Received: 14 June 2024 | Revised: 1 September 2024 | Accepted: 18 September 2024

DOI: 10.63029/jrgvn456

MANAGERIAL INSIGHT

Empowering the exercise of responsibility in the design, implementation and use of AI-based systems

Valérie Chapuis[1] | **Delphine Guégan**[2]

[1]Researcher, Orange Innovation, Châtillon, France

[2]UX Designer, Orange Innovation, Cesson-Sévigné, France

Correspondence
Corresponding author Valérie Chapuis.
Email: Valerie.Chapuis@orange.com

Abstract

Objective - This editorial aims to explore the critical role of responsibility in the design, implementation, and use of AI-based systems. It emphasizes the ethical challenges posed by the unique characteristics of Artificial Intelligence (AI) and advocates for the development of a culture that promotes individual and collective responsibility among those involved with AI.

Approach - The article outlines a conceptual approach based on three theoretical frameworks—Edgar Morin's complex thinking, Jacques Ardoino's multi-referential approach, and Yves Clot's clinic of activity. These frameworks are applied to enhance the responsible management of AI systems through reflective practices and responsible design methods.

Results - The editorial highlights that AI is not inherently responsible or ethical, and that true responsibility lies with the individuals and teams who design, implement, and use AI systems (AIS). A structured methodology involving targeted training, cultural adaptation, and reflective practice is essential for fostering responsibility and addressing the societal and ethical issues posed by AI.

Practical Implications - Organizations must foster a culture of responsibility through continuous education, reflective practice, and ethical discussions at all levels. By equipping teams with the tools and knowledge necessary to navigate AI's ethical landscape, companies can ensure that AI systems are designed and used in a manner that is socially and environmentally responsible.

Originality - This article contributes to the ongoing discourse on ethical AI by presenting a holistic and systemic perspective on responsibility. It offers practical methods for integrating responsibility into the AI design and deployment process, providing valuable insights for organizations seeking to align technological innovation with ethical and societal expectations.

KEYWORDS

Artificial intelligence; responsibility; ethics; reflective practice; responsible design.

1 | INTRODUCTION

This article is part of a corporate research program aimed at enabling, encouraging and supporting the responsible design, implementation and use of AI (Brynjolfsson and Mcafee 2017), as it continues to spread across a multitude of fields with an ever-increasing number of applications (Wei 2024). To achieve this, part of the work integrates methods that support professional gestures and responsible decision-making in the development and deployment of AI systems. The aim is to meet the challenges in terms of responsibility and ethics that arise from the unique characteristics of AIS and the teams that design, implement and use them (Chapuis and Guégan 2023).

To contribute to a declining mood in AI research (Wei 2024), this article describes the entire methodological approach, which is grounded in three theoretical (or conceptual) frameworks that emphasise a resolutely systemic perspective: Edgar Morin's concept of complex thinking (Montuori 2008), Jacques Ardoino's multi-referential approach (Ardoino and Mialaret 1995), and Yves Clot's clinic of activity (Clot 2005). The articulation of two complementary toolboxes makes up the experimental trunk, in the form of a method for supporting teams: responsible design and reflective practice. The ultimate goal is to deploy and disseminate a culture of individual and collective responsibility, equipping teams to

question and tackle the social, societal and environmental challenges posed by AI technologies (Clarke 2019).

2 | PREAMBLE ON RESPONSIBILITY AND ETHICS

Responsibility, in the broadest sense, is the dual capacity to account for one's actions and their impact and to answer the questions asked about one's actions and decisions, motivations and resulting consequences. In other words, regardless of the domain – legal, moral or parental – being responsible, or response-able, means being able to understand, account for, and report on one's actions. Moreover, the degree of responsibility attributed to an individual or an organisation is proportional to the power, ascendancy or control exerted over their entourage or ecosystem. Responsibility can therefore be seen as the just and beneficial, or at least non-harmful, application of this power from social, societal and environmental perspectives. In contrast, the misuse – whether intentional or not – of such power introduces the notion of risk. From a legal point of view, responsibility only applies to those who are of legal age and sound mind, etc., which presupposes a sufficient capacity for awareness and understanding and, therefore, necessitates education and support (Wright 2003).

Ethical principles, meanwhile, translate individual and collective responsibilities into practical terms. These common rules form "voluntary safeguards" that enable us to practically assume the responsibilities of which we are aware. Thus, ethics mobilises reflection and enables self-regulation by questioning circumstances, posing problems, identifying conflicting values, exploring possible alternatives and guiding judgment. Therefore, ethics can only serve both individual and collective good when it is rooted in responsibility, rather than irresponsibility (Jonas 1984).

Hence, we regard responsibility as the fundamental ferment that determines and motivates ethics, while ethics, in turn, acts as the process that questions responsibility and translates it into a voluntary, shared and evolving code of conduct.

3 | STRENGTHENING INDIVIDUAL AND COLLECTIVE CAPACITIES TO ADDRESS THE COMPLEX ISSUES RAISED BY AI

Terms like "Trusted AI," "Ethical AI," and "Responsible AI" have become common in efforts to define AI systems that would be respectful of the rights of people and the planet. However, this anthropomorphism reveals a core issue and its complexity: the true responsibility lies not with AI itself but with the individuals who design, train, implement, and use these systems. In essence, AI is only as "trustworthy" as the people behind it even if trust can be conceptualized according to three dimensions, namely interpersonal trust, competence trust, and systems trust (Lissillour and Sahