

MODULE 5: META-COGNITION, ETHICAL DECISION MAKING, ETHICAL THEORIES

Meta-cognition in the context of ethical decision making as is the conscious awareness and control of one's thinking processes. It involves thinking about thinking, understanding how you learn, and making adjustments to improve learning outcomes. It's the foundation for effective study skills and problem-solving, allowing individuals to adapt their strategies and monitor their own comprehension and learning progress.

Ethical decision-making is the process of evaluating and choosing between moral dilemmas. It's a vital skill for IT professionals facing complex ethical choices. This process involves considering the ethical principles and values that guide one's actions, as well as the consequences of those actions on individuals and society. Ethical decision-making helps individuals navigate difficult moral issues with integrity and responsibility.

Ethical theories provide structured frameworks for understanding what is morally right or wrong. These theories offer different approaches to evaluating ethical questions. Understanding these ethical theories enables individuals to engage in informed ethical debates and make well-reasoned moral judgments. Each theory offers unique perspectives on how to address ethical dilemmas and make principled decisions.

5.1 HOW DO WE DEFINE ETHICS?

Ethics is the general name for the branch of moral philosophy that deals with behaviour that increases people's well-being. Ethics in the context of this book is therefore about how technologists should behave to increase people's well-being. Ethics is not about religion or being slavishly law-abiding, nor is it about going along with the majority view held by the people around you. Ethics is having your own moral compass.

Technology is ethical when it helps people reach their fullest potential; when it improves their quality of life, makes them happier and more fulfilled, and gives them the freedom to choose what they want to be. We consider the interests of people living here and now, but also the interests of future generations, other living creatures, and the preservation of the environment.

Technology is unethical when it dehumanises; when it makes a person less human than they were. It forces people to engage in behaviour that diminishes them or the environment in some way or creates a problem for future generations. Simply put, ethics is a guide to how to live well, how to be in the world in a way that creates benefit and minimises harm.

WHY DO WE NEED ETHICS?

Ethics allows us to live in harmony and cooperation with others. When people are ethical, we can trust one another. We can build communities and organisations that can achieve outcomes that a single, self-interested individual would be incapable of.

Without ethics, we would lack loyalty and be unable to trust others and form cooperative communities of interest. Long-term relationships would be difficult if not impossible. We could not have the economies that now exist in the developed world where wealth and a high standard of living are enjoyed by most. Life without ethics would likely be nasty, short, and brutish.

LEVELS OF ETHICS

Ethics or right behaviour has three broad levels of application:

Personal ethics guide how you live, what you do, and how you interact with others. It helps you to develop a sense of personal responsibility by making you think, both before and after you act. It considers how your behaviour impacts on others. As a rational being with free will, you choose how you behave on a day-to-day basis with full awareness of the consequences of your actions.

Organisational ethics is an aspect of organisational culture. It is how the organisation behaves and how it interacts with people. This level of ethics has explicit and implicit components. The explicit is clearly stated by management, written down and understood to be 'correct' behaviour. The implicit is not written down but is nonetheless understood to be the 'way things are done'.

As with personal ethics, this middle level should cultivate a sense of responsibility for how the organisation's actions impact on the world.

System ethics is concerned with how the overall economic and social systems behave, how it interacts with people. Ethics at the system level is codified into laws and codes of acceptable conduct; cultural practices that by consensus are widely understood and practices. As with the previous two levels, systems ethics cultivates a sense of responsibility for how the system impacts on the world in general. System ethics tries to create a system that best serves the interests of the greatest number of people.

As a citizen, you have a right to vote and to have your voice heard. You are free to argue for a more humane society.

VALUES & ETHICS

Values feed into ethics in four broad ways; (a) how to get along with each other, (b) what is a 'good life', (c) what are our obligations to each other, and (d) what are my rights?

If ethics is about behaviour, values are about what you believe to be important, and what you would like to see more of by means of more ethical behaviour.

For example, in western-style democracies, values are codified into 'rights'. Freedom of speech, freedom of religious worship, the pursuit of happiness and many other values are all considered to be our birth right as human beings.

Values come *before* ethics. The ethical standards of a society will reflect these pre-existing values. Values come from many sources; one's family, the media, religion, the community, one's education and life experiences.

Values change over time with the evolution of societies and culture. While it is true that much of our value system is created through our childhood experiences, they can nonetheless be changed through a process of conscious self-reflection and external influences.

ROLES & ETHICS

The *roles* we play have a strong determining effect on our ethics and on our behaviour generally. A role is simply a set of relationship responsibilities and expectations that we have adopted either voluntarily, or because they have been placed upon us through circumstance.

The first experience of roles for many is within the early environment where a child has a role in relation to their parent(s) or carers. Later we adopt a variety of roles by choice; we choose to get married, have children, and enter an occupation or profession. We might join a faith community and attend worship. We might become a volunteer for a worthy cause, or indeed any number of possible roles.

Each role has a set of responsibilities and expectations that belong to it and which we must fulfil if we are not to be sanctioned in some way. Roles can come into conflict with each other, for example a member of a religious community might find a role conflict if s/he were to perform military service.

The obligations that go along with a role can form the basis of ethical conduct for that person.

5.2 ETHICS IS META-CONSCIOUSNESS

Ethics is the general name for the branch of moral philosophy that deals with behaviour that increases people's well-being. Ethics in the context of this book is therefore about how technologists should behave to increase people's well-being. Ethics is not about religion or being slavishly law-abiding, nor is it about going along with the majority view held by the people around you. Ethics is having your own moral compass.

Meta-cognition involves actively engaging the recently evolved parts of the brain, the places where higher, rational thought occurs, the place where you can recognise the causal links.

This state of mind contrasts with the semi-conscious autopilot that people commonly use as their default setting. Reacting to situations in a habit-driven, stimulus-response manner based on prior learning. Conditioned responses to specific situations have been acquired over time from social learning. Thus, a poorly programmed autopilot is why people continue to make the same mistakes time and again. Meta-cognition is the only remedy to lift oneself out of this semi-conscious mode into a fully conscious state where a person responds to situations in a rational way that is based on the needs of the situation at hand.

This rational, meta-cognitive ability is what sets humans apart from intelligent animals. The neural infrastructure of the evolved human brain is said by neuroscience to be the most complex biological structure ever to have existed on this planet. Our brains and the abstract thinking that it is capable of is what has made humanity the most adaptable creature living on this planet.

5.3 CODES OF ETHICAL CONDUCT

Computer societies are working towards *licensing* its members so that like doctors, lawyers, teachers, accountants and other professions, it is not lawful to work unless you are licensed. To be licensed, a practitioner must have completed an approved study program that includes instruction on professional ethics. They must agree to abide by the code of conduct.

This chapter presents typical code of conduct, based on the Australian Computer Society's (ACS) code. This code is used since The *Ethical Technologist* is the textbook used in an ethics course at an Australian university. We might just as well have a code from North America, the United Kingdom, Japan, Germany, France or any country in the developed world. The underlying code is the same.

The ACS Code is summarised into six core ethical values that it expects its members to always practice in their professional life (source ACS):

1. **The Primacy of the Public Interest.** You will place the interests of the public above those of personal, business or sectional interests.
2. **The Enhancement of Quality of Life.** You will strive to enhance the quality of life of those affected by your work.
3. **Honesty.** You will be honest in your representation of skills, knowledge, services and products.
4. **Competence.** You will work competently and diligently for your stakeholders.
5. **Professional Development.** You will enhance your own professional development, and that of your staff.

6. **Professionalism.** You will enhance the integrity of the ACS and the respect of its members for each other.

THE PRIMACY OF THE PUBLIC INTEREST

The term 'Primacy' indicates that this is the core ethical value that takes precedence over any personal, private, or sectional interests that you might have. Where a conflict exists, it must be resolved in favour of the public interest. There is no room for self-interest, looking after 'number one'.

As you go about your work, you act in the interests of your employer so long as this does not conflict with your duty to the public interest. This means that you should not be developing technology that will adversely affect public health, public safety and the natural or built environment.

You identify those who will be impacted by your work and actively consider their interests to avoid harming them.

If you become aware of conflicts between your professional work and any legal or social factors, you work with the stakeholders to resolve the conflict before the problem becomes more serious. These can include problems the stakeholder(s) might have with what you are doing, or any conscientious objections you yourself might have.

Your duty to the public interest includes preserving the integrity and public image of the profession, respect for other people's intellectual property and for the confidentiality of any information that might come into your possession.

THE ENHANCEMENT OF QUALITY OF LIFE

Information and Communication Technology (ICT) has the potential to create both harm and benefit. The ethical technologist considers the impact that technology has on society and individuals and actively works to minimise the negative effects while maximising the positive.

The ethical technologist cultivates an equity of access attitude that gives the under-privileged members of society the same access that the more privileged already have.

As an ethical technologist, you develop an awareness of the many ways that ICT can enhance people's quality of life, particularly those less advantaged people in society and the world generally (for example in the developing world).

The technology you develop should promote the health and safety of the people who use it or are affected by it. At the very least it should not harm anyone.

At a more abstract level, the use of technology should create a positive perception and a deeper sense of personal satisfaction in people. It should help people

become a fuller expression of their human potential by allowing them to do what they were previously unable to do, and which gives them great satisfaction to do.

This core ethical value is an extension of the Public Interest value discussed in the previous section.

HONESTY

It is imperative that you do nothing to undermine public trust in the profession, or the trust of the stakeholders in a situation (i.e., your employer, the users etc.). Trust is a valuable but fragile commodity. It requires much time and effort to build, and yet it can be destroyed the moment deception is detected.

Trust can only be maintained in the long-term by being consistently honest in your dealings with people. You must be perceived as a person who can be relied upon to act with integrity, someone who avoids deception even when there is little risk of discovery.

As an ethical technologist, you therefore avoid offering or receiving inducements (favours, bribes, gratuities) or place yourself in a position where you can be coerced. Any situation intended to bring favour to one stakeholder at the expense of another.

Neither shall you mislead anyone as to the suitability of a product or service. You keep your professional life separate from your personal or sectional interests. It is not uncommon for IT practitioners to act as agents for a commercial organisation without disclosing that conflict of interest to their employer or customer.

Any estimates you give will be accurate and unbiased, you qualify a professional opinion that is based on limited expertise, you give credit where credit is due for the work of others, nor do you attempt to build your own reputation at the expense of other(s).

COMPETENCE

Given the complex nature of technology as a global industry, no single technologist can possibly know everything about everything. Yet it is common for IT practitioners to pretend to know more than they do and knowingly accept work that they are unqualified to perform. This is done on the assumption that they can learn the required skill at short notice or as they go along. In this they are little more than a trainee masquerading as a competent professional. It is a practice commonly seen when people “pad their CV’s” with skills they do not possess.

The client has a right to know that the technologist they engage is competent to perform the work, so as an ethical technologist you only accept work that you know you are competent to perform and avoid over-stating your skills and capabilities.

You deliver products and services that meet your clients' operational needs and respect their proprietary interests. If you are aware of issues in relation to a project that are not in the clients' interests, you make the client aware of these issues even if it might be in your personal interests to say nothing (for example, allowing you to stay employed on a project for longer).

Competency also means taking responsibility for your work, avoiding putting the blame on others when things go wrong.

PROFESSIONAL DEVELOPMENT

In the age of exponentially advancing technology, finding the time to stay up to date in your field can be a major challenge. It is tempting to let recent developments slip by when you realise that the work you did to learn the latest technology not so long ago is now redundant. The instinct we all must conserve energy suggests 'don't bother'. You must resist this 'economy of effort' mind-set, it is a major contributing factor to the burn-out and cynicism of mid and late-career members of the profession.

Professional development for the ethical technologist means taking the time and making the effort to not only stay abreast of the latest developments, but also to pass on your knowledge and experience to colleagues, particularly those in more junior roles. In the spirit of win-win, you understand that by helping others advance, you are ultimately benefiting everyone, including yourself. Win-win thinking benefits the profession.

So, the ethical technologist makes it their business to acquaint themselves with the technological issues having impact on the world, they encourage their colleagues and subordinates to do the same, and support educational initiatives aimed at the professional development of themselves and others.

PROFESSIONALISM

The computer industry, while global, is relatively new and does not yet have an established set of ethical standards. It takes time for the profession to mature. As an ethical technologist, you can help to establish these standards by always being professional and so improving the perception and image of the profession in the eyes of the public. The challenge is to build public confidence in the profession, particularly in the workplace.

The public has mixed feelings about computer technology; on the one hand they enjoy the convenience that it affords them, but on the other they do not understand it and sometimes fear that it might do them harm.

To dispel this fear, the ethical technologist takes a calm, objective and well-informed approach to their professional work.

As an ethical technologist, you encourage other practitioners to behave in accordance with the code and do nothing to tarnish the image of the profession. This includes ensuring that properly qualified people are not excluded from employment through unfair discrimination.

You also do what you can to extend public knowledge and appreciation of ICT, taking pride in being an IT professional.

A FINAL WORD

Professional societies around the world provide real assistance to practitioners in time of need. The excerpt below is from the Australian Computer Society, though every society will be offering the same service, should you need it:

‘All people have a right to be treated with dignity and respect. Discrimination is unprofessional behaviour, as is any form of harassment. Members should be aware that the ACS can help them resolve ethical dilemmas. It can also provide support for taking appropriate action, including whistleblowing, if you discover an ACS member engaging in unethical behaviour’.

For more detail, visit: www.acs.org.au or the equivalent society in your country.

5.4 ETHICAL DECISION MODEL (EDM)

For the purpose of resolving ethical dilemmas, we define a *dilemma as a complex problem for which there is no obvious solution*. A solution exists but is obscured by the complexity. Common sense would suggest that the best way to deal with a complex problem is to *simplify* it. You can do this by breaking it down into more comprehensible pieces.

Here we outline the *Ethical Decision Model* (EDM), a general-purpose model for analysing complex situations in a range of domains including IT. It helps you to reveal optimal solution(s), ones that might be described as ethical, and be defended as such.

Appendix A is an example of how the EDM can be applied to an IT-related case study. The solutions in the example are indicative, not definitive.

The model has three main stages: *analysis, prioritisation, decision*.

1. **Analysis** is getting the facts and categorising them into *extrinsic* factors (legal, professional, employment, social, personal) and *intrinsic* (a person's individual attributes).
2. **Prioritisation** involves ranking the elements into order of importance by means of a priority table.
3. A **Decision** is made by rationally weighing up the relative importance of the elements.

No two people who approach a complex situation will perceive the various factors the same. Their perceptions are filtered through the lens of their personal experience and intrinsic leanings. The precise nature of what reaches their cognitive centre will be different for every person and might even differ for the same person on different occasions.

Applying the defined process of the EDM helps to remove the subjectivity from the situation and gives us an objective, process-based approach to the solving of ethical dilemmas.

STEP 1: ANALYSIS

In preparing for the ethical analysis, there are some questions that you should ask:

- *What are the relevant facts of this case?*
- *What do we know, what do we not know that we need to know before deciding?*
- *Who are the stakeholders?*
- *Is this a legal matter for which a prescribed course of action already exists?*

Every effort must be made to obtain satisfactory answers to these questions before proceeding.

It is the nature of ethical dilemmas that they are a complex mix of factors for which there is no obvious solution. Maybe there are two or more obligations that conflict with each other, or the outcomes of anything you do will be undesirable, or even that the cost of doing the right thing is too high.

The factors that comprise a given situation can be broadly categorised as *Extrinsic* and *Intrinsic*; those that exist in the outside world, and those that exist within the individual. The Extrinsic factors include Legal, Professional, Employment, Social and Personal. The Intrinsic factors have been grouped together under a single heading.

Extrinsic factors

1. **Legal** factors take precedence over the others since breaking the law will get you into serious trouble, even loss of liberty. There will be no conflict between Legal and Professional factors since professional bodies are in the business of creating a solid, respectable public image for its members and will never advocate acting illegally.
2. **Professional** factors are the obligations you have to the profession, as prescribed in their code of practice. These take precedence over the obligations you have to your employer since it is possible that your employer will ask or demand that you do something unprofessional (unethical) in the profitability interests of the employer. Many dilemmas stem from this source.
3. **Employment** factors. Most employers have their own code of ethical conduct, as prescribed in their mission statement and other documents that define the values of the organisation. This code sets standards of ethical conduct. These will be generally compatible with the legal, professional and social standards, since no organisation, particularly commercial ones, will want to be seen as deviating from the standards of society. There will be some exceptions to this in the case of organisations on the periphery of society, ones that do not share its mainstream ideals, one's with an extreme political agenda.
4. **Social Factors**. The society in which the employer operates will have its inherent standards that are reinforced by the family, at school, in the community generally and in the media and other institutions. All the ways a society communicates with itself. Society is complex, so standards will not always be unanimously agreed upon. Some members of society will agree, and others disagree on the rightness of various issues. We see this often in polarised political debate. Legal, professional and employment factors take precedence over social factors where there is disagreement.
5. **Personal** factors include those aspects of your make-up that psychologists categorise as coming from the 'Nurture' side of the 'Nature-Nurture' theory (of what makes us what we are). These are the factors that you acquire from the environment, your family, close friends and associates, your peer group, sporting association or faith community. While these are undoubtedly within you, they have their origin from outside of you. Personal factors account for much of a person's ethics, their morality. When there is variance between one's personal morality and that of the Social, Professional and Legal environments, a person will have the greatest difficulty resolving this ethical

conflict. How does one remain true to oneself and still behave ethically in a professional sense? The unpleasant truth for some is that one's professional obligations must take precedence over any personal qualms you might have about what is ethical. To be a member of a profession means to accept its standards and practice them. To act otherwise will exclude you from the profession.

Intrinsic Factors

Intrinsic factors include what psychologists categorise as the 'Nature' side of 'Nature-Nurture'. It is your set of innate qualities, the behavioural disposition with which you were born, the disposition that your genetic make-up has equipped you with. People are born with differing degrees of a wide variety of personality traits. These are summed up in *The Big Five Personality Traits* which are *extraversion, agreeableness, openness, conscientiousness, and neuroticism*. Each trait represents a continuum. Individuals fall anywhere on the continuum for each trait. For example, with the extraversion/introversion continuum you can be anywhere on the bell curve from very extraverted to very introverted or somewhere in between. The Big Five remain relatively stable throughout most of one's lifetime.

So, people's *Nature* can vary widely within the broad definition of being human. This is a complex area well beyond the scope of this chapter and this book. In addition to The Big Five, you might also google the Myers-Briggs personality profile to learn more on this fascinating subject.

Jonathan Haidt's Moral Foundation Theory. On a more general level, Haidt's Moral Foundation Theory suggests that there are six innate moral foundations that all humans are born with, the innate moral code that we all share:

- *Care/harm,*
- *Fairness/cheating,*
- *Liberty/oppression,*
- *Loyalty/betrayal,*
- *Authority/subversion, and*
- *Sanctity/degradation* (discussed in later chapter).

Personal factors (previous section) and intrinsic attributes often exert the strongest yet most idiosyncratic influence on the process of ethical decision-making. While this is a potential problem, someone with the kind of personal and intrinsic attributes that makes them uncomfortable with what is generally accepted for an IT developer is unlikely to last long working in this capacity.

Applying the analysis to an example. Consider the case of the market research company that collects demographic information from the broader community and sells contact lists to interested parties who want to do targeted direct marketing.

The market research company obtains people's informed consent to collect and store this information. But now the company changes hands and the new owner

wants to increase profits. The owner instructs their web programmers to implement deceptive strategies aimed at gathering information for which they have no informed consent. This instruction contravenes privacy legislation, and the professional code of conduct. It is also contrary to community expectations on privacy. In this instance, when we prioritise the factors, it is clear what the developers should do – refuse to comply, even at the risk of losing their job.

For example, if the new owner agreed to supply a gay hate group with the names and addresses of people known to have an interest in gay culture. While it is clearly wrong from a legal, professional, and social perspective, if an IT developer working there is intrinsically homophobic, their disposition will influence their thinking on whether it is right to supply the names. The developer may well perceive this as an ethical dilemma, when the developer sitting in the next cubicle clearly sees it as a wrong act.

STEP 2: PRIORITISATION

Prioritisation is most easily performed by the making of a list that shows each factor in descending order of importance. It can be helpful to include a column that outlines related matters beside each factor. This has the same common-sense value as the Ben Franklin decision-making method of listing Pro's and Con's on a sheet of paper with a single vertical line drawn down the middle. The format with the EDM is somewhat different, but the principle is the same.

As a rule, the Legal and Professional factors take precedence since there is an obligation on everyone to abide by the law with no exceptions. This is a long-standing principle that was established for the benefit of the greatest number. The rule recognises personal freedom but says that there is a point where personal freedom ends and the public interest begins. A person can have their personal freedom curtailed by society if it is believed that such freedom is not in the public interest or the greater good.

Related to the obligation to abide by the law is the obligation to know the rules laid down by law. Ignorance of the law is not a defence in court for breaking the law.

Within the legal framework that governs society we have the various professions, medicine, law, accounting for example. All professions have a *Code of Professional Conduct*. It is always incumbent upon members to know it and practice it. Membership of that profession is conditional on a sincere undertaking that as a member you will do your utmost to follow the code.

Codes of Professional Conduct have relevance to professional standards legislation that exists in many jurisdictions. Breaches of the Code can be used as grounds for a claim of professional negligence. In legal proceedings, the Code can be quoted by an expert witness giving an assessment of professional conduct.

The Australian Computer Society's code of ethics can be summarised as follows; *always act in the public interest, your work should enhance people's quality of life, you should be honest, hard-working, competent and stay current with the latest developments, and finally do what you can to enhance the reputation of the profession.* If a conflict occurs between these values, the deciding factor is what is in the public interest, otherwise known as the 'greater good'.

Codes of conduct of professional computer societies in other countries will not be much different. The way in which they are expressed may be outwardly different but the essential, underlying meaning will be similar.

Codes of professional conduct and the larger laws of society are certain to be consistent with each other. For example, the first item in the ACS code clearly states that you should always act in accordance with the public interest, which by default is governed by law. Professional groups will never advocate behaviour that even hints at being unlawful or not consistent with the values of the society in which it operates. They want to establish a respectable place for themselves in society.

Social factors will also be broadly consistent with legal and professional factors. There is room for disagreement here because as society evolves, its values change, but the law, which is inherently conservative, does not change as quickly. There may be some gap between the two, with the legal taking precedence over the social. The process of law reform will take its course in time and the law will come to reflect community values.

Professional codes maintain a safe legal position. Extended debate within professional forums will perform the same role as the law reform bodies in larger society.

Prioritising the factors inherent in a situation should always have the legal, professional, and social factors at the top of the list. Most likely to conflict with these are *Work* factors. The goals, policies and culture of an organisation are at the discretion of the owners who may well perceive their first responsibility as being to their own financial interests and those of the shareholders. It is not being overly cynical to suggest that some business owners are more concerned with the question of whether they will get caught, not whether something is legal. Beyond the question of being caught, there is also the issue of how likely it is the state will prosecute, given that the law lags the pace of technological change. And given the expense of legal proceedings, Prosecutors will usually only pursue cases of significance that are likely to result in a conviction.

A commercial organisation's reason to exist is to make a profit or at least to survive and continue to trade. Despite outward appearances, many companies operate on the verge of collapse, delaying payment of their debts for as long as possible while trying to extract payment from debtors as quickly as possible. In desperate circumstances even a normally honest business owner has been known

to resort to unethical if not illegal strategies if they can get away with it. Most organisations are honest and ethical, but it is not difficult to see how a technologist working in *some* organisations are going to find themselves told to do 'questionable' things.

STEP 3: DECISION

Having drawn up a prioritised list that shows each factor in order of importance, you are now able to decide, based on rational choice, what will be the most ethical course of action.

In deciding, you might take into consideration which course of action: *does the most good or the least harm, respects stakeholder rights, treats people justly, best serves the public interest (not just some members), and which allows me to be the best kind of person I can be?*

If called upon, you should be able to make a strong argument, citing evidence as to why you chose as you did. Imagine that you have been called to explain yourself to the board of directors, or the ethics committee of a professional society or even the police/prosecutor. Your case should be strong enough that you could deliver it with confidence and a clear conscience.

5.5 THEORIES OF ETHICAL BEHAVIOUR

This section summarises the major philosophical theories that have bearing on ethics, the branch of philosophy that deals with morality. The list is a representative sample, not exhaustive. This level of detail is appropriate for a discussion on ethics in IT. For balance, the list covers both the philosophies of the West, starting with the classical Greeks, and then those of the East, including Buddhist, Confucist and Taoist philosophies. It should be noted that Buddhism, Confucism and Taoism are rightly called philosophies not religions since they concern themselves with how to think and behave correctly and recognize no deity. These Eastern philosophies are a kind of applied psychology which might explain their popularity in contemporary Western culture.

Each philosophy is useful, yet none are complete all the time in every situation. No one philosophy can be all things to all people. Therefore, the rational course of action is to consider them together and look for underlying common factors that may be present. We make allowances for superficial differences in the way they are expressed, since each is a product of the culture that created it.

Some discretion and judgment are required to know how best to apply them. As you will see, they can contradict each other, for example moral relativism and universalism. On the one hand Relativism says that right action is determined by circumstances, while Universalism says that right action is determined by principle, regardless of circumstance.

RELATIVISM

Relativism holds that moral or ethical propositions do not reflect objective and/or universal moral truths, but instead make claims relative to social, cultural, historical or personal circumstances. Right action is determined on a case-by-case basis, being dependent on who is involved and a host of situational factors.

Relativism is differentiated into *subjective* and *cultural* relativism.

SUBJECTIVE RELATIVISM

A personal and subjective moral core lies or ought to lie at the foundation of a person's moral acts. This is essentially an inward-looking approach to morality, with each person being their own ultimate authority on what is right action.

In the subjective view, public morality is merely a reflection of social convention. Only personal, subjective morality expresses true authenticity. The French philosopher Jean-Paul Sartre is a foremost exponent of this approach to morality.

CULTURAL RELATIVISM

In contrast to the subjective approach, in Cultural Relativism a person's beliefs and activities are understood in the context of his or her culture. Right action is

defined by cultural convention and exists as a commonly understood principle in that culture.

Since morality varies from culture to culture, with each culture having an equal claim as to what constitutes right action. This approach to morality grew out of the work of anthropologist Franz Boaz in the early 20th century. Anthropologists, if they are to properly understand a culture must not impose moral judgments on their practices even if they differ from the anthropologist's own cultural beliefs.

A criticism of both subjective and cultural relativism is that they differ fundamentally and take no account of the other. Arguably, both approaches have merit, and both deserve to be recognised, but not to the exclusion of the other. A blended approach that could simply be called Relativism is proposed, which takes both subjective and objective factors into account and tries to reconcile them. This would lead to a more balanced understanding of a given situation.

KANTIANISM

Immanuel Kant (1724–1804) was a notable German philosopher who argued with good reason that morality be based on a standard of rationality that he dubbed the *Categorical Imperative* (CI). Immorality is therefore a violation of the CI and is irrational.

The importance of being rational is a consistent theme in Western philosophy. The Stoic philosophers of classical Greece emphasises the use of logic and rationality to overcome the tendency to act emotionally and irrationally.

Kant's position can be summed in his categorical imperatives which form the foundation of his work.

Categorical Imperative (First Formulation): Act only according to that maxim whereby you can at the same time will that it should become a universal law. Ask yourself, *if I do this, would be all right if everyone did it?*

Categorical Imperative (Second Formulation): Act so that you always treat both yourself and other people as ends in themselves, and never only to an end. Ask yourself, *am I exploiting someone to get what I want?*

The first formulation is the foundation of the *Universalist* view of morality that if something is right, then it is always right, all the time. To make a special case exception is little more than a sense of selfish entitlement.

The second formulation lies at the heart of much of what the Ethical Technologist is about; the importance of helping people to come to a fuller expression of their potential. This position maintains that whatever you do must not harm other people or diminish them by treating them to an end.

Kant's theory belongs to the broader category of non-Consequentialist theories that determines whether an action is right or wrong by considering the underlying

rule or principle that motivates the action. Social Contract theory is another member of this category.

UTILITARIANISM

Utilitarianism asserts that moral behaviour is that which promotes happiness or pleasure; that which creates the greatest good and/or does the least harm.

A wrong act is one which produces unhappiness or suffering. The degree of 'wrongness' is determined by how much harm the act has caused. Therefore, the guiding principle in Utilitarianism is to do the thing that brings the greatest good to the greatest number.

Utilitarianism is sometimes known as a Consequentialist approach; if the outcome or consequence of an act is good, then the act itself is good. It is often used in the world of business and politics to achieve desired ends, sometimes incurring damage along the way. *The ends justify the means*. Though, the ends do *not* justify the means if significant harm is caused by doing so.

ACT UTILITARIANISM

With Act-utilitarianism the principle of utility is applied directly to each alternative act in a situation of choice. The right act is defined as the one which brings about the best results, or the least amount of harm.

Criticisms of this viewpoint to the difficulty of having full knowledge of the consequences of our actions.

Act-utilitarianism has been used to justify barbaric acts, for example suppose you could end a war by torturing children whose fathers are enemy soldiers to find out where the fathers are hiding.

Act utilitarianism is supremely pragmatic as it confines itself to a simple moral calculus; for example, if I can save 10,000 lives by killing one innocent person, the killing is a moral act.

RULE UTILITARIANISM

With Rule-utilitarianism the principle of utility is used to determine the validity of the *rules* of conduct, the moral principles that underlie.

For example, if we have a rule about keeping promises, it is because we have considered what the world would be like if people broke promises when they feel like it, compared with a world where people keep their promises. Moral behaviour is therefore defined by whether we follow the rules.

There are limits to how far Rule utilitarianism can be applied. When more and more exceptions to the rule are applied, it collapses into Act utilitarianism.

More general criticisms of this view argue that it is possible to generate *unjust rules* by resorting to the principle of utility. For example, slavery in ancient Greece might have been right if it led to an overall achievement of cultivated happiness at the expense of some mistreated individuals.

SOCIAL CONTRACT THEORY

Philosopher Thomas Hobbes argued that everybody living in a civilised society has implicitly agreed to (a) establish a set of moral rules to govern relations among citizens, and (b) establish a government capable of enforcing these rules. This is called the social contract.

In practical terms, Social Contract theory might also be construed to be a kind of reciprocal social obligation, society to the individual, and the individual to society. When individuals live in a society and enjoy the benefits of doing so (a place to live, meaningful work, the chance to raise a family in safety and so on), they have a reciprocal obligation to contribute to that society in whatever way they are best able to do. A person who takes and refuses to give according to their ability is little more than a parasite.

Social Contract theory belongs to the broader category of non-Consequentialist theories that determines whether an action is right or wrong by considering the underlying rule or principle motivating the action. Kant's theory is another member of this category.

MARCUS AURELIUS AND THE STOICS

Marcus Aurelius (full name Marcus Aurelius Antoninus Augustus, 121 – 180 AD) was an exceedingly rare individual; a genuine philosopher-king. His leadership is based on the often-misunderstood Stoic philosophy. The power and relevance of this philosophy is as potent today as it was when he was Roman Emperor (161 to 180AD).

Marcus Aurelius might have been a Roman, but his thinking had been shaped by the classical period of ancient Greece. Even today, classical Greek thinking is still at the foundations of Western civilisation.

Influenced by the earlier work of Socrates and Diogenes of Sinope, the Stoic school of philosophy was founded around 300 BC by Zeno of Citium. Speaking from beneath a painted portico (*Stoa Poikilē*) in Athens, signifying openness to anyone passing by Zeno taught that a wise person should not allow their emotions to rule them; instead, they should master their emotions and use logic to think rationally about how to behave in life. He urged his followers to carefully study the laws of Nature and to live in harmony with them. In this respect his ideas coincide with those of far distant Lao Tzu, the ancient Chinese philosopher who wrote the Tao Te Ching.

A central point in Stoic philosophy is the active relationship between the laws of Nature that rule the Cosmos, and human free will. A wise person derives maximum benefit and happiness in life by bringing his or her will into harmony with Nature. They come to know themselves, recognising that their inner nature (microcosm) is a representation of the outer macrocosm, or universe; it the same nature in both, only differing in scale.

Stoics conceived of the universe as being governed by Logos, what we today would think of as the Laws of Physics. Pure, abstract, these laws pervade the universe and make it behave in the way it does. The same informing principle resides in humans. Virtue is therefore gained by recognising this and working to harmonise one's inner self with the qualitatively similar outer world.

The Greek founders of Stoicism conceived of three interrelated elements that collectively make Philosophy. These are *logic*, *physics*, and *ethics*. Logic allows us to recognise truth when we see it, and to avoid making mistakes. Logic allows us to understand Physics, which is the way the world operates, the laws of Nature. Together, Logic and Physics allows us to practice Ethics, or moral behaviour that brings benefit.

Ethical behaviour is that which is in harmony with the unfolding laws of Nature. This unfolding is the cause of both pleasure and suffering in people. If we are to stay in accord with it, we must discipline our minds to become indifferent to suffering, accepting with grace that it is necessary and inevitable to suffer sometimes. This state of mind is called *apatheia*. Likewise, we must not become so attached to pleasure that we cannot relinquish it when it passes. The goal is to become self-sufficient, or *autarcheia*.

The Stoic therefore becomes equally indifferent to good fortune or bad, whether they are rich or poor, well-respected or despised. They understand that the approval or disapproval of others can exert undue influence to conform to values that may not be true. The Stoic does his or her duty in accordance with Nature as revealed by careful observation and logical enquiry. They do their duty regardless of whether it is easy or hard.

With its emphasis on duty and right action, Stoicism is therefore well-suited to the needs of those who would lead. It was used as a guide by the ruling class of Rome for centuries.

BUDDHISM & THE FOUR NOBLE TRUTHS

About the same time as the classical Greek philosophers were formulating their ideas a revolution in thought was taking place in northern India. Siddhartha Gautama, the man who would become the Buddha, or Awakened One, was formulating some ideas of his own. It is remarkable how similar in structure and meaning the philosophies of East and West at this time were. It is almost as if it was a good idea whose time had come to be brought into the world.

Buddhism is thought by many to be a religion, yet it recognizes no deity. In its basic form is an applied psychology expressed in the language of the time. It outlined a formula for how to become self-actualized. The foundation of Buddhist philosophy is the so-called Four Noble Truths and the Noble Eight-Fold Path. The eight-fold path aims to improve your (a) *Wisdom* by practicing right view and intention, (b) *Ethical conduct* by practicing right speech, action and livelihood, and (c) *Mental capabilities* by practicing right effort, mindfulness and concentration. We shall examine more closely the three aspects of ethical conduct.

RIGHT SPEECH

Words are powerful. Words can make or break a person's life, start wars or bring peace. Words can indeed be mightier than the sword, as great orators through the ages have proven. Right speech (including written words) is therefore the principle of expressing oneself in a way that enhances the quality of people's lives and does no harm. It means to refrain from (a) lies and deceit, (b) malicious language (including slander), (c) angry or offensive language, and (d) idle chit-chat (including gossip). Notice the correspondence between this principle and the prime ethical value in the ACS code of conduct to act in ways that improves people's quality of life.

Therefore, tell the truth, speak with warm gentleness when you do speak, and refrain from speaking when you have nothing important to say.

RIGHT ACTION

Right action can be defined open-endedly by prescribing what a person should *not* do. That then leaves the field wide open for choice. Broadly, right action means refraining from (a) harming any sentient creature, (b) stealing, and (c) sexual misconduct. Doing no harm to others covers a very broad range of behaviours. The worst a person can do is to take the life of sentient creatures, hence many Buddhists are vegetarians. Not stealing includes all forms of robbery, theft, deceit and fraud; essentially taking what you have not earned the right to have.

The ethical person is therefore kind and compassionate in their dealings with the world. They respect other people's property, and do not engage in sexual behaviour that harms another either at a physical or emotional level.

RIGHT LIVELIHOOD

Right livelihood is about earning one's living in ways that does no harm to others. Of all the possible ways a person might earn money, they should avoid those that exploit people's weaknesses.

Right livelihood means one should refrain from any employment that is contrary to the principles of right action and right speech, including but not limited to (a) trading in weapons, (b) trading in living beings, including slavery, prostitution and raising animals for slaughter, (c) butchery and meat processing, and (d) trading in drugs and poisons, including alcohol and recreational drugs.

LAO TZU & THE TAO TE CHING

The *Tao Te Ching* is said to have been written by Lao Tzu (604 – 531 BCE), the philosopher and Custodian of the Imperial Archives in the time of the Chou Dynasty in ancient China. It is uncertain when Lao Tzu was born or died, but he is said to be a contemporary of Confucius (551–479 BCE).

Central to Taoist philosophy is the avoidance of extremes, to always seek the middle way on our journey through life. Find the middle ground between the extremes and occupy that space and in doing so have the fewest consequences to deal with. The principle at work here is that extreme action always results in an equal and opposite reaction. As a pendulum swings to one extreme, it will always swing to the other extreme in equal measure. Following the middle path reduces the “swing” to a minimum. Only through this practice can harmony in society be achieved.

We are encouraged to sense the world around us directly and to contemplate our impressions deeply. It advises against relying on the structures and belief systems that have been created by others and put forward as orthodox truth. Such ideologies remove us from a *direct* experience of life and effectively cut us off from our intuition.

The middle path requires us to develop an awareness of the physical forces that shape our world. Such forces operate uniformly at all levels from the largest to the smallest. They operate in the universe as a whole and in the minds and lives of individual people. An understanding of these natural laws and the forces they direct give us the power to influence events in the world without force. Influence is achieved through guiding rather than coercion. The objective is always to avoid taking action that will elicit strong counter-reactions. In Nature, an excessive force in a particular direction always triggers the growth of an opposing force, and therefore the use of force cannot be the basis for establishing an enduring social condition.

We come to understand that everything in the universe is impermanent, in a state of change. The emotional and intellectual structures that we build for ourselves to feel secure are likewise subject to change by external forces that are largely beyond our control. The challenge is to accept the inevitability of change and not waste our energy trying to prop up these impermanent structures, defending them against criticisms, and trying to convince others to believe in them so that they might become recognised as permanent truth.

Lao Tzu wrote the Tao Te Ching from the point of view of the “superior man”, the person who is transcending their base nature by consciously improving their lives through wise choices.

THE ETHICS OF CONFUCIUS

Confucius (551 BC – 479 BC) established a system of personal and governmental morality that has endured for 2,500 years. It concerns itself with correctness in

social relations during a time of great disturbance. The work of Confucius and Lao Tzu are both aimed at achieving social harmony and coherence to remedy the rampant chaos of the times.

Three key principles are emphasized in Confucius' teachings: the principles of *Li*, *Jen* and *Chun-Tzu*.

The term *Li* has several meanings; it is often translated as propriety, reverence, courtesy, ritual or ideal conduct. It is what Confucius believed to be the ideal standard of religious, moral, and social conduct.

The second principal *Jen* is the fundamental virtue of Confucian teaching, the virtue of goodness and benevolence. It is expressed through recognition of value and concern for others, no matter their rank or class. *Jen* is summarised as the Silver Rule: *Do not do to others what you would not like them to do to you*. (Analects 15:23) *Li* provides the structure for social interaction. *Jen* makes it a moral system.

The third principle, *Chun-Tzu* describes the idea of the true gentleman (should not be seen as gender-specific). This is the person who lives according to the highest ethical standards. The gentleman displays five virtues: *self-respect, generosity, sincerity, persistence, and benevolence*.

As a son, he is always loyal; as a father, he is just and kind; as an official, he is loyal and faithful; as a husband, he is righteous and just; and as a friend, he is faithful and tactful. In today's world, the words *she, mother* and *wife* could be substituted for he, father and husband.

THE UNIVERSAL MORAL CODE

To identify underlying moral principles across cultures, Kent W. Keith puts forward these two lists, one expressed in 'do this' form and the other in the 'do not do this' form. These principles are found embedded in the moral codes of diverse cultures. The first list, *do no harm*, essentially says, whatever you do, do not do these. The list can be seen as the foundation upon which a positive set of behaviours can be based, the *do-good* list.

Do no harm. *Do not do to others what you would not like them to do to you, do not lie, do not steal, do not cheat, do not falsely accuse others, do not commit adultery, do not commit incest, do not physically or verbally abuse others, do not murder, do not destroy the natural environment upon which all life depends.*

Do good. *Do to others what you would like them to do to you, be honest and fair, be generous, be faithful to your family and friends, take care of your children when they are young, take care of your parents when they are old, take care of those who cannot take care of themselves, be kind to strangers, respect all life.*

THE GOLDEN RULE

Perhaps the most often quoted moral absolute is the so-called *Golden Rule*. Beyond the religious or even the philosophical, this principle is recognisable in Physics as Newton's second law of motion; *the mutual forces of action and reaction between two bodies are equal, opposite and collinear*. What we do elicits an equal and opposite reaction. As humans, we are not separate from the laws of Physics. If we take the position that we are not masochists and we want good things to happen to us, then we have the Golden Rule:

Christianity. Therefore, all things whatsoever ye would that men should do to you, do ye even so to them: for this is the law and the prophets. *Matthew 7:12*

Confucianism. Do not do to others what you would not like yourself. Then there will be no resentment against you, either in the family or in the state. *Analects 12:2*

Buddhism. Hurt not others in ways that you yourself would find hurtful. *Udana-Varga 5,1*

Hinduism. This is the sum of duty; do naught onto others what you would not have them do unto you. *Mahabharata 5,1517*

Islam. No one of you is a believer until he desires for his brother that which he desires for himself. *Sunnah*

Judaism. What is hateful to you, do not do to your fellowman. This is the entire Law; all the rest is commentary. *Talmud, Shabbat 31d*

Taoism. Regard your neighbour's gain as your gain, and your neighbour's loss as your own loss. *Tai Shang Kan Yin P'ien*

Zoroastrianism. That nature alone is good which refrains from doing another whatsoever is not good for itself. *Dadisten-I-dinik, 94,5*

COMPARISON OF KNIGHTS' CODES

The Japanese Samurai and the chivalric knights of medieval Europe were separated by a great distance, and likely had no contact with each other. Yet independently they arrived at noticeably similar codes of ethical conduct as seen below. Interestingly, there is correspondence with the Australian Computer Society's code of professional conduct too.

Samurai Code	Knight's Code	ACS Code of Prof Conduct
Courage	Courage	Objectivity and Independence Integrity

Loyalty	Loyalty	Confidentiality
Honor	Nobility	Subordinates Responsibility to your Client
Honesty/ Trust	Defense Justice	The Public Interest The Image of the Profession Promoting Information Technology
	Prowess Franchise / replicate	Competence Keeping Up To Date
Rectitude	Faith	Right action
Respect	Humility	Respect for stakeholders
Benevolence	Generosity	Do what is in best interests of client and public

DIGITAL ETHICS & RESPONSIBLE AI

Artificial intelligence (AI) is transforming the world in many ways, from improving health care and education to enhancing productivity and innovation. However, AI also poses significant challenges and risks, such as potential bias, discrimination, privacy breaches, security threats, and ethical dilemmas.

How can we ensure that AI is used for good and not evil? How can we design and implement AI systems that are fair, transparent, accountable, reliable, and respectful of human values?

Follow the AI Ethics Principles

Many countries and organizations have developed ethical principles or guidelines for AI, such as [Australia's 8 AI Ethics Principles](#), the [IEEE's Ethically Aligned Design](#), or the [Berkman Klein Centre's report on ethical principles in eight categories](#). These principles provide a common framework and a shared language for understanding and addressing the ethical issues of AI. They also help to build public trust and consumer loyalty in AI-enabled services.

The principles cover various aspects of AI, such as human wellbeing, human-centred values, fairness, privacy protection and security, reliability and safety, transparency and explainability, contestability, and accountability. By following these principles and committing to ethical AI practices, you can achieve safer, more reliable and fairer outcomes for all stakeholders.

5.6. ETHICAL AI & ALGORITHMIC BIAS

ETHICAL AI & ALGORITHMIC BIAS

Artificial intelligence (AI) is a powerful technology that can enhance decision-making, optimize processes, and create new value in various domains.

However, AI also poses ethical challenges that need to be addressed by IT professionals who design, develop, deploy, or use AI systems. One of the most pressing ethical issues in AI is algorithm bias, which is a kind of error or unfairness that can arise from the use of AI.

WHAT IS ALGORITHM BIAS AND WHY DOES IT MATTER?

Algorithm bias is a situation where an AI system produces outcomes that are systematically skewed or inaccurate, often resulting in unfair or discriminatory treatment of individuals or groups based on their characteristics, such as race, gender, age, or disability. Algorithm bias can have negative impacts on human rights, such as the right to equality, privacy, dignity, and justice.

Algorithm bias can occur for several reasons, such as:

- The data used to train or test the AI system is not representative of the target population or context, leading to overfitting or underfitting.
- The algorithm design or implementation is flawed or contains hidden assumptions or preferences that favour certain outcomes or groups over others.
- The interpretation or application of the AI results is influenced by human biases or prejudices, either intentionally or unintentionally.

Some examples of algorithm bias in real-world scenarios are:

- A facial recognition system that performs poorly on people of colour, resulting in false positives or negatives that can affect security, access, or identification.
- A hiring system that screens candidates based on their resumes but excludes qualified applicants who have non-traditional backgrounds or names that indicate their ethnicity or gender.
- A credit scoring system that assigns lower scores to people who live in certain neighbourhoods or have certain occupations, affecting their access to loans or insurance.

HOW CAN IT PROFESSIONALS ADDRESS ALGORITHM BIAS?

As IT professionals who are involved in the development or use of AI systems, we have a responsibility to ensure that our AI systems are ethical and aligned with human rights principles. We can do this by following some best practices, such as:

- Conducting a thorough analysis of the data sources, algorithms, and outcomes of the AI system, and identifying potential sources and impacts of bias.
- Applying appropriate methods and tools to mitigate or reduce bias in the data collection, processing, analysis, and validation stages of the AI system.
- Implementing transparency and accountability mechanisms to explain how the AI system works, what data it uses, what assumptions it makes, and what results it produces.
- Engaging with relevant stakeholders, such as users, customers, regulators, and experts, to solicit feedback, address concerns, and ensure compliance with ethical standards and legal requirements.
- Monitoring and evaluating the performance and impact of the AI system on an ongoing basis and updating or correcting it as needed.

Algorithm bias is a serious ethical challenge that can undermine the trustworthiness and value of AI systems. IT professionals have a key role to play in ensuring that our AI systems are ethical and respect human rights. By following some best practices, we can create AI systems that are fair, accurate, and beneficial for all.

THE IMPORTANCE OF ETHICAL AI POLICIES

AI poses significant challenges and risks, such as potential bias, discrimination, privacy breaches, and accountability gaps. Therefore, it is essential to develop and implement ethical AI policies that can ensure the safe, secure, and responsible use of AI for the benefit of individuals, society, and the environment.

WHAT ARE ETHICAL AI POLICIES?

Ethical AI policies are guidelines or principles that aim to align the design, development, and deployment of AI systems with human values and rights. Ethical AI policies can help to:

- Achieve safer, more reliable, and fairer outcomes for all stakeholders affected by AI applications.
- Reduce the risk of negative impacts or harms caused by AI systems.
- Build public trust and confidence in AI systems and their providers.
- Encourage innovation and competitiveness in the AI sector.
- Comply with existing laws and regulations related to AI.

Ethical AI policies can be developed and implemented by various actors, such as governments, businesses, researchers, civil society, and international organizations. Ethical AI policies can also vary in their scope, level of detail, and enforceability.

EXAMPLES OF ETHICAL AI POLICIES

Several countries and regions have developed or are developing ethical AI policies to guide their AI strategies and initiatives. For example:

Australia has published its AI Ethics Framework, which includes eight voluntary AI Ethics Principles that cover human, social, and environmental wellbeing; human-centred values; fairness; privacy protection and security; reliability and safety; transparency and explainability; contestability; and accountability.

The **European Union** has proposed its Artificial Intelligence Act, which is a comprehensive legal framework that aims to regulate high-risk AI systems and promote trustworthy AI based on four ethical principles: respect for human dignity and autonomy; prevention of harm; fairness; and democratic values.

The **United States** has issued its Executive Order on Maintaining American Leadership in Artificial Intelligence, which directs federal agencies to foster public trust and confidence in AI technologies by promoting reliable, robust, trustworthy, secure, portable, and interoperable AI systems.

In addition to governments, many private sector companies have also adopted their own ethical AI policies or principles to demonstrate their commitment to responsible AI practices. For example:

Microsoft has established its *Responsible AI Standard*, which is a set of requirements and processes that help its teams design, develop, deploy, and operate AI systems in a manner consistent with its six ethical principles: fairness; reliability and safety; privacy and security; inclusiveness; transparency; and accountability.

Google has published its *Responsible AI Practices*, which is a collection of best practices and tools that help its engineers build AI systems that are aligned with its seven principles: socially beneficial; avoid creating or reinforcing unfair bias; be built and tested for safety; be accountable to people; incorporate privacy design principles; uphold high standards of scientific excellence; and be made available for uses that accord with these principles.

MITIGATING BIAS

IDENTIFY AND ASSESS POTENTIAL SOURCES OF BIAS

The first step to mitigate bias is to identify and assess the potential sources of bias in the IT system or decision. This can be done by conducting a thorough analysis of the data, algorithms, processes and outcomes involved in the system or decision. Some questions to ask are:

- What are the objectives and criteria of the system or decision?
- What are the data sources, methods and quality of the data collection and processing?

- What are the assumptions, limitations and trade-offs of the algorithms and models used?
- How are the results interpreted, communicated and acted upon?
- Who are the stakeholders, beneficiaries and potential victims of the system or decision?
- What are the ethical, legal and social implications of the system or decision?

Some tools that can help with this step are:

- IBM's AI Fairness 360 toolkit, which provides a set of metrics, algorithms and visualizations to detect and mitigate bias in datasets and machine learning models.
- IBM's AI Factsheets, which provide a standardized way to document the characteristics, capabilities and limitations of AI systems.
- IBM Watson OpenScale, which provides a platform to monitor, explain and improve AI performance, fairness and compliance.

IMPLEMENT BIAS MITIGATION STRATEGIES

The second step is to implement bias mitigation strategies that address the identified sources of bias. This can be done by applying various techniques, such as:

- Data augmentation, transformation or sampling to improve the representativeness, diversity and balance of the data.
- Algorithm selection, modification or regularization to reduce the complexity, opacity or sensitivity of the models.
- Human review, feedback or intervention to provide oversight, validation or correction of the results.
- Stakeholder engagement, consultation or participation to ensure transparency, accountability and inclusiveness of the system or decision.

Some examples of bias mitigation strategies are:

Conflicts and Biases in the Boardroom, which provides guidance on how to address conflicts of interest and common biases that impact board decisions.

Algorithmic bias detection and mitigation: Best practices ... - Brookings, which provides policy recommendations on how to detect and mitigate algorithmic bias in consumer harms.

AI Ethics Part 2: Mitigating bias in our algorithms - CMO, which provides best practices on how to build fairness and bias metrics and run a model governance process.

EVALUATE AND MONITOR BIAS MITIGATION OUTCOMES

The third step is to evaluate and monitor the outcomes of the bias mitigation strategies. This can be done by measuring, testing and reporting on the performance, fairness and trustworthiness of the system or decision. Some questions to ask are:

- How effective are the bias mitigation strategies in achieving the objectives and criteria of the system or decision?
- How fair are the system or decision outcomes for different groups of stakeholders?
- How trustworthy are the system or decision processes and results for different audiences?
- How robust are the system or decision against changes in data, algorithms or contexts?
- How adaptable are the system or decision to new requirements, feedback or challenges?

Some tools that can help with this step are:

- IBM Watson OpenScale, which provides a platform to monitor, explain and improve AI performance, fairness and compliance.
- IBM Watson Discovery, which provides a service to analyse text data for sentiment, emotion, tone and personality insights.
- IBM Watson Assistant, which provides a service to build conversational agents that can interact with users and provide feedback.

Mitigating bias in IT governance is a complex and ongoing challenge that requires a holistic and proactive approach. By following these three steps - identify and assess potential sources of bias, implement bias mitigation strategies, and evaluate and monitor bias mitigation outcomes - IT leaders can ensure that their systems and decisions are more ethical, fair and trustworthy.

ETHICAL AI IN CRITICAL DOMAINS

Certain domains, such as criminal justice and healthcare, hold significant ethical ramifications for AI usage. Biased algorithms in predictive policing can lead to unjust targeting, while healthcare AI biased against certain demographics might exacerbate health disparities. Ethical AI policies should emphasize thorough evaluation and validation of algorithms in these critical contexts.

IDENTIFY THE ETHICAL PRINCIPLES FOR AI

The first step to build ethical AI is to identify the ethical principles that should guide its development and use. There are many sources of ethical principles for AI, such as the OECD Principles on AI, the World Economic Forum's 9 Ethical AI Principles for Organizations, or the Ethics of Artificial Intelligence course by Coursera. These principles usually include values such as fairness, transparency, accountability, privacy, security, human oversight, and social good.

However, these principles are not enough by themselves. They need to be translated into concrete norms and practices that can be implemented and governed in specific contexts and domains. For example, what does fairness mean for an AI system that diagnoses diseases or recommends treatments? How can transparency be achieved for an AI system that predicts criminal behaviour or assesses legal risks? How can accountability be ensured for an AI system that controls autonomous vehicles or drones?

To answer these questions, we need to conduct a thorough ethical analysis of the AI system and its impacts and implications for the stakeholders involved.

CONDUCT AN ETHICAL ANALYSIS OF THE AI SYSTEM

The second step to build ethical AI is to conduct an ethical analysis of the AI system and its impacts and implications for the stakeholders involved. This analysis should consider the following aspects:

- **The purpose and goals of the AI system.** What problem does it aim to solve? What benefits does it provide? What risks does it entail?
- **The data and algorithms of the AI system.** What data is used to train and test the AI system? How is it collected, processed, stored, and shared? What algorithms are used to analyse the data and generate outputs? How are they designed, validated, and updated?
- **The outputs and outcomes of the AI system.** What outputs does the AI system produce? How are they interpreted and used? What outcomes do they lead to? How are they measured and evaluated?
- **The stakeholders of the AI system.** Who are the stakeholders of the AI system? How are they affected by its outputs and outcomes? What are their needs, preferences, values, and expectations?
- **The ethical issues of the AI system.** What ethical issues arise from the AI system's purpose, data, algorithms, outputs, outcomes, and stakeholders? How can they be identified, prioritized, and addressed?

To conduct this analysis, we need to use critical skills and methods that can help us clarify and ethically evaluate the AI system in different domains of life. We also need to consult with relevant experts and stakeholders to ensure that we capture their perspectives and concerns.

IMPLEMENT ETHICAL SOLUTIONS FOR THE AI SYSTEM

The third step is to implement ethical solutions for the AI system that can address the ethical issues identified in the previous step. These solutions may include:

- **Ethical design.** Applying ethical principles and values in the design process of the AI system, such as user-centered design or value-sensitive design.

- **Ethical development.** Applying ethical standards and guidelines in the development process of the AI system, such as code of ethics or best practices.
- **Ethical testing.** Applying ethical criteria and methods in the testing process of the AI system, such as audits or impact assessments.
- **Ethical deployment.** Applying ethical rules and regulations in the deployment process of the AI system, such as policies or laws.
- **Ethical governance.** Applying ethical mechanisms and structures in the governance process of the AI system, such as oversight boards or ethics committees.

To implement these solutions, we need to use appropriate tools and techniques that can help us operationalize ethics in practice. We also need to monitor and evaluate the impacts and outcomes of the AI system on a regular basis.

Ethical AI is not only a moral duty but also a strategic advantage for organizations that want to create value and trust with their customers, employees, partners, regulators, and