

# Data and Privacy

---

## Ethics in Computing

Computing is a fast growing and ever changing field that has brought incalculable benefits to the world. Today computers are so ubiquitous, they touch almost every single aspect of our lives, but at what cost? While there is no doubt that computing is changing the way we do everything, it also raises some serious ethical questions that we as humans have never had to deal with before.

In this module we will be looking at data, specifically:

- What do we do with other people's data?
- What do companies do with our data?
- How is data used to shape the world?

# Privacy and Data Protection

## Privacy and data protection

Script Editors: Cathleen Berger, Aditi Gupta, Jonathan Jacobs  
Director/Producer: Meghna Gupta  
Animation and Design: Gat Powell  
Contributors: Cathleen Berger, Ad...

**VIEW ON YOUTUBE >**

## Learning Objectives

At the end of this module you will be able to:

- 1 Differentiate ethics and laws
- 2 Identify ethical problems in computing
- 3 Define copyright and piracy
- 4 Explain algorithmic bias

## Ethics Vs. Laws

Before we dive into those specific ethical dilemmas, we need to understand the difference between ethics and laws.

---

### Ethics

Ethics is the study of right and wrong, often applied to a particular field, like ethics in computing for example. It is a branch of moral philosophy that helps to guide the conduct of people during their day-to-day lives. Furthermore, failure to adhere to a set of ethics is not punishable, although the people around you may not like it!

### Laws

Laws, while they may be informed by ethics, are not a set of guidelines. Laws are a concrete set of rules and regulations made by governments meant to keep people in line. Furthermore, adherence to laws is not optional, and failure to do so is punishable.

The key difference is that laws have to be codified by the government before they take effect, and that takes time. This means that often times ethics outpace laws, especially in a rapidly changing field like computing, leaving a discrepancy between what is legal and what is ethical.

Can you think of something that is legal, but not ethical?

Computing, specifically, raises many ethical concerns, due to the fact that computing is a powerful new tool that allows us to do many things that were simply not possible before its invention. Here are some ethical considerations in computing:

- Vast quantities of data can be stored indefinitely
- Computers can process this data quickly and make inferences
- Computers can make perfect copies of existing data
- Information can be disseminated instantly across the world through the internet
- Computers are being used to make important decisions

Take some time and reflect on how these considerations could cause concerns.

## What Do We Do With Other People's Data

Whenever you do something with a computer, phone or tablet you create data. That data could be a picture, note, text, video, internet history, or even the location that you traveled to. We create lots of data every day; in fact, in 2018, more than 2.5 quintillion bytes of data were created every day! That's 2.5 million terabytes *every day*. To put it into perspective, that's the amount of data needed for over 800 million hours of high-resolution YouTube videos!

That's a staggering amount of "stuff". The question is though, what do we do with all that data? This is where ethics come into the picture. Even though we have access to an unbelievable amount of data to use, the ethical thing is not to use other people's data without their permission in order to respect their privacy.

**Copyright**

Copyright is a set of evolving laws that gives authors and owners the exclusive right to copy and distribute creative work. The idea of copyright emerged to stop the unregulated copying of literary works. Today, copyright law protects the authors of content from having their works stolen and reproduced by others without the owner's permission.

If you create anything, you own the copyright to it, simply by the virtue of creating it. That means you own anything that you create! In the US, you have to register your copyright with the government before you can claim damages in court. If you do decide to register your work, an official US copyright lasts for 70 years after the death of the author if a person copyrighted the item, or 95 years after publication if a corporation copyrighted the item.

## Piracy

Piracy, otherwise known as copyright infringement, is when an unauthorized person uses someone else's creative property without their permission or in an unauthorized way. It's essentially just stealing, but instead of taking physical items, pirates take intellectual property. Piracy is illegal in the US and you can face legal consequences if you use someone else's data. Furthermore, if the person you steal from feels that they lost money, they can sue you for damages!

