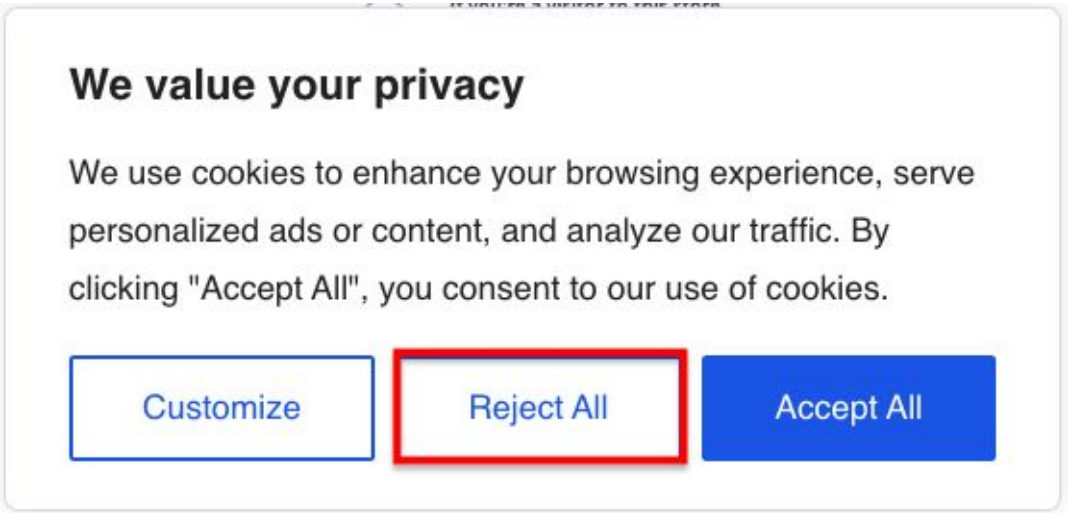
Privacy Protection Through Individual Authorization

Opt out data Policy

In an opt-out data policy, the data will be collected about people unless they specifically say they do not want that. Letting people opt out is better than nothing.

Opt in data policy

No data is collected until a person has given express permission. You often see this implemented with checkboxes that the user can either check or leave empty.



Privacy Protection Through Data Management

Data management strategies can protect privacy by making it difficult to identify individuals from personal data.

De-identification is the process used to prevent unique identity being revealed in a database.

Preventing identification requires careful consideration of **both direct and indirect** identifiers. **Direct identifiers** can be used alone to recognize a specific person, while **indirect identifiers** can be combined with other data to determine a unique individual.

Pseudonymization involves replacing direct identifiers such as real names with a temporary code. There is a re-identification risk, but it is minimal. Indirect identifiers remain in the data set, so it might still be possible for data points to be combined to determine individual identities either within the data set, or by linking it to an outside database.

With **anonymization**, direct and indirect identifiers have been removed or mathematically manipulated so that the data can never be re-identified. The personal data is not substituted but destroyed.

Privacy by Design

Privacy by Design is a holistic approach to privacy that integrates privacy considerations into all organizational operations.

There are seven foundational principles in Privacy by Design that should be considered in the design and operation of information and communication technologies.

Privacy by Design

01 The first principle is **proactive prevention**, which focuses on preventing privacy risks from ever occurring rather than reacting to them.

02 Privacy is the **default setting**, which means that privacy protection is automatically embedded in system or business practices without requiring any action from the user.

03 Privacy **embedded into design**, which emphasizes that privacy protection should be an essential component of technical systems and business practices, not an add-on.

04 The fourth principle is **full functionality**, **positive-sum, not zero-sum**, which states that Privacy by Design can accommodate all of an organization's objectives.

Privacy by Design

05 The fifth principle is **end-to-end security**, which prioritizes security at all stages of the data lifecycle, from collection to destruction.

06 The sixth principle is **visibility and transparency**, which ensures that all stakeholders are informed about the business practices and technologies involved.

07 The seventh and final principle is **respect for user privacy**, which keeps the user at the core of all decisions to protect their data from misuse.

More Details

Differential Privacy

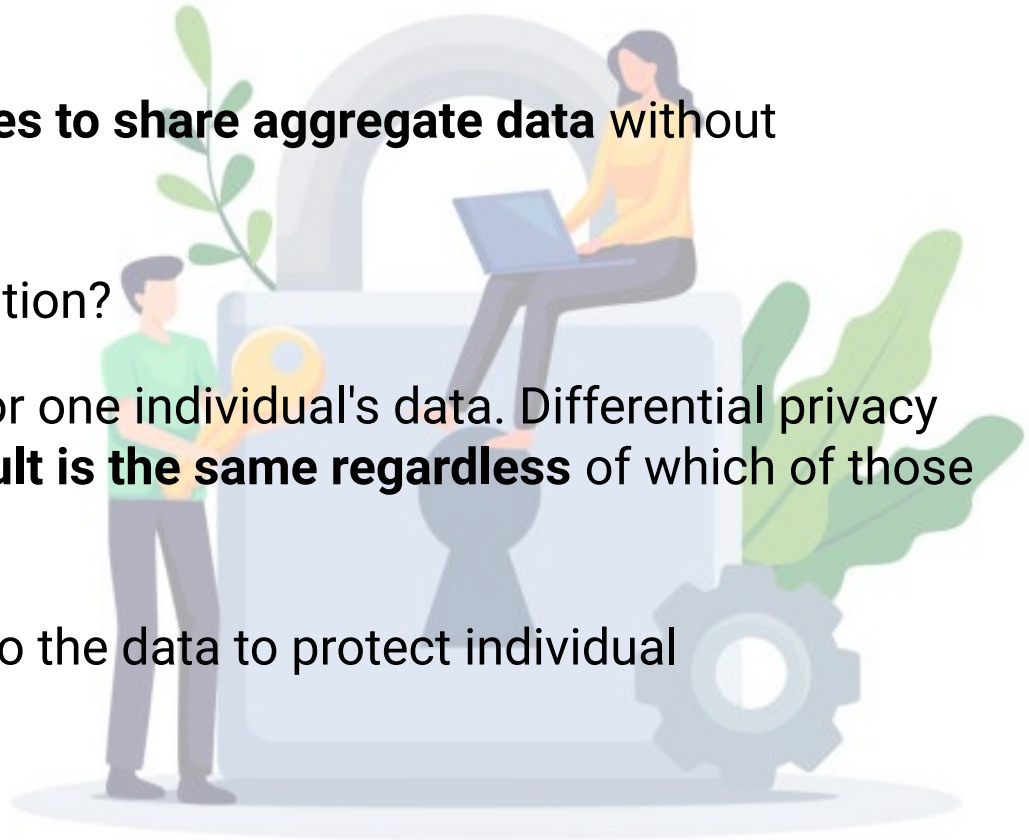
Differential privacy is a mathematical definition of privacy that provides the **probability that a query will not reveal** whether any one person is present in the data, let alone what data belongs to them.

Differential privacy makes it **possible for companies to share aggregate data** without violating the privacy of individuals.

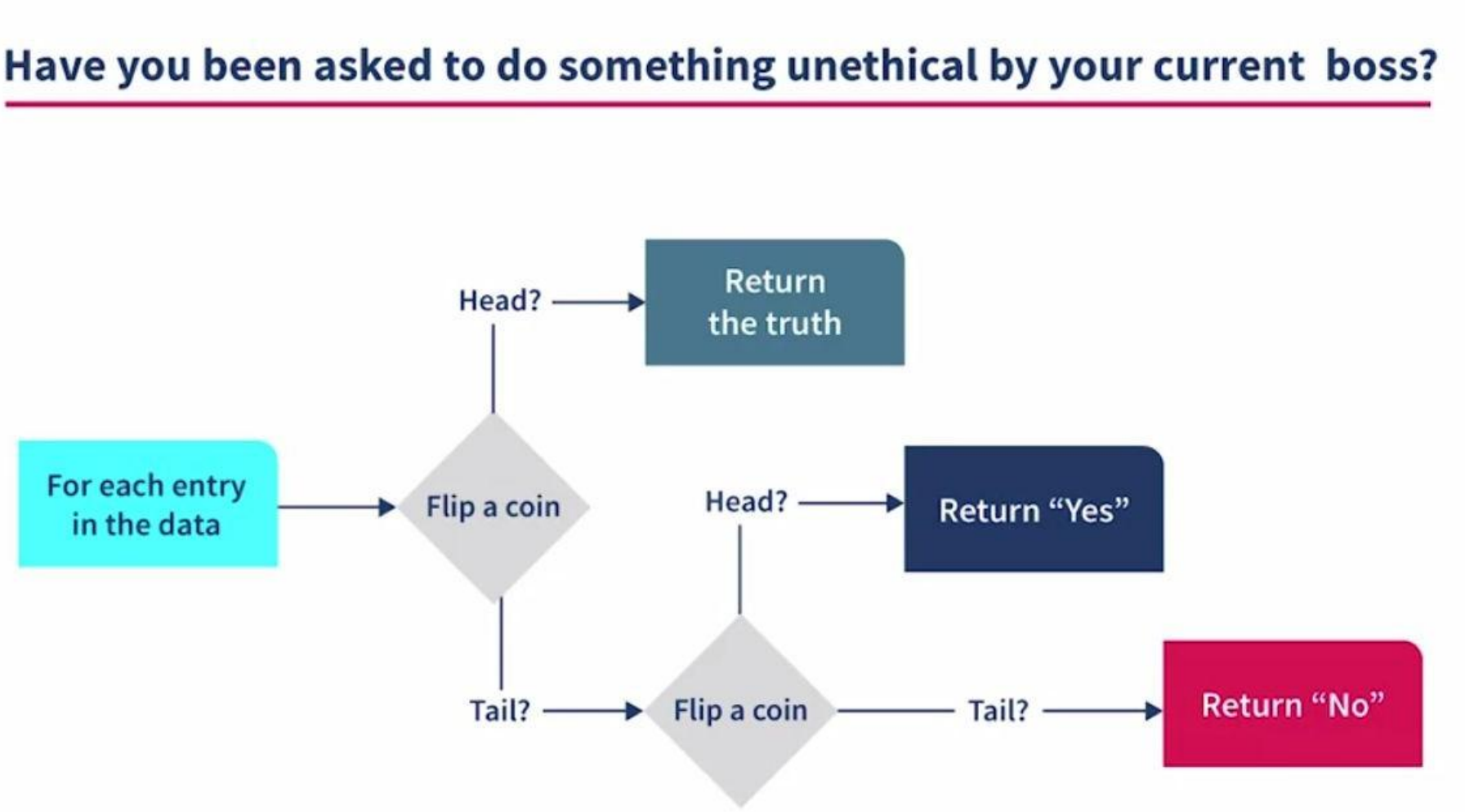
Why it is needed? Encryption, Anonymization, Curation?

Imagine you have two data sets identical except for one individual's data. Differential privacy ensures that the **probability of getting a given result is the same regardless** of which of those two data sets it is run on.

The Curator or Privacy Guard **adds random noise** to the data to protect individual answers and maintain privacy.



Differential Privacy



How Anonymous Is Anonymous?

Anonymizing personal data is an admirable goal for protecting the privacy of that data, but sometimes claims of anonymity can be misleading. A recent study showed that researchers were able to correctly identify 99.98% of individuals in datasets that were considered anonymized.

Read this TechCrunch article about the study to learn how anonymization can be difficult to guarantee, and why users should be skeptical about claims of anonymity.

There is a web app for testing "anonymous" data. You can use this app to see how data you thought was anonymous can actually be used to identify you.

Legal Concepts Related to Data-Driven Technologies

Legal Terminology

Responsibility

Responsibility is the duty to take action. Responsibility can also be shared. Taking responsibility is something you or a group chooses to do, not something an authority assigns to you.

Accountability

Being accountable means that some authority can hold you answerable for your actions after the task or situation is over. Only one person can be accountable.

Liability

Legal responsibility for one's actions and being accountable to the justice system. In addition to the fact you need to explain things to the courts, there is also the possibility of a sanction or punishment.

Technology Contract Types

The purchase and use of computer software whether it is out of a box or cloud based software as a service requires a contract.

01

End User Legal Agreement (EULA)

Contract between **software company and license**. Establishes user's right to use software in specific ways described in license terms. It protects the intellectual property of the software company while allowing users the right to use it.

02

Terms of Service

Legal agreement between **service provider and user**. Designed to **protect provider,** includes disclaimer and limitation of liability. Details user's rights and responsibilities, proper usage, privacy policy etc. If the user fails to follow those rules, the service provider can deny them service.

03

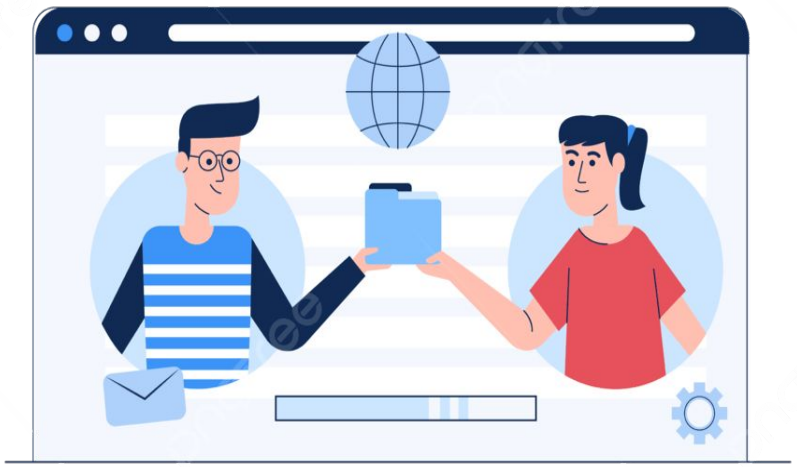
Service Level Agreement (SLA)

Defines **exactly what one party will receive from another party**. Can be external agreement between company and customers or internal agreement between departments. It establishes the goals of both groups and what is needed by both to reach their goals as well as who is accountable for making sure all goals are met.

Smart Contracts

- A smart contract is a **self-executing contract written as code**. A computer program with **if-then loops**.
- It is stored in a **blockchain**, which is a secure distributed network.
- Smart contracts provide the details of an agreement, like traditional contracts.
- Smart contracts are **completely digital** and remove the need for a **central authority**.
- Trusted central mediator can supervise the transaction if there is one.
- Smart contracts **eliminate the need for a central authority**, as the code controls the administration of the transactions.
- Example: A crowdfunding platform

Data Sharing Agreement



A data sharing agreement is a **formal contract** that specifies what data is being shared between two organizations and how that data can be used. It protects the organization providing the data.

It usually includes details about:

- Period of agreement
- Number of parties participating
- Type of data being shared
- Timing and frequency of updates
- Intended use of the data
- Constraints on use of the data
- Confidentiality practices
- Data security practices
- Circumstances for termination

Click-through Agreement

- Click through agreements are **digital prompts** that require an individual to accept or decline a digital policy before they can download something or use an online service.
- They are often used for EULAs, privacy policies, terms of service, and other user policies.
- They are a contract of adhesion, which means it **only benefits one party** and forces the other into a take it or leave it situation. Often the actual policy is on a separate page than the “Agree” button.
- The average person would have to spend 76 working days reading all of the digital privacy policies they agree to in the span of a year. Reading Amazon’s terms and conditions alone out loud takes approximately nine hours.