
# Ethics and Privacy

# 9

**Abstract**

This chapter discusses ethics and privacy where professional ethics are a code of conduct that governs how members of a profession deal with each other and with third parties. It expresses ideals of human behaviour, and the fundamental values of the organization, and is an indication of its professionalism. Privacy is defined as "the right to be left alone", and specifies there should be no intrusion upon seclusion, and no public disclosure of private facts or false information.

**Keywords**

Business ethics · Computer ethics · Privacy and the law · GDPR · Security · AI · Internet of things · Social media

## 9.1 Introduction

Ethics is a practical branch of philosophy that deals with moral questions such as the nature of what is right or wrong, as well as how a person should behave in a particular situation in a complex world. Ethics explore what actions are right or wrong within a specific context or within a certain society and seek to find satisfactory answers to moral questions. It is a search for moral principles to guide the behaviour of individuals or groups, and ethical issues occur when a conflict arises between an individual's moral compass, and the values or moral principles held by the society that the individual belongs to. The origin of the word "ethics" is from the Greek word ἠθικός, which means habit or custom.

There are various schools of ethics such as the *relativist* position (as defined by Protagoras), which argues that each person decides on what is right or wrong for them; *cultural relativism* argues that the particular society determines what is right



G. O'Regan, *Concise Guide to Software Engineering*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-031-07816-3_9

or wrong based upon its cultural values; *deontological ethics* (as defined by Kant) argues that there are moral laws to guide people in deciding what is right or wrong; and *utilitarianism* which argues that an action is right if its overall effect is to produce more happiness than unhappiness in society.

*Professional ethics* are a code of conduct that governs how members of a profession deal with each other and with third parties. A professional code of ethics expresses ideals of human behaviour, and it defines the fundamental principles of the organization, and is an indication of its professionalism. Several organizations such as the Association Computing Machinery (ACM), the Institute of Electrical and Electronic Engineers (IEEE) and the British Computer Society (BCS) have developed a code of conduct for their members, and violations of the code by members are taken seriously and are subject to investigations and disciplinary procedures (see Chap. 2).

Business ethics define the core values of the business and are used to guide employee behaviour. Should an employee accept gifts from a supplier to a company as this could lead to a conflict of interest? A company may face ethical questions on the use of technology. For example, should the use of a new technology be restricted because people can use it for illegal or harmful actions as well as beneficial ones? How can we balance the rights of a business to sell products that benefit society and the rights of citizens to be protected from harm from any unintended consequences of the technology?

Consider mobile phone technology, which has transformed communication between people, and thus is highly beneficial to society. What about mobile phones with cameras? On the one hand, they provide useful functionality in combining a phone and a camera. On the other hand, they may be employed to take indiscreet photos without permission of others, which may then be placed on inappropriate sites. In other words, how can citizens be protected from inappropriate use of such technology?

Professional responsibility in the computing and software engineering fields refer to the responsibility of computer professionals to carry out their work professionally to the highest standards, and to use sound judgement in the exercise of their duties. Engineers are accountable to themselves and others for their actions, and they must be willing to accept professional responsibility when performance does not meet professional standards.

Professional engineers have a duty to their clients to ensure that they are solving the real problem of the client. They need to precisely state the problem before working on its solution. Engineers need to be honest about current capabilities when asked to work on problems that have no appropriate technical solution, rather than accepting a contract for something that cannot be done. That is, engineers have a professional responsibility and are required to behave ethically with their clients. The membership of the professional engineering body requires the member to adhere to the code of ethics of the profession.

## 9.2 Business Ethics

Business ethics (also called corporate ethics) is concerned with ethical principles and moral problems that arise in a business environment (Fig. 9.1). They refer to the core principles and values of the organization and apply throughout the organization. They guide individual employees in carrying out their roles, and ethical issues include the rights and duties of a company, its employees, customers, and suppliers.

Many corporation and professional organizations have a written "*code of ethics*" that defines the professional standards expected of all employees in the company. Unfortunately, sometimes the code of ethics of an organization are window dressing, where they give the impression that these are the core values of the organization, whereas in reality they have not been properly implemented on the ground or are not being followed rigorously by employees in their day-to-day work practices and are not ingrained in the organization culture.

All employees are expected to adhere to the core values in the code whenever they represent the company. The human resource function in a company plays an important role in promoting ethics, and in putting internal HR policies in place relating to the ethical conduct of the employees, as well as addressing discrimination, sexual harassment and ensuring that employees are treated appropriately (including cultural sensitivities in a multi-cultural business environment). HR has a responsibility to provide training and awareness to staff on its core values.


**Fig. 9.1** Corrupt legislation. 1896. Public domain

Companies are expected to behave ethically and not to exploit its workers. There was a case of employee exploitation at the Foxconn plant (an Apple supplier of the *i*Phone) in Shenzhen in China in 2006, where conditions at the plant were so dreadful (long hours, low pay, unreasonable workload, and crammed accommodation) that several employees committed suicide. The scandal raised questions on the extent to which a large corporation such as Apple should protect the safety and health of the factory workers of its suppliers. Further, given the profits that Apple makes from the *i*Phone, is it ethical for Apple to allow such workers to be exploited?

Today, the area of *corporate social responsibility* (CSR) has become applicable to the corporate world, and it requires the corporation to be an ethical and responsible citizen in the communities in which it operates (even at a cost to its profits). It is therefore reasonable to expect a responsible corporation to pay its fair share of tax, and to refrain from using tax loopholes to avoid paying billions in taxes on international sales. Today, environment ethics has become topical, and it is concerned with the responsibility of business in protecting the environment in which it operates. It is reasonable to expect a responsible corporation to make protection of the environment and sustainability part of its business practices, even if this has an impact on its profitability.

Unethical business practices refer to those business actions that don't meet the standard of acceptable business operations, and they give the company a bad reputation. It may be that the entire business culture is corrupt, or it may be result of the unethical actions of an employee. It is important that such practices be exposed, and this may place an employee in an ethical dilemma (i.e., the loyalty of the employee to the employer versus doing the right thing such as becoming a *whistle-blower* and exposing the unethical or unsafe business practices). There are dangers that a whistle-blower could suffer career suicide following her exposure of unethical practices, and organizations need to create an effective structure or mechanism, where employees can raise serious ethical issues so that these may be resolved without fear of negative consequences to their career.

Some accepted business practices in the workplace might cause ethical concerns. For example, in many companies it is normal for the employer to monitor email and Internet use to ensure that employees do not abuse it, and so there may be grounds for privacy concerns. On the one hand, the employer is paying the employee's salary, and has a reasonable expectation that the employee does not abuse email and the Internet. On the other hand, the employee has reasonable rights of privacy provided that the computer resources are not abused.

The nature of privacy is relevant in the business models of several technology companies. For example, Google specializes in Internet based services and products, and its many products include Google Search (the world's largest search engine); Gmail for email; and Google Maps (a web mapping application that offers satellite images and street views). Google's products gather a lot of personal data, and create revealing profiles of the users, which can then be used for commercial purposes.

A Google search leaves traces on both the computer and in records kept by Google, which has raised privacy concerns as such information may be obtained by a forensic examination of the computer, or in records obtained from Google or the Internet Service Providers (ISP). Gmail automatically scans the contents of emails to add context sensitive advertisements to them and to filter spam, which raises privacy concerns, as it means that all emails sent or received are scanned and read by some computer. Google has argued that the automated scanning of emails is done to enhance the user experience, as it provides customized search results, tailored advertisements, and the prevention of spam and viruses. Google maps provides location information which may be used for targeted advertisements, and smartphones with Google maps may be used for the surveillance of users by tracking the places that they visit as well as the times and duration that they visit.

## 9.3 What is Computer Ethics?

Computer ethics is a set of principles that guide the behaviour of individuals when using computer resources. Several ethical issues that may arise include intellectual property rights, privacy concerns, as well as the impacts of computer technology on wider society.

The Computer Ethics Institute (CEI) is an American organization that examines ethical issues that arise in the information technology field. It published the *ten commandments on computer ethics* (Table 9.1) in the early 1990s [1], which attempted to outline principles and standards of behaviour to guide people in the ethical use of computers.

**Table 9.1** Ten commandments on computer ethics

| No. | Description |
|---|---|
| 1 | Thou shalt not use a computer to harm other people |
| 2 | Thou shalt not interfere with other people's computer work |
| 3 | Thou shalt not snoop around in other people's computer files |
| 4 | Thou shalt not use a computer to steal |
| 5 | Thou shalt not use a computer to bear false witness |
| 6 | Thou shalt not copy or use proprietary software for which you have not paid |
| 7 | Thou shalt not use other people's computer resources without authorization or proper compensation |
| 8 | Thou shalt not appropriate other people's intellectual output |
| 9 | Thou shalt think about the social consequences of the program you are writing or the system you are designing |
| 10 | Thou shalt always use a computer in ways that ensure consideration and respect for your fellow humans |

The first commandment says that it is unethical to use a computer to harm another user (e.g., destroy their files or steal their personal data), or to write a program that on execution does so. That is, activities such as spamming, malware, spyware, phishing, ransomware, and cyberbullying are unethical. The second commandment is related and may be interpreted that malicious software and viruses that disrupt the functioning of computer systems are unethical. The third commandment says that it is unethical (with some exceptions such as dealing with cybercrime and international terrorism) to read another person's emails, files, and personal data, as this is an invasion of their privacy.

The fourth commandment argues that the theft or leaking of confidential electronic personal information is unethical (computer technology has made it easier to commit fraud from the theft of personal information). The fifth commandment states that it is unethical to spread false or incorrect information (e.g., fake news or misinformation spread via email or social media). The sixth commandment states that it is unethical to obtain illegal copies of copyrighted software, as software is considered an artistic or literary work that is subject to copyright or license. All copies should be obtained legally.

The seventh commandment states that it is unethical to break into a computer system with another user's id and password (without their permission), or to gain unauthorized access to the data on another computer by hacking into the computer system. The eight commandment states that it is unethical to claim ownership of an intellectual creation that does not belong to you (e.g., to claim ownership of a program that was written by another, or to use an invention that is protected by a patent without proper authorization).

The ninth commandment states that it is important for companies and individuals to think about the social impacts of the software that is being created, and to create software only if it is beneficial to society (i.e., it is unethical to create malicious software or addictive software). That is, individual and companies need to consider the common good as well as profitability. The tenth commandment states that communication over computers and the Internet should be courteous and users should show courtesy and respect for others (e.g., there should be no use of abusive language or spreading of false information).

### 9.3.1 Ethical Problems in Computing

The ten commandments of computer ethics outline various principles to guide ethical behaviour in the information technology field. The computing field has introduced a unique set of ethical problems such as the unauthorized use of computer resources, the problem of hacking and theft of personal data, the problem of computer viruses, the professional responsibility of computer professionals in their work, the protection of personal data and privacy, and computer crime. Some ethical problems that arise in the computing field are summarized in Table 9.2.

**Table 9.2** Some ethical problems in computing

| Type | Description |
|---|---|
| Privacy | The use of computer technology raises concerns on data protection and privacy, as sensitive data may be compromised |
| Computer Crime | This may involve the theft of funds using a computer, or the theft of confidential information through unauthorized access of computer resources |
| Viruses | A virus is malicious code that an individual places on a network, and it is designed to spread and infect other machines. The virus may have destructive behaviour such as destroying data |
| Hacking | This is where a hacker who uses his (or her) computer skills to gain unauthorized access to computer files or networks (to cause damage or steal confidential information) |
| Cyberbullying | This is where an individual is bullied by others online, and it may lead to deep emotional distress to the individual |
| Professional responsibility | The development of a software product is a professional activity, and software engineers have a professional responsibility to ensure that the software product adheres to the highest possible standards. Software engineers must be accountable for their decisions and must ensure that the software is safe to use |
| Fake news | This refers to the systematic spreading of false or misleading information in traditional media or social media |

### 9.3.2 The Ethical Software Engineer

Software engineers have a professional responsibility to create ethical designs that satisfy the requirements, and to ensure that their designs are robust and protect the safety of the public. Software designers need to follow best practice in privacy and security in collecting, processing, and protecting data. The ethical design of a software system should give an open and accurate account of the system and should satisfy all relevant legal and regulatory requirements.

Ethical software designers need to be conscious of the algorithms that they create to ensure that they are unbiased, and do not discriminate against minority groups in society. This is especially important in machine learning algorithms based on pattern matching that are employed in the AI field, where *biased algorithms* may lead to discrimination against minorities.

Software engineers should consider the ultimate purpose of the project including its benefits to society as well as harm of the technology. Social media and various other apps are deliberately designed to be *addictive* to their users, where the software captures the attention of the human at a primal level, and the company reaps financial gain from the addiction of the users. This poses questions on the ethics of this addictive design, and whether the consequences of design as well as the product should be considered in ethical decision making.

The system needs to be designed for security, as it is difficult to add security after the system has been implemented. Security engineering is concerned with the

development of systems that can prevent malicious attacks and recover from them. Software developers need to be aware of the threats facing a system and develop solutions to manage them. Security loopholes may be introduced in the development of the system, and so care needs to be taken to prevent these as well as preventing hackers from exploiting security vulnerabilities.

Software testers need to always behave ethically during the development and testing of the software. The ISTQB Code of Ethics for test professionals is based on the IEEE and ACM code of ethics, and it states that software testers should act in the public interest and in the best interest of their client and employer. They ensure that their deliverables meet the highest standards, and they are independent in their professional judgements. They are required to be ethical and to be supportive of their colleagues, and to work closely with software developers. Software testers need to keep their knowledge up to date with lifelong learning.

Ethical issues may arise during testing if the project is behind schedule, and when there is pressure applied to the test team to stay with the original project delivery schedule. This could lead to the quality of the released software being compromised, and the test manager needs to resist any pressure that poses risks to quality.

### 9.3.3 Ethics in Data Science

Information is power in the digital age, and the collection, processing and use of information needs to be regulated. Data science is a multi-disciplinary field that extracts knowledge from data sets that consist of structured and unstructured data, and large data sets (*big data*[1]) may be analysed to extract useful information. The field has great power to harm and to help, and data scientists have a responsibility to use this power wisely. Data science may be regarded as a branch of statistics as it uses many concepts from the field, and it is essential that both the data and models are valid to prevent errors occurring during data analysis.

Personal data is collected about individuals from their use of computer resources such as their use of email, their Google searches, their Internet, and Social media use to build up revealing profiles of the user that may be targeted to advertisers. Modern technology has allowed governments to conduct mass surveillance on its citizens, with face recognition software allowing citizens to be recognized at demonstrations or other mass assemblies.

Further, smartphones provide location data that allows the location of the user to be tracked. It is important that such technologies are regulated and not abused by the state. Privacy has become more important in the information age, and it is the way in which we separate ourselves from other people and is the right to be left alone. The European GDPR law has become an important protector of privacy and personal data, and both European and other countries have adapted it.

[1] Big data involves combining data from lots of sources such as bar codes, cctv, shopping data, drivers license, and so on.

Companies collect lots of personal data about individuals, and so the question is how should a company respond to a request for personal information on users? Does it have a policy to deal with that situation? What happens to the personal data that a bankrupt company has gathered? Is the personal data part of the assets of the bankrupt company and sold on with the remainder of the company? How does this affect privacy agreements and compliance to them or does the agreement cease on termination of business activities?

The consequence of an error in data collection or processing could result in harm to an individual, and so the data collection and processing needs to be accurate. Decisions may be made based on public and private data, and often individuals are unaware as to what data was collected about them, whether the data is accurate, and whether it is possible to correct errors in the data.

Further, the conclusions from the analysis may be invalid due to errors in incorrect or biased algorithms, and so a reasonable question is how to keep algorithmically driven systems from harming people? Data scientists have a responsibility to ensure that the algorithm is fit for purpose and uses the right training data, and as far as practical to detect and eliminate unintentional discrimination in algorithms against individuals or groups.

That is, problems may arise when the algorithm uses criteria tuned to fit the majority, as this may be unfair to minorities. Another words, the results are correct, but presented in an over simplistic manner. This could involve presenting the correct aggregate outcome but ignoring the differences within the population, and so leading to the suppression of diversity, and discriminating against the minority group. Another example is where the data may be correct but presented in a misleading way (e.g., the scales of the axis may be used to present the results visually in an exaggerated way).

The ownership of personal data is important, for example, if I take a picture of another individual does the picture belong to me (as owner of the camera and the collector of the data)? Or does it belong to the individual who is the subject of the image? Most reasonable people would say that the image is my property, and if so what responsibilities or obligations do I have (if any) to the other individual?

That is, although I may technically be the owner of the image, the fact that it contains the personal data (or image) of another should indicate that I have an ethical responsibility or obligation to ensure that the image (or personal data) is not misused in any way to harm that individual. Further, if I misuse the image in any way then I may be open to a lawsuit from the individual.

Ethical rules are shared values that are followed voluntarily to make the world a better place, whereas legal rules are used to enforce social values. Often, the benefits of following the rules outweigh the costs of following them. For example, following the defined rules of the road leads to safe and predictable travel, whereas the cost of obeying the rules is that an individual must drive under the speed limit on the correct side of the road.

There has been a phenomenal growth in the use of digital data in information technology, with vast amounts of data collected, processed, and used, and so the ethics of data science has become important. There are social consequences to the

use of data, and the ethics of data science aims to investigate what is fair and ethical in data science, and what should or should not be done with data.

A fundamental principle of ethics in data science refers to *informed consent*, and this has its origins in the ethics of medical experiments on individuals. The concept of informed consent in medical ethics is where the individual is informed about the experiment and gives their *consent voluntarily*. The individual has the right to withdraw consent at any time during the experiment. Such experiments are generally conducted to benefit society, and often there is a board that approves the study and oversees it to ensure that all participants have given their informed consent and attempts to balance the benefits to society with any potential harm to individuals. Once individuals have given their informed consent data may be collected about them.

The principle of informed consent is part of information technology, in the sense that individuals accept the terms and conditions before they may use software applications, and these terms state that data may be collected, processed, and shared. However, it is important to note that generally users do not give informed consent in the sense of medical experiments, as the details of the data collection and processing is hidden in the small print of the terms and condition, and this is generally a long and largely unreadable document. Further, the consent is not given voluntarily, in the sense that if a user wishes to use the software, then he or she has no choice but to click acceptance of the terms and conditions of use for the site. Otherwise, they are unable to access the site, and so for many software applications (apps) consent is essentially coerced rather than freely given.

There was some early research done on user behaviour by Facebook in 2012, where they conducted an experiment to determine if they could influence the mood of users by posing happy or sad stories to their news feed. The experiment was done without the consent of the users, and while the study indicated that happy or sad stories did influence the user's mood and postings, it led to controversy and major dissatisfaction with Facebook when users became aware that they were the *subject of a psychological experiment without their consent*.

The dating site OKCupid uses an algorithm to find compatibility matches for its users based on their profiles, and two people are assigned a match rating based on the extent to which the algorithm judges them to be compatible. OKCupid also conducted psychological experiments on its users without their knowledge, with the first experiment being a "love is blind" day where all images were removed from the site, and so compatibilities were determined without the use of images.

Another experiment was controversial and unethical, as the site lied to the users on their match ratings (e.g., two people with a compatibility rating of 90% were given a rating of 30%, and vice versa). The site was trying to determine the extent that two people would get along irrespective of the rating that they were given, and it showed that two people talked more when falsely told that the algorithm matched them, and vice versa. The controversy arose once users became aware of the deception by the company, and it provides a case study on the *socially unacceptable manipulation of user data* by an Internet company.

Data collection is not a new phenomenon as devices such as cameras and telephones have been around for some time. People have reasonable expectations on privacy, and do not expect their phone calls to be monitored and eavesdropped by others, or they do not expect to be recorded in a changing room or in their home. Individuals will wish to avoid the harm that could occur due to data about them being collected, processed, and shared. The question is whether reasonable rules can be defined and agreed, and whether trade-offs may be made to balance the conflicting rights and to protect the individual as far as is possible.

The consequence of an error in the data analysis or with the analysis method could result in harm to the individual. There are many sources of error such as the sample chosen, which may not be representative of the entire population. Other problems arise with knowledge acquisition by machine learning, where the learning algorithm has used incomplete training data for pattern (or other knowledge) recognition. Training data may also be incomplete if the future population differs from the past population.

The data collection needs to decide on the data and attributes to be collected, and often the attributes chosen are limited to what is available, and the data scientist will also need to decide what to do with missing attributes. Often errors arise in data processing tasks such as analysing text information or recognizing faces from photos. There may be human errors in the data (e.g., spelling errors or where the data field was misunderstood), and errors may lead to poor results and possible harm to the user. The problem with such errors is that often decisions are made based on public and private data, and often individuals are unaware as to what data was collected and whether there is a method to correct it.

Even with perfect data the conclusions from the analysis may be invalid due to errors in the model, and there are many ways in which the model may be incorrect. Many machine-learning algorithms just estimate parameters to fit a pre-determined model, without knowing whether the model is appropriate or not (e.g., the model may be attempting to fit a linear model to a non-linear reality). This becomes problematic when estimating (or extrapolating) values outside of the given data unless there is confidence in the correctness of the model.

Further, care is required before assigning results to an individual from an analysis of group data, as there may be other explanations (e.g., Simpson's paradox in probability/statistics is where a trend that appears in several groups of data disappears or reverses when these groups are combined). It is important to think about the population that you are studying, and to make sure that you are collecting data on the right population, and whether to segment it into population groups, as well as how best to do the segmentation.

It may seem reasonable to assume that data-driven analysis is fair and neutral, but unfortunately the problem is that humans may unintentionally introduce bias, as they set the boundary conditions. The bias may be through their choice of the model, the use of training data that may not be representative of the population, or the past population may not be representative of the future population, and so on. This may potentially lead to algorithmic decisions that are unfair (e.g., the Amazon hiring algorithm discriminated against female applicants, and so the question is how

to be confident that the algorithms are fair and unbiased. Data scientists have a responsibility to ensure that the algorithm is fit for purpose and uses the right training data, and as far as practical to detect and eliminate unintentional discrimination (individual or target group).

Another problem that may arise is data that is correct but presented in a misleading way. One simple way to do this is to manipulate the scales of the axis to present the results visually in an exaggerated way. Another example is where the results are correct, but presented in an over simplistic manner (e.g., there may be two or more groups in the population with distinct behaviour where one group is the dominant), where the correct aggregate outcome is presented but this is misleading due to the differences within the population, and by suppressing diversity there may be discrimination against the minority group. In other words, the algorithm may use criteria tuned to fit the majority and may be unfair to minorities.

Exploration is the first phase in data analysis, and a hypothesis may be devised to fit the observed data (this is the opposite of traditional approaches where the starting point is the hypothesis, and the data is used to confirm or reject the hypothesis based on the data from the control and target groups, and so this approach needs to be used carefully to ensure the validity of the results).

## 9.4 Privacy

Privacy is a fundamental concept in modern society, and it has become an important area in the computing field. In Greek mythology there was a giant called Argus Panoptes, who was an all-seeing giant with one hundred eyes looking in every direction, and he would always have some eyes open even when asleep. That is, he was always watching and monitoring the world around him, and so was the perfect guardian. He was later slain by Hermes (the messenger of the gods).

Jeremy Bentham designed a circular prison in the eighteenth century termed the Panopticon, where a single guard in the centre of the complex could observe all the prisoners. His idea was that although individual prisoners did not know if they were being watched or not at a given time instant (as this depended on the direction that the guard was facing), that they would behave as if they were being watched, and so they would behave all the time (Fig. 9.2).

The modern version of the Panopticon is a set of security cameras that is watching people, or websites that are monitoring the behaviour of individuals who are visiting the site, or the entire Internet, which is working out everything that we are doing by watching us. The question is whether we as individuals should be concerned about this, and whether it matters if we as individuals are doing nothing wrong. Some have argued that everyone would be completely honest due to zero privacy, and where everyone could know what everyone else is doing. Others respond by saying that privacy is a basic human right, and that it is needed for freedoms to be exercised in society.

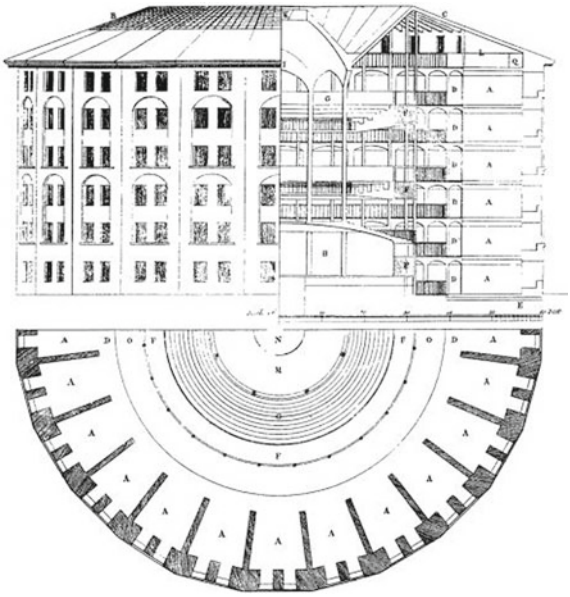

**Fig. 9.2** Bentham's panopticon prison

The "*Right to have privacy*" was an influential legal article written by Louis Brandeis and Samuel Warren and published in the Harvard Law Review in 1890 [2]. The article advocates for *privacy* as "*the right to be left alone*". William Prosser wrote an article in the Californian Law Review in 1960 in which he outlined four types of privacy torts [3]:

- Intrusion upon seclusion
- Public disclosure of private facts
- Publishing objectionable, false information
- Misappropriation of name or likeness.

There has traditionally been a difference between rural and urban living, where in a small-town people know everything about every other person in the town, and there is essentially very little privacy from all the gossip (*pueblo pequeño infierno grande*). In a larger city, people are anonymous, and nobody knows or cares about what others are doing, and so there is a greater sense of privacy. Further, an individual living in a small town has a choice of moving to a new town for a fresh start or waiting in the town for the community to forget a particular event, whereas in a large city this problem is a lot less relevant due to the anonymous nature of city living.

There are some parallels of the Internet being like the small village, except that the relationship is asymmetric. Another words, in a small town everyone knows as much about another as vice versa (i.e., it is a symmetric relationship), whereas the relationship is asymmetric for the Internet. This makes it a very unequal

relationship, with one party gathering lots of information and building up a profile about all other parties and using that information for commercial purposes. The other parties are not actively gathering information and have a very limited picture of what is going on with all the data that is gathered.

Further, while events and information may be forgotten in a village over time this does not happen with the Internet: i.e., it is very difficult to forget things on the Internet with web pages surviving forever in some archive even if taken down. Another issue is that once a page is put up many copies of it are made, and even if page is taken down there may still be many copies remaining elsewhere, and so there is no way of really deleting something once it has been published on the web. This could create major problems for individuals who pose indiscreet content online, as that content may be there in perpetuity.

People need an understanding of how their personal information and data is collected, shared, and used across the many computer platforms that they use, and the extent to which they have control over their personal information. New technology has led to major changes in the way in which privacy is experienced by society, and so it is important to understand the nature of privacy, and to consider the problems and risks that exist, as well as privacy laws and rules that are available to protect individuals from its abuse. The main sources of personal data that are collected include (Table 9.3).

These sources of information can collect vast amounts of data, and the collected data may potentially result in harm to an individual. The collected data is commercially valuable, especially when data about individuals are linked from several

**Table 9.3** Sources of information

| Source | Answers |
|---|---|
| Data collected by merchants and service providers | This includes personal data entered for the purchase of products and services such as name, address, date of birth, products and services purchased, etc. |
| Activity tracking | This involves monitoring the user's activity on the site (or app), and recording the user's searches, and the products browsed and purchased<br>It may involve recording the user's interests, their activities, and their interactions and communications with others on the site |
| Search Profile | The history of a person's searches over a period of time on a search engine such as Google reveals information about the individual and their interests |
| Sensors from devices | There are many sensors in the world around us such as personal devices as part of the Internet of Things that may record information such as health data or what the individual is eating<br>Third party devices such as security cameras may be conducting public or private surveillance<br>GPS technology on smart phones may be tracking the user's location |

sources. *Data brokers* are companies that aggregate and link information from multiple sources to create more complete and valuable information products (i.e., profiles of individuals) that may then be sold on to interested parties. Meta data (i.e., data about the data such as the time of a phone call or who the call is made to) also provides useful information that may be collected and shared.

For example, suppose that the probability of an individual buying a pair of hiking boots is very low (say 1 in 5000 probability). Next, that individual starts scanning a website (say Amazon) for boots then that individual is now viewed as being more likely to buy a pair of hiking boots (say a 1 in 100 probability). This large increase in probability will mean that the individual is now of interest to advertisers and sellers, and various targeted (popup) advertisements will appear advertising different hiking boots to the individual. This may become quite tedious and annoying to the user, who may have been just browsing, and is now subject to an invasion of advertisements, but many apps are free and often the source of their revenue is from advertisements, and so they gather data about the user that is then sold on to advertisers.

De-identification is the removal of identifiable information from data and includes the removal of fields (or attributes) such as name, address, and phone number so that no personally identifiable attributes remain in the dataset. This means that the identity of the person is not immediately identifiable, and so it provides some safeguards to the individual. However, it is possible that the individual's identity may be determined from the other retained fields, and this means that care must be taken if public records are to be released. That is, it may still be possible even if de-identification has taken place to identity individuals. Anonymity is limited or virtually impossible given the extent of public and private information that is available about individuals, and facial recognition technology allows the rapid identification of individuals from the images of their faces.

Privacy is important, and individuals should be able to express themselves without worrying about who may be watching. Individuals naturally desire rights such as the right to be left alone, for secret or intimate information to be kept secure from others, and for *control over personal information* where individuals can decide what information will be shared, when it will be shared, and how it will be communicated and shared with others (Fig. 9.3).

That is, users should be in control of how their data is used, and most user agreements are "all-or-nothing" in the sense that a user must give up control of their data to use the application, and so essentially the user has no control. That is, a user must click acceptance of the terms and conditions to use the services of a web application. Clearly, users would be happier and feel that they are in control if they were offered graduated choices by the vendor, to allow them to make trade-offs, and to choose a level of privacy that they are comfortable with.

Privacy has become quite topical with recent developments in the information age, and especially with the rise of social media, the Internet of Things and Artificial Intelligence. However, privacy concerns are not a new phenomenon, and they initially grew out of the development of early technologies such as the first cameras, microphones and telephones, where indiscreet or unauthorized images or recordings

**Fig. 9.3** Cardinals eavesdropping in the Vatican

could be made leading to concerns of an invasion of privacy by a prying media or others.

The early concerns over privacy were often with maintaining the security and confidentiality of a message, and so this led to some people and groups to communicate with each other using ciphers. For example, Julius Caesar communicated important messages using an alphabetic cipher during his campaign in Gaul in the first century B.C., and the emperor Augustus used a similar approach for communication. Further, some of the leaders during the American War of Independence used codes and pseudonyms to protect their identity during sensitive communication.

Societies vary in terms of their political systems, with democracies offering a peaceful way of replacing an unpopular government, whereas totalitarian states are often ruthless in their control of the population. Some autocratic societies run by dictators employ a culture of surveillance on the population, and this may include identifying individuals who pose a potential threat to the regime and removing such individuals from society either by imprisonment or political assassination. Often, these societies are characterized by mass surveillance of individuals, police searches and seizure of private property, police brutality, and so on. In democratic societies there are usually laws to protect the citizen against unreasonable police searches and behaviour.

The importance of privacy in the information technology field became apparent in the early 1970s with the introduction of databases. These could hold private information about individuals, and there was a need for a set of rules to protect how information should be collected and used. This led to the development of a set of

fair information processing principles (FIPPs) that was concerned with the way that data is used, collected and privacy. This was developed by the US Secretary's Advisory Committee on Automated Personal Data Systems, and published in their 1973 report on Records, Computers, and the Rights of Citizens [4]. This led to the Privacy Act in 1974, and this act remains the basis on which data collection is governed in the United States. The report outlined several principles such as:

- Transparency of collection and storage of information,
- Accessibility of personal information,
- Purpose limitations (consent),
- Correction of personal data,
- Personal data safeguards and accountability.

That is, the organization that is collecting personal data must be doing so openly (i.e., it is not secretly or covertly collecting data), and an individual must be able to access any data that the organization has about her. There must be a way for an individual to prevent information that was gathered for one purpose to be used for another purpose without their consent. Further, there must be a way for an individual to correct or amend information about him. Finally, any organization that is creating, maintaining, or disseminating personal must ensure the reliability of the data for the identified use, and take reasonable precautions to prevent against any misuse of the data.

Computing technology has evolved in a major way from the mainframes and databases of the early 1970s, and today modern society has embraced a plethora of leading-edge technologies such as smart phones, social media, the Internet of Things, and Artificial Intelligence. It is reasonable to ask what privacy means in the modern digital world and whether there is privacy anymore? Users of social media share large parts of their lives with a massive on-line audience as well as with large corporations, and social media companies gather lots of data about its users that may be used to determine patterns, and to generate profiles that may be targeted to advertisers. So much data is being collected about individuals, and the question is where does it go? Who controls it? Are companies adequately managing risks of data breaches? What happens when data privacy is breached or data is not secured properly? Is there transparency? Is user data encrypted? Is confidentiality and authenticity maintained?

The *Internet of Things* (IoT) is not a single technology as such, and instead it is a collection of devices, sensors and services that capture data to monitor and control the world around them. It refers to interconnected technology that is now an integral part of modern society, where computation and data communication are embedded in the environment. It allows everyday devices to connect to other devices or people over the Internet, and this may include smart phone to smart phone communication, vehicle to vehicle communication, connected cameras, GPS tracking, the smart grid, and so on. It allows a vast amount of data to be gathered and transmitted to and processed by companies. It means that information processing is now an integral part of people's lives, and IoT connects many devices to the Internet.

The level of interconnectivity and data gathered with IoT means that security and privacy have become important concerns, and it is essential to control both the devices and the data. For example, control could be lost if someone hacks into the smart phone, as the smart phone often links to bank accounts, email accounts and even household appliances. A lot of user data is potentially gathered painting a profile of individual users through their online activities as well as their searches, and the data gathered is used to improve the user experience, and the profile of users may be sold on to advertisers. Data should only be gathered with user consent, and there are risks of hacking or eavesdropping.

There has been a major growth in AI technology in recent years, and AI has been applied to self-driving cars, facial recognition, machine translation and so on. Facial recognition technology may be used to unlock phones to authenticate identity, and it has also been applied to read facial expression during job interviews, as well as following the movement of individuals.

A vehicle may contain several on-board computers for processing various vehicle controls as well as for entertainment systems. Vehicles that connect to the Internet are potentially at risk of being hacked, where a hacker may potentially commandeer vehicle controls such as steering and the brakes.

It is often unclear who is collecting personal information, the type of information that they are collecting, what is being done with the data, and who the data is being shared with. Information privacy refers to control over information and is a value that in a sense protects from certain kinds of harm. For example, if others have information about a particular individual, they may be able to use it against the individual. For example, if the individual has been the victim of phishing or identity theft where their personal financial information such as credit cards are stolen, then the perpetrators have power over the individual since they have personal and sensitive information about the individual.

### 9.4.1 Social Media

Social media involves the use of computer technology for the creation and exchange of user-generated content. These web-based technologies allow users to discuss and modify the created content, and it has led to major changes in communication between individuals, communities, and organizations (Fig. 9.4).

Social media is designed to have the individual share as much information as possible, and to continue to do so while they are on the site, and with every disclosure (or post) the individual reveals a little bit more about himself or herself. It is very easy to post photos and information on social media sites such as Facebook or Twitter, and social media is designed in such a way that it is addictive and poses risks to the privacy of an individual.

There is a danger that both social media companies and other users could harm the individual's privacy. The harm from other users may arise when a piece of the user's information is shared with the wrong audience, and this later leads to problems for the user. There are two distinct audiences for the individual's

**Fig. 9.4** Young peoples on smart phones and social media. Public domain

information namely other users and the platform itself. The social media platform maintains a vast quantity of electronic information consisting of immense databases, which can collect a vast amount of data on the individual and other users.

There is a power imbalance between the platform and the user, with the platform designed to have the individual share as much as possible, and people may potentially pose risks in social interaction. An individual's information may be viewed by friends, family, employer, work colleagues and nameless others, and so everyone in the individual's network as well as others could be an unwanted audience.

Users often may not realize the full extent of their audience when they post, and the people who are authorized in an individual's network may not be the desired recipients of certain posts (disclosures). It is difficult to delete online messages, and destructive posts may last long after an incident. Therefore, it is very much in the interests of users to keep their Social Media posts discreet, as both friends and outsiders of their social media network pose risks to their privacy.

Another words, it is difficult for an individual to protect herself from the risks of social media, and there are several threats such as:

- Manufactured disclosures
- Extracting consent
- Overexposure
- Faithless friends
- Online harassment

*Manufactured disclosures* refer to how a social media site gets people to disclose more and more information, and this is similar in a way to surveillance. Traditional surveillance involves watching people to learn something about them, whereas modern surveillance as in social media has less to do with this, and it generally involves getting people to disclose something more about themselves and so in effect to learn something new about the person.

*Extracting consent* refers to how a social media site obtains consent from its users on the various practices employed on the site. A user must click acceptance of the associated terms and conditions to use the site, and often this involves accepting invasive practices described in a long, dense, and largely unreadable terms of use document. The social media site may also request access to the camera, location, and address book of the individual. Often, users just accept the terms and conditions and permission requests because they are so worn down with so many requests from different apps, and they have no choice but to accept the terms of use and invasive practices so that they may use the site.

Social media sites constantly introduce new features to make user data more visible, more searchable, and more complete to others and may result in *over exposure* of their information. *Faithless friends* refer to when information that has been shared in the individual's network is shared more widely by one of the "friends" of the individual. This may lead to embarrassment or harm to the individual. Finally, *online harassment* is where repeated insults or bullying of an individual takes place online, which may even include threats of violence, posting of indiscreet images or even revenge porn.

#### 9.4.1.1 Data Analytics for Social Media

Data analytics provides a quantitative insight into human behaviour on a social media website and is a way to understand users and how to communicate with them better. It enables the business to understand its audience better, to improve the user experience, and to create content that will be of interest to them. Data analytics consist of a collection of data that says something about the social media conversation, and it involves the collection, monitoring, analysis, summarization, and a graph to visualize insight into the behaviour of users.

Another words, *data analytics* involves learning to read a social media community through data, and the interpretations of the quantifiable data (or metrics) gives information on the activities, events, and conversations. This includes what users like when they are online, but other important information such as their opinions and emotions need to be gathered through *social listening*. Social listening involves monitoring keywords and mentions in social media conversations in the target audience and industry, to understand and analyse what the audience is saying about the business and allows the business to engage with its audience.

Social media companies use data analytics to gain an insight into customers, and elementary data such as the number of likes, the number of followers, the number of times a video is played on YouTube, and so on are gathered to obtain a quantified understanding of a conversation. This data is valuable in judging the effectiveness of a social media campaign, where the focus is to determine how effective the campaign has been in meeting its goals. The goals may be to increase the number of users or to build a brand, and data analytics combined with social listening help in understanding how people are interacting, as well as what they are interacting about and how successful the interactions has been.

Facebook and Twitter maintain a comprehensive set of measurements for data analytics, with Facebook maintaining several metrics such as the number of page

views and the number of likes and reach of posts (i.e., the number of people who saw posts at least once). Twitter includes a dashboard view to summarize how successful tweet activity has been, as well as the interests and locations of the user's followers. Social listening considers user opinions, emotions, views, evaluations, and attitude, and social media data contains rich collection of human emotions.

The design of a social media campaign is often an iterative process, with the first step being to determine the objective of the campaign and designing the campaign to meet the requirements. The effectiveness of a campaign is judged by a combination of social media analytics and social listening, with the campaign refined appropriately to meet its goals and the cycle repeating. The key performance indicators (KPI) may include increased followers/subscribers or an increase in the content shared, and so on.

### 9.4.2 Internet of Things

The Internet of Things is a collection of devices, sensors and services that capture data to monitor and control the world around them, and these include cars, clothing, fridges, fitness monitors, and many of the things that are in a person's day to day life have potential as an internet device. An individual may be continuously connected to multiple home devices with sensors (e.g., microphones and cameras), and connection and access to these devices increases the risk to data security (Fig. 9.5).

The fact that there are many devices with sensors connected to the Internet means that there are, in effect, more eyes watching the individual and gathering data about her, and there are also more points of failure. This means that IoT poses increased risks to the safety of individuals than when using basic computers, and the risks include:

- Security risks,
- Privacy risks.

**Fig. 9.5** Fitbit Surge. Smart-watch activity tracker. Creative commons

The fact that these devices consist of both hardware and software means that there are now two points of failure: i.e., hardware failure and software failure. Hardware is generally more reliable than software, and hardware failures tend to be because of components wearing out over time. Software failures are often due to design issues, and software often requires regular updates to correct problems or to deal with security vulnerabilities. The fact that these devices are connected to the Internet means that software upgrades are possible but being connected to the Internet means that the device may be targeted by hackers in a similar way to which a computer is hacked.

Further, these Internet devices contain sensors that gather a lot of personal data about the individual, and they collect, use, and share this data, and so the IoT devices pose similar data security risks as laptops or smart phones. Many IoT devices are inexpensive and have serious security vulnerabilities, with some Internet devices failing to encrypt data when transmitting data or images to the cloud. This means that that an eavesdropper could intercept this Internet traffic, and cause harm to the individual.

IoT has serious implications for privacy in that the IoT devices can produce granular personal data such as when the individual is at home, what the individual eats, and so on. They gather a lot of personal data the individual, and the data may be shared with other devices or platforms thereby posing risks to the privacy of the individual.

### 9.4.3 AI and Facial Recognition

There has been a major growth in the AI field in recent years, and facial recognition is a new AI technology that offers the ability to unlock phones to authenticate identity, and so it may be used to protect the individual. Facial recognition has advanced in sophistication to allow individuals to be recognized at demonstrations and street protests, and this means that some authoritarian governments could potentially use facial recognition technology as a tool for authoritarian control. That is, surveillance combined with facial recognition could be oppressive to individuals and society and lead to a totalitarian state.

A society that adopts a paradigm of constant surveillance, where individuals are living in a world with technology monitoring their activities, learning to recognize patterns, and drawing inferences is moving towards totalitarianism. Facial recognition is a potentially dangerous technology that may challenge civil liberties, and it could severely impact marginalized groups in society.

- Faces are hard to hide,
- Faces are central to identity,
- Existing face and name databases,
- Facial recognition is widespread.

Facial recognition is a biometric technology that analyses visual data from social media and other sources, and they can detect facial features and to essentially reduce each face to a mathematical equation using factors such as the distance between the individual's eyes, the width of the nose, and so on, and the patterns in the visual data are compared to patterns in facial recognition databases to confirm identity.

Some companies have applied facial recognition technology to read facial expression during job interviews, and this provides a mechanism for the company to obtain data that they may not otherwise receive. Deep fakes are an AI technology that allows convincing images and videos to be created of individuals doing things that they never did or said, and this disruptive technology has been applied to misrepresent individuals in a variety of ways. An individual may be seen to make false or even preposterous claims, and this is achieved from content taken from social media and other media that is then manipulated and edited in various ways to achieve the desire effect. This technology could potentially show an individual committing a crime or present the individual in a very negative way. The technology has at this time mainly been applied to humour as in political satire, but as the technology improves it may become difficult to distinguish the real from the fake with serious consequences for society.

### 9.4.4 Privacy and the Law

Data collection laws focus on how data is collected, used, and shared, and data protection includes the right to information self-determination. The web is full of privacy policies that specify what type of personal data will be collected, how it will be processed and used, how it is shared, and what can be done about it. Further, individuals may take a lawsuit against another for a tort, for example, when someone pries or stalks them, or publishes a defamatory article, or violates their privacy. There are three main areas that impact upon an individual's privacy namely:

- The Media,
- Surveillance,
- Personal Data.

Media laws protect an individual against intrusion, where another party may be held liable for the invasion of the individual's privacy (e.g., phone tapping, snooping, examining a person's bank account, and so on). The tort of the public disclosure of private facts is part of the legal system in many states, and its goal is to prevent the public disclosure of private facts concerning the private life of an individual, where the matter is not of legitimate concern to the public. That is, others are prevented from widely spreading private facts such as the individual's face or identity for their own benefit, and there are slander and libel laws to protect an individual's good name and reputation, and to prevent defamation of character.

There are laws and rights to regulate surveillance with search warrants required in most countries to search the home of a private individual, as well as the right to seize personal property. Warrants are generally required to obtain personal electronic records held by telecommunication companies (e.g., the calls made and received as well as meta data such as geo-location data), and warrants may be required to obtain records held by Internet technology companies (e.g., emails, web sites visited, searches, and other electronic messages).

Countries vary in their laws for the protection of security and privacy, but many countries recognize that the security and privacy commitments made by a company in their policies should be fully implemented. Further, companies should be held accountable for any security breaches that occur that lead to data security or privacy being compromised, and the company may be liable for any losses suffered by individuals resulting from the breach.

Further, people must not be misled about the functionality of a website or mobile app that places their security or privacy at risk, and users must give their consent to any changes to the privacy policy that would allow for the collection of additional personal data, and users must be informed about the extensiveness of tracking and data collection. The collection and use of personal information of Facebook[2] users by Cambridge Analytica was a factor in the victory of Donald Trump over Hilary Clinton in the 2016 presidential election in the United States.

### 9.4.5 EU GDPR Privacy Law

Europe has been active in the development of data protection regulation, and the European General Data Protection Regulation (EU GDPR 2016/679) is a comprehensive data protection framework that became operational in 2018. The importance of both privacy and data protection has been recognized in Europe, and these are regarded as fundamental human rights in the EU. The goal is to give individuals control over their personal data, and it has had a huge impact on privacy laws of other countries around the world, with other countries using it to develop similar laws (e.g., Japan and the state of California in the US). GDPR also addresses the transfer of personal data outside of the EU, and it prohibits the transfer of personal data outside of the EU to countries that do not provide an equivalent or adequate data protection framework as GDPR (Fig. 9.6).

GDPR consists of a data governance framework that attempts to place privacy on a par with other laws. It creates protections that follow the data, and it places responsibilities on companies in managing privacy and information. GDPR applies whenever personal data is processed, and it starts from the presumption that the processing of the personal data is illegitimate. This means that companies carry the burden of legitimizing their actions, and they must be able to show that they have a legitimate basis for processing data. That is, they must be able to show that they

[2] Facebook was fined $5 billion in 2019 for deceptive and unfair trade practices related to Facebook's user interface.

**Fig. 9.6** EU GDPR 2016/679

have the consent of the data subject, or that the processing is necessary because of the contract that exists between them and the data subject, or where they have a legitimate interest, and where the interest of the data controller prevails over that of the data subject. The company must be able to demonstrate adherence to the fair information practice below:

- Standards for data quality
- Standards for transparency
- Special protections for sensitive data
- Standards of enforcement.

This means that data must be obtained legitimately and is used in the manner of the purpose for which it was acquired, and there must be openness and transparency so that individuals will know how their data will be used. There should be special protections for sensitive data with the ability to opt in for consent (e.g., race, sexual orientation, political beliefs), and there must be standards for enforcement to ensure compliance to the standards. The *Data Privacy Impact Assessment* (DPIA) is mentioned in GDPR, and it is needed if the processing of personal information is likely to result in a high risk to the rights and freedoms of individuals. This assessment helps to ensure that companies are complying with privacy requirements.

The standard for informed consent is very high which means that it is freely given and informed. GDPR also gives very strong data subject rights, including the right to access data, data portability, the right to rectify data, the right to erase data, the right to object to processing, and the right to restrict processing. These rights provide a powerful tool for data subjects to exercise control over their personal information.

## 9.5 Review Questions

1. What is data science?
2. What is the role of the data scientist?
3. What is privacy? Why is it important?

4. What are the main sources of personal data collected on line?
5. What are the main risks to an individual using social media?
6. What are the main risks to an individual using a fitness device (as part of the Internet of Things)?
7. What are the main risks with AI facial recognition technology?
8. Explain the importance of the EU GDPR law.
9. What is a digital privacy impact assessment?

## 9.6 Summary

Ethics is a practical branch of philosophy that deals with moral questions such as the nature of what is right or wrong, as well as how a person should behave in a particular situation in a complex world. Computer ethics is a set of principles that guide the behaviour of individuals when using computer resources. Business ethics (also called corporate ethics) is concerned with ethical principles and moral problems that arise in a business environment. Ethics guide individual employees in carrying out their roles, and ethical issues include the rights and duties of a company, its employees, customers, and suppliers. Several ethical issues that may arise include intellectual property rights, privacy concerns, as well as the impacts of computer technology on wider society.

Companies collect lots of personal data about individuals from their use of computer resources such as email, search engines, their Internet and Social media use, and the data is processed to build up revealing profiles of the user that may be targeted to advertisers. Modern technology allows mass surveillance to be conducted by governments on its citizens, with face recognition software allowing citizens to be recognized at demonstrations or other mass assemblies.

Privacy is important in the information age, and it is the way in which we separate ourselves from other people and is the right to be left alone. The European GDPR law has become an important protector of privacy and personal data, and both European and other countries have adapted it.

## References

1. R.C. Barquin, In Pursuit of a ‘ten commandments’ for computer ethics. Computer Ethics Institute (1992)
2. L. Brandeis, S. Warren, BrW:90 the right to have privacy. Harvard Law Rev., (1890)
3. W. Prosser, Privacy. Californian Law Rev., (1960)
4. Records, Computers and the Rights of Citizens, US Secretary’s Advisory Committee on Automated Personal Data Systems (1973). https://aspe.hhs.gov/report/records-computers-and-rights-citizens