## Streaming

In the early 2000's, piracy mainly took the form of file-sharing tools like Napster. Napster was a tool where people could post movie and music files for other people to download illegally. Eventually, Napster was shut down, but piracy took on a new form by moving on to a peer-to-peer approach called BitTorrent. BitTorrent is a protocol (method) that allows users to download movies and music illegally, but due to its decentralized nature, it is impossible to shut down.

Fortunately for movie companies and record labels, torrenting files is significantly more difficult than downloading stolen files from Napster which discourages many people from using it. Furthermore, the proliferation of streaming services has further dissuaded the average user from piracy.

Currently, the big question regarding piracy is whether or not sharing paid accounts on sites like Hulu, Netflix, Spotify, and others counts as piracy. In most cases it violates the terms of service, but whether

or not it is criminal is still being figured out.

What do you think?

## Creative Commons

Creative Commons is an organization founded in 2001. It believes that many people want to share data but copyright makes it difficult to get permission. Creative Commons provides licenses that anyone can use to mark data they created as freely usable by others. Depending on the license, Creative Commons lets people specify:

- Whether their work can be changed or remixed
- Whether their work can be used for commercial purposes

Commercial sites like Flickr and Google allow you to search for Creative Commons licensed images. Many sites are dedicated to Creative Commons content, here's a few examples:

- Sounds: <https://freesound.org/>
- Music: <https://freemusicarchive.org/>

## Plagiarism

In an academic setting, a common problem is representing data that belongs to someone else as data *you* created. This misuse of data is often called plagiarism. In many but not all cases, data can be used for educational purposes, even without getting prior permission, as long as credit is given to the creator or owner.

Plagiarism is a serious issue, and the consequences in academic settings reflect that serious nature. Possible consequences of misusing data at school include:

- Lowered grades on assignments

- Suspension or expulsion
- Loss of a degree if the plagiarism is discovered after the degree is granted

## Wikipedia

Although [Wikipedia.org](https://www.wikipedia.org) is only a single, free website, it deserves special attention because of its popularity.

When working on academic projects, it's inappropriate to cite Wikipedia as a source. While Wikipedia does have a very strong moderating system, the problem is that Wikipedia is created by a collection of anonymous volunteers who cannot be credited and have no accountability. However, Wikipedia pages usually list references used to develop each page, and these references are often excellent places to begin further research.

## What Companies Do With Our Data

Not all the data we create is as intentional as taking pictures or writing messages. We can actually unintentionally create data by doing every day things with our devices like visiting websites or using apps. Even our smart phones gather data on where we are and when we use apps. All of this data is extremely valuable for many companies.

Companies use this data for many different applications. A few of them are:

- Creating targeted advertisements
- Selling it to other companies, usually for marketing purposes
- To change how their products work, often to get the users to use a product more

Companies use this data to construct elaborate consumer profiles that build a picture of what we like, who our friends are and how we spend our time. This information allows them to create hyper targeted advertisements and leverage your behavior to drive up user engagement.

## Terms of Service

While all of this feels quite illegal, it's actually completely legal, and in most cases, users unknowingly give their permission for companies to do this.

Websites and apps have terms of service or end user license agreements (EULAs) where users sign away their rights to privacy. Most of the time users do this unknowingly because the privacy information is buried deep within thousands of words worth of legal text. Furthermore, even if you do find the clause, you have to accept the terms, or else you won't be allowed to use the website or app. Often times, merely using a website or app is seen as legally agreeing to the terms.

## What Can You Do About Your Data And Privacy?

Truthfully, your options are limited. Services like Meta (Facebook) and Google are so widely used that it's extremely difficult to avoid them. Furthermore, not having a smartphone in today's society would also be quite difficult.

Fortunately many apps and services have data privacy settings that allow you to reduce the amount of information you share with these companies. Take some time and see if you can find information about your favorite app's privacy settings.

It's important to remember that while the government may be restricted in the data it can collect about you and how it can go about doing so, companies have very few restrictions. This is a case where laws have yet to catch up to ethics. Clearly, most people are very uncomfortable with this invasive tracking, but legally speaking, companies face no repercussions.





"If you are not paying for it, you're not the customer; you're the product being sold"

Andrew Lewis

## How Data Shapes The World

Data itself is neither positive or negative. It's just information. What is positive and negative is what people decide to do with it. That same technology that allows Google to track your every move is the same technology that allows Google Maps to monitor traffic jams and suggest better routes. This is an example of machine learning.

Machine learning is a technology that uses data to model the world, with the model improving as it processes more data. As time goes on and these tools improve, machine learning and other algorithms are being used for more and more important decisions. Here are a few examples:

- Screening resumes for job applicants
- Recognizing faces for security purposes
- Deciding who can borrow money
- Predicting a person's future risk for committing crimes

Depending on who you are this could sound like a technocratic paradise, or a cyberpunk dystopia. If you fall into the latter category, unfortunately, machine learning is still prone to bias. Due to the fact that these programs are trained on data, biased data can lead to biased results. Some examples of this looks like:

- Training data that only represents a portion of the population

- Training an algorithm to mimic historical decisions, which may have been sexist, racist, homophobic, or subject to other biases
- Errors in programming or mathematics

## Addressing Algorithmic Bias

As with privacy, social and legal changes are necessary. It is important to remember that decisions made by computers are really just decisions made by the people who coded or trained the algorithms. Biased people make biased inputs and that leads to biased outputs.

We should also demand more transparency about how such important algorithms work, since their details are often secret. Lastly, advocating for diversity within computer science may help to broaden the perspectives in the field.

## Activities

Here are a few activities that can help you learn more about Data and Privacy:

---

### Activity 1

- Think of one of your favorite movies, books, or songs.
- When was it created, and who holds the copyright for it?

- When will it enter the public domain and no longer be protected by copyright?

## Activity 2

- If you use a streaming service like Netflix, Hulu, Spotify, or Pandora, look up the terms of service.
- Are there any restrictions on sharing the account information with other people?
- What repercussions could you face if you violate the terms of service?

## Activity 3

- Pick a topic you're interested in, whether from history, popular culture, science, or anything else
- Visit the Wikipedia page for that topic, and look through the references at the bottom
- If possible:
  - Borrow a book from the library that's listed in the references
  - Find the original source of some information given on the page

## Activity 4

- Read through the Google Privacy Policy.
- What data does Google collect about you?
- Was there anything that Google collects that you were surprised by?
- Consider reading similar privacy policies at other sites you use frequently.
- Note that some use legalistic language that can be difficult to read.

## Activity 5

- If you use any Google services, consider doing their Privacy Checkup
- This tool allows you some controls over the information that Google keeps
  - How long before your web activity will be deleted?
  - How long before your YouTube history is deleted?
  - What happens to your data if you stop using your account?
- If you use social media services like Facebook, consider exploring their privacy settings

## Activity 6

- Have you done any of the following activities recently?
  - Bought a house or applied for a mortgage
  - Applied for a credit card
  - Applied for a small business loan
  - Applied for a job
- If you have, try to look through the associated documentation or relevant journalism to determine if an algorithm may have played a part in the decision-making process
- Whether or not an algorithm was involved, list any potential sources of bias for the decision in your case

*Lesson 2 of 2*

# Knowledge Check

---

A quick knowledge check on Data and Privacy

---

Question

01/03

Match the terms with the correct definitions

---

☰ Laws

A set of rules laid out by a government that carries punishment if not followed

☰ Ethics

A set of guidelines meant to show people how to conduct themselves

*Question*

**02/03**

Choose all that could cause ethical problems in computing

---

- Computers let us store vast quantities of data indefinitely
- Computers are loud
- Computers let us play video games
- Computers can make perfect copies of someone else's data
- Computers are being used to make important, nuanced decisions

*Question*

**03/03**

Select the best definition of "copyright"

---

- Copyright is a set of laws that protect the authors of content from having their intellectual property stolen or reproduced without their consent
- Copyright is the constitutional right to be able to reproduce any content that you find on the internet
- Copyright is the computational process of the repetition of data on a hard drive