History & Ethics

Ethics of Computer Science

CS51 – Spring 2026

Ethics

- Ethics can be organized in three levels:
  - **Personal ethics:** an individual's moral principles and values guiding their behavior. It's shaped by family, culture, and personal experiences.
  - **Professional ethics:** the ethical standards and principles specific to a particular profession or field. It often includes codes of conduct and guidelines designed to ensure responsible and ethical practice within that profession.
  - **Societal ethics:** the ethical principles that govern how a society functions, including laws, customs, and social norms. It reflects the collective values and expectations of a community or culture.
- When we move from the theoretical concepts of computer science to applying those theories in real life, the decisions we make have consequences.

Activity: ethics for computing professionals

- *What ethics should govern the computing profession?*
- *Do you think that the computing profession has a formal way of codifying its professional ethics?*

ACM Code of Ethics

- ACM has established a Code of Ethics and Professional Conduct (known as "the Code") that expresses the conscience of the computing profession.
- The latest Code was adopted in 2018 and is split into four sections.
  - Section 1: fundamental ethical principles that form the basis for the remainder of the Code.
  - Section 2: additional, more specific considerations of professional responsibility.
  - Section 3: guides individuals who have a leadership role, whether in the workplace or in a volunteer professional capacity.
  - Section 4: principles involving compliance with the Code. Commitment to ethical conduct is required of every ACM member and award recipient.

1. General Ethical Principles

A computing professional should…

1. Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.
2. Avoid harm.
3. Be honest and trustworthy.
4. Be fair and take action not to discriminate.
5. Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.
6. Respect privacy.
7. Honor confidentiality.

2. Professional Responsibilities

A computing professional should...

1. Strive to achieve high quality in both the processes and products of professional work.
2. Maintain high standards of professional competence, conduct, and ethical practice.
3. Know and respect existing rules pertaining to professional work.
4. Accept and provide appropriate professional review.
5. Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.
6. Perform work only in areas of competence.
7. Foster public awareness and understanding of computing, related technologies, and their consequences.
8. Access computing and communication resources only when authorized or when compelled by the public good.
9. Design and implement systems that are robustly and usably secure.

3. Professional Leadership Principles

A computing professional, especially one acting as a leader, should…

1. Ensure that the public good is the central concern during all professional computing work.
2. Articulate, encourage acceptance of, and evaluate fulfillment of social responsibilities by members of the organization or group.
3. Manage personnel and resources to enhance the quality of working life.
4. Articulate, apply, and support policies and processes that reflect the principles of the Code.
5. Create opportunities for members of the organization or group to grow as professionals.
6. Use care when modifying or retiring systems.
7. Recognize and take special care of systems that become integrated into the infrastructure of society.

4. Compliance with the Code

A computing professional should…

1. Uphold, promote, and respect the principles of the Code.
2. Treat violations of the Code as inconsistent with membership in the ACM.

Case study: malware disruption

Before we proceed, some vocabulary:

- **Malware**: software that is intentionally designed to cause disruption to computers, leak private information, gain unauthorized access to computer systems, or deprive access to information.
  - Common subtypes: computer viruses, botnets, **worms**, Trojan horses, ransomware, spyware, etc.
- **Spam**: unsolicited messages sent to a large numbers of recipients for the purpose of commercial advertising, non-commercial proselytizing, or any illegal purpose.
- **ISPs** (Internet Service Providers): organizations that provide services related to accessing the Internet. You might know some: Verizon, Frontier, Spectrum, Cox, AT&T, etc.

Activity: malware disruption

- *Rogue Services* advertised its web hosting services as *“cheap, guaranteed uptime, no matter what.”* While some of Rogue’s clients were legitimate web-based retailers, the majority were focused on **malware** and **spam** and used Rogue Services’ reliability guarantees to protect their illegal operations.
- Despite repeated requests from major ISPs and international organizations, Rogue Services refused to intervene with these services, citing their “no matter what” pledge to their customers. Furthermore, international pressure from other governments failed to induce national-level intervention, as Rogue Services was based in a country whose laws did not adequately proscribe such hosting activities.
- Ultimately, Rogue Services was forcibly taken offline through a coordinated effort from multiple security vendors working with several government organizations. This effort consisted of a targeted worm that spread through Rogue Services’ network. All of Rogue Services’ clients were affected and much of the data stored with the ISP in the process. No other ISPs reported any impact as it was designed to not spread further. As a result of this action, malware circulation decreased.

Analysis: malware disruption – for Rogue Services

- Rogue Services' actions include violations of several principles of the ACM Code of Ethics.
- By allowing for the hosting of malware, they facilitated the harm caused by their clients, violating both Principles 1.1 (Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing) and 1.2 (Avoid Harm).
- Additionally, they were complicit in violating Principle 2.8 (Access computing and communication resources only when authorized or when compelled by the public good), as the ISP was aware that their machines were hosting code that caused infections that were clearly not authorized.
- Finally, Rogue failed to consider the public good, violating Principle 3.1 (Ensure that the public good is the central concern during all professional computing work).

Analysis: malware disruption – for worm authors

- Key nuance of Principle 1.2 (Avoid Harm). Given that the worm was designed to cause harm to Rogue Services' systems, the authors were obligated to ensure the harm was ethically justified. The worm aimed to shut down services that were harmful and malicious, an intent consistent with the moral obligations identified in Principle 1.1 (Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing).
- Additionally, the worm included mechanisms to limit itself solely to Rogue Services' systems, thus demonstrating an attempt to minimize unintended harm. Rogue's retailer clients could rightfully object to the deletion of their data, so a better solution would have included additional precautions to avoid this unintentional harm.
- The worm also highlights the guidance in Principle 2.8 (Access computing and communication resources only when authorized or when compelled by the public good.) The worm accessed Rogue Services' systems without authorization, destroying data in the process. However, the goal of targeting malware demonstrates the service disruption was consistent with the public good.

Professional ethics in our field

- Software engineers don’t have general licensing requirements in contrast to professions like lawyers, doctors, and engineers.
- Being a computer science researcher and analyzing data generally does not require approval by a review board (IRB), unless it includes humans.
- There is no single regulatory body and consequences are limited.
- We will next see examples of Ethics in Computer Science split across four categories: we will mostly focus on algorithmic bias and data collection and privacy, and touch upon AI and autonomous systems, and impact on environment.

Algorithmic Bias

Algorithmic decision-making and bias

- Algorithmic bias describes systematic and repeatable harmful tendency to create unfair outcomes, such as "privileging" one category over another in ways different from the intended function of the algorithm.
- Sometimes, bias is due to the algorithm design, the unintended or unanticipated ways the data were collected and used to train the algorithm.
- Algorithmic bias can reinforce social biases of race/ethnicity, gender, sexuality, ethnicity, age, religion, socioeconomic background, disabilities.
- It can also result to privacy violations.
- Algorithms are often seen as neutral and unbiased which can make them falsely appear as more unbiased than humans and more authoritative.
- We will see a number of examples of bias in public and private settings.

Activity: group work

- Form four groups. Each group will read an article on topics related to algorithmic bias.
- **Trigger warning**: One article mentions suicide
- Set a five minute timer to read through the handout
- Come back as a group and for the next 10 minutes:
  - Summarize the article
  - Were you aware of such examples of algorithmic bias?
  - What surprised you?
  - What can be done?
- Each group will provide an overview of what they read to the rest of the class. What would you want your peers to know about it? Encourage a conversation.

Wrongfully accused by an algorithm

- In January 2020, Robert Julian-Borchak Williams was in his office when he got a call from the Detroit Police Department telling him to come to the station to be arrested. He thought it is a prank.
- He was later arrested at home and drove to a detention center where he had his mug shot, fingerprints and DNA taken, and was held overnight.
- He was accused of shop-lifting five watches from an upscale boutique, being shown a surveillance video as "proof" that it was him who committed the crime.
- Mr. Williams was wrongfully arrested based on a flawed match from a **facial recognition** algorithm.
- Studies have shown that while the technology works relatively well on white men, the results are less accurate for other demographics, partly because of a lack of diversity in the images in the databases.
- Mr. Williams, who is Black, was held in custody for hours although the match with the suspect was obviously wrong. When the case was called, the prosecutor moved to dismiss, but "without prejudice," meaning Mr. Williams could later be charged again.
- NY Times ran a popular story which resulted in the case and fingerprint data expunged.

Machine bias in risk assessments in criminal sentencing

- In 2016, ProPublica released an analysis on how the software COMPAS, Correctional Offender Management Profiling for Alternative Sanctions, assigned recidivism scores.
- Scores like this-known as **risk assessments**-are increasingly common in U.S. courtrooms. They are used to inform guiding judges during criminal sentencing.
- The journalists obtained the COMPAS risk scores assigned to more than 7,000 people in Florida and checked how many were charged with new crimes over the next two years.
- The scores proved remarkably unreliable in forecasting violent crime: Only 20% of the people predicted to commit violent crimes actually went on to do so.
- The formula was particularly likely to falsely flag Black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.
- White defendants were mislabeled as low risk more often than Black defendants.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Search engines

- Social scientist Dr. Safiya Noble showed in her 2018 book, Algorithms of Oppression: How Search Engines Reinforce Racism, examples of algorithmic bias when she used search engines like Google.
    - Terms like "black girls" returned pornography and "Jew" returned anti-Semitic pages.
    - Google claimed it was unable to erase those pages unless they were considered unlawful.
- Google similarly received heat when queries related to occupations were found to propagate sexist stereotypes, e.g., the term CEO or doctor would return pictures with men, while the term nurse would return women.
- Although Google claimed that they fixed the problem, researchers from UW showed that simple tweaks, like quering “CEO + United States” returned fewer photos of cis-female presenting people. Taking a ‘whack-a-mole’ approach doesn’t seem to be fixing the problem.

AI and hiring bias

- Similar biases of racism, sexism, and ageism have been identified in how automatic job application screening tools would rank job applicants' names according to perceived race and gender.
- The boom of large-language models like ChatGPT has led to a "tsunami" of AI-generated resumes that inundate employers.
- Companies have responded with more automation, using AI-ran chats or video interviews.
- But candidates can also use AI to cheat in these interviews.
- Do we live in the AI vs AI era?
- https://www.nytimes.com/2025/10/07/business/ai-chatbot-prompts-resumes.html

AI hallucinations and human hallucinations

- The arrival of ChatGPT and similar tools has transformed not just the hiring process.
- Reasoning systems from OpenAI, Google, and DeepSeek are generating more errors, not fewer and as the sources of data are exhausted, synthetic data are needed to train them.
- Since these AI tools learn from using complex mathematical systems to analyze enormous amounts of digital data and they cannot distinguish what is true or not, they often make things up, a phenomenon known as **AI hallucinations**.
- Beyond the dangers of inventing false information, they also spread conspiracies and become confidants which can harp on people’s mental health when blindly trusting information from seemingly authoritative systems. This is known as **algorithmic appreciation.**

https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html

https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html

https://www.nytimes.com/2025/06/13/technology/chatgpt-ai-chatbots-conspiracies.html

Data Collection and Privacy

Data collection

- Many websites are funded by advertising. They provide content or functionality to consumers for free, and the cost of creating and maintaining the website is covered by advertisers, who pay to have the website show ads to the consumers.
- Advertisers want to show ads only to users who are most likely to buy the product or service.
- Therefore, data about consumers are valuable, because they let advertisers direct ads most effectively.

Profiles

- Advertisers want profiles of consumers. A profile is a collection of facts, including:
  - Demographics: age, gender, race, etc.
  - Market participation: income level, location
  - Preferences: family information, taste in books and movies, favorite restaurants or travel destinations, etc.

Data Economy – Data Collection

- Websites have a strong incentive to get the best data possible on their users, so that they get paid more for advertisements. This has led to **hyper-targeting** in ads, with ads attempting to reach more niche populations.
- Check out the different categories popular websites have identified as relevant for you:
  - Google: https://myadcenter.google.com/home
  - Instagram: https://accountscenter.instagram.com/ad_preferences/ad_topics
  - TikTok: Settings and Privacy > Ads > How your ads are personalized

- *What kinds of data are you comfortable having collected by companies?*
- *Where might you draw a line?*

Third parties

- When we talk about data collection and privacy, we're usually really talking about **data sharing**.
- When a consumer interacts with a company, and a different company learns something about the interaction, we say that the data has been shared with a **third party**.
- This kind of data sharing is easy if both companies know a unique piece of information about the user, like the email address they used when creating an account.
- But people don't always make accounts, and sometimes they use different email addresses.
- This means that the core of data sharing is **tracking**: finding a way to link together all the actions that a person takes on different websites and at different times.

Cookies

- Tracking often relies on **cookies** to identify the people visiting a website. A cookie is a piece of data that a website asks a browser to remember. For example:

1. You visit a website and select the “dark mode” view. Next time you visit it, your browser sends the cookie to the website and it shows you the dark mode view automatically.
2. You visit a shopping website, and, behind the scenes, it assigns you an ID number and tracks which products you looked at. It doesn’t know your name or email address, so it uses the ID in place. Weeks later, you visit the site again and enter your email address to place an order. Your browser sends the cookie, and the company links your email address to the products you looked at the first time you were on the site.
3. **Third-party cookie**: You visit a website that displays an ad. When you visit a different website that displays ads from the same company, your browser sends the cookie, and the ad company updates their records with the new information. If lots of websites show ads from the same ad company, they can see a lot of what you do online.

Fingerprinting

- First-party cookies are usually used for preferences, remembering that a user is logged in, and so on. Third-party cookies are mostly used for tracking, so some browsers allow users to block third-party cookies.
- In response, data aggregators began using **fingerprinting** to track users across websites.
- Fingerprinting is a technique that gathers lots of little pieces of information about the computer visiting a website. With enough little pieces, the website can uniquely identify the computer next time it visits.
- This is possible because your browser makes lots of information about your computer (or phone) visible to the websites you visit.
- This is not done maliciously – services can use this information for good, e.g., knowing the size of your screen lets a website show you the mobile or desktop version of the site.
- Check out the data your browser shares here: https://webkay.robinlinus.com/

Trackers

- **Trackers** are pieces of code created and published by data aggregators: companies that want to collect lots of data on lots of people.
- Other companies use the tracker's code while building their website. The tracker adds some functionality to the website, like letting the company see which pages of their website get the most traffic.
- At the same time, the tracker reports back to the data aggregator that made it with information about the people visiting the website. Trackers can also be included in emails.
- Trackers use a combination of cookies and fingerprinting to identify the website's visitors.

Consent

- Regulators around the world have tried a variety of methods to balance the desire of some consumers for privacy with the desires of companies.
- In the US, a model called **notice and choice** has historically been dominant.
- Under the notice and choice model, a company can legally use and share data about a consumer if:
- They **notify** the consumer, usually in a privacy policy or terms of service document, and
- The consumer **chooses** to agree, either by checking an “I agree” box or by continuing to use the website.

- *When a website shows you a privacy policy or terms of service document, what do you do?*

Data privacy regulation

- In the European Union, the **GDPR** (General Data Protection Regulation) similarly restricts some data sharing and gives consumers additional rights.
- The US has no nationwide data privacy law, but legislators, regulators, and civil society groups have all shown interest in possible future legislation.
- Instead, individual states, about 20 including California which has the **CCPA** (California Consumer Privacy Act), give consumers rights to know about data collection and reject some collection.

Facebook–Cambridge Analytica data scandal

- In the 2010s, personal data belonging to millions of Facebook users was collected by British consulting firm Cambridge Analytica for political advertising without informed consent.
- The data was collected through an app called "This Is Your Digital Life", developed by data scientist Aleksandr Kogan, a data scientist at the University of Cambridge in 2013.
- The app consisted of a series of questions to build psychological profiles on users, and collected the personal data of the users’ Facebook friends via Facebook's Open Graph platform.
- The app harvested the data of up to 87 million Facebook profiles.
- Cambridge Analytica used the data to analytically assist the 2016 presidential campaigns of Ted Cruz and Donald Trump.

https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal

Protecting your data

- If you want to protect your data online, you have options! Most browsers let you block cookies and can request that websites do not track you. You can also restrict permissions given to websites and applications on your devices.
- You can check what kinds of trackers your browser stops and what your fingerprint looks like here: https://coveryourtracks.eff.org/
- You can see what trackers and fingerprinting techniques a website is using by entering it here: https://themarkup.org/blacklight
- One factor that fingerprinting uses is your IP address. You can hide your IP address from the websites you visit using a VPN. Pomona has a VPN (though then Pomona will know which websites you're accessing): https://www.pomona.edu/administration/its/services

AI and autonomous systems

Deepfakes

- **Deepfakes** are media (images, videos, audio) that have been altered or fabricated using AI to appear as they are real.
- They can have some benign applications in entertainment. BUT!
- They have been used to spread disinformation, incite political divisions, in child sexual abuse material, revenge porn, bullying, and financial fraud.
- Some states are moving to ban their malicious use. E.g., https://www.nytimes.com/2025/05/22/business/media/deepfakes-laws-free-speech.html

Advanced Driver Assistance Systems (ADAS)

- **Level 0 - No automation**. Driver has 100% control
- **Level 1 – Driven assistance**. At least one driver support system that provides assistance, for example, adaptive cruise control.
- **Level 2 – Partial automation**. Can take control over steering, acceleration, breaking but driver should remain active and supervise.
- **Level 3 – Conditional automation**. Uses various driver assistance functions and AI to make decisions but driver must be present, alert, and should be able to take over.
- **Level 4 – High automation**. No human supervision is needed. Does not require a steering wheel or pedals and is designed to stop at system failure. Companies like Waymo or Cruise.
- **Level 5 – Full automation.** Vehicle can drive independently without any restriction. Not available yet.
- Level 3-5 are known also as Automated Driving Systems (ADS) and distinguished by 1-2 (ADAS).

Accidents and fatalities (2019-2024)

- Tesla (ADAS) has had the highest number of incidents (2146). Waymo (ADS) follows next with 415. California is the state with the most self-driving incidents, followed by Texas and Arizona, respectively.
- 10% of autonomous vehicle accidents have resulted in injury, and 2% have resulted in a fatality.
- 83 fatalities related to autonomous vehicle accidents as of June 17, 2024. Below are a few select incidents. For example,
  - In 2019, Walter Huang dropped his child off at school and then engaged the autopilot feature of his Tesla Model X. The car veered out of the lane and began to accelerate, crashing into a barrier at 70mph and killing the driver.
  - In March 2018, an Uber self-driving test vehicle struck and killed Elaine Herzberg, a pedestrian, in Tempe, Arizona. This was the first recorded case of a pedestrian fatality involving a fully autonomous vehicle. The backup driver of the vehicle was later charged with negligent homicide.

https://www.craftlawfirm.com/autonomous-vehicle-accidents-2019-2024-crash-data/

Impact on Environment

Mining for technology

- The Interior’s U.S. Geological Survey published a list of 35 mineral commodities considered critical to the economic and national security of the United States.
- These minerals are used virtually in every sector of the economy and type of technology.
- They require mining the Earth, often leading to both environmental and geopolitical problems for the neighboring communities.
- **Conflict minerals**, that is natural resources extracted in conflict zones and then sold to fund the conflict are inextricably linked with the production of computers.
- But even locally, mining can be controversial. For example, lithium mines exist in Nevada and tension exists between environmentalists and green-energy advocates.
- The average smartphone life span is less than 5 years!

https://www.usgs.gov/news/national-news-release/interior-releases-2018s-final-list-35-minerals-deemed-critical-us
https://www.nytimes.com/2025/01/24/magazine/nevada-lithium-mines.html

Powering technology

- A **data center** is a temperature-controlled building that houses computing infrastructure, such as servers, data storage drives, and network equipment.
- Scientists have estimated data centers consumption is at the same level with whole countries, being in the top 10 consumers!
- Data centers are used to train and run the deep learning models behind popular tools like ChatGPT and DALL-E. While not all data center computation involves generative AI, the technology has been a major driver of increasing energy demands.
- In 2021, it was estimated that to train GPT-3, the power used was equivalent to 120 average U.S. homes for a year and that it generated about 552 tons of carbon dioxide.
- The strain on water access to local communities near data centers is real.
- The cost of running an AI vs traditional query can be 10-50x higher!

https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

https://www.nytimes.com/2025/07/14/technology/meta-data-center-water.html

Further reading

- Books
  - Atlas of AI : power, politics, and the planetary costs of artificial intelligence by Kate Crawford