LUISS

Master degree in Strategic Management

Chair of: “Digital transformation and innovation”

“Corporate ethics, innovation and AI”

Edoardo Guidi

778181

Prof. Alessandro Mechelli

CANDIDATE

SUPERVISOR

Academic year 2024/2025

TABLE OF CONTENTS

INTRODUCTION

Over time, the link between business, ethics, and innovation has become increasingly decisive. Companies can no longer focus solely on profit, as rapid technological change and global challenges, such as sustainability, inequality, and digital disruption, demand a broader responsibility. By embedding ethical principles into innovation, businesses not only create competitive advantage but also build trust and contribute to long-term social progress, proving that economic growth and societal well-being must go hand in hand.
The traditional idea of the firm as an institution solely aimed at the maximization of profits has become more and more detached in the last two decades: social legitimacy is more correlated with the way companies innovate, manage technological change, and cope with the ethical issues that the ensuing processes bring.
Against this backdrop, company ethics and innovation no longer constitute separate domains, but two aspects increasingly entwined. Thinking ethically has moved out of the domain of the compliance field or altruism in order to be part of strategic management. At the same time, innovation, in the past viewed only as a tool of competitiveness and expansion, now requires convergence with sustainability, fairness and transparency principles. The formation of this relationship is particularly clear in the digital economy, with new technologies, especially artificial intelligence (AI), raising intense ethical and social issues.
The breakthrough's speedup intensified the challenge. Artificial intelligence is far from a neutral instrument: it is an emerging socio-technical system whose design and use embed choices, values, and priorities. Algorithms may solidify or dampen discrimination, increase efficiency but also cultivate opacity, maintain the environment or, instead, extend inequalities. For multinationals, the problem of the governance of AI is that of the ultimate frontier: how can innovation contribute to competitiveness while being sensitive to core rights, protecting trust, and constructing shared value with stakeholders.
This question is central to the current thesis, whose research question is:
"How can ethical principles be implemented in practice through the introduction and use of Artificial Intelligence by international companies?"
The research objective is twofold. Firstly, the study intends to follow the historical and conceptual evolution of company ethics from the origins in business responsibility until the interface with innovation and digital transformation. Secondly, the study seeks to identify how companies can translate general ethical values into actionable management tools, policies and practices, and particularly in the field of artificial intelligence.

Following this objective, the thesis is structured in four parts. The first part is the presentation of the conceptual framework of business ethics, which analyzes its development and integration in company strategy. The second part is the analysis of the link between ethics and innovation with special reference to the challenges posed by responsible innovation, intellectual property, sustainability and long-term competitiveness. The third part is the analysis of the ethical challenges of artificial intelligence, with special reference to definitions, regulation, zones at risk, and management tools of governance. The fourth part is the presentation of the case study dedicated to Microsoft, an international giant that has implemented one of the most developed responsible AI frameworks, both spotting the best practices as well as the open tensions.

The relevance of this thesis is that it succeeds in linking the analytical examination of ethical theories, the technological innovation dynamics, the patterns of regulation and the practical business practices. The thesis contributes to the current debate regarding how businesses should contribute responsibly in the future development of technology. The long-term perspective is to prove that ethics is no external constraint but can be a creative strategic instrument, a competitiveness advantage as well as digital world legitimacy.

PART ONE: Business ethics

1.1. Definition of ethics and ethical theories

Ethics originated as a branch of philosophy, deriving from the Greek word *ethos*, a term attributable to Aristotle (384-322 BC) who coined the expression *ethike theoria* to identify the type of knowledge that investigates human practice (*praxis* as action). Therefore, ethics can, at first approximation, be defined as the branch of philosophy that deals with conduct[1]. Ethics is therefore a tool for reflecting on human action that should allow us to discern between good and bad behaviour and, therefore, good and bad choices. However, as we will see later, ethics is not only a criterion, it is also the goal towards which human behaviour should strive.

Ethics as a philosophical science can be divided into two broad categories:

- Normative ethics: focuses on the study of what is right or wrong, good or bad, correct or incorrect. It is also known as prescriptive ethics as it provides general guidelines and principles on human behaviour.
- Metaethics: deals mainly with the study of moral language[2].

Normative ethics, which is the one we are interested in, can, in turn, be broken down into two subgroups:

- Ethical theory: proposes, examines and compares all the theories that have been proposed over time with reference to "how and why we should behave in a certain way" or "what kind of action/behaviour we should take/have". In essence, it elaborates ethical theories.
- Practical ethics: examines the various "operational applications" that have been implemented in various fields of science. (Examples of practical ethics are: business ethics, bioethics, environmental ethics, etc.)

Ethical theory and practical ethics interact and influence each other continuously.

Ethical theory can then be divided into two groups, within which the main ethical theories are classified:

- The ethics of duties or conduct or behaviourism.
- The ethics of character or virtues.

[1] Aristotele, 2005. *Etica Nicomachea*. Milano: Bompiani.
[2] Moore, G. E., 1903. *Principia ethica*. London: Cambridge University Press.

We have said that ethics is that part of philosophy that deals with human behaviour, but it is not easy to understand if and when a behaviour is good or bad.

Over time, various theories have developed in the field of ethics, proposing different solutions:

- Consequentialist ethical theories: it is claimed that one must choose the action that produces the greatest good. There are two types:
  - Utilitarian theories: the end (the good) is for everyone
  - Egoistic theories: the end (the good) is reserved only for those who perform the action.
- Deontological ethical theories: these state that we must choose the action that is most correct in itself.

The birth of utilitarianism and its systematization originated in the 1700s thanks to the work of J. Bentham (1749-1832) and J. S. Mill (1806-1873).

Utilitarianism is a consequentialist or teleological approach, rather than an a priori or end-state oriented one, in the sense that it judges and chooses what to do or not to do based on the consequences (results) of the action itself[3].

For Bentham and Mill "The ultimate goal of ethics is the good of humanity, of the community. This means that an action is morally right (correct) if it allows the good of the community to be achieved more than any other action that could be taken"[4].

Selfishness also judges an action or behaviour based on its consequences, butjudges them solely from the point of view of the person performing the action (moral agent) or, in any case, from the point of view of a small group of people (their family, class, race, etc.).

The egoist is focused on himself, on his own good, on his own happiness (personal interest): his "ego" prevails over everything and everyone, and he comes to believe that not only is his behaviour guided by his own interests, but that the behaviour of others should also be guided by his own interests. The egoist therefore decides and acts according to a purely selfish calculation.

Deontology derives from the Greek word '*δέον*', which means 'duty'. The first formulation of deontological theories can be attributed to I. Kant (1724-1804) and is contained in three volumes[5].

[3] Deigh, J., 2012. *Etica. Un'introduzione*. Milano: Apogeo.
[4] Mill, J. S., 1861. *Utilitarianism*. London: Parker, Son, and Bourn.
[5] Kant, I., *Groundwork of the Metaphysics of Morals*, 1785., Kant, I., *Critique of Practical Reason*, 1788., Kant, I., *The Metaphysics of Morals*, 1797.

According to Kant, moral behaviour cannot be defined in terms of the consequences it produces, but must be objective and universal and, as such, valid at all times and in all places. In his view, the end does not justify the means[6].

The ethics of character or virtue answers the question: "What kind of people should we be or become?", or "How should we live?".

What matters is not the goodness or rightness of the action or what results from the action, but the goodness and rightness of the person themselves[7].

Ethics and economics originated in the same field, both falling within the realm of moral philosophy, but the two sciences subsequently separated. This separation became concrete with the birth of modern economics[8], as the importance of the ethical approach weakened substantially as modern economics evolved.

Economics can be defined as the science that studies human behaviour when, given a hierarchy of objectives, choices must be made about scarce resources that can be used for alternative purposes[9]. It therefore deals with the problem of choosing how to adapt the few resources available to humans (including time) to the multiple ends they wish to achieve.

B. Mandeville, a Dutch physician and philosopher (1670-1733), stated that "in order to achieve social well-being, it was necessary for individuals not to behave ethically"[10].

A capitalist society, such as the English society of the 18th century in which he lived, needs individuals to behave selfishly in order to be and remain rich.

To explain these concepts, Mandeville represented society as a vast beehive full of bees[11], whose numbers also contributed to the common prosperity. They lived like humans (characterized by all the private vices and contradictions that distinguish human beings) and performed, on a small scale, all our actions.

It was these vices that allowed the hive to flourish; in other words, selfish action promoted the public good.

Nevertheless, some bees, who had gained great fortune from private vice, began to complain about these vices and injustices and invoked the gods to change everything towards more correct private behaviour.

---

[6] Kerstein, S. J., 2002. *Kant's search for the supreme principle of morality*. Cambridge: Cambridge University Press.
[7] Aristotele, *Etica Nicomachea*, cit.
[8] Robbins, L., 1947. *Saggio sulla natura e l'importanza della scienza economica*. Torino: UTET.
[9] Robbins, L., *Saggio sulla natura e l'importanza della scienza economica*, cit.
[10] Mandeville, B., 1988. *The fable of the bees: or, private vices, publick benefits*. (a cura di F. B. Kaye, 2 voll.). Indianapolis: Liberty Fund.
[11] Mandeville, B., 1968. *La favola delle api*. In *Grande antologia filosofica*, vol. XIV. Milano: Marzorati.

Indignant, Jupiter heard this complaint and cried out: "From this moment on, honesty will take hold of all their hearts"[12]; however, the consequences of these changes towards virtuous private behaviour were sad: the court was depopulated, lawyers lost their jobs, blacksmiths who produced chains, locks and prison bars were no longer needed, and the same happened to jailers, prison guards and their assistants.

The abandonment of private vices quickly destroyed the most prolific of hives.

This teaches us that it is wrong to indulge in moral behaviour if we want to live in a prosperous world; a society must indulge in vice because satisfying the vices of the wealthier classes facilitates the circulation of money and therefore consumption, investment and employment, bringing wealth.

Pure economics hypothesizes that the decision-maker, *homo economicus*, directs behaviour towards maximizing their utility function under a constraint of resource availability and by ordering their preferences hierarchically. In other words, this means that modern economic theory bases its model on the ethical theory of utilitarianism, pushing it far beyond the intentions of utilitarian moralists. For pure economists, in fact, it is selfishness that guides choice and, therefore, the utility that is to be maximized is only that of the person acting[13].

Modern economics is therefore characterized by a diabolical combination of individual selfishness and philosophical utilitarianism, in which the only parameter that matters for decision-making is the value achieved in terms of utility, whether individual or collective, the latter being understood as the sum of individual selfishness[14].

According to A. Sen, the model of rationality adopted by modern economics is based on assumptions that, paradoxically, may not be consistent with rationality itself; it is not certain that rational behaviour is true, i.e. consistent with actual behaviour, or, at the very least, it is not certain that rational behaviour is the only conceivable behaviour; therefore, it is not true that choosing not to maximize one's personal interest exclusively is irrational behaviour[15].

The rationality hypothesized by economists erroneously disregards all those moral aspects that characterize human beings and their decision-making process. The real world constantly demonstrates how human choices are often not dictated by pure and selfish rationality[16].

Individuals who, in addition to being motivated by self-interest, also act on the basis of what Sen calls *sympathy* (feelings of compassion for others, discomfort caused by the suffering of others) and *commitment* (inner commitment, obligation, sense of duty).

[12] Mandeville, B., *La favola delle api*, cit., p. 138
[13] Simon, H., 2001. *Il comportamento amministrativo*. Bologna: Il Mulino.
[14] Caltagirone, C., 2017. *Amartya K. Sen. Tra economia ed etica*. Roma: Studium.
[15] Sen, A., 2006. *Etica ed economia*. Bari: Laterza.
[16] Sen, A., *Etica ed economia*, cit.

This could lead the decision-maker to make choices that, in the awareness of the other, could be maximizing at all.

It therefore becomes necessary to develop a different, more correct, more just approach, i.e. an approach that tends to reconcile economics with ethics.

What matters is not economic or material well-being (*wel-fare economics*) but well-being (*well-being*), people's quality of life[17].

What matters, therefore, is what people who are free to choose can do and be with the resources available to them; in other words, opportunities.

1.2. Stakeholder theory

The definition of stakeholder can be attributed to R. E. Freeman, professor of philosophy and holder of the chair of business administration. In his well-known text *Strategic Management*, he writes: "The stakeholder of an organization is, by definition, a group or individual who can influence or be influenced by the achievement of the company's objectives"[18].

The term stakeholder therefore refers to any person or group that has a particular '*stake*' in the company, i.e. that has a stake in the company.

Freeman's definition is typically a general one, and it has been followed by many other definitions. In particular, the literature distinguishes between narrow definitions, which consider only some (the main, vital) stakeholders, and broad definitions, which tend to include all categories that may have an interest in the company (e.g. the environment, competitors, the media)[19].

The introduction of the concept of stakeholders is intended to convey the idea that the economic entity of a company (which is also a stakeholder) must take into account not only the owners (shareholders) in its actions, but also other entities that are more or less directly involved in the company's activities.

Freeman's definition of stakeholder is contained in a text dedicated to strategic management; therefore, it is clear that, in such a strategic context, the stakeholders that must be taken into consideration by the company are only those that are "useful" for the implementation of the business strategy in a dynamic environment.

[17] Caltagirone, C., Amartya K. Sen., *Tra economia ed etica*, cit.
[18] Freeman, R. E., 1984. *Strategic management: A stakeholder approach*. Boston: Pitman.
[19] Freeman, R. E., *Strategic Management: A Stakeholder Approach*. cit

These are, in particular, stakeholders who are functional to the survival of the company and the achievement of an adequate rate of profit and long-term growth; the reference to stakeholders seems opportunistic rather than ethical.

Clarkson, unlike Freeman, distinguishes between two categories of stakeholders: the first group is that of primary stakeholders, i.e. those without whom the company could not survive; the second group is that of secondary stakeholders, i.e. those groups that influence or are influenced by the company but do not trade with it; they can cause damage, even serious damage, to the company, but are not essential to its survival.

The objective of the manager and the company must be to ensure the participation of primary stakeholders.

Subsequently, thanks to R.K. Mitchell, B.R. Agle and D.J. Wood, various classes of stakeholders are identified, based on the fact that the individuals or groups involved in the company's activities possess one or more of the following attributes[20]:

a) the power of stakeholders to influence the company;

b) the legitimacy of the relationship between the stakeholder and the company;

c) the urgency of the stakeholder's expectations of the company.

The various combinations of the three attributes identify different types of stakeholders (Figure 1).

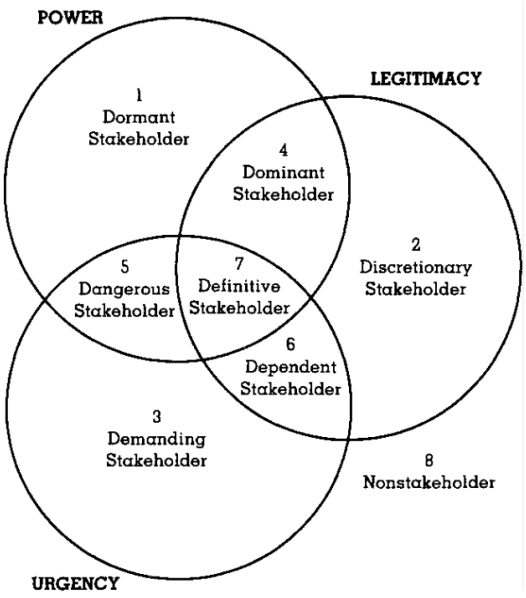

Figure 1: *R.K. Mitchell, B.R. Agle D.J. Wood: Defining the Principle of Who and What Really Counts, in Academy of Management Review*

[20] Mitchell, R. K., Agle, B. R., & Wood, D. J., 1997. *Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts*. Academy of Management Review.

According to R. Phillips, a stakeholder is someone who is considered to be legitimately worthy of the attention of managers and the organization.

There are two types of legitimacy and, therefore, two types of stakeholders:

- Normative legitimacy, which allows us to identify the normative stakeholders, i.e. those towards whom the organization has the greatest moral obligations. These are the stakeholders for whose benefit the company must be managed, i.e. the relevant stakeholders.
- Derivative legitimacy, which identifies derivative stakeholders, i.e. those groups whose actions and expectations must be taken into account by managers because of the potential effects they may have on the organization, the achievement of its goals and its normative stakeholders[21].

The concept of stakeholders is refined when business economists define the company as an open and purposeful system[22]:

- Open system: a system that exchanges energy, materials and information with the wider external environment (composed of a series of markets and other external forces).
- Purposeful system: oriented towards a mission and general objectives.

In order to cope with all these opposing internal (employees) and external (independent) forces and keep the company focused on its long-term goals, management must necessarily and appropriately manage all the parties involved, namely the stakeholders.

According to Goodpaster, introducing stakeholder analysis into corporate decision-making (from a strategic business management perspective) is not the same as introducing ethics into those decisions.

In particular, the author cites two reasons to support his thesis[23]:

- The first: the decisions that management makes in favour of stakeholders are only aimed at preventing them from hindering the achievement of long-term objectives. These are therefore opportunistic decisions rather than ethical ones.

[21] Phillips, R,. *La legittimità degli stakeholder*, cit.
[22] Zanda, G., 1974. *La grande impresa. Caratteristiche strutturali e di comportamento*. Milano: Giuffrè.
[23] Goodpaster, K. E., 1991. *Business ethics and stakeholder analysis*. Business Ethics Quarterly.

- The second: managers have an exclusive fiduciary relationship with shareholders and must be primarily concerned with them, as other stakeholders are already sufficiently protected by market forces and the law.

1.3. Corporate social responsibility (CSR)

We have said that stakeholder analysis in corporate decision-making (from a strategic business management perspective) is not the same as introducing ethics into those decisions. A step forward, in a more deontological sense, has been taken with the emergence and evolution of what is known as corporate social responsibility (CSR).

CSR consists of the voluntary integration of social and environmental concerns into companies' commercial operations and relationships with stakeholders[24].

CSR, therefore, goes beyond compliance with legal requirements and identifies practices and behaviours that a company adopts on a voluntary basis, in the belief that it will achieve results that can bring benefits and advantages to itself and the context in which it operates.

Although CSR only began to spread more widely in the 1980s, its roots can be traced back to fifty years earlier, when these concepts were obviously still in their infancy.

Howard Bowen, an American economist and university professor, published the essay "*Social Responsibilities of the Businessman*"[25] in which he stated that the obligations of the businessman must inevitably go "beyond" the mere business context and profit alone. These obligations must also include responsibility for the effects of his behaviour on all those involved. The businessman/manager has a responsibility (social, which Bowen also calls public responsibility, social obligation or business morality) towards society as a whole.

Social responsibility encompasses two aspects:

- Socio-economic responsibility: this is achieved by creating value (economic welfare), i.e. maintaining adequate employment levels, remaining competitive, etc.
- Responsibility for human values: this consists of the obligation, again on the part of the businessman, to satisfy and develop human capital.

There are three reasons drive the businessman to be socially responsible[26]:

1) The risk of losing the social power that characterizes him, which consists of the businessman's ability to influence the community through his behaviour.

[24] Bowie, N. E., & Werhane, P. H., 2005. *Management ethics*. Malden, MA: Blackwell.
[25] Bowen, H., 1953. *Social responsibilities of the businessman*. New York: Harper & Brothers.
[26] Bowen, H., *Social Responsibilities of the Businessman*, cit.

2) The need to manage human capital that is not satisfied with economic welfare alone but also expects an environment that allows it to satisfy its dignity, creativity and potential.
3) The community has expectations of businesses that go beyond economic expectations alone.

With Carroll, social responsibility becomes a real managerial action in response to the needs of society. A business is socially responsible when it meets the economic, legal, ethical and philanthropic expectations of society.
Carroll illustrates this in an imaginary pyramid (Figure 2):

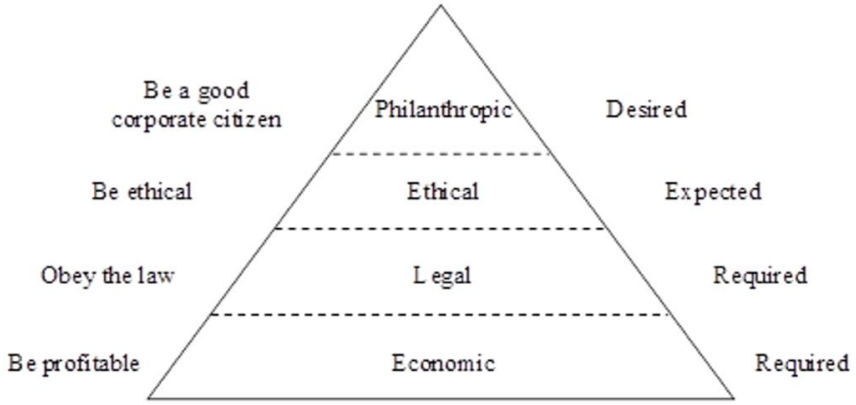

Figure 2: *A.B. Carroll, The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders, Business Horizon.*

At the base of the pyramid are economic responsibilities associated with the phrase "be profitable". This means that the company must be able to produce goods and services that meet consumer needs so that they are purchased and allow for the realization of an adequate profit.
The second level consists of legal responsibilities associated with the formula "obey the law". Legal responsibility consists of the company's required compliance with legislation.
Next, we find the (*expected*) ethical responsibilities associated with the phrase "be ethical". A company is ethically responsible when its behaviour follows standards that are considered correct and desirable by society but are not (yet) codified by law or, in any case, for which the law provides for lower levels than those expected.
Finally, there are the philanthropic responsibilities associated with the adage "be a good corporate citizen". These are not required or expected by society in the same way as ethical responsibilities; they are socially desired responsibilities and, as such, are discretionary, i.e. left to the free will of the company.

Management that bases its decisions on the interests of all stakeholders is defined by the author as "*moral management*". Although success-oriented, it takes the law into account when making decisions and goes beyond the law, which is seen as the minimum standard to be respected. It does not make any decisions (even if profitable) that are detrimental to ethics[27].

Unethical management is focused exclusively on success and profit in open opposition to moral rules, resents the law and takes any action to avoid it in order to achieve personal and corporate success.

There are two limitations to the pyramid: firstly, the pyramid shape is not capable of representing the actual relationships between the four components that make up a company's CSR, and secondly, the shape could suggest that there is a hierarchy between the various components[28].

It considers philanthropy as a category in its own right, but if it is carried out for economic reasons, then it falls within the scope of economic responsibilities, whereas if it is carried out for "non-economic" reasons, then it may well fall within the scope of ethical responsibilities.

To overcome these limitations, it proposes a three-domain approach[29] (Figure 3).

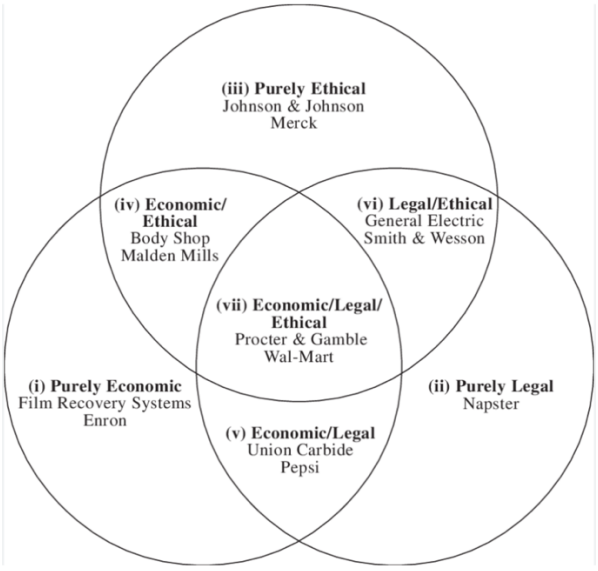

Figure 3: *M.S Schwart, A.B. Carroll, Corporate social responsibility: a three-domain approach.*

[27] Carroll, A. B., 1987. *In search of the moral manager*. Business Horizons, 30(2), 7–15.
[28] Carroll, A. B., *The Pyramid of Corporate Social Responsibility*, cit.
[29] Schwartz, M. S., & Carroll, A. B., 2003. *Corporate social responsibility: A three-domain approach*. Business Ethics Quarterly,

P. Onida (1902-1982), Italian economist and professor emeritus at the University of Rome "La Sapienza", stated: the supposed contrast between the economic nature of a business and its social nature (understood as concern for the good of society or the common good) derives from "an insufficient understanding of the economic and social nature of a business or from a misleading assessment of their real needs"[30].

If the two concepts are interpreted correctly, it is easy to reach a convergence:

Economic efficiency should not be judged exclusively as a measure of profit but should be understood as the lasting existence and convenient development of the enterprise, which, therefore, cannot ignore the creation of jobs and wealth for society.

Sociality should also be understood in the sense that the mortification of the enterprise and its downfall are still bad for society.

From the concept of the company as a social institution that "serves to enhance human well-being, promote the development of the individual and achieve the goals of human life in society", it is clear that ethical rather than purely economic objectives prevail. Economic actors who are aware of the ethical nature of the company must ensure that these ethical objectives are achieved[31].

Sergio Sciarelli, analyzing the evolution of corporate social responsibility, highlights how in the initial phase it focused mainly on externalities produced in favour of employees through, for example, housing, cultural and sports centres, assistance organizations for the disabled, childcare facilities, etc., and subsequently (today) has evolved into a true global social responsibility.

This is structured in four conceptual components[32]:

The first component is the assumption that coincides with the economic purpose, i.e. the creation of value, without which the company ceases to exist.

The second component is the corollary of the economic purpose, which consists in the balanced distribution of value among the various stakeholders. The corollary therefore has a social value.

This is followed by the complement, which takes the form of community social responsibility, i.e. integrating the company into its reference context.

Finally, we have the option, i.e. the possibility (not the obligation) for the company to carry out (discretionary) interventions for the benefit of the community. This last component is also known as strategic philanthropy.

[30] Onida, P., 1960. *Economia d'azienda*. Torino: UTET.
[31] Onida, P., *Economia d'azienda*. cit.
[32] Sciarelli, S., 1967. *Il processo decisorio nell'impresa*. Padova: CEDAM.

Milton Friedman (1912 - 2006), an American economist, believed that in a free economy, social responsibility cannot apply to companies as abstract entities but can only refer to individuals. He distinguishes and classifies two types of companies: widely owned companies and individual companies[33].

In the former, managers are employees of the owners-shareholders and, as such, have a direct responsibility to the latter who have chosen and hired them; this responsibility is realized by creating as much value (or profit) as possible.

Profit has the advantage of being an easily identifiable parameter and, therefore, allows for an immediate assessment of the manager's performance. This profit must be achieved in compliance with the law and ethical standards.

The manager's behaviour in his private life may be different; in this case, he is his own boss and does not act on behalf of others; therefore, he can do as he pleases or devote himself to social responsibility.

Similarly, owner-shareholders (stockholders), as individuals, can spend the money they earn, thanks to managers, in whatever way they prefer.

If managers were to spend the company's money to promote social responsibility initiatives, they would be spending money that is not theirs to achieve something that does not relate to the role for which they were hired.

In essence, by doing so, managers would no longer be performing an economic role but would be taking on a political role. Managers are not selected to play a political role but to act as agents of the principal-shareholder, by virtue of a procedure that is specific to company executives[34].

Furthermore, arguing that it is good for companies to replace the state in certain cases, which is often characterized by slowness and inefficiency, is wrong because it would go against the principles of democracy, the free market and the primacy of the state.

Peter Drucker (1909 - 2005), an Austrian-born American economist and essayist, argues that social responsibilities concern two aspects[35]:

- The responsibility that derives from the impact of the company's activities on society (e.g. pollution);
- Responsibility for social problems in the community (e.g. racism). In this case, these are dysfunctions that do not originate within the company.

[33] Friedman, M., 1970. *The social responsibility of business is to increase its profits*. New York: The New York Times Magazine.
[34] Friedman, M., *The Social Responsibility of Business is to Increase its Profits*. cit.
[35] Drucker, P. F., 2000. *Manuale di management. Compiti, responsabilità, metodi*. Milano: Rizzoli.

The company must take responsibility for the impacts because management, like others, is responsible for the (negative) effects of its actions. Moreover, eliminating the impacts could translate into a business opportunity for the company.
Sometimes, reducing or eliminating the impacts of the company's actions could entail very high costs which, if not borne equally by all companies in the same sector, would result in a loss of competitiveness: the virtuous company would undermine its own survival.
In this case, the only viable solution to reduce or eliminate the impact would be public intervention, i.e. legislation that would affect all companies equally and solve the problem without creating distortions in competition. In this case, management must take action by collaborating with the public in order to enable the best form of regulation[36].
Management, aware that a healthy business needs a healthy society, cannot ignore social problems and must check whether anyone has already addressed and, in some way, resolved these problems.
However, management's ability to address such problems faces certain objective limitations in its operations:

- First limitation: survival of the business; the management team has the ultimate responsibility towards the business.
- Second limitation: expertise; management must only deal with matters in which it has knowledge and experience.
- Third limitation: the authority that management can assume; if a company, through its management team, takes responsibility for a social problem, it automatically assumes authority over such issues. In doing so, the company takes the right to make decisions on issues that would otherwise be the responsibility of the state or other entities.

Management, while not replacing the government, must nevertheless take charge of problems, studying them, examining them in depth and proposing useful solutions to the competent authorities[37].

In conclusion, analysis of the evolution of business ethics, from its philosophical roots to the affirmation of social responsibility and stakeholder theory, highlights how businesses can no longer be considered solely as profit-oriented economic entities, but rather as social actors

[36] Drucker, P.F., *Manuale di Management. Compiti, responsabilità, metodi*. cit.
[37] Drucker, P.F., *Manuale di Management. Compiti, responsabilità, metodi*. cit.

with responsibilities towards a variety of stakeholders. However, while in the past such considerations focused mainly on the balance between economic and social aspects, in the current context the emergence of digital transformation and innovation processes makes it necessary to rethink traditional categories.

New technologies not only generate competitive advantages and growth opportunities, but also introduce unprecedented ethical risks that affect the relationships between businesses, stakeholders and society as a whole. In this perspective, social responsibility can no longer be limited to a reactive or voluntary dimension, but must evolve towards a proactive and strategic approach capable of guiding innovation in a sustainable and responsible manner.

This transition marks the point of contact with the following chapters: while this chapter has illustrated the historical and theoretical foundations of business ethics, the next chapter will analyse how it intertwines with innovation processes, introducing the concept of *business ethics of innovation*, and then delving deeper into the emerging challenges related to artificial intelligence.

PART TWO: The ethics of innovation

2.1. Ethics – innovation: introductory framework

The concept of innovation has accompanied human existence since the beginning. The ability to innovate, to change and improve processes and instruments, has probably been one of the most unique features of the human race and the prime driver of major economic and social revolutions. Innovating does not only mean the development of new technologies, but the innovation of new concepts capable of radically changing the lives of people, work relationships, production systems, and even political and cultural institutions. It has been described by various economic and historical literature as witnessing several eras of acceleration in innovation at breakneck speeds, called industrial revolutions.

- The beginning of the first industrial revolution (18th - 19th century) had the most profound turning point in the form of the development of James Watt's steam engine. Cities like Manchester, the symbol of new industrial capitalism, witnessed the emergence of textile factories that completely altered the production pattern. The progress had a cost in the form of poor work conditions, exploitation in the form of child labor, and the growth of the protest movements like the *Luddites* workers who broke machinery in an attempt to save jobs[38].
- The second industrial revolution (mid-19th - early 20th century) introduced electricity, chemistry, the internal combustion engine, and new transport infrastructures. The scale expanded when the large corporation form developed (e.g., General Electric), and social tensions grew and became more widespread (e.g., labour alienation and colonial exploitation as bases for supply of raw materials)[39].
- The third industrial revolution (second half of the 20th century) coincided with digitalization: telecommunications and the Internet transformed production and communication radically. Such companies as IBM, Microsoft, and Apple emerged and re-arranged the world economy. At the same time new moral challenges arose: technological dependence, the concentration of power by Big Tech giants, and the loss of privacy.
- Finally, the fourth industrial revolution or the *Industry 4.0* that we are experiencing today is one whereby artificial intelligence, robots, the Internet of Things and big data are converging. Klaus Schwab described the times as “a fusion of technologies

[38] Landes, D.S., 1969. *The Unbound Prometheus*. Cambridge: Cambridge University Press.
[39] Rifkin, J., 2011. *The Third Industrial Revolution*. New York: Palgrave Macmillan.

that is blurring the lines between the physical, digital, and biological"[40]. The promise is phenomenal - tailor-made medicine to clean energy - while the risk is equally immense: mass surveillance and joblessness by technology and global inequalities.

Every industrial revolution then had two faces: one social development and economic growth and the other injustices and conflicts and moral questions. Innovation then then has a double-faced nature that has brought unprecedented, good only to engender risk and trouble.

It is for this reason that understanding the connection between innovation and ethics is crucial. Ethics should not be seen as a limitation but as a compass for innovation processes directed at sustainable and collective outcomes, such that the fruits of progress be not mixed with the wrong consequences.

2.1.1. Innovation as a competitive lever

The innovation-competitiveness linkage has been studied by numerous economists and management thinkers. It may be the case that the most significant contribution belongs to Joseph Schumpeter. As early as in his classical *The Theory of Economic Development* (1934), Schumpeter says: "Capitalism is by nature a form or method of economic change, and not only is it never stationary, but it can never be so"[41]. Innovation to Schumpeter means the drive behind such change through the process of *creative destruction* by which old equilibriums are upset and eliminated to accommodate new better models.

Schumpeter identified five major kinds of innovation:

1. New products: smartphones that changed people's means of communication and e-vehicles that altered how people travel.
2. Innovation in production methods: the best-known example is the assembly line introduced by Henry Ford at the start of the 20th century. This enabled production to multiply explosively and the product to become accessible to a large section of the population.
3. Creating new markets: e-commerce, by virtue of the rise of Amazon, has shown how innovation can create whole value chains of production and consumption.
4. Novel natural-resource sources: from oil, the driver of the 20th-century economy, through lithium and cobalt today to be the foundation of the energy transition.

[40] Schwab, K., 2016. *The Fourth Industrial Revolution*. Geneva: World Economic Forum.
[41] Schumpeter, J.A., 1934/1983. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press.

5. New modes of organization: new digital-platform business models (e.g., Uber, Airbnb) have reshaped entire industries.

Innovation has also been highlighted by Michael Porter as the key to the competitive advantage at the company and nation level. Introducing the *diamond model* in the work T*he Competitive Advantage of Nations* (1990), innovative ability emerges through the interaction among four factors including production factor conditions, home demand specifications, industry-related industries and competition strength[42]. The fashion district in Italy demonstrates how innovation and quality are constantly stimulated by a culture of craftsmanship, advanced and internal competition.

Finally, Clayton Christensen introduced the concept of *disruptive innovation*, i.e. innovation that not only improves the already dominant product offerings, but re-engineers the markets and disrupts established giants who do not innovate[43]. Kodak failing to rise to the challenge of digital photography provided you the best illustration of how a company gets swamped by disruption. Blockbuster failing to react to streaming got soon overtaken by Netflix. Even the telephony world leader, Nokia, failed to meet the challenge created by the smartphone.

These cases demonstrate how innovation is a central competitive lever: those who do not innovate risk the penalty of marginalization. However the disruptive nature of innovation also raises questions about ethics: what are the social costs of such transformation? Who are the exclusions or sanctions in the process of creative destruction?

2.1.2. Innovation and ethical challenges

If progress has a motor in innovation, it too forms a world of obstacles and obligations. Every development in technology actually alters social and cultural balances and allows new horizons while forging new and unforeseen threats. Philosophical and moral thinking has tried to predict such troubles and provide the means of judging them.

The German philosopher Hans Jonas in *The Imperative of Responsibility* (1979) highlighted the fact that the ethics of technology must take a new position, which will be based on a responsibility to future people as follows: “Act so that the consequences of your action are compatible with the permanence of an authentically human life on Earth”[44]. Jonas thereby

[42] Porter, M.E., 1990. *The Competitive Advantage of Nations*. New York: Free Press.
[43] Christensen, C.M., 1997. *The Innovator’s Dilemma*. Boston: Harvard Business School Press.
[44] Jonas, H., 1984. *The Imperative of Responsibility*. Chicago: University of Chicago Press.

anticipated the questions of the future–sustainability and intergenerational responsibility– central to the current debates on artificial intelligence and the climate change challenge.

Martin Heidegger argued in *The Question Concerning Technology* (1954) that technology is not neutral at all, but instead represents a kind of "revealing" of the actual world and the threat it presents to turn human beings into objects to exploit and calculate[45]. Technology then takes on the form of destiny and preconditions human liberty and questions about how much control will be able to be maintained over it.

Jacques Ellul, in *La technique ou l'enjeu du siècle* (1954), radicalized this stance even further and argued that technology develops through the force of its own internal rationality and not by the instrumentality or volition of humankind: "Every technical advance is adopted because it is possible, not because it is necessary"[46]. This view, plausible in the 1950s, seems even more plausible today in the computerized and biotechnological age and the issue at question is not should we develop new solutions but when should we implement them.

These insights are verified by recent cases.

- Tesla embodies the values of clean energy and electric vehicles in sustainable innovation. The model has a few major issues, however: the Tesla battery production requires lithium and cobalt mined in places that do not provide good environmental and social protection.
- The world has been revolutionized by Google and Facebook (now Meta) by giving billions of individuals free services. But doing so, they have centralized unprecedented economic and informational power and created severe questions about privacy and information manipulation and technological dependency.
- Big Pharma best symbolizes the innovation-related ethical dilemma. It has introduced life-preserving vaccines and new treatments while rigidly maintaining the right to patents and refusing to yield access to the medicines in the majority of low-resource countries, just as it did for the COVID-19 vaccines.
- Volkswagen and Dieselgate, on the other hand, offer indications of "unethical" innovation: the company developed software that could cheat emission testing, reflecting how the imperative for competitiveness can yield fraudulent innovation[47].

These situations demonstrate how innovation represents at all times an "ethical challenge". It cannot be studied solely through the prism of the technique or the economy and must be

[45] Heidegger, M., 1954. *La questione della tecnica*. Milano: Mursia (trad. it. 1976).
[46] Ellul, J., 1954. *La technique ou l'enjeu du siècle*. Paris: Armand Colin.
[47] Hotten, R., 2015. Volkswagen: The scandal explained. *BBC News*, 10 December.

taken in account considering the social, environmental and political repercussions. The difficulty lying ahead for companies and decision-makers is the building of governance systems capable of steering innovation in the long term and reduce the risk of exclusion and inequalities.

2.1.3. Objectives and connection with corporate ethics

The analysis undertaken already prompts the important point that innovation cannot be regarded as a purely technical or economic process. It always takes place in a social, political and cultural context, and has consequences extending far beyond the company domain.

This view has a direct connection with the question of business ethics. The *stakeholder theory* proposed by Edward Freeman holds that businesses must take the interests of all who are implicated in the process of business activities - consumers, employees, providers, local society - and not only the interests of shareholders[48]. When applied to the innovative process, this implies not only the evaluation of economic returns, but the social effect on employees and consumers and society at large.

Archie Carroll built on that theory with his CSR (Corporate Social Responsibility) pyramid to explain four different levels of company responsibility:

1. economic (profitable)
2. legal (following the law),
3. ethical (acting correctly even beyond legal obligation),
4. philanthropic (giving voluntarily to the community's welfare)[49].

Innovation goes hand-in-hand in this regard: innovation by itself if not accompanied by responsibility can bring about short-term economic returns while long-term environmental and social maladies become the byproduct.

Finally, Michael Porter and Mark Kramer introduced the concept of Creating Shared Value (CSV), underscoring the fact that innovation can be the key tool for the development of both economic and social value: “Companies can create shared value by developing products and services that meet the needs of society”[50].

[48] Freeman, R.E., 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.
[49] Carroll, A.B., 1999. Corporate social responsibility: Evolution of a definitional construct. *Business & Society*, 38(3), pp.268–295.
[50] Porter, M.E. and Kramer, M.R., 2011. Creating shared value. *Harvard Business Review*, 89(1/2), pp.62–77.

So innovation through one of the firm and economy's major competitive levers and resulting in dilemmas that do call for the framing of innovation by the perspective of an ethical lens, the imperative has generated the concept of responsible innovation. This does not view ethics as the exogenous constraint but understands it as inherent to innovation in and by itself. The succeeding section (2.2) will then explore the innovation and responsibility theme and how public policies, businesses and civil society institutions can lead change to sustainable and shared objectives.

2.2. Innovation and responsibility

The concept of responsible innovation has emerged over the last two decades following the growing awareness that not all innovation yields positive consequences. Even as they are the chief driver of economic growth and competitiveness, they are also the major source of severe hazards to society, the natural environment and democratic politics.

In a globalized and interconnected world, where new technologies spread rapidly, the issue of responsibility has become central. It is no longer enough to ask whether an innovation is technically possible or economically viable: we must also question its sustainability, fairness and long-term impacts.

This necessity has spurred the model of Responsible Innovation (RI), whose objective is to integrate the ethical, social and environmental dimensions at the very origin of the development process of new technologies. As Stilgoe, Owen and Macnaghten (2013) clarify: "Responsibility should not be an accessory to innovation, but an intrinsic part of it"[51].

2.2.1. The concept of Responsible Innovation

The concept of responsible innovation came into existence in the 2000s, especially in the European context. As quoted by René von Schomberg (2013), the term can be defined as "a transparent, interactive and inclusive process that takes into account the expectations and values of society to ensure that the resulting products and services are socially desirable and sustainable"[52].

The key aspects of RI are:

[51] Stilgoe, J., Owen, R. and Macnaghten, P., 2013. Developing a framework for responsible innovation. *Research Policy*, 42(9), pp.1568–1580.

[52] Von Schomberg, R., 2013. A vision of responsible research and innovation. In: Owen, R., Bessant, J. and Heintz, M. (eds). *Responsible Innovation*. London: Wiley, pp.51–74.

1. Anticipation: assessment of the potential beneficial and harmful consequences of innovation beforehand.
2. Inclusion: involvement by various social actors (citizens, stakeholders, institutions) in the decision-making process.
3. Reflection: constantly questioning the goals and values that motivate innovation.
4. Responsiveness: adaptation of innovation strategy based on social feedback.

This view is completely different from the classical view that innovation comes only through the market or technology. RI asserts that innovation and ethics cannot be two different dimensions by any means, but two factors that must be present together right at the starting point.

2.2.2. Responsible Research & Innovation (RRI) in European policies

The RRI approach has had particular use for European research and innovation policies. The Horizon 2020 programme (2014–2020) was the largest research investment in the European Union's history by far (with over €80 billion). As part of this, the concept of Responsible Research and Innovation (RRI) emerged as a transversal orienting principle: this meant that all financed projects had to be able not only to demonstrate scientific legitimacy, but also to be mindful of social inclusion and gender equality and committed to open science and to citizens' involvement[53].

The upcoming Horizon Europe programme (2021–2027), even larger-budget-wise, ratified this orientation further by giving central priority to the need for research and innovation to be aligned with the Sustainable Development Goals (SDGs). The European Commission reiterated the point again that innovation needs to be aimed not only at economic growth, but also at sustainability, social inclusion and respect for the fundamental rights[54].

Meanwhile, the UN has also pushed the notion of responsible innovation in direct connection to the 2030 Agenda. Innovations in fields such as renewable energy (SDG 7), resilient infrastructures (SDG 9) and climate change-fighting technologies (SDG 13) are viewed as key instruments in ensuring a just ecological transition. The example thereof has been the creation of mRNA vaccines in the pandemic context of COVID-19: a technological

[53] European Commission, 2014. *Responsible Research and Innovation: Europe's ability to respond to societal challenges*. Brussels.
[54] European Commission, 2021. *Horizon Europe Strategic Plan 2021–2024*. Brussels.

development that had a direct effect on global health, as well as created questions of a different order connected to accessibility and fair distribution[55].

The OECD, in the paper *Responsible Innovation in a Global Context* (2018), also suggested the adoption of structured forms of ethical governance in companies, such as systems for the assessment of impacts, codes of conduct, global norms and social reporting. The organization also placed the relationship between the responsible innovation and the parameters of the ESG (Environmental, Social, Governance) criteria at the centre, which become increasingly relevant among third parties and investors[56].

One such sensitive area is the bioethics domain. The European Union has proceeded with caution in GMO application and established restrictive legislations that do not merely hold scientific assessment in regard but also social sentiments and environmental risk. Embryonic stem cell research has also proceeded under stringent controls to preserve the European people's ethical feelings[57].

Another important case in point refers to artificial intelligence. The 2021 draft European AI Act contains the risk classification of AI systems as follows:

- unacceptable (e.g. mass surveillance),
- high risk (e.g. uses in health or legal contexts),
- limited risk.

This method shows how the EU tries to balance innovation and the protection of the fundamental rights while maintaining security and transparency[58].

Even by this achievement, critical challenges persist: Regulation in Europe continues to be viewed by many as too slow and cumbersome relative to the pace at which technologies are advancing. Additionally, there continues to be no single global regulatory standard in place, which gives rise to the threat of fragmentation and Regulatory Competition to the downside[59].

[55] United Nations, 2015. *Transforming our world: the 2030 Agenda for Sustainable Development*. New York: UN.
[56] OECD, 2018. *Responsible Innovation in a Global Context*. Paris: OECD Publishing.
[57] European Commission, 2010. *Directive 2001/18/EC on the deliberate release into the environment of genetically modified organisms*. Brussels.
[58] European Commission, 2021. *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (AI Act)*. Brussels.
[59] Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133), pp.1–13.

2.2.3. Technological innovation and social innovation

Innovation has traditionally been familiar almost exclusively as a techno-technical fact: innovation in new products, processes or informational technologies capable of boosting the competitiveness and productive capacities of businesses[60]. However, in the past few decades, scientific publications and public policies have awakened to the potential value of another dimension: social innovation, i.e. the ability to introduce new answers to social questions, to improve the quality of life and to reinforce social cohesion.

Technological innovation remains the biggest driver of economic development. Just think about the impact of the Internet, mobile telephony, large data and machine intelligence, which reshaped entire tracts of the industrial economy and created new types of economic value. Massive multinationals such as Google, Apple, Microsoft and Tesla are quintessential representatives of the power of technology to remake markets and patterns of consumption.

Innovation in technology alone cannot offer sustainable and inclusive development. The very digital technologies that brought global connectivity also brought the threat of mass surveillance, digital exclusion and the centralization of power among a very small number of economic actors. This is why the technological aspect must be supplemented by the social aspect in a vision of responsibility and inclusion.

Social innovation focuses on solving social needs through new organization models, participatory processes or new forms of inter- and multi-actor collaboration at the local level among private and public actors and local citizens. It differs from technological innovation because the latter does not aspire principally to economic returns, but social value creation.

Just think about the microcredit developed by Muhammad Yunus and the Grameen Bank which enabled millions of people in emerging economies to get credit and start businesses[61]. Other significant cases regard the sharing economy through platforms like Couchsurfing and BlaBlaCar meant to induce the social sharing of goods[62]. Even if some of them afterwards had taken more or less commercial forms (think about the case of Airbnb or Uber), they too demonstrate how innovation doesn't solely proceed just in the laboratories of technology and how it may be born even in civil society.

Another area of social innovation is environmental sustainability: ranging from local energy communities that foster self-production and sharing of renewable energy to circular economy projects that reduce waste and recycle materials.

---

[60] Murray, R., Caulier-Grice, J. and Mulgan, G., 2010. *The Open Book of Social Innovation*. London: Nesta.
[61] Yunus, M., 2007. *Creating a World Without Poverty*. New York: PublicAffairs.
[62] Botsman, R. and Rogers, R., 2010. *What's Mine Is Yours: The Rise of Collaborative Consumption*. New York: Harper Business.

Technological innovation and social innovation are not typically alternatives, but complements. Digital technologies, for example, can become social innovation instruments when used to achieve inclusiveness and participation. Consider platforms for civic crowdfunding that allow the local community financing of initiatives aimed at culture or the environment or digital solutions aimed at the accessibility of health care services in distant regions[63].

It is not just a matter of technology (more effective solar panels, lithium batteries), but also social innovation: behavioral change among consumers, new forms of governance and ways of collaboration among citizens and institutions.

The research on responsible innovation has shown how it is possible to steer processes of social and technological change towards sustainable and inclusive outcomes through the use of governance, participation and transparency tools. However, in order to understand the entire set of innovation's ethical outcomes, it is necessary to shift the focus beyond public policies and broad orientations and direct the attention to the companies as the leading change actors.

This perspective has spurred the construction of the Business Ethics of Innovation research field, which explores how companies can and should integrate ethical values in business models, research and development priorities, and relationships with stakeholders. The section that follows (2.3) will then provide the fundamentals of this approach and the main challenges and practices for the company-level management of ethical innovation.

2.3. Business Ethics of Innovation

2.3.1. Definition and origins of the concept

Innovation Business Ethics is the crossing point of two long-established and formerly separate research lines: the business ethics tradition, namely the analysis of the ethical values that should inform businesses' governance; and technological and organizational innovation research, intended as the driving factor for the economy's competitiveness.

Business ethics used to focus mainly on social responsibility questions for corporations, finance transparency, respect for employees' rights and the fight against corruption. Innovation research, on the other hand, applied mainly the economic-technical method and

[63] Nicholls, A., Simon, J. and Gabriel, M., 2015. *New Frontiers in Social Innovation Research*. London: Palgrave Macmillan.

examined processes, strategies and consequences for the markets. It is only since the 2000s that specialized reflection on the innovation ethics has developed and endeavored to explain how the two fields can be related[64].

As proposed by von Schomberg and Blok (2019), innovation business ethics can be defined as "the science that researches the moral challenges and the ethical commitments produced by processes of innovation, both by the agents economic and by the social effects"[65]. It goes beyond just the question of whether the company is "ethical" in its ordinary business activities, and goes beyond whether and how the launched innovations are socially beneficial and ethically viable.

It is particularly relevant today, in the world defined by the fourth industrial revolution. Novel technologies – from artificial intelligence to robots through the Internet of Things and beyond to biotechnology – promise the extraordinary but carry unprecedented risk. When faced by innovation capable of radically changing society, it cannot be sufficient merely to call for generic social responsibility criteria: innovation must be defined by a unique innovation ethics.

The conceptual roots of this orientation lie in three broad traditions:

1. Classical business ethics: scholars such as Freeman (1984), and his *stakeholder theory*, have put the point that businesses must regard the interests not just of the shareholders, but also of all the other stakeholders. Applied to innovation, this means considering the future impacts on consumers, employees and local people of new technologies[66].
2. The technology philosophy: philosophers like Hans Jonas, and his *rule of responsibility*, have highlighted the responsibility the current generations owe the future ones and provided the basis for an evaluation of such innovations using ethics that span the generations[67].
3. The theory of responsible innovation: conceived in the 2010s by scientists like Stilgoe, Owen and Macnaghten, it has pinpointed anticipating, reflecting, including and responding as innovation design process key elements[68]. The model has linked

[64] Werhane, P.H. and Freeman, R.E., 1999. Business ethics and the origins of stakeholder theory. *Business Ethics Quarterly*, 9(1), pp.169–180.
[65] von Schomberg, R. and Blok, V., 2019. Technology in the Age of Innovation: Responsible Innovation as a New Subdomain of Philosophy of Technology. *Philosophy & Technology*, 32(1), pp.1–12.
[66] Freeman, R.E., 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.
[67] Jonas, H., 1984. *The Imperative of Responsibility*. Chicago: University of Chicago Press.
[68] Stilgoe, J., Owen, R. and Macnaghten, P., 2013. Developing a framework for responsible innovation. *Research Policy*, 42(9), pp.1568–1580.

scientific research at the academic level and public policy research and has also become the business standard.

One real-world example of the adoption of such an approach is the discussion on genomics and genetic editing. While such technologies as CRISPR-Cas9 can potentially heal devastating diseases, they pose tremendous questions regarding the control and editing of human DNA. The question in such a scenario isn't merely the question of whether the biotech firm adheres to the laws in place, but whether the innovation in question should be undertaken at all and how.

Just as in the digital economy and the platforms such as Facebook (Meta today), the exchange and the communication have innovated at a radical level in the world of the digital economy. However the innovation model for the aggregation and the monetization of the user data created severe privacy, security and centralization of power questions. This highlights the challenge for a business ethics innovation able to judge not only profits, but the long-term consequences of the technology choice decision.

The business ethics of innovation can be generally defined as the set of principles, tools and practices through which innovative processes can be oriented towards objectives not incompatible with sustainability, social justice and regard for human rights. It derives from the integration of established ethical heritages and new requirements related to the acceleration of technological innovation, and today represents one of the most flourishing fields of research in the field of management and organizational studies.

2.3.2. The trade-off between profit, competitiveness and ethical values

One of the most sensitive areas in innovation business ethics pertains to the balance among goals and values that are economic and moral. Innovation often has its roots in the imperative to increase profits, reduce costs or obtain competitive advantage, but innovative processes can pose controversial impacts on employees, consumers and society in general.

Classical thinking by Milton Friedman since 1970 believes that "the only social responsibility of business is to increase its profits, while respecting the rules of the game"[69]. This dominant thinking that has stood the test for decades considers ethics and responsibility as exogenous constraints. This stance has in the past decades faced criticism as reductionist. Edward Freeman challenged it by proposing the consideration of business activity not only

[69] Friedman, M., 1970. The Social Responsibility of Business is to Increase its Profits. *New York Times Magazine*, September 13.

in comparison to the stockholders but to all the stakeholders through the theory of the stakeholder. This introduced the economic perspective to a new dimension by incorporating social and environmental considerations[70].

The profit-ethics trade-off assumes various forms depending upon the sector. There exist profitable inventions that are troublesome to society at large: the tobacco industry, gambling and some digital platforms patterned after the mass consolidation of personal data have generated humongous profits at the cost of dire health, privacy and personal liberty impacts[71]. Other inventions are sustainable ones whose profits appear less immediate: clean energy or eco-friendly products entail upfront costs and profits extended in the long term and sometimes posing challenge for certain businesses to adopt them unequivocally[72]. Finally, social inventions such as microcredit result in titanic social returns for marginalized people but provide lesser profits than conventional finance[73].

The global aspect makes the picture even more complex. Companies in Europe that meet the region's tough environmental standards may be placed at a competitive disadvantage versus other businesses operating in nations whose rules are less stringent. The scenario raises the temptation to relax standards or shift production to less rigorous settings[74].

To answer such dilemma, several scholars and leaders in policy have suggested ways that put profit and ethics together. Michael Porter and Mark Kramer, in the concept of Creating Shared Value (CSV), hold the view that the two objectives do not actually contradict each other: companies can innovate by creating goods and services that respond to social and environmental needs while staying ahead in the competition[75]. The food industry, for example, can bring in profits and image by investing in health-improving foods and reducing the level of sugar and fat and thus resolving the concern of consumers.

The profit vs. ethical values compromise cannot be said to be an insurmountable barrier, but a challenge in the realm of governance. Innovation ethics does not call for sacrificing competitiveness at the altar of moral values nor converting ethics into merely a marketing gimmick. It aims at developing parameters, tools and decision-making processes by which corporations may innovate sustainably and responsibly and still hold social legitimacy and trust among the stakeholders.

---

[70] Freeman, R.E., 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.
[71] Zuboff, S., 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.
[72] Schaltegger, S., Lüdeke-Freund, F. and Hansen, E.G., 2012. Business cases for sustainability. *International Journal of Innovation and Sustainable Development*, 6(2), pp.95–119.
[73] Yunus, M., Moingeon, B. and Lehmann-Ortega, L., 2010. Building social business models. *Long Range Planning*, 43(2–3), pp.308–325.
[74] Vogel, D., 2005. *The Market for Virtue*. Washington: Brookings Institution Press.
[75] Porter, M.E. and Kramer, M.R., 2011. Creating shared value. *Harvard Business Review*, 89(1–2), pp.62–77.

2.3.3. Ethical innovation and unethical innovation: illustrative examples

It may be easier to get a sense of the business ethics of innovation by examining real-life cases of corporations that innovated in alignment with values generally regarded as ethical, as well as other examples in which innovation has actually been employed in a warped or damaging manner. These only serve to illustrate the point that ethics is not something remote and theoretical, but something that impacts corporations' reputations, stakeholders' trust and long-term viability directly.

One interesting example of ethical innovation occurs in the case of Tesla, which has taken electric cars and not only made them environmentally viable, but viable in design and performance too. Tesla has done what it can to reduce fossil-fuel dependence and move the entire auto industry closer to sustainable transport[76]. Of course, there are many dodgy elements to the production of the raw materials needed for the batteries, but the overall effect has been to push the environmental transition.

Another symbolic case is that of Patagonia, which has built its brand on environmental responsibility. Using recycled materials, through supply chain visibility and awareness campaigns, the company demonstrates that innovation and competitiveness and social responsibility cannot be incompatible[77].

Novo Nordisk in the pharmaceutical sector has developed solutions not only to improve drugs for the treatment of diabetes, but also to deliver increased fairness in the distribution of drugs among poor countries. It has also utilized policies incorporating science researches, environmental protection and social responsibility[78].

Alongside such virtuous uses, innovation has sometimes been distorted to selfish or unfair ends. The most prominent example to date probably occurs in the Volkswagen Dieselgate scandal: emission tests had been manipulated by software to portray vehicles as environmentally friendlier than they were. Technological innovation had thereby been used to cheat consumers and regulators and cause ecological and reputational damage[79].

The case of Facebook (Meta) raises significant ethical concerns too. The platform has changed world communication fundamentally, but the monetization model based on the aggregation and commodification of personal data has yielded scandals such as Cambridge

[76] Vance, A., 2015. *Elon Musk: Tesla, SpaceX, and the Quest for a Fantastic Future*. New York: HarperCollins.
[77] Chouinard, Y. and Stanley, V., 2012. *The Responsible Company*. Patagonia Books.
[78] Novo Nordisk, 2020. *Annual Report 2020*. Bagsværd: Novo Nordisk.
[79] Hotten, R., 2015. Volkswagen: The scandal explained. *BBC News*, 10 December.

Analytica, the case that placed the risk of political information interference and transgressions of privacy at the epicentre[80].

The case in the pharmaceutical industry involving Martin Shkreli, who had earned the nickname "Pharma Bro," has become the symbol of unethical innovation. The Shkreli company increased the price of the drug Daraprim by more than 5000% in 2015. Innovation in the context of the markets has in this case become warped in the form of speculative rationality at the expense of patients[81].

2.3.4. Ethical codes and governance tools to support innovation

To translate innovation ethics not abstruse principles but living practices, companies have developed over the years a whole set of practical tools to support them in deciding. The starting point is the code of ethics: a document outlining rules on transparency, protection of privacy, security, sustainability and respect for human rights. Increasingly the codes dedicate reserved paragraphs on digital technologies and artificial intelligence, recognizing the need to establish clear criteria on the exploitation of data and on the management of threats such as algorithmic discrimination or security breaches. The majority take inspiration from international benchmarks such as the United Nations Global Compact's Ten Principles or the OECD Guidelines for Multinational Enterprises and represent common references and permit maintaining consistency in the various contexts of regulation[82].

Codes are complemented by standards and management systems as well because they can help translate principles into feasible processes. ISO 56002 provides guidance on how innovation may be ingrained in firm culture and strategy, for example. ISO 31000 provides norms for dealing with the risk, including the ethical risk, in a systematic manner, and ISO 26000 handles social responsibility. The Corporate Sustainability Reporting Directive (CSRD) in the EU has made the reporting by most companies of environmental, social and governance impacts in the form of common standards compulsory and encourages them to make visible the ethical judgments connected with innovative processes as well[83].

Functional tools capable of having a more concentrated impact became widespread in recent decades. *Impact assessment* takes the first place among them, as it upfront estimates the social, environmental or fundamental rights impacts stemming from innovation; then there

[80] Cadwalladr, C. and Graham-Harrison, E., 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica. *The Guardian*, 17 March.
[81] Pollack, A., 2015. Drug Goes From $13.50 a Tablet to $750, Overnight. *The New York Times*, 20 September.
[82] UN Global Compact, 2000. *The Ten Principles of the UN Global Compact*. New York: United Nations.
[83] European Union, 2022. *Directive (EU) 2022/2464 (CSRD)*. Brussels.

are *ethics by design* techniques, which integrate values and principles into the design process of products and services; and documentation practices such as *model cards* and *datasheets on datasets*, born in the artificial intelligence community to make the behavior of complex systems more evident and easier to understand[84]. Most companies also established internal ethics committees, composed of members who belong to different functions of the company (research and development, legal department, compliance, human resources, sustainability), whose task is to decide on the most sensitive projects and take corrective measures when necessary.

Such tools only function when they are supported by committed and ingrained leadership in the mainstream organization processes. When isolated or loosely related to strategy, they are at risk of becoming exercises in formality or comms tools. By comparison, when regularly utilized, they engender ethics as a genuine organizational competency: an internal ability that reduces risk to a minimum, buttresses the trust of the stakeholders and, in the long term, yields a competitive advantage. However, there are complexities: piecemeal regulations, differential standards by country and the impulse in some cases to limit one's efforts merely to "ethical washing" make ethical governance a challenging area to manage.

The development of the business ethics of innovation has shown how companies can equip both themselves and innovation processes with values, tools and practices to guide them in a responsible way. From codes of ethics to global standards, from *ethics by design* approaches to impact assessment tools, we've seen that ethics can be a real organizational capacity capable of increasing the competitiveness and social legitimacy of companies.

But the adoption of such instruments doesn't eliminate the tensions inherent in innovation completely. Far from it, in the majority of cases, quandaries arise more clearly precisely when trying to match economic objectives and values of morality. Innovation by definition brings abrupt and radical change affecting work, health, the natural environment and relations among powers: spheres challenging to control through abstruse rules alone.

Hence, the paragraph that follows will solely be devoted to the ethical issues concerned with innovation and managed through four major perspectives: social consequences of disruptive innovation, disagreements concerned with intellectual property and distributive justice, the dilemma concerned with sustainability and finally the controversies that have transpired in the various industry domains.

[84] Mitchell, M. et al., 2019. Model cards for model reporting. Proceedings of the FAT* Conference; Gebru, T. et al., 2021. Datasheets for datasets. *Communications of the ACM*.

2.4. Ethical dilemmas related to innovation

2.4.1. Disruptive innovation and social impacts

Clayton Christensen's theory of disruptive innovation refers to those statistical patterns whereby new technologies or economic models immediately render obsolete old products, economic models and entire branches of economic activity[85]. Economically speaking, this mechanism is often seen as a sign of vitality and progress, but its social and ethical implications are less clear.

Technological revolutions over the centuries offer clear examples of this. The arrival of mechanical looms in the 19th century made the labour of thousands of artisans redundant, triggering violent reactions from the Luddites. More than a hundred years later, electrification and mass production increased productivity and the sense of alienation at work, giving rise to new social conflicts[86]. More recently, digitalization has created new professions such as computer scientists and data analysts, but it has radically reduced the demand for basic repetitive jobs, thereby exacerbating polarization[87].

Today, automation and machine learning emphasize this dialectic. From this point on, amid the optimism of experts and investors, millions of medium-skilled jobs are set to be replaced by algorithms and robots, and new jobs require highly specialized skills that are not always available to everyone[88]. This creates an ever-widening gap between those who possess these skills and those who do not, amplifying inequality.

The influences, however, extend beyond the workplace. New technologies such as social media have transformed the means by which people communicate, inform themselves and participate in public life. While on the one hand they have facilitated broader democratic participation, on the other hand they have stimulated processes of disinformation, polarization and digital dependency[89]. The challenge in question then refers to the capacity to draw a balance between the strength of progress and the social expenses. Without active accompanying policies, such as lifelong training programmes or inclusive welfare protection systems, innovation could aggravate social cleavages[90].

---

[85] Christensen, C.M., 1997. *The Innovator's Dilemma*. Boston: Harvard Business School Press

[86] Landes, D.S., 1969. *The Unbound Prometheus*. Cambridge: Cambridge University Press.

[87] Autor, D.H., 2015. Why Are There Still So Many Jobs? *Journal of Economic Perspectives*, 29(3), pp.3–30

[88] Frey, C.B., Osborne, M., 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, pp.254–280.

[89] Pariser, E., 2011. *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin Press.

[90] Acemoglu, D., Restrepo, P., 2018. Artificial Intelligence, Automation and Work. *NBER Working Paper* No. 24196.

2.4.2. Licenses, intellectual property and equitable access

Intellectual property is one of the most important tools for encouraging research and innovation. Patents grant innovators temporary exclusive rights, allowing companies to recoup research and development expenses and make a profit[91]. Without the protection provided by the legal system, most ventures would not be undertaken because they are not profitable. When patents concern essential areas such as health or the environment, very serious ethical questions arise.

The case of the pharmaceutical industry is typical. Companies justify the excessive cost of drugs by appealing to the hugely expensive research and extended periods of testing. But this justification has the effect of pushing prices up to such a level that drugs become inaccessible in poor nations[92]. The COVID-19 pandemic has sharply highlighted the problem: vaccines were dispatched promptly to the richest nations, while the poorest were kept at arm's length for many months. The World Trade Organization discussion on waiving patents has brought to the surface the contradiction between the imperative of maintaining protection for business investment and the global right to health[93]. Meanwhile, the WHO has urged the *COVID-19 Technology Access Pool (C-TAP)* to help catalyze the sharing of knowledge and technology, but industry participation has again remained very modest[94].

The same question arises in the green technology industry. Intellectual property rights on central items for solar panels, wind turbines, or new-generation batteries can slow the global implementation of the very solutions the planet needs to transition its energy systems. Intellectual property rights here again risk being at odds with the common right to a sustainable world.

Digital technology adds further complexity. The most advanced AI algorithms are patented or protected as trade secrets. Although reasonable in order to maintain competitive advantage, such secrecy prevents the ability to verify the existence of bias, discrimination, or security flaws. Secrecy thus threatens the very existence of consumer trust and the possibility of independent control by regulators[95].

To grapple with such dilemmas, organizations have attempted to utilize solutions as diverse as open-source licenses, *tiered pricing* programmes that divide prices by purchasing power,

[91] Stiglitz, J.E., 2008. Economic Foundations of Intellectual Property Rights. *Duke Law Journal*, 57(6), pp.1693–1724.
[92] Kapczynski, A., 2013. Innovation and Access to Medicines. *Yale Law Journal*, 121, pp.101–203.
[93] Hoen, E.F.M., 2021. *The TRIPS Waiver Proposal: An Urgent Measure to Expand Access to COVID-19 Vaccines*. Health Action International.
[94] World Health Organization, 2021. *COVID-19 Technology Access Pool (C-TAP)*. Geneva: WHO.
[95] Perel, M., Elkin-Koren, N., 2017. Black Box Tinkering: Beyond Transparency in Algorithmic Accountability. *Florida Law Review*, 69(1), pp.1–52.

or public-private partnerships that allow research risks to be shared. Nevertheless, the tension between profit and fairness of access persists as one of the most troublesome ethical challenges in the innovation system today.

2.4.3. Innovation and sustainability: SDGs, ESG and green innovation

Sustainability has become a key moral challenge for innovative companies in recent years. Companies are not anymore only requested to deliver economic value but should support the United Nations Sustainable Development Goals (SDGs) and align themselves to the ESG (Environmental, Social, Governance)[96] criteria. It isn't always simple to put such ideals into practice by means of real decisions.

The first challenge is the return on investment. Sustainable initiatives, such as the integration of recyclable products, the shift to renewable energy or the adoption of circular economy practices, are expensive at the start and only provide economic returns in the long or the medium term[97]. Listed companies, who are subject to the short-term perspective of shareholders' expectations, are not typically willing to invest in sustainable projects even if aligned with the expectations of society. This creates the challenge between competitive survival and moral leadership.

Another concern is the risk of *greenwashing*. Most companies declare sustainability efforts that actually do very little to change business practices. Some look for obscure emissions offsets, while others launch "green" product lines whose impact on their core business is peripheral[98]. As a result, sustainability becomes a public relations tool and not a meaningful change at the organizational level, damaging investor and customer trust.

Thirdly, there is the issue of international competition. The European companies, already weighed down by stringent environmental protection and reporting obligations (e.g., introduced by the Corporate Sustainability Reporting Directive), pay more than other companies in less highly regulated nations[99]. The effect is unfair competition and the risk of tempting offshoring that undermines some of the environmental successes.

Even if such challenges are present, sustainability can be a competitive advantage. Michael Porter and Claas van der Linde argued that stringent environmental rules, if well crafted, can

[96] United Nations, 2015. *Transforming our world: the 2030 Agenda for Sustainable Development*. New York: UN.
[97] Geissdoerfer, M., Savaget, P., Bocken, N.M.P., Hultink, E.J., 2017. The Circular Economy – A new sustainability paradigm. *Journal of Cleaner Production*, 143, pp.757–768.
[98] Delmas, M.A., Burbano, V.C., 2011. The Drivers of Greenwashing. *California Management Review*, 54(1), pp.64–87.
[99] European Union, 2022. *Directive (EU) 2022/2464 (CSRD) on corporate sustainability reporting*. Brussels.

stimulate efficiency and innovation, opening new markets and lower long-run risk[100]. John Elkington's *triple bottom line* model advises including the assessment of business performance not only by means of economic measures, but also by social and environmental ones as well[101]. Moreover, the texts on the circular economy outlined how the reduction-, reuse- and recycling-oriented models can generate value and strength in the perspectives of scarcity of resources[102].

Finally, the moral question is less the imperative to innovate sustainably than the ability of companies to do so routinely and in good conscience. Without verifiable measures, common norms and genuine manager commitment, the risk is that sustainability will be merely a rhetorical token. Innovation ethics faces the challenge of providing benchmarks to separate the innovators who do so responsibly and the innovators who, to the contrary, appeal the term sustainability as a pure and simple communication tool.

The important question then is who must be responsible for overcoming such dilemmas. This identifies the role of strategic management. It is the managers who get to decide what projects to finance, what technologies to develop, what markets to serve, and how to strike a balance between profit and responsibility. Basing the innovation strategies on the concept of ethics is not something dependent solely on external rules or the prodding of the stakeholders only, but on the leadership and guidance of the individuals who run the company.

The subsequent section (2.5) will then explore strategic management's contribution to the innovation-related ethical management and how organizational decisions, leadership models and the involvement of stakeholders can transform challenges into opportunities for collective value achievement.

2.5. The role of strategic management

2.5.1. The enterprise as a promoter of responsible innovation

Contemporary society, characterized by rapid technological development and global challenges related to the issue of sustainability, views businesses not only as economic actors anymore, but as actors who bear greater responsibilities towards society. Unlike when the

[100] Porter, M.E., van der Linde, C., 1995. Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, 9(4), pp.97–118.
[101] Elkington, J., 1997. *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Oxford: Capstone.
[102] OECD, 2020. *Green Innovation and Sustainable Growth*. Paris: OECD Publishing.

sole function of the company was to induce profit and employment, social legitimacy for the company today accompanies the manner in which it conducts innovation responsibly[103].

This function identifies expression at different layers. First, businesses show responsible innovation when they integrate ethical and sustainability considerations at the very core of strategic thinking, and not merely in peripheral functions or in marketing campaigns. A company that decides to introduce a new product or service cannot only investigate the profitability thereof, but should also account for the long-term social, environmental and culture effects thereof[104]. The strategic orientation then emerges as the determining factor: if the leadership incorporates such values in the decision-making process when it comes to investments, then the innovation ethics do not represent a foreign constraint anymore, but a competitiveness driver.

Secondly, companies are champions of responsible innovation when they are social laboratories: they experiment models, technologies and practices that can be scaled at the systemic level. This is also true for companies that put in place circular economy systems and reduced garbage and natural resources generation and consumption or companies that created inclusive access products for goods and services. For such companies, innovation push accompanies the creation of shared value by the client, the community and civil society[105].

Finally, corporations can take on a political role through the explicit participation in public debate and the shaping of the rules of the game. More and more companies are taking pledges to collaborate with institutions, NGOs and stakeholders to prepare standards and guidelines that spell out what responsibility in the complex fields of artificial intelligence, biotechnology and energy entails[106]. Such a move is not risk-free: the line between successful advocacy and mere lobbying opportunism is a thin one. The take-up by corporations of public responsibility has become a requisite complement to their strategic activism.

As such, the firm as the standard bearer for responsible innovation is not only a theory of how things ought to be but a necessity for managers. The firms that successfully incorporate innovation, sustainability and ethics have social legitimacy and reinforce the strength when crises arise. The firms innovating for the short term or for opportunistic motives stand to lose trust and reputational capital and long-term competitiveness.

---

[103] Freeman, R.E., 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.
[104] Jonas, H., 1984. *The Imperative of Responsibility*. Chicago: University of Chicago Press.
[105] Porter, M.E., Kramer, M.R., 2011. Creating Shared Value. *Harvard Business Review*, 89(1–2), pp.62–77.
[106] Owen, R., Bessant, J., Heintz, M. (eds.), 2013. *Responsible Innovation*. London: Wiley.

2.5.2. Management tools: codes of ethics, sustainability reports, impact assessments

When strategic management acts as the engine driving the company towards the path of responsible innovation, managerial tools constitute the effective levers translating the orientation into practice. The most spread and relevant among them are codes of ethics, sustainability reporting and impact evaluation. The three tools, if combined, allow the values to be transformed into the daily practices and the consequences be communicated to the stakeholders.

Codes of ethics are probably the most ancient and diffused tool. Initially written as checks lists of internal rules aimed at inspiring the behavior of the employees, they progressively became strategic tools translating the identity and key-values company[107]. Nowadays they hold not only prohibition and obligation lists, but general behavior principles: respect for the human rights, clarity in the relation with the client, fairness when dealing with the supply company, environmental commitment. More and more they hold articles on the utilization of new technologies and data treatment, indicative that the innovation ethics became company agenda. The most important limitation to this tool, then, is that the code risk becoming a statement of intent if not accompanied by control systems, internal sanctions, and leadership capable to set the example.

The sustainability report is a reporting and communication instrument. Through non-financial reporting, businesses release the environmental, social and governance effect they incur and introduce forth the efforts at integrating sustainability in the company development to the notice of the concerned parties[108]. During the last years, reporting has become mandatory for increasing numbers of businesses, and especially in Europe, through the implementation of the Corporate Sustainability Reporting Directive (CSRD) and the European Sustainability Reporting Standards (ESRS). It transformed reporting from a voluntary to a statutory obligation and worked to increase the level of transparent reporting and comparability at both company and industry level. But the ethical challenges remain the same in this respect: the risk of *reporting fatigue*, the many different standards and the risk of *greenwashing* when businesses announce successes and mask the major challenges.

Impact assessments represent the most developed frontier among management tools. They represent processes whereby the innovation consequences are evaluated ahead not only

[107] Kaptein, M., Schwartz, M.S., 2008. The effectiveness of business codes: A critical examination. *Journal of Business Ethics*, 77(2), pp.111–127.
[108] European Union, 2022. *Directive (EU) 2022/2464 (CSRD)* on corporate sustainability reporting. Brussels.

economically, but socially and environmentally and on grounds of fundamental rights too[109]. The European GDPR introduced Data Protection Impact Assessments (DPIAs) as a condition for high-risk processing and the same trend applies to other fields, for environmental sustain-ability and artificial intelligence too. Impact assessments commit the managers to think ahead the consequences of decisions taken and reduce the risk of causes which cannot be reversed and increase the level of the trust among the stakeholders. However the process effectiveness relies on the quality of the data collected, the autonomy of the assessment and the ability of the administration to take the findings into decision- and policy-making processes.

When used complementarily, the instruments permit one to develop a balanced system: codes of ethics establish precepts, sustainability reports communicate outcomes, and impact assessments sidestep pitfalls. Without complementarity, each risks failure: a code lacking follow-up is rhetoric; a report lacking clearly established goals is marketing; an impact assessment lacking implementation effects is a routine exercise.

Lastly, codes, reports and evaluation should not be regarded as separate tools, but parts of a single governance ecosystem. It is only when part of firm strategy and supported by committed leadership that they can be real transformation levers for translating innovation into a responsible and ethical process.

2.5.3. Creating Shared Value

Creating Shared Value (CSV) theory by Mark Kramer and Michael Porter is among the most prominent contributions to the last couple of decades' company responsibility discussion. According to the authors, the company may achieve sustainable competitive advantage by creating such strategies that concurrently develop economic and social value[110]. The result is that companies don't have to be confronted by the dilemma of decision-making when it comes to profit versus values and that the two elements may be inseparable and become each other's complement.

The CSV approach differs both from traditional corporate philanthropy and even from corporate social responsibility (CSR) by itself. Philanthropy implies the giving of only a portion of the profits for the social good, and CSR implies only the neutralization of the negative consequences of the business activity, while CSV implies the direct reassessment

[109] Stilgoe, J., Owen, R., Macnaghten, P., 2013. Developing a framework for responsible innovation. *Research Policy*, 42(9), pp.1568–1580.
[110] Porter, M.E., Kramer, M.R., 2011. Creating Shared Value. *Harvard Business Review*, 89(1–2), pp.62–77.

of the business model in social and environmental aspects[111]. The mission is not to do good after the production of profits, but to create shared value in the very business activity.

Porter and Kramer outline three areas to implement the theory. The first refers to the redesigning of markets and products: innovation through the development of goods and services addressing social or environmental needs. Practical case studies are health-oriented food products to assist in the control of obesity or low-carbon energy products to reduce the company climate footprint[112]. The second refers to the replatforming of value chain productivity: the cost savings through the reduction of wastage, energy or negative effect in the supply chain. The third refers to the development of local clusters through investments in the territory and community that can improve the economic and social context in which the company operates[113].

The past years have witnessed increasing numbers of corporations adopt CSV-like initiatives. Nestle, for example, has supported initiatives for improved nutrition and sustainable agriculture in the hope of striking a balance between profit growth and the reduction of malnutrition. Unilever has defined the *Sustainable Living Plan* to reduce the environmental footprint of product and improve the living condition for millions[114]. Even in finance, as with the case of microcredit, new models can marry accessibility and profitableness and generate shared value for corporations and societies[115].

CSV has similarly played a significant role in the realm of academia and management, but it also has detractors. Some scholars insist that it can only become a notion that cannot be measured objectively[116]. Others refer to the fact that corporations will adopt it opportunistically and call practices that represent the standard measures of effectiveness or social marketing campaigns the outcome of the creation of the "shared value"[117]. It then continues to be necessary to distinguish the rhetorical commitments and real business model shifts.

---

[111] Crane, A., Palazzo, G., Spence, L.J., Matten, D., 2014. Contesting the Value of "Creating Shared Value". *California Management Review*, 56(2), pp.130–153.
[112] Porter, M.E., Kramer, M.R., 2006. Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility. *Harvard Business Review*, 84(12), pp.78–92.
[113] Nidumolu, R., Prahalad, C.K., Rangaswami, M.R., 2009. Why sustainability is now the key driver of innovation. *Harvard Business Review*, 87(9), pp.56–64.
[114] Unilever, 2020. *Sustainable Living Plan Progress Report*. London: Unilever.
[115] Yunus, M., Moingeon, B., Lehmann-Ortega, L., 2010. Building Social Business Models: Lessons from the Grameen Experience. *Long Range Planning*, 43(2–3), pp.308–325.
[116] Beschorner, T., 2013. Creating Shared Value: The one-trick pony approach. *Business Ethics Journal Review*, 1(17), pp.106–112.
[117] Dembek, K., Singh, P., Bhakoo, V., 2016. Literature review of shared value: A theoretical concept or a management buzzword? *Journal of Business Ethics*, 137(2), pp.231–267.

Despite such limitations, the CSV model presents a useful point of reference in refreshing our understanding of the strategic management approach to innovation ethics. It suggests that value creation can never again be conceived as a shareholder-stakeholder conflict but as a comprehensive process whereby competitiveness and responsibility reinforce each other.

2.5.4. Innovation, competitiveness and long-term responsibility

The long-term responsibility question represents the central axis for the identification of the strategic management function in innovation orientation. In the short-term, the companies are evaluated mainly by the economic and financial performance, while the survival in the medium-long term relies on the ability to balance competitiveness and sustainability[118]. The balance represents the transition from a model focusing only on shareholder value to a perspective focusing on the value created for all the stakeholders and for society.

- One among the earliest questions pertains to time horizons. Innovations that are culpable generally entail mass-scale investments whose fruits arise only after years. The imperatives of the financial markets, looking for returns on a four-quarterly basis, mandate short-run solutions[119]. The dilemma for the management then is that of defending long-run consistency, rationalizing long-run gains to be had through sustainable channels to the investor and other stakeholders and countering shortsighted biases.
- Another driver is organizational resilience. Companies who are only concerned about profit maximization in the short term may appear competitive in the short term, but more exposed to crises, scandals or exogenous shocks. Companies who integrate long term-oriented ESG considerations intrinsically have better relations of trust with customers, providers and society in general and develop greater ability to adapt. The pandemic of COVID-19 showed that the organizations who integrate long term-oriented considerations of the ESG provided more stable performance and larger capacity to recover[120].
- Third among the factors is social license to do business. Controversial technologies such as genes and robots risk encountering hostility if perceived as secretive or harmful. Social license isn't a one-off award, but something that must be regularly

[118] Freeman, R.E., Harrison, J.S., Wicks, A.C., 2007. *Managing for Stakeholders: Survival, Reputation, and Success*. Yale University Press.
[119] Laverty, K.J., 1996. Economic "short-termism": The debate, the unresolved issues, and the implications for management practice and research. *Academy of Management Review*, 21(3), pp.825–860.
[120] Morningstar, 2021. *ESG Funds Outperformed During COVID-19*. Chicago: Morningstar Research.

earned through transparency, consultation and accountability[121]. The ethical then ceases to be a burden and instead becomes an investment that guarantees social acceptance and reduces the risk of conflict and limiting legislation.

The concept of strategic sustainability then entails the ability of the management to convert challenges in opportunities. Integrating ethics at the very heart of innovative strategies enables companies to differentiate themselves, to attract talent and capital and strengthen their reputation and build long-term value[122]. The challenge is to change the defensive mentality through which the responsibility occurs as a constraint and embrace the proactive mentality through which the responsibility constitutes an entire competitive advantage.

Finally, this whole chapter has revealed how innovation and ethics are two indissoluble dimensions. From the past histories of industrial revolutions to the most recent management techniques, the common feature is the following one: innovation only generates value if coupled with values and instruments able to steer it in the right direction and bring it to the common good. Management faces the challenge of converting this vision into operative decisions, weighing profits and responsibility, short- and long-term horizons, rationality and social legitimacy.

This is precisely what the next chapter on the ethics of artificial intelligence is all about. AI represents the frontier par excellence on which the challenges investigated in the following chapter seem most prominently: work and automation, data ownership, risk of discrimination, sustainability and governance. Agreeing on how the deployment and design of AI can be rendered compatible with the use of the most basic ethical principles involves addressing the most imperative challenge of the hour: ensuring innovation yielded by technology not only builds competitiveness, but also safeguards rights, inclusion and social trust.

[121] Suchman, M.C., 1995. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), pp.571–610.
[122] Eccles, R.G., Ioannou, I., Serafeim, G., 2014. The impact of corporate sustainability on organizational processes and performance. *Management Science*, 60(11), pp.2835–2857.

PART THREE: Ethics and AI (artificial intelligence)

3.1. Why AI challenges corporate ethics

3.1.1. Definitions and main characteristics (AI, ML, deep learning, socio-technical systems)

In recent years, artificial intelligence (AI) has been one of the most discussed and explored technologies, both economically and socially and ethically. Providing a definition of AI is a pretty difficult task, as it is an interdisciplinary area including computer science, mathematics, neuroscience, philosophy and social science. One among the most well-known definitions is the one provided by Russell and Norvig, according to whom AI is "the study of agents that receive perceptions from their environment and take actions that maximize their chances of success"[123]. That is, an AI system is designed and developed in order to automatically solve problems or make decisions, in an attempt at replicating a certain capacity of cognition typical in human beings, namely reasoning, learning and interaction.

The AI methods encompass a variety of approaches and methodologies. Among the most relevant is machine learning (ML), or, in particular, the set of methods that allow algorithms to "learn" from data and progressively perfect their outputs without being explicitly programmed. This capacity in machine learning has provided the foundation for a wide range of practical applications from voice and face recognition and market demand modeling up to financial risk analysis[124].

In ML, one sub-category that is essential to note is deep learning, based on multilayered artificial neural networks emulating how the human brain works. Deep neural networks allow complex data patterns to be characterized, and thus enabling data recognition in images, machine translation, and the creation of written and visual content. Over the recent few years, through deep learning, there are remarkable advances in the field of artificial intelligence, and the beneficiaries have been end-users as diverse as self-driving systems and massive-scale language models[125].

To be in a position fully to understand the phenomenon, a useful distinction runs between weak AI and strong AI. The first are systems built for a particular end, capable in a narrow domain of surpassing human abilities (e.g., a computer programme competent at playing a game of chess better than a world champion). The second are systems with general, true

[123] Russell, S., Norvig, P., 2021. *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
[124] Mitchell, T., 1997. *Machine Learning*. McGraw-Hill.
[125] Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.

intelligence, capable in a dynamic way of absorbing and adapting in a multiplicity of situations, as in humans[126]. Today, all practical applications are within the domain of weak AI, but debate on the chance and threat of eventual strong AI continues scientifically and philosophically.

Another issue to be considered concerns the non-neutrality of technology. AI is generally thought as a set of neutral tools, whose impact depends only on how it is utilized. Actually, as a variety of scholars in the philosophy of technology point out, design choices encapsulate values, prioritizations and perceptions on the world. AI therefore has to be considered as a socio-technical system, in which the technological aspects (algorithms, data, infrastructure) cross organizational, cultural and regulatory ones[127]. This means that algorithm design, dataset selection and business objectives are never technical choices, but ones with fundamental ethical and social consequences.

A specific example relates to the choice of training data. When historical data contains bias or discrimination, the algorithm will mirror and reinforce it. An AI system designed to process job applications could inadvertently discriminate against specific groups of individuals, duplicating existing forms of prejudice in the labour market. In that regard, technology is itself not neutral but a reflection of societies it emerges from[128].

These are the reasons why we are reminded of certain intrinsic attributes that differentiate AI from previous innovations. The first is scalability: a product developed in a specific context can be replicated on a large scale in a really short period, reaching millions of individuals. The second is transversality: it does not refer to a specific sector, as all the rest technologies do, but goes beyond really different sectors such as healthcare, finance, mobility, public administration and security. The third attribute is opacity: most of the models, and deep learning models in particular, are "black boxes" that are difficult to interpret, also by their own designers. This lack of transparency leads to issues on responsibility and how it can be possible to ensure accountability[129].

[126] Searle, J.R., 1980. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), pp.417–424.
[127] Floridi, L., Cowls, J., 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1).
[128] Barocas, S., Selbst, A.D., 2016. Big data's disparate impact. *California Law Review*, 104(3), pp.671–732.
[129] Burrell, J., 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

3.1.2. Historical evolution and stages of development (from the 1950s to generative AI)

To understand why artificial intelligence represents such a significant challenge for business ethics today, it is necessary to reconstruct its historical stages. AI is not a recent invention: its roots date back to the 1950s, when mathematicians, computer scientists and philosophers began to question the possibility of simulating human thought mechanisms through machines.

The first fundamental reference is Alan Turing's essay, *Computing Machinery and Intelligence* (1950), in which the scholar proposed the famous "Turing test" as a criterion for assessing whether a machine could be considered intelligent[130]. Shortly afterwards, in 1956, the Dartmouth Conference organized by John McCarthy, Marvin Minsky, Claude Shannon and others marked the official birth of the field of research known as "artificial intelligence"[131]. In those early days, there was almost unlimited optimism: many scientists believed that within a few decades, machines would achieve intelligence comparable to that of humans.

The 1960s and 1970s were characterized by significant advances, especially in symbolic AI. The first logical problem-solving programmes, such as *Logic Theorist and General Problem Solver*, were developed. In medicine, the MYCIN system was one of the first examples of an *expert system*, capable of suggesting diagnoses and therapies based on predefined rules[132]. However, these systems were fragile and limited to very narrow domains, unable to generalize outside the programmed rules.

Over time, technical difficulties and a lack of concrete results led to the so-called *AI winter*, a period of sharp decline in investment and confidence. The first occurred in the mid-1970s, when the scientific community began to doubt the possibility of building truly intelligent machines with the means available at the time. A second "AI winter" occurred between the late 1980s and early 1990s, when enthusiasm for expert systems gave way to disillusionment due to the high costs and lack of flexibility of the solutions developed[133].

The real renaissance of AI began in the 2000s, with the emergence of machine learning and, subsequently, deep learning. Three factors were decisive: the availability of enormous amounts of digital data (*big data*), the increasing power of computers, and advances in

[130] Turing, A.M., 1950. Computing machinery and intelligence. *Mind*, 59(236), pp.433–460.
[131] McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 1955. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.
[132] Shortliffe, E.H., 1976. *Computer-Based Medical Consultations: MYCIN*. New York: Elsevier.
[133] Crevier, D., 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.

learning algorithms. Algorithms such as *support vector machines* and, later, deep neural networks enabled impressive results to be achieved in fields such as image recognition, machine translation, and natural language processing[134].

A symbolic moment came in 1997, when IBM's Deep Blue supercomputer defeated world chess champion Garry Kasparov, demonstrating the superiority of machines in highly specialized tasks. Even more impressive was DeepMind's AlphaGo's victory in 2016 against the world champion of Go, a game considered for decades too complex to be mastered by a computer[135].

In recent years, attention has shifted to generative AI systems, capable not only of recognizing patterns, but also of creating new content. Large language models, such as GPT developed by OpenAI and integrated by Microsoft, or image generation tools such as DALL·E and Stable Diffusion, have brought AI into the hands of millions of users around the world. These technologies open up extraordinary prospects – from the possibility of automating text writing to the creation of multimedia content – but at the same time raise unprecedented dilemmas about intellectual property, the reliability of information and the manipulation of public opinion[136].

Today, AI is considered a general-purpose technology (GPT), on a par with electricity or the Internet. Its ability to be applied across almost all productive sectors, from transport to healthcare, finance to agriculture, means that its effects are not limited but systemic. This characteristic makes it, on the one hand, an unprecedented driver of innovation; on the other, a source of risks and ethical dilemmas that cut across the entire social and economic fabric[137].

The history of artificial intelligence has not been linear, but one of enthusiasm, disappointment and rebirth. From the initial ambitions to replicate human thought, through the failures of expert systems, to the current successes of deep learning and generative AI, a common thread emerges: each phase has expanded the technological potential but also the ethical and social dilemmas.

3.1.3. Concrete impacts on business ethics (neutrality, dilemmas, key sectors)

Artificial intelligence is no impartial technology, but a set of tools and practices infused with worldviews, priorities and values. This element gives its influence on business ethics a

[134] Mitchell, T., 1997. *Machine Learning*. McGraw-Hill.
[135] Silver, D. et al., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, pp.484–489.
[136] Benaich, I., Hogarth, N., 2023. *State of AI Report*.
[137] Brynjolfsson, E., McAfee, A., 2017. *Machine, Platform, Crowd*. New York: Norton.

certain fateful dimension. Every element in the AI lifecycle – data collection, algorithm development, release, and monitoring – involves decisions with social and ethical consequences. That is why AI operates to multiply and redefine challenges innovation ethics already faces and obliges business companies to assume additional responsibilities.

The first argument concerns the perceived neutrality of technologies. Managers and builders tend to have a vision in mind of a conception of AI as a neutral tool, capable of producing equitable decisions based on data. But reality is quite different: the data sets on which the algorithms are trained are reflective of the prevailing levels of bias and social inequity. Experiments have been shown, for example, that face recognition systems are errorprone on people with dark skin or on females, betraying algorithmic discriminatory nature directly stemming from the quality of data collected[138]. Or, in a similar vein, recruitment algorithms available can punish female job seekers if trained on the history of business displaying uneven recruitment behaviors[139].

The second ethical issue is the issue of balancing responsibility and profits. Some AI-driven online platforms' business models are centered on user-time maximization, usually by whatever means, including profiting from their psychological well-being. Recommendation bots, designed as user interaction devices, in fact end up rewarding sensational or divisive content, propagating digital addiction and disinformation[140]. The organization in question has a straightforward trade-off here: make as many profits as possible in the short term or adopt a more responsible approach creating long-term sustainability and trust.

Another issue concerns transparency and responsibility. Some AI systems, especially deep neural networks, are opaque: both developers and end-users do not fully understand how the model operates internally. Such incomprehensibility prevents the assignment of responsibility in case the system makes a mistake or inflicts injury. Who is responsible if an algorithm unfairly denies a loan to a client? The company that uses it? The programmers who created it? The data provider? The "accountability gap" stretches usual business ethics models, based on availability of blame and credit assignment to well-defined decisions[141].

The ethical impacts are most prominently illuminated in a few exemplary industries. In medicine, AI heralds rapid diagnosis and customized therapies, but raises concerns of fairness: systems trained on narrow populations are less effective on different groups,

[138] Buolamwini, J., Gebru, T., 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, pp.1–15.
[139] Dastin, J., 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10 October.
[140] Zuboff, S., 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.
[141] Burrell, J., 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

generating unintended discriminatory impacts[142]. In human resources, the use of algorithms in selection or evaluation of staff risks importing unintended biases difficult to perceive, affecting workers' fundamental rights[143]. In finance, statistical models make forecasts that increase risk management but enforce exclusion by denying lines of credit to entire segments of clients on statistical relationships rather than individual judgment[144]. In public administration and security, AI is used in surveillance and facial recognition systems, and the question is raised concerning privacy and individual freedom[145].

These cases demonstrate that AI defies business ethics' first principles. This does not involve establishing whether a business is “ethical” in its customary practices, but whether and how what innovations it delivers are in harmony with principles of justice, fairness and sustainability. Where there is a proliferation and all-encompassing application of technology, managers are challenged to consider, in addition to applications' profitability, their overall social impact.

Therefore, firms cannot afford to limit themselves to simple compliance with prevailing regulations or release of boilerplate codes of ethics, but instead, need to equip themselves with effective governance tools capable of promoting transparency, inclusion and responsibility. That way, and that way only, it shall be possible to prevent the dangers that AI can cause in the form of fresh inequities and instead transform it into a driver of shared value.

The research conducted has shown that AI is no neutral tool, but it is a socio-technical system with values, reflects prioritizations and produces ethical consequences shared by companies and society. With the historic reconstruction and definitions, we come to understand that AI does nothing more than make bigger the challenges already seen with innovation in general: bias, inequality, lack of transparency, conflicts between financial gain and responsibility. These points confirm that business ethics cannot be considered anymore as an external or complementary constraint, but it must be integrated as part of management strategy.

[142] Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), pp.447–453.
[143] Ajunwa, I., 2020. The paradox of automation as anti-bias intervention. *Cardozo Law Review*, 41(4), pp.1671–1714.
[144] Citron, D.K., Pasquale, F., 2014. The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), pp.1–33.
[145] Andrejevic, M., 2020. Automated facial recognition in public: A comparative analysis. *Surveillance & Society*, 18(2), pp.212–226.

However, voluntary action by individual companies is not enough to ensure responsible application of AI. The cross-cutting diffusion pace and reach of the technology necessitated regulatory processes capable of establishing joint minimum conditions, protecting fundamental rights and ensuring trust among users. That is why, both internationally and in Europe, debate on how to regulate artificial intelligence intensified significantly in recent years.

The next section will therefore be devoted to regulatory issues: we will pit the European approach against the AI Act and the GDPR, the requirements of international organizations such as UNESCO and the OECD, and non-EU requirements such as the American and Chinese ones. Finally, we will insist on the frailness and flawed nature of those approaches in an attempt to be clear on how effective they are in practice in aligning firms towards responsible and trustworthy AI.

3.2. Regulatory challenges

3.2.1. The European approach: AI Act, GDPR and fundamental rights

The European Union made itself the international leading player in the regulation of artificial intelligence, seeking to build a model based on the protection of fundamental rights and public trust. The EU chose not to limit itself to general principles, but instead to propose binding regulatory tools, in an effort to make "*trustworthy AI*" the hallmark of the European model[146].

A primary reference point is the General Data Protection Regulation (GDPR), enforceable in 2018, which, while not a direct regulation on AI, has been a determining force. The GDPR articulates basic principles such as a limitation on data collection, transparency on processing and a right to explanation in the event of automated decisions[147]. Article 22 in particular enacts that individuals have a right not to be subjected to decisions made solely on an automated processing conferring important legal effects. This Article has been a primary reference point for AI systems in sensitive sectors such as finance, human resources and healthcare.

Then there was the 2021 introduction by the European Union of the draft proposal on the European Regulation on Artificial Intelligence (AI Act), which shall be the first worldwide

[146] European Commission, 2021. *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*. Brussels.
[147] European Union, 2016. *General Data Protection Regulation (GDPR)*, Regulation (EU) 2016/679.

exclusively on AI[148]. The AI Act adopts a risk-based approach, whereby it categorizes AI systems into four levels:

- Unacceptable risk: prohibited (i.e., Chinese-style social scoring systems or forms of cognitive manipulation).
- High risk: permitted only with strict conditions that demand transparency, documentation, human monitoring and risk management (e.g. AI in personnel selection, credit, critical infrastructure).
- Minimal risk: less transparency requirements (i.e., chatbots that don't need to declare that they are not human).
- Low risk: low cost, in-house solutions, e.g., software packages or virus protection.

This strategy aligns with Europe's need to find a balance between innovation and the safeguarding of rights. The AI Act does not seek to put a brake on technological progress, but on creating conditions of trust in order to facilitate the responsible uptake of AI. Contrary to other situations, the EU stresses that the legitimacy of innovation cannot be severed from a regulatory framework meant to forestall its misuse.

Another feature of the European method is highlighting basic rights. The Parliament and the Commission have been adamant, time and again, in making sure that the regulation of AI is in line with the Charter of Fundamental Rights of the European Union, and in securing rights such as human dignity, privacy, equality and non-discrimination[149]. The same line is followed by the *Ethics Guidelines for Trustworthy AI*, endorsed in 2019 by the High-Level Expert Group on AI, and which laid down the requirements for "trustworthy" AI: legality, ethics, technical and social robustness.

However, there are a couple of fundamental concerns yet. Some are worried that the GDPR and the AI Act would generate compliance costs that are problematic in small and medium-sized enterprises, and it could only favor big players who can sustain the regulatory expenses. Some emphasize the point that, though a first in regulating, the EU yet hasn't developed an AI industrial ecosystem comparable with the United States and China, and it could yet put the innovative potential of Europe on hold[150].

[148] Veale, M., Borgesius, F.Z., 2021. Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), pp.97–112.
[149] European Commission, 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence.
[150] Wachter, S., Mittelstadt, B., Floridi, L., 2017. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), pp.76–99.

The European approach remains a global point of reference in any case. As the GDPR itself has been a blueprint replicated or modeled by all nations outside the European Union, the AI Act is set to influence international law in the years ahead, too. To multinationals, it signifies that Europe is a major marketplace, and a test bed, too, whose consequences could end up having resonances worldwide.

3.2.2. International guidelines and approaches outside the EU (UNESCO, OECD, NIST, China)

Besides the European Union, there are several international institutions and nation-states that have published guidelines and regulatory regimes on AI, aiming at a balance between technological development and the enforcement of rights. This regulatory patchwork reflects different political, cultural, and economic priorities, and shows how there is no universal consensus on how to regulate AI globally yet.

An important first reference point is UNESCO, which in 2021 approved the *Recommendation on the Ethics of Artificial Intelligence*, adopted by 193 Member States[151]. This is the first universal regulatory instrument on AI, enshrining principles of transparency, fairness, protection of privacy, protection of human rights and respect for cultural diversity. The recommendation places a specific focus on ensuring that AI does no harm and does not increase existing inequalities, rather social inclusion and sustainable development. Despite the fact that it is not legally binding, the text of UNESCO has particular political and symbolic value since it enshrines a consensus shared by almost all the world's nations on aligning AI with ethical purposes.

Another significant actor is the Organization for Economic Co-operation and Development (OECD), which in 2019 adopted the *Principles on Artificial Intelligence*, approved by the G20 in turn[152]. These principles are based on five pillars: inclusive and sustainable growth, protection for human rights, transparency and explainability, robustness and security, and responsibility of the actors involved. The OECD notes that AI need be innovative and *trustworthy*. The strength point of this approach is the ability it has to influence national policies and international standards, as well as be a reference framework both for governments and companies.

The American side takes a more diffuse and market-oriented approach to AI regulation. The National Institute of Standards and Technology published in 2023 the *AI Risk Management*

[151] UNESCO, 2021. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.
[152] OECD, 2019. *Principles on Artificial Intelligence*. Paris: OECD Publishing.

Framework[153], a voluntary tool that provides principles and guidelines by which institutions and firms can find, assess, and manage risks associated with AI. The framework is voluntary, but it reflects the American preference for soft law and self-regulation over rigid laws. The White House, in the meantime, published the *Blueprint for an AI Bill of Rights*, which provides basic user rights – protection from discriminatory systems and transparency and the right to opt out – but has yet to issue specific federal laws. Essentially, the United States is giving a balance between technological first mover advantage, the capability to respond to social pressures and ethical expectations, without, however, damaging the competitiveness of firms.

The Chinese approach is radically different. China considers AI as a tool in support of economic competitiveness and political power. In 2017, it published the *Next Generation Artificial Intelligence Development Plan*, and it aims by 2030 to be world leader on the issue[154]. Chinese regulation is more interested in national security, social stability and contents control, than in individual rights protection. For example, in 2022, the Cyberspace Administration of China published guidelines on online recommendation algorithm monitoring, with registration and transparency requirements, in order to prevent diffusion of contents considered dangerous or destabilizing[155]. China published specific rules on generative AI in 2023, and it requires contents generated to be conform with the "core socialist values". This approach expresses a vision on AI ethics radically different from the West's: Whilst in the West and in some areas in the United States individual rights protection is the focus, in China there is a general perception that ethics is a social order and political stability protection issue.

Comparison of the models shows a lack of a prevailing international consensus, but complementarity among different approaches. UNESCO and the OECD offer universal frameworks based on shared, but non-binding principles. The most advanced model of legally binding regulation is the EU in the form of the AI Act. The United States takes a market-focused and soft law approach, and China uses AI as a tool of economic development and political governance. This creates a need on the part of worldwide companies to operate in a fragmented world, where compliance is a question of adjustment in conformity with substantially different rules, principles and expectations in function of the targeted market.

[153] National Institute of Standards and Technology (NIST), 2023. *AI Risk Management Framework*. Washington, DC: U.S. Department of Commerce.
[154] Government of China, 2017. *New Generation Artificial Intelligence Development Plan*. Beijing: State Council.
[155] Cyberspace Administration of China, 2022. *Provisions on the Administration of Algorithmic Recommendation in Internet Information Services*. Beijing.

3.2.3. Limitations and critical issues of regulatory frameworks

In spite of numerous recent regulatory initiatives on artificial intelligence both globally and at the national levels, the regulatory picture remains a spotty and uneven one with significant gaps. This result begets uncertainty and induces potential danger in the efficacy of regulations as a collective force in nudging AI towards responsible and sustainable use.

The first limitation is geographical and cultural fragmentation. The European Union, the USA, China and the other big powers have chosen very different approaches, differing both in political priority and in value system. The EU favors the safeguarding of fundamental rights and the establishment of a binding regulatory framework, the USA a self-regulation and soft law strategy, and China the use of AI as a development instrument as well as a political governance tool. Such diversity has the potential to build up a "regulatory mosaic", whose lack of coherence troubles multinationals in aligning their activity[156].

Another problem is the speed of technological innovation relative to the speed of regulation. The AI technologies are advancing at a literally staggering pace: consider the unforeseen development of generative AI that in a couple of months has created unprecedented copyright, disinformation, and liability problems. Politicians, on the other hand, work with slow and laborious procedures that take a few years before enacting laws. The time lag creates a threat of "ex post regulation", in which the rules come too late, following the distortion or damage that has been created[157].

The third limitation concerns the voluntary nature of international standards. Charters such as UNESCO's or OECD's establish shared principles of great symbolic weight, yet no binding instruments necessitating effective application by companies. Absent sanctions, firms are generally in a position to exploit such standards as communication or reputational justification instruments rather than changing business models. Such a practice, termed "*ethics washing*", has the potential to turn ethical principles into aphorisms rather than effective governance criteria[158].

The other fundamental issue is doubt over enforcement mechanisms. Even in jurisdictions that have binding rules, as in the European AI Act or the GDPR, it remains to be seen whether effective control and effective sanctions are possible to ensure. Supervisory authorities are underfunded and are stretched by cumbersome and unclear systems. Also, lack of mutual

[156] Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133).
[157] Bostrom, N., Yudkowsky, E., 2014. The ethics of artificial intelligence. In: *Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.
[158] Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), pp.501–507.

technical standards makes it difficult to verify compliance with requirements such as transparency, robustness or fairness[159].

Ultimately, there is a lack of symmetry in big firms and small firms. The former, via the economic and legal resources in their disposal, handle intricate rules and turn them into a competitiveness advantage. The latter are put in danger by penalties via high levels of compliance costs, and their ability to innovate and reach the AI marketplace decreases. This in turn could stimulate power concentration in a small number of Big Tech companies' hands, and competition and innovative ecosystem diversity are impaired[160].

Analysis of regulatory regimes has indicated how the European Union, international institutions and nation-states are seeking to build principles and standards in order to guide the development and application of artificial intelligence in a responsible and ethical way. But it has been made equally clear that regulation, while a prerequisite, falls short in addressing all the dilemmas. Geographical fragmentation and contrasting value sets in Europe, the United States and China beget a regulatory patchwork inducing disarray in globally operating businesses, and meanwhile, technological progress is overwhelming institutions' capacity for response.

This, in turn, means operating a multifaceted business, whereby legal compliance is no more than part of the problem. Even with increasingly evolved regulations, there are a host of tangible risks in relation to the mere fact of running AI systems. Algorithmic bias, transparency, privacy breaches, security and impact on work and democracy are issues which cannot be addressed by abstract rules only: they require individual company policies and in-house governance tools.

The next section shall therefore focus on the business's most important areas of risk, analyzed in three main points: issues on bias, fairness, and transparency of algorithms; data handling, privacy, and security concerns; and finally, work, employment, and democratic life consequences of AI.

[159] Veale, M., Borgesius, F.Z., 2021. Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), pp.97–112.
[160] Ebers, M., 2022. Standardisation of AI and EU law: Setting the right framework for SMEs. *European Business Law Review*, 33(2), pp.259–280.

3.3. Critical risk areas for businesses

3.3.1. Bias, fairness e black box problem

Perhaps the most discussed and controversial danger associated with artificial intelligence is the issue of algorithmic unfairness and bias. Contrary to popular perception that algorithms are value-neutral and objective tools, in reality, they tend to mirror and enshrine prevailing bias in data or in design choices of programme developers. This is not merely harmful in the sense that it undermines user trust, but it directly affects business ethics, and jeopardizes the organization's reputation, social legitimacy and, in some cases, legal correctness.

Bias can manifest in many different forms. The first is data. When data sets used in training a model are non-generalizable to the true population, the developed algorithms will be likely to make discriminatory decisions. There are well-documented results that were seen in facial recognition systems having substantially larger error rates recognizing darker-skinned individuals and females compared with white males, showing a systematic bias in keeping with the composition of data sets used in training[161].

Another concern is algorithm design. Even with a balanced data set, designers' choices when designing an algorithm can be discriminatory. Deciding which variables are significant in regard to a customer as potentially creditworthy, as an example, involves value judgements never objective. Some credit-scoring models based on AI have been penalizing certain user groups, e.g., young individuals or individuals living in socio-economically distressed neighborhoods, recreating social exclusion forms already present in the economic system[162].

The third issue regards the application context. An algorithm developed with a specific function in consideration can be detrimental if put in a different context. Forecasting tools developed to forecast the chances of recidivism in the US legal process, as an example, were discovered to be automatically harsher on African-American suspects and thereby acted to entrench existing historical disparities[163].

The issue of fairness – i.e. the fairness of the results produced by algorithms – is therefore central to the ethics of AI. But no general definition exists of fairness: it can be equal treatment, equal chances or proportionality of outcome. This plurality of senses is expressive of the difficulties in translating subtle ideas of morality into mathematical and technical

[161] Buolamwini, J., Gebru, T., 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, pp.1–15.
[162] Hurley, M., Adebayo, J., 2016. Credit scoring in the era of big data. *Yale Journal on Regulation*, 103.
[163] Angwin, J., Larson, J., Mattu, S., Kirchner, L., 2016. Machine bias. *ProPublica*, May 23.

constraints. For companies, it therefore no longer involves a question of formally demonstrating the correctness of an algorithm, but also examining its real effect on actors[164].

Beyond the issue of bias, a larger issue is that the workings are opaque, or "*black box problem*". Many machine-learning systems, particularly deep neural networks, are black boxes: they make good predictions, but no clear line of logic exists as how they got there. This hinders transparency, explainability and a sense of accountable action. Deploying an opaque algorithm in business is a reputational and legal liability as how can a customer be justified in rejecting a loan if managers themselves cannot say what was the logic in the model?[165]

The black box problem is most relevant in highly regulated sectors, like healthcare and finance, where automated decisions impact fundamental rights. That is why we are designing mechanisms like explainable AI (XAI), aiming to make algorithmic procedures for making decisions more interpretable. But here as well, there are trade-offs: increased explainability decreases the accuracy of models, and companies are left facing a tough choice between transparency and effectiveness[166].

The consequences for companies that do not consider these risks can be severe. Alongside possible legal sanctions, examples of algorithmic bias have consistently involved severe reputational damage. In 2018, Amazon was compelled to abandon an automated recruitment system that automatically excluded female job applicants, causing widespread outcry and calling into question the legitimacy of its HR procedures[167]. In financial services, a variety of studies have established that credit rating models designed on the basis of AI were discriminatory against ethnic minorities, exposing banks to allegations of discriminatory practices and loss of customer confidence[168].

3.3.2. Privacy, data governance and security

Another essential area in companies' implementation of systems based on artificial intelligence revolves around data administration. Current AI, and machine learning in general, relies on access to vast levels of data, gathered, accumulated and processed on a

[164] Barocas, S., Hardt, M., Narayanan, A., 2019. *Fairness and Machine Learning*. Cambridge, MA: fairmlbook.org.
[165] Burrell, J., 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).
[166] Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
[167] Dastin, J., 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10 October.
[168] Citron, D.K., Pasquale, F., 2014. The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), pp.1–33.

hitherto unprecedented scale. The data-centric paradigm has fundamental implications for privacy at the individual level, data governance practices at the company level and security in IT systems.

The first issue concerns privacy. The AI systems are trained through big data sets, typically including personal and sensitive information: biometric, behavioral, clinical and financial data. Even if data are anonymized, in fact, there is a possibility of re-identification through cross-matching different sources of information. Owing to the application of new profiling techniques, it is possible to reconstruct people's habits, preferences and political views as well, with the potential violation of the right to privacy[169]. Some recent news cases have highlighted how social media and mobile apps retain a broad range of data on people, sometimes without informed consent.

This relationship concerns the issue of informed consent. Precisely, the user is confronted with long and obscure privacy policies under which he approves conditions without real knowledge. This produces an "information asymmetry" among persons and companies: the former lack effective defence mechanisms, the latter use data as a decisive element, generating economic value via targeted advertising, personalization or resale on the secondary market[170]. The issue is legal as well as ethical on the companies' part, as data processing is a point in trust-building with customers and stakeholders.

The second issue concerns data governance, i.e. the procedures, rules and tools through which an organization manages its data. Effective governance involves assuring information quality, integrity and security, making decisions as to who has the right access to what, why and with what protections. In reality, however, many companies do not implement proper policies. The main concerns are issues of data traceability (where it is coming from and under what licence), how to handle bias intrinsic in information sources and protection from illegal use[171]. Without efficient data governance, AI systems are vulnerable to producing outputs that are unreliable and biased and, in addition, companies are exposing themselves to legal sanctions and reputational damage.

The closest example we have to a practical case is the data on which generative AI models are trained. Some are generated by scraping text and images from the web without particular authority, and so they create copyright controversies and worries over whether their use is legal. Some lawsuits, as in artists and photographers suing companies producing software

[169] Solove, D.J., 2021. *Understanding Privacy*. Harvard University Press.
[170] Zuboff, S., 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.
[171] Khatri, V., Brown, C.V., 2010. Designing data governance. *Communications of the ACM*, 53(1), pp.148–152.

generating images, demonstrate how data collection activities are precarious and companies are subject to contentious lawsuits[172].

The third fundamental area is security. AI systems are no exception to cyber attacks and present new specific vulnerabilities. *Adversarial attack* techniques, e.g., allow fooling recognition algorithms in pictures through slight changes impossible to detect by eye, leading to serious errors. Similarly, an automated driving car can be tricked by slight modifications in a traffic sign, with possible disastrous consequences[173]. For companies, AI security is both a technical requirement and an ethical commitment: system reliability implies protecting the user and stakeholders from material and intangible risks.

Another dimension is liability in the event of breaches or accidents. When a cyberattack leverages vulnerabilities in an algorithm and causes economic or physical damage, who is responsible? The developer, the user, or both? This dilemma goes along with the greater question of responsibility in AI. Where no specific rules apply, firms are subject to lawsuits and loss of trust. Some regulatory regimes, the GDPR and the European AI Act among them, are beginning to detail responsibility requirements, yet the question remains lingering and particularly troublesome on the world stage[174].

3.3.3. Work, employment and democracy

Impacts on the world of work are among the most contentious and hotly discussed issues, economically and socially, and also on an ethical and social level, due to the impact of artificial intelligence. Artificial intelligence, in fact, does not merely put into practice new techniques of production: it redefines the sense of competence, responsibility and participation themselves, and in a sense goes well beyond the world of work and affects democratic institutions' operation.

Economically, AI raises the traditional dichotomy of technology and manual labor. Firms in the United States could automate as many as 47% of the most vulnerable jobs, namely the ones entailing repetitive or codable tasks[175]. Robots and programmes on machine learning are increasingly replacing manual, administrative and routine work and creating a lower demand for medium-skilled jobs. This induces labour market polarization: on the positive,

[172] Vincent, J., 2023. Getty Images sues Stability AI over copyright infringement. *The Verge*, 17 January.
[173] Goodfellow, I., Shlens, J., Szegedy, C., 2015. Explaining and harnessing adversarial examples. *International Conference on Learning Representations (ICLR)*.
[174] Wachter, S., Mittelstadt, B., Floridi, L., 2017. Why a right to explanation of automated decision-making does not exist in the GDPR. *International Data Privacy Law*, 7(2), pp.76–99.
[175] Frey, C.B., Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, pp.254–280.

skilled professions linked with designing, administering and commanding AI systems; on the negative, precarious or low-skilled work that cannot be quickly automated but is less secure and lower paid[176].

The impact is uneven. Some sectors, such as manufacturing, supply and financial services, are most susceptible to displacement threats. Others, such as hospitals and schools, see AI as a possibility of complementing, as opposed to replacing, humans: virtual nurses, predictive diagnosis and customized pedagogy can support professionals without replacing them. The "replacement" vs. "complementarity" binary thus sits at the heart of how AI shall redefine the future of work[177].

One basic issue concerns work quality. The universal application of AI-driven monitoring and evaluation systems – from worker productivity monitoring to apps in the sharing economy – can enhance efficiency yet create new forms of alienation. Some companies have applied software capable of monitoring workers' activity in real time, monitoring mouse movements, time spent in front of a computer or tone of voice during communication with clients. Such practices, as they make workers more efficient, also give birth to grave ethical issues associated with surveillance and work dignity[178].

Ethical work governance by AI therefore requires both business policy and public initiative. Continuing education is essential in order that workers be prepared for the new skill sets required. *Upskilling* and *reskilling* programs have been established in most multinational corporations yet are still insufficient in the face of the scale of transformation. The lack of transition plans risks denying broad segments of the population access to the fruits of innovation, fueling social tensions and unfairness already present[179].

The effects of AI are not only work-related, but democracy as well. Algorithmic political profiling, diffusion of opinion manipulation and targeted communications have already yielded real-world results in a variety of election environments. The Cambridge Analytica scandal, by misusing data from millions of Facebook users in an effort to influence election campaigns in the Unites States and in Great Britain, shows how AI can be used in manipulating democratic processes[180].

---

[176] Autor, D.H., 2015. Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), pp.3–30.
[177] Brynjolfsson, E., McAfee, A., 2014. *The Second Machine Age*. New York: Norton.
[178] Moore, P.V., Akhtar, P., Upchurch, M., 2018. Digitalisation, automation and the future of work in Europe. *Work Organisation, Labour & Globalisation*, 12(2), pp.1–12.
[179] Arntz, M., Gregory, T., Zierahn, U., 2016. The risk of automation for jobs in OECD countries. *OECD Social, Employment and Migration Working Papers*, No. 189.
[180] Cadwalladr, C., Graham-Harrison, E., 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica. *The Guardian*, 17 March.

AI also aggravates the problem of disinformation. After the spreading of generative models, it has been cheap and easy to generate *deepfakes* and believable fake media. This increases the threat of enormous manipulation campaigns that are capable of breaking down citizens' trust in institutions and in traditional media. In that sense, the line between information and propaganda increasingly blurs, and the quality of public debate is significantly impacted[181].

Finally, the strongly centralized concentration in the hands of colossal technological platforms provides a cause for concern in democracy-market relations. Companies controlling world AI infrastructures – search engines, social networks or online storage – dispose over a hitherto unprecedented influence on information circulation, trade and even political decisions. The concentration of economic and information power can be converted into a form of "private technocracy", constraining the leeway in making decisions by democratic institutions and posing a new threat in world governance[182].

These points suggest that while AI is a transformation which goes well beyond a technical or managerial issue, it has a significant impact on the social and political organization. Such firms, therefore, need to be mindful that decisions have consequences which extend well beyond the market, and extend into workers' rights, social relations and the character of democratic life.

The risk area study has verified that, in addition to offering never before seen opportunities in competitiveness and innovation, artificial intelligence also poses gigantic ethical, social and political challenges to business. Bias and discrimination, data privacy infringement, security threats, labour shift and democracy-related threats are no longer theoretical scenarios, but already reality concerns which influence organizational legitimacy and trust among stakeholders.

The question is no longer whether, but how, in a systematic and efficient manner, multinational corporations can address such challenges. Management comes into the picture here: it is only through certain instruments of governance that ethics in AI can be converted from a general principle into a day-to-day practice. Management is challenged to integrate responsibility into the process of decision-making, detail procedures of control, establish criteria for evaluation and construct organisational conditions propitious to innovation without watering down fundamental values.

[181] Chesney, R., Citron, D., 2019. Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), pp.1753–1820.
[182] Zuboff, S., 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.

The next part will therefore be dedicated to practical tools for managers, and how companies can apply methodologies as varied as AI Ethics by Design, impact and risk management systems, audits and certifications, and up to the release of in-house policies, ethics committees and responsible procurement practices.

3.4. Practical tools for managers

3.4.1. AI Ethics by Design e impact assessment

Among the most discussed and promising approaches in placing ethical principles in the building of artificial intelligence, there is AI Ethics by Design. The basic proposal is that ethics cannot be considered as a peripheral or adjunct element on which business processes come afterwards, but a principle that is incorporated from the outset in the design, development and implementation of AI systems[183]. This approach takes up and generalizes the *privacy by design* model put forth by the GDPR and enacts it on all the ethical issues: fairness, transparency, responsibility, inclusion and sustainability.

The by design approach is based on the idea that all the problematic elements in relation to AI – bias, discrimination, lack of transparency, security vulnerabilities – can be prevented if they are considered in the design stage. This means, in particular, that datasets utilized in training are diversified and impartial in a systematic manner, that models are challenged in relation to robustness and interpretability, and human control processes are implemented in relation to decisions with significant effect. That is, it is a question of transferring ethical principles into particular design requirements translatable into technical and operative specs[184].

Those organizations that adopt AI Ethics by Design are driven by two main reasons. The first is legal and reputational risk mitigation: it is cheaper by a long way either to prevent bias or a privacy incident than it is to react after something has happened. The second is building stakeholder trust: customers, employees, investors and regulators are increasingly interested in business social responsibility. AI that is perceived as unfair or obscure challenges business legitimacy, but trustworthy AI can be a long-term competitiveness driver[185].

---

[183] Cavoukian, A., 2011. Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario.

[184] Floridi, L., Cowls, J., 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1).

[185] Jobin, A., Ienca, M., Vayena, E., 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), pp.389–399.

After by design, a second valuable tool is impact assessment, or thinking ahead about the potential effects of an AI system before it goes into use. The tool borrows techniques created in other areas, e.g. environmental or social impact assessment, and is a rough approximation of the ethical, legal and reputational effects a technology is going to have.

In the European Union, the General Data Protection Regulation ushered in Data Protection Impact Assessments (DPIAs) as a condition precedent before a high-risk process, and the same has been utilized in the field of AI, e.g., in the *High-Level Expert Group on AI* guidelines[186]. Impact assessments can be in diverse forms. Some organizations utilize structured questionnaires to determine the risks of bias, discrimination, privacy violation or socially undesirable effects.

Some use multidisciplinary teams made up of engineers, lawyers, ethic experts and stakeholder representatives. Some designs permit the issuance of results in the form of reporting back to institutions and the public. In any case, the intention is to make the managers think ahead and consider the social consequences of their decisions, so that innovation does not proceed without consideration[187].

Microsoft, as an example, developed its own *Responsible AI Standard* and integrates ethical impact evaluations in its product development processes. The guide includes working instruments, such as checklists and guidance, and requires all AI high-risk projects to be assessed by a company's multidisciplinary committee. IBM has developed audit and evaluation processes, ensuring systems developed by AI are aligned with transparency and nondiscrimination principles and contribute to universal standards through joint efforts with institutions and universities[188].

Even if by design and impact assessment methods are promising, there are, nevertheless, several critical points with them. First, there is a risk of formalism: turning ethics into a bureaucratic checklist to be duly ticked off a form can lower its true effectiveness. Then there is no common benchmark: all companies do their own models, and comparing and externally validating are then unthinkable. Finally, integrating ethics in design procedures implies new, multidisciplinary skills that are not shared by all companies, and in particular by small and medium ones[189].

---

[186] European Commission, 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence.
[187] Reisman, D., Schultz, J., Crawford, K., Whittaker, M., 2018. Algorithmic impact assessments: A practical framework for public agency accountability. *AI Now Institute Report*.
[188] Microsoft, 2022. *Responsible AI Standard*. Redmond: Microsoft.
[189] Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), pp.501–507.

In spite of all these limitations, AI Ethics by Design and impact evaluations are the first practical tools that enable managers to turn principles into real-world practices in business. They illustrate that AI responsibility cannot be entrusted into the domain of external rules or principles, but it must be integrated into business processes and everyday design choices.

3.4.2. Audits, certifications and metrics

Even if *by design* and impact assessment are preventive tools, audits, certifications and metrics are ex-post control mechanisms in order to make sure that systems of artificial intelligence actually comply with ethical principles and prevailing regulations. These tools, taken from accounting and industrial quality, are more and more implemented in AI as well, as a reaction to the need for establishing transparency, responsibility and comparability among firms.

Algorithmic audits include systematic verification processes, conducted on an external or in-house expert level, designed to ascertain whether or not an AI system operates in adherence to expectations of fairness, safety and nondiscrimination. An audit can range along the entire algorithm's lifespan – from data gathering to application – or be more specific in nature, covering matters such as dataset quality, control of bias or technological robustness. Already, a small number of firms have experimented with independent external audits, as a point in case in financial markets, in which transparency on credit-scoring algorithms is essential in maintaining trust among customers[190].

But the concept of an audit is thwarted in a variety of ways. First, there is the unintelligibility of intricate models, as in deep-learning models, which prevents knowledge of how the algorithm functions. The second is an inability to share standards, so all audits are created in reference to different criteria, which decreases comparability by company.

Thirdly, there is the issue of independence: an audit by accountants chosen by the company itself runs the risk of having no credibility and raising concerns about *ethics washing*[191]. Parallel to audits, there are fresh plans on the drawing board to certify AI systems, as there are in sectors such as food safety or industrial quality.

[190] Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P., 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

[191] Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), pp.501–507.

Such standards as ISO/IEC 42001:2023, targeting artificial intelligence management systems, establish requirements for the responsible implementation of AI in organizations[192]. Some countries are already pilot-testing compliance labeling in AI deployments assessed as "high risk": the idea is to establish a trust scheme, as in the case of quality certifications, which shall make all stakeholders--consumers, investors and authorities--feel confident.

The question of metrics is no less important. Defining objective indices for evaluating ideas as intangible as fairness, transparency or accountability is a vastly complicated undertaking, but it is essential in order to turn broad principles into measurable parameters. A range of metrics has been put forward as a means of evaluating the fairness of algorithms, say by comparing error rates across various groups (statistical fairness), assessing equality of opportunity or checking for the lack of undue correlations between sensitive features and choice outputs[193]. But no metric applies universally, the selection rests on the application scenario and on the company's intended value priorities.

For companies, accepting audits, certifications and measurements involves tackling a double challenge. First, on complying with regulatory requirements and controlling legal risks; second, in gaining stakeholders' trust by demonstrating a concrete commitment to responsible governance. In sectors as delicate as finance and healthcare, being in a position to demonstrate through certifications and audits that an algorithm is trustworthy can be both a competitiveness factor and a regulatory requirement.

But there are limits and possibilities of harm. The audits are ritualistic proceedings, certifications are effective deterrents to entry by small business, and application of measurement, if done mechanically, can reduce rich ideas into nothing but numbers, which are senseless. All this must be prevented by a corresponding sincere managerial commitment to use all these instruments, and not as bureaucratic pressures, for raising the quality and consistency in innovation.

3.4.3. Internal policies, ethics committees and supply chain

Concurrent with technical instruments by nature, impact evaluations and audits, organizational procedures and in-house policies fashioned by companies as a part of regulating the use of AI themselves are important in ensuring the ethicality of AI. In a world where exogenous regulations are still in development, many multinational companies have

[192] ISO/IEC, 2023. *ISO/IEC 42001: Artificial Intelligence Management System Standard*. Geneva: International Organization for Standardization.
[193] Barocas, S., Hardt, M., Narayanan, A., 2019. *Fairness and Machine Learning*. Cambridge, MA: fairmlbook.org.

instead chosen to adopt in-house rules, codes of conduct and ethics committees as a part of inscribing principles of responsibility within company governance itself.

Internal policies are the first form of governance. These are policies and procedures that lay down how AI is to be developed, employed and regulated in the company. Some corporations have put in place actual codes of AI ethics, which lay down principles like transparency, non-discrimination, fairness, protection of privacy and sustainability. Microsoft, as an example, has put in place functional guidelines in light of its *AI Principles*, which are fairness, reliability, inclusion and accountability[194]. IBM, on its part, has produced its *Principles for Trust and Transparency*, under which it undertakes to ensure that there is the explainability of algorithms and that the use of AI, if non-compliant, is restrained[195].

Along with paper, governance also requires certain organizational settings. The AI ethics committee format, consisting of both external and in-house experts, has in recent years propagated widely. The most sensitive work is scored by the committees, and recomendations are provided by them alongside ensuring coherence among principles and procedures. Google once instituted the *Advanced Technology External Advisory Council* (ATEAC), but its experience was troubled by outcry and was short-lived. Microsoft instituted the *Office of Responsible AI* and the AETHER (AI, Ethics and Effects in Engineering and Research) committee, and both evaluate high-risk work and support development teams. These tests demonstrate that engineers cannot be solely left in charge of ethics, but it requires cross-functioning processes by engaging management, lawyers, philosophers, sociologists and external groups[196].

But there are certain fundamental issues in the introduction of ethics committees. The first potential threat is a lack of independence: if the committees are internal and report directly and immediately to a superior level of management, they are likely to be viewed as instruments of legitimization and not effective control mechanisms. That is why there are already moves by certain companies to experiment with hybrid models, with independent experts or collaboration with universities and NGOs. The second issue is the speed of decision-making: in highly competitive industries, committees are likely either to be skipped or formally convened if they are viewed as a brake on the pace of innovation[197].

---

[194] Microsoft, 2022. *Responsible AI Principles*. Redmond: Microsoft.
[195] IBM, 2018. *Principles for Trust and Transparency*. Armonk: IBM.
[196] Whittlestone, J., Nyrup, R., Alexandrova, A., Cave, S., 2019. The role and limits of principles in AI ethics. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*.
[197] Bietti, E., 2020. From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*.

One often underexamined and increasingly important factor concerns supply chain and supplier relationships. Fewer and fewer companies develop AI systems in-house, but instead purchase or integrate them from external suppliers. This raises the question of joint liability: if a partner-supplied algorithm is discriminatory or flawed, the supplier is responsible in the eyes of customers and enforcers. To mitigate this threat, some organizations are bringing contractual provisions and ethics thresholds into their purchase procedures, requiring suppliers themselves to reach minimum thresholds in fairness, transparency and security[198]. In the European Union, the AI Act already sets specific requirements all along the value chain, subjecting manufacturers and end-users themselves to strict documentation and monitoring requirements.

Large companies are now applying these principles in practical ways. Salesforce, as an example, has established an *Ethical Use Advisory Council* that oversees not merely internally developed goods, but partnerships and acquisitions as well. Unilever released ethical requirements on human resources AI use in a set of principles on AI use, specifying fairness and transparency requirements in personnel selection and appraisal processes. The public sector example is Amsterdam city, which established a public register of administration-employed algorithms, themselves imposing transparency on the private suppliers involved in city activities[199].

These advances verify that in-house policies, ethics committees and ethical supply management are no adjuncts, but fundamental tools in turning principles into everyday practice. Their effectiveness, however, depends on three conditions: the honest involvement by the executive management, the independence and interdisciplinarity of control mechanisms, and the extension of ethical standards along the value chain.

The analysis developed in the chapter has clarified how AI is the most complex and delicate frontier in contemporary business ethics. Like no other technology, AI is both a technical tool, on the one hand, and a socio-technical system made up of data, algorithms, infrastructure and managerial choices, producing implications vastly transcending organizational frontiers, on the other.

The path opted for has proved, as a first aspect, that challenges arise for business ethics' established business patterns from AI (3.1). Its characteristics – scalability, transversality and

[198] European Commission, 2021. *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*. Brussels.
[199] Algorithm Register of Amsterdam, 2021. *City of Amsterdam Algorithm Register*.

opacity – make it a tool capable of reinforcing bias, generating unexpected dilemmas and raising responsibility issues difficult to address under classical instruments.

Secondly, it seemed that regulatory systems (3.2), as a condition, are incompletely structured. The European approach, via the GDPR and the AI Act, sets the protection of basic rights as a main priority, and codes published by UNESCO and the OECD offer universally shared principles. However, the discrepancy among regulatory schemes – European, US, Chinese – and the pace of technological progress indicate that regulation on its own cannot be enough in order to control the changes in progress.

The third area of reflection was on critical points at risk for firms (3.3): bias and discrimination, privacy and security, work and democracy impact. These are real and existing challenges and cannot be discounted as potential future issues. All firms that adopt AI systems find themselves solving puzzles directly affecting customer trust, reputation, social sustainability and, in the end, its own legitimacy.

Finally, the chapter illustrated practical tools for managers (3.4), how ethics could be translated into organizational practices. Such practices as AI Ethics by Design, impact reviews, audits and certifications, and in-house policies, ethics committees and responsible procurement practices, are all practical moves towards alignment of innovation and responsibility. Their efficacy, however, depends on the willingness on the part of senior management first and foremost, in treating them as something other than bureaucratic formality, but as a tool in building trust and long-term competitiveness.

Finally, this chapter has shown that AI cannot be considered purely technically or economically, since its success depends on the capability to locate it within a strong and integrated ethical and regulatory context that takes account of social and political effects. Ethics, in this regard, is no constraint but a condition of sustainability and legitimacy for companies that hope to operate on a world-wide basis. This is precisely the subject of Chapter 4, which is devoted to a case study. Through the observation of a big multinational firm, it will be possible to understand how the principles and instruments illustrated in this chapter are implemented in practical AI governance practices. The case will thus be an occasion to verify how much firms are able to implement the ethics of artificial intelligence, transforming it from a statement of principles into an integral part of everyday managerial practice.

PART FOUR: Microsoft case study

4.1. Introduction to the case study

The study performed in the previous chapters has evidenced how artificial intelligence currently represents one of the most complex challenges that business ethics and strategic management are called on to address. The intersection between innovation potential and ethical risks has compelled recourse to regulatory frameworks, organizational tools, and corporate policies designed to transform principles into operative procedures. Yet, to fully grasp how ethical principles are effectively put into practice, theoretical reflection alone is insufficient: it becomes necessary to examine how leading global corporations translate abstract commitments into managerial routines and governance mechanisms.

For this purpose, the present chapter develops a case study dedicated to Microsoft, one of the most influential and advanced technology corporations in the application of ethical policies towards AI. The choice of Microsoft is not accidental. Founded in 1975 and today a global leader in software and cloud services, Microsoft was among the first Big Tech companies to adopt an official Responsible AI policy, articulating principles of fairness, transparency, accountability, and inclusion[200]. At the same time, Microsoft provides a paradigmatic example of the tensions between competitiveness and ethics: despite its advanced tools and policies, the company has faced criticism and controversy regarding the use of its AI systems, showing how difficult it is to reconcile commitments with daily practice.

The case study will be conducted through a qualitative approach, relying primarily on secondary sources such as company reports, policy documents, academic literature, and media investigations. Particular attention will be given to Microsoft's *AI Principles* and *Responsible AI Standard*, the functioning of internal structures like the Office of Responsible AI and the AETHER Committee, and the company's external initiatives in transparency and stakeholder engagement. This analysis will be complemented by a critical review of controversies such as the debates around facial recognition technologies and generative AI which illustrate the unresolved dilemmas faced by the corporation.

Microsoft's strategic transformation in recent years provides an ideal setting for this analysis. The firm has heavily invested in cloud computing and AI-based services, incorporating algorithms into products such as Office 365, Azure, and LinkedIn. Its partnership with OpenAI and the integration of generative models like GPT into its services have further

[200] Microsoft, 2018. *Microsoft AI Principles*. Redmond: Microsoft.

cemented Microsoft's role as a global frontrunner in the diffusion of AI. This leadership position makes the company a privileged observatory for understanding how a corporation responds to the challenge of applying AI ethics at scale.

Ultimately, the Microsoft case offers the opportunity to address the central research question of this thesis: *how can abstract ethical principles be transformed into practical instruments of governance?* By examining both the company's organizational structures and its external initiatives, as well as the controversies that highlight persistent tensions, the case illustrates that AI ethics is not a static set of principles but a dynamic process, emerging through compromises, trade-offs, and organizational decisions that balance values and economic imperatives.

4.2. Microsoft's ethical framework on AI

The software giant was the first large tech player to lay out a specific ethical scheme on artificial intelligence, recognizing that the success of its innovations could never be severed from user trust and support for fundamental rights. Already in 2016, in a New York speech, then President Brad Smith said that "AI must be guided by human values", expressing a policy which was later set into form in the Microsoft AI Principles published in 2018[201].

The company's ethical system is built on six basic principles:

1. Fairness: The AI systems shall treat all humans equitably and never discriminate.
2. Reliability and safety: The AI shall be robust, reliable, and safe in all operating conditions.
3. Security and privacy: data need to be protected and processed in a clear and confidential fashion.
4. Inclusiveness: Technologies must be accessible and inclusive, and all barriers removed from individuals with disabilities or disadvantage groups.
5. Transparency: systems need to be interpretable and comprehensible, such that users are aware if they are engaging with an algorithm.
6. Accountability: Developers and deployers of AI, as well as organizations and individuals, must be responsible for the decisions made by machines[202].

Such principles were never abstract statements but were translated into organizational structures and development instruments. Microsoft created in 2019 the Office of Responsible

[201] Smith, B., 2016. The need for a Hippocratic Oath for AI. Speech, New York.
[202] Microsoft, 2018. *Microsoft AI Principles*. Redmond: Microsoft.

AI (ORA), having the mission of defining internal policies, tech standards and development principles. The ORA works in close connection with the AETHER (AI, Ethics, and Effects in Engineering and Research) committee, a cross-function working team comprised by engineers, lawyers, ethicists, and civil society leaders, and it examines the higher risk projects and provides binding recommendations[203].

One of the distinctive features of the system is the presence of binding documents for workers. The most important among them is the 2022-launched Responsible AI Standard, which outlines necessary procedures in designing and deploying AI systems. This over-100-page-long paper outlines in detail data handling, human monitoring and risk evaluation requirements. All the teams designing a high-impact AI system are mandated to make technical documentation demonstrating adherence to the principles and submit it in front of the inner committee[204] to get approved.

Microsoft has also put in place procedures for ethical review in the most sensitive developments. Applications of AI in healthcare or face recognition, for example, have to be reviewed in detail, as well as tested on bias, social effect, and scenario simulation on risks. This approach relies on the “principle of responsible precaution": it is better to suspend or modify a development in progress than to encounter negative consequences in circulation[205].

At the same time, the company has tried to stimulate an in-house sense of responsibility. With compulsory education courses, awareness programmes and online tools, employees are encouraged to report potential problems in ethics in job tasks. This focus on the culture dimension arises from the reality that ethics cannot be conveyed in a set of principles, but rather, it must infuse the entire organizational ecosystem[206].

One area is Microsoft's external transparency. The organization publishes biannual reports on responsible AI efforts and collaborates with educational institutions and NGOs in the development of international standards. The organization has also been a driving force behind initiatives such as AI for Good, a programme through which it aligns itself with AI-driven initiatives in environmental sustainability, accessibility and world health. These initiatives are intended to support Microsoft's claim to be a responsible AI leader, both by having processes in check internally and by participating in public discourse actively[207].

---

[203] Microsoft, 2019. *Office of Responsible AI and AETHER Committee*. Company Report.
[204] Microsoft, 2022. *Responsible AI Standard v2*. Redmond: Microsoft.
[205] Whittaker, M., Crawford, K., Dobbe, R. et al., 2018. *AI Now Report 2018*. AI Now Institute, New York.
[206] Raji, I.D., Bender, E.M., Paullada, A., Denton, E., Hanna, A., 2021. AI and the Everything in the Whole Wide World Benchmark. *FAccT Conference Proceedings*.
[207] Microsoft, 2021. *AI for Good Annual Report*. Redmond: Microsoft.

However, there are strains and challenges. Some have been critical of the fact that, despite its professed principles, Microsoft continues to be a purveyor of contentious technologies, including facial recognition systems. In 2020, in reaction to outrage over the use by law enforcement officers, the company issued a moratorium on sales to American police departments, waiting on federal regulation. This episode illustrates how an evolved framework is nevertheless compelled to navigate pressures from the marketplace and reputational challenges[208].

Overall, Microsoft's ethical scheme is perhaps the most explicit and ambitious venture thus far in bringing ethical ideals into organizational practices. Its value statements, organizational arrangements, binding instruments and plans of transparency are drawn together into a package that is increasingly being copied by other multinational corporations. Its success, however, depends less on the quality of the policy itself, but on effective implementation in everyday processes and along the value-chain as well.

4.3. Concrete governance tools

Microsoft's ethical AI framework does not restrict itself to broad principles and expression of intentions: the firm has translated these principles into practical instruments of governance aimed at guiding technological development and application in operative fashion. This conclusion arises from recognition that, in the absence of operative procedures, principles are a dead letter.

Among the primary instruments is the Responsible AI Standard, issued in 2022 and revised periodically. The binding document sets up required procedures in all AI system development teams. The requirements are determining data management plans, determining risks of bias, determining safety and robustness of models, fully documenting the lifecycle of the algorithm, and needing human review on high-impact systems[209]. In order to be in line with the rules, every AI project has to finish a *Responsible AI Impact Assessment* (RAIA), a kind of technical-ethical audit that substantiates how a system adheres to principles set by the company.

Concurrently with these procedures, Microsoft itself has created an in-house system of ethical review. Projects that are classified as “high risk” – i.e., involving face recognition, healthcare use, or sensitive data processing – are considered by the *Office of Responsible AI* and by the AETHER committee. This is a system that is never purely bureaucratic: in a few

[208] The Guardian, 2020. Microsoft bans police use of its facial recognition technology. *The Guardian*, 11 June.
[209] Microsoft, 2022. *Responsible AI Standard v2*. Redmond: Microsoft.

cases, Microsoft has opted to suspend or modify projects that did not meet approved standards. In 2020, as part of a response to potential abuses in the field of civil rights, the company publicly announced limits on facial recognition technology sales to law enforcement departments[210].

Another working instrument is the AI Ethics Champions, staff trained to be ethical points of reference in the various development groups. These are accountable in making sure potential risks are considered, reporting serious issues and maintaining a culture of day-to-day responsibility. This is a new approach that attempts to diffuse ethical responsibility throughout the entire organization, making sure it does not remain in a headquarters' central office[211].

Microsoft has also invested in tech instruments in favor of governance. These are, on the one hand, the FairLearn toolkit, in open source, through which it is possible to measure and counteract bias in machine learning models, and on the other, InterpretML, offering methods by which predictive models can be made more interpretable. Together with assuring inner consistency, the latter instruments also allow the scientific community and the other actors in the industry to be involved, in hopes of transparency and collaboration[212].

The relevant dimension concerns responsible management of the algorithmic supply chain. Microsoft, recognizing that all AI applications are either developed by third-party partners or through integration of third-party modules, has established responsible procurement provisions. That is, the suppliers of AI software and services are obligated to meet minimum security, transparency, and no-discrimination levels. Such contractual provisions evidence recognition that AI ethics cannot be left within organizational boundaries but must extend along the entire value supply chain[213].

Ultimately, Microsoft has made efforts to promote transparency among end-users through instruments such as datasheets on datasets and model cards, papers that describe the characteristics of the datasets and AI models utilized, outlining their constraints, possible bias and conditions of use. This initiative, following the scientific practice model, aims at

[210] The Guardian, 2020. Microsoft bans police use of its facial recognition technology. *The Guardian*, 11 June.
[211] Microsoft, 2021. *AI Ethics Champions: Building a culture of responsible AI*. Company Report.
[212] Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K., 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft Research*.
[213] European Commission, 2021. *Proposal for a Regulation on Artificial Intelligence (Artificial Intelligence Act)*. Brussels.

disseminating more knowledge among stakeholders and permitting external independent control[214].

These tools confirm as well that Microsoft does not perceive ethical AI governance as something abstract, but as a set of practices codified and incorporated into business processes. Their effectiveness is dependent on something greater than the existence of policies, namely on the ability to implement it into practice in everyday work, among engineers, managers, and external interlocutors.

4.4. Successes and best practices

The most striking aspect in the Microsoft case is how the company implemented principles of ethics in practical forms that have provided a constructive contribution both internally and externally. These learned lessons, despite their share of limitations, are a demonstration on how technological innovation and social responsibility can be combined and are good case models on how corporations elsewhere can follow suit.

First among first successes is commitment to fairness and inclusion. Microsoft has developed tools aimed at making AI accessible to otherwise excluded groups in the innovation processes. The *AI for Accessibility* initiative, which was established in 2018, aids initiatives applying artificial intelligence to make the lives of people with disabilities improved, e.g., through voice recognition systems in support of the blind or through real-time translation systems in support of the hearing impaired[215]. Such apps illustrate how the inclusion principles can be turned into innovations reinforcing the social legitimacy of the business as well as shared value creation.

Another field of triumph is environmental sustainability. Through the *AI for Earth* programme, which was launched in 2017, Microsoft sponsors efforts that use AI in addressing challenges as diverse as climate change, preserving biodiversity and use of natural resources. Among the funded efforts are systems monitoring satellites in a bid to prevent deforestation, predictive analytics by which energy efficiency is improved, and solutions in sustainable farming[216]. This initiative, in addition to helping towards the Sustainable Development Goals by the United Nations, also complements Microsoft's brand as a responsible global player.

---

[214] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T., 2019. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT)*.

[215] Microsoft, 2018. *AI for Accessibility Program*. Redmond: Microsoft.

[216] Microsoft, 2017. *AI for Earth Initiative*. Redmond: Microsoft.

Internally, there has been a big success in creating a culture of responsibility within the firm. With systematic courses and through the initiative of the *AI Ethics Champions*, Microsoft has made efforts at spreading ethical awareness among employees. More than 50,000 employees were trained by 2021 in modules on the Responsible AI Standard, a broad commitment which goes far outside the technical departments[217]. That demonstrates that AI ethics cannot be left in a narrow circle of experts, but it has to permeate the entire organization.

Microsoft has also been hailed for its transparency initiatives. The launch of tools such as *dataset datasheets and model cards* was a step in the right direction in algorithm reporting, and it made other companies follow the same policy. Furthermore, the company was also among the first ones in publishing publicly accessible toolkits in the form of *Fairlearn* and *InterpretML*, through which the interpretability and fairness of AI models can be determined. These tools, along with making Microsoft's integrity more robust, lead towards creating international standards[218].

Another example of good practice is judicious facial recognition management. Facing concerns over the potentially intrusive use of the technology, Microsoft made a voluntary decision in 2020 to suspend, temporarily, the sale of its facial recognition systems to American police forces, pending more certain federal regulation. This was a strong signal: a company willing to miss out on short-term advantage in a bid to safeguard its ethical integrity and trust among stakeholders[219].

Less well-known but no less important is the aspect concerning collaboration with institutions and civil society. Microsoft is involved in working groups with the OECD, the Partnership on AI and other institutions in an effort to formulate universal standards of responsibility. It has also been a sponsor of educational and academic programs in a move aimed at facilitating research on the ethics of AI, in aid of formulating new abilities and propagating a culture of technological responsibility[220].

These initiatives have provided Microsoft with a position as a founder in the craft of ethical AI, standing out from competitors by a combination of innovation and responsibility. Criticism and scandal remain, but the portfolio of practice and project work thus developed demonstrates that ethics can be a competitiveness driver, not a brake.

---

[217] Microsoft, 2021. *Responsible AI Annual Report*. Redmond: Microsoft.
[218] Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K., 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft Research*.
[219] The Guardian, 2020. Microsoft bans police use of its facial recognition technology. *The Guardian*, 11 June.
[220] Jobin, A., Ienca, M., Vayena, E., 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), pp.389–399.

4.5. Critical issues and controversies

Even as Microsoft has made itself a front-runner company in formulating principles and tools for responsible AI, its track record also shows the limitation and contradiction in linking ethics with universal market rationale and competitiveness. The debate on critical issues in no way intends to criticize efforts made but indicate how difficult it is to apply principles in a systematic and rational manner.

First, there is a concern regarding the application of facial recognition. Although a 2020 moratorium on resale by Microsoft and other corporations of the systems to US law enforcers was proclaimed, the company develops and sells identical or similar tools to other institutional and private customers. Some have commented that the moratorium did not extend in every marketplace and was adopted first and foremost in response to political and reputational pressures, rather than as a long-term business policy[221]. This episode illuminates the dilemma between business interests and ethical ideals: on the one hand, the company claims it wants to uphold civil rights, but on the other, it maintains facial recognition as a viable live business sector, if with greater caution.

The other area of concern relates to its collaboration with OpenAI. Its nine-figure investment in the start-up, culminating in infusing models like GPT-4 into its software and cloud services, has been questioned in the context of whether the company was in a position to hold in check the dangers of generative AI. Some scholars have mentioned the danger of misinformation, bias and copyright infringement associated with the tools. Despite Microsoft having security mechanisms and filters in place, it has been criticized for accelerating the uptake of a formative technology, on the basis of a need to give a head start in market position over Google and other competitors[222]. This case shows a trade-off typical in Big Tech: running ahead in innovation in a quest to be ahead in the market, including despite the dangers of opening itself up to ethical risks that remain partly uncontrollable.

The third basic issue is consistency in ethical principle application internally. Some former staff have mentioned the difficulty in balancing pressure for innovation and compliance with the rules of the *Responsible AI Standard*. Particularly, there has been evidence that development teams perceive ethical review procedures as bureaucratic obstacles that cause *time-to-market* lag. The allegation has stoked accusations of "ethics washing", namely application of ethical principles as communication tools more than as effective constraints

[221] The Guardian, 2020. Microsoft bans police use of its facial recognition technology. *The Guardian*, 11 June.
[222] Financial Times, 2023. Microsoft bets on generative AI with OpenAI partnership. *FT*, January 23.

on business practice[223]. The question raised by the issue is fundamental: how far are companies willing to sacrifice speed and profits in the name of ethics?

The second point of contention is within relations with government institutions and the public. Microsoft has provided cloud infrastructure and AI solutions to government ministries in various nations, including ones with narrow democratic standards. Human rights organizations have been critical of the use of technologies in the promotion of bulk surveillance efforts, and whether the company has significant autonomy in the selection of its clients[224]. This highlights how, as a multinational, the ethical issue is both intrinsic and geopolitical: operating in various markets means engaging with regimes whose views on digital rights and privacy are vastly different.

Finally, public opinion cannot be ignored. Despite its transparency efforts, Microsoft belongs to the Big Tech club and is usually accused of enjoying too much economic and informative power. Trust in technology companies is fragile and can be derailed by a mere incident, be it malfunction, data breach or scandal through misuse of data. In 2016, as an example, experimental chatbot Tay was removed a few hours after it was launched as it was ever hijacked by the public and used as a tool in spewing racist and offensive commentary. This was an experiment, but the incident highlighted the dilemma in regulating unexpected algorithmics' behaviour and left a bad mark on the brand of the company[225].

These controversial matters demonstrate that ethical governance in AI is never definitive, but a dynamic and on-going process. The companies can equip themselves with the latest frameworks and the most sophisticated tools, but they are still in danger from both external and internal forces that relentlessly test their consistency. Both successes and scandals coexist in Microsoft, and a contradictory portrait remains: on the one hand, the company is a frontrunner in AI ethics, and on the other, it still has to face the fact of certain kinds of compromises within a globalized and competitive market.

4.6. Summary and managerial implications

The Microsoft case allows a privileged vantage point on how to answer the research question driving this thesis, namely how ethical principles can be translated concretely into practice by multinational companies in the development and release into production of artificial intelligence. The analysis conducted shows there exists no universal recipe, but rather a

[223] Bietti, E., 2020. From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy.
[224] Human Rights Watch, 2021. *World Report 2021*. New York: HRW.
[225] Vincent, J., 2016. Twitter taught Microsoft's AI chatbot to be a racist jerk in less than a day. *The Verge*, 24 March.

variety of practices and circumstances, if deployed, which make it possible to turn ethics as a philosophical statement into a working governance tool.

One of the earliest basic points that emerges is the need to institutionalize ethics. Microsoft has taught us that principles such as fairness, transparency and accountability derive no real significance if they are not translated into enforceable policy and organizational mechanisms. The establishing of the Office of Responsible AI and the cross-disciplinary AETHER committee was a determining step in this regard: no more empty value expression, but processes of control and review embedded in the organizational design. This tells us that ethics cannot remain a personal discretion of individual development teams in a multinational, but needs to be an integral part of governance, with well-charted roles, responsibilities and procedures.

A second point concerns how ethics and trust are a strategic resource. The Microsoft case shows that responsible development of AI is as much a response to reputational risks or regulatory pressures, as it is a competitiveness driver. Programs such as AI for Accessibility or AI for Earth have created tangible social value, but also solidified Microsoft's brand as a global pioneer of responsible innovation. The implication is that, in a world in which Big Tech is coming under greater and greater distrust, the ability to build and maintain trust is a real source of strategic advantage. In contrast, scandals associated with facial recognition or generative AI illustrate that trust is fragile: a dramatic incident is all it takes to reverse a long train of reputational investment. The manager implication is simple: ethics is no external constraint, but an inner prerequisite of long-term sustainability.

The third lesson is on the systemic dimension of responsibility. AI is a system that goes beyond organizational, sectoral and geographical boundaries: as a result, corporate responsibility cannot be kept within internally created applications, but goes significantly beyond along the entire value chain. Microsoft has added ethical provisions in procurement contracts, published model cards and datasheets publicly in order to stimulate model transparency, and initiated joint efforts with universities and NGOs in order to build shared standards. This approach accepts the fact that effective ethical governance cannot be developed in a vacuum, but requires suppliers, technology partners as well as civil society on board. This, in other multinational companies' cases, implies a paradigm shift: ethics as a domain no longer an internally oriented domain, but a networked collective sense of responsibility.

The fourth point is the recognition that AI ethics is a process in motion, and not an end point. The controversies reviewed – from the moratorium on facial recognition to collaborating

with OpenAI – demonstrate how even a cutting-edge giant as Microsoft is forced periodically to renegotiate the line drawn between professed values and marketplace imperatives. Multinational corporations operate in a world characterized by patchwork rules, culturally different environments and cutthroat competition: in a situation like that, ethics cannot be a fixed set of ideals, but a dynamic process oriented towards adapting and necessitating flexibility, transparency and an inclination towards dialogue with external and internal actors.

These are assumptions for a bigger conclusion: principles of ethics can be implemented by large business corporations only if they are transformed into organizational designs, assets for strategy, webs of accountability and procedures of dynamic review. The case of Microsoft shows it is possible, but in a translation of ethics into practice, there is a need for political motivation within, investment in culture and education, sensitivity towards civil society and ability to manage the inevitable trade-offs.

To business leaders, it is a simple message: AI ethics is no reputational nicety nor a cost of compliance, but a foundation of business strategy in the era of the digital economy. And only by infusing responsibility, transparency and inclusion into their organizational decisions, multinationals can be sure that artificial intelligence is a driver of innovation and growth, and a force shaping equitable and sustainable progress.

CONCLUSIONS

The research conducted in this thesis identified how the confluence of company ethics, innovation, and artificial intelligence defines one among the greatest challenges in contemporary management. No longer marginal or decorative, ethics in the contemporary world is rather an strategic dimension that defines competitiveness, social legitimacy, and long-term survivability of transnational firms. Now that the research question that informs this work - *How can ethical principles be concretely implemented by global corporations in the development and use of Artificial Intelligence?* - be responded through the comparison among the most significant insights that have been derived from the debate at the theoretical level as well as from the case study.

First, ethical values may be introduced through institutionalization within the framework of company governance. The history of the study of the analysis of business ethics and the debate on responsible innovation proved that universal values such as fairness, transparency, and sustainability acquire meaning only if they translate into rules, processes, and special structures. The example of the Microsoft case proves the point: the creation of the Office of Responsible AI, the AETHER Committee, and the release of the document on the Responsible AI Standard illustrate the idea that ethics may be codified in organizational architecture and become the series of binding procedures rather than aspirational statements. For multinationals, this implies that governance devices such as committees, standards, and risk analysis are inescapable in order to give concrete shape to ethical will.

Second, ethics is strategic when thought of as a strategic resource, as opposed to a constraint. The chapters on innovation demonstrated how companies could produce not only economic but shared value and be legitimate and trustworthy players in the stakeholders' perception. The activities of Microsoft, as represented in its AI for Accessibility and in its AI for Earth initiatives, provide illustrations of how responsible innovation is capable of improving reputation, consolidating the trust of stakeholders, and unveiling new business opportunities. The message to managers is that competitiveness and ethics are no opposites but complements: including ethical thinking in the strategies of innovation can become an enabler of differentiation and resilience, especially in the context of crisis or reputational exposure.

Third, ethical implementation requires a relational and systemic approach. Artificial intelligence as general-purpose technology is implemented in large global networks in which responsibility is diffused across numerous players. The thesis showed how issues such as data governance, algorithmic bias, and disinformation cannot be addressed in lone-handed

fashion by companies. Microsoft's efforts at extending ethical norms along the chain of its supply chain, through clauses in responsible procurements and use of tools such as datasheets and model cards, illustrate the extent of the engagement with suppliers, partners, regulators, and civil society that is required. For multinationals, this means that the practical application of ethics in the creation of AI involves cooperation across sectors and geography, as well as engagement with public debate around standards and regulation.

Fourth, the ethics of AI is provisional, not absolute. The scandals reviewed—in the scope of the moratorium versus the application of face recognition, the rapid commercial breakthrough of the generative AI through OpenAI partnerships—are typical of the inadequacy of company declarations and the ongoing negotiation of values and commercial emphases. What this establishes is that the ethics of AI cannot be finally determined: they require frequent adaptation, repeated re-assessment, and subject to external review. For managers, that means they should be embracing the new and experimenting with adaptable governance frameworks that evolve with the shifts in technology and society.

By integrating the above factors, the research question can be answered as follows: Global firms can concretely enact ethical values in the development and application of AI through institutionalizing responsibility within the governance system, treating ethics as a strategic tool, adopting the system thinking that pervades the entire value chain, and treating ethics as an iterative dynamic unfolding that remains open to ongoing revision.

The finding has both managerial and theoretical importance. Theoretically, it contributes to the business ethics and innovations scholarship in illustrating how competitiveness and ethicality may be aligned in tandem, particularly in the digital age. Managerially, the study provides practical guidance in concrete terms: create special structures and standards, invest in stakeholder trust, be responsible beyond organizational boundaries, and institutionalize frequent ethical review. Lastly, the thesis underpins the larger argument: in the era of the artificial intelligence, business ethics is no add-on or apparatchik function but the prerequisite of survival and legitimacy. The businesses that will succeed in embedding ethical principle within strategies and innovations will be the ones that will be able to build sustainable trust, achieve sustainable competitiveness, and create value within society. The businesses that embrace ethics as an add-on or optional function or resort to using an optional function at will will run the risk of not only reputational losses but strategic marginalization in an age becoming increasingly sensitive to questions of responsibility, sustainability, and justice.

REFERENCES

Acemoglu, D., Restrepo, P., 2018. *Artificial Intelligence, Automation and Work*. NBER Working Paper No. 24196.

Ajunwa, I., 2020. *The paradox of automation as anti-bias intervention*. Cardozo Law Review, 41(4), pp.1671–1714.

Andrejevic, M., 2020. *Automated facial recognition in public: A comparative analysis*. Surveillance & Society, 18(2), pp.212–226.

Angwin, J., Larson, J., Mattu, S., Kirchner, L., 2016. *Machine bias*. ProPublica, May 23.

Aristotele, 2005. *Etica Nicomachea*. Milano: Bompiani.

Arntz, M., Gregory, T., Zierahn, U., 2016. *The risk of automation for jobs in OECD countries*. OECD Social, Employment and Migration Working Papers, No. 189.

Autor, D.H., 2015. *Why Are There Still So Many Jobs?* Journal of Economic Perspectives, 29(3), pp.3–30.

Barocas, S., Hardt, M., Narayanan, A., 2019. *Fairness and Machine Learning*. Cambridge, MA: fairmlbook.org.

Barocas, S., Selbst, A.D., 2016. *Big data's disparate impact*. California Law Review, 104(3), pp.671–732.

Benaich, I., Hogarth, N., 2023. *State of AI Report*.

Bentham, J., 1789. *An introduction to the principles of morals and legislation*. London: T. Payne.

Beschorner, T., 2013. *Creating Shared Value: The one-trick pony approach*. Business Ethics Journal Review, 1(17), pp.106–112.

Bietti, E., 2020. *From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy*. AAAI/ACM Conference on AI, Ethics, and Society.

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K., 2020. *Fairlearn: A toolkit for assessing and improving fairness in AI*. Microsoft Research.

Bostrom, N., Yudkowsky, E., 2014. *The ethics of artificial intelligence*. In: Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press.

Botsman, R., Rogers, R., 2010. *What's Mine Is Yours: The Rise of Collaborative Consumption*. New York: Harper Business.

Bowen, H., 1953. *Social responsibilities of the businessman*. New York: Harper & Brothers.

Bowie, N.E., Werhane, P.H., 2005. *Management ethics*. Malden, MA: Blackwell.

Brynjolfsson, E., McAfee, A., 2014. *The Second Machine Age*. New York: Norton.

Brynjolfsson, E., McAfee, A., 2017. *Machine, Platform, Crowd*. New York: Norton.

Buolamwini, J., Gebru, T., 2018. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of Machine Learning Research, 81, pp.1–15.

Caltagirone, C., 2017. *Amartya K. Sen. Tra economia ed etica*. Roma: Studium.

Carroll, A.B., 1987. *In search of the moral manager*. Business Horizons, 30(2), 7–15.

Carroll, A.B., 1991. *The pyramid of corporate social responsibility*. Business Horizons, 34(4), 39–48.

Carroll, A.B., 1999. *Corporate social responsibility: Evolution of a definitional construct*. Business & Society, 38(3), pp.268–295.

Cath, C., 2018. *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*. Philosophical Transactions of the Royal Society A, 376(2133).

Cavoukian, A., 2011. *Privacy by Design: The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario.

Chouinard, Y., Stanley, V., 2012. *The Responsible Company*. Patagonia Books.

Christensen, C.M., 1997. *The Innovator's Dilemma*. Boston: Harvard Business School Press.

Crane, A., Palazzo, G., Spence, L.J., Matten, D., 2014. *Contesting the Value of "Creating Shared Value"*. California Management Review, 56(2), pp.130–153.

Crevier, D., 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.

Deigh, J., 2012. *Etica. Un'introduzione*. Milano: Apogeo.

Drucker, P.F., 2000. *Manuale di management. Compiti, responsabilità, metodi*. Milano: Rizzoli.

Elkington, J., 1997. *Cannibals with forks: The triple bottom line of 21st century business*. Oxford: Capstone.

Freeman, R.E., 1984. *Strategic management: A stakeholder approach*. Boston: Pitman.

Friedman, M., 1970. *The social responsibility of business is to increase its profits*. The New York Times Magazine, New York.

Goodpaster, K.E., 1991. *Business ethics and stakeholder analysis*. Business Ethics Quarterly, 1(1), 53–73.

Jonas, H., 1979. *Das Prinzip Verantwortung*. Frankfurt am Main: Suhrkamp.

Jonas, H., 1990. *Il principio responsabilità*. Torino: Einaudi.

Kerstein, S.J., 2002. *Kant's search for the supreme principle of morality*. Cambridge: Cambridge University Press.

Mandeville, B., 1968. *La favola delle api*. In *Grande antologia filosofica*, vol. XIV. Milano: Marzorati.

Mandeville, B., 1988. *The fable of the bees*. Indianapolis: Liberty Fund.

Mandeville, B., 2000. *La favola delle api*. Bari: Laterza.

Mill, J.S., 1861. *Utilitarianism*. London: Parker, Son, and Bourn.

Mitchell, R.K., Agle, B.R., Wood, D.J., 1997. *Toward a theory of stakeholder identification and salience*. Academy of Management Review, 22(4), 853–886.

Moore, G.E., 1903. *Principia ethica*. London: Cambridge University Press.

Onida, P., 1960. *Economia d'azienda*. Torino: UTET.

Phillips, R., Freeman, R.E., Wicks, A.C., 2003. *What stakeholder theory is not*. Business Ethics Quarterly, 13(4), 479–502.

Porter, M.E., 1990. *The competitive advantage of nations*. New York: Free Press.

Porter, M.E., Kramer, M.R., 2011. *Creating shared value*. Harvard Business Review, 89(1/2), 62–77.

Robbins, L., 1947. *Saggio sulla natura e l'importanza della scienza economica*. Torino: UTET.

Schumpeter, J.A., 1934. *The theory of economic development*. Cambridge, MA: Harvard University Press.

Schwartz, M.S., Carroll, A.B., 2003. *Corporate social responsibility: A three-domain approach*. Business Ethics Quarterly, 13(4), 503–530.

Sciarelli, S., 1967. *Il processo decisorio nell'impresa*. Padova: CEDAM.

Sen, A., 2006. *Etica ed economia*. Bari: Laterza.

Simon, H., 2001. *Il comportamento amministrativo*. Bologna: Il Mulino.

Zanda, G., 1974. *La grande impresa. Caratteristiche strutturali e di comportamento*. Milano: Giuffrè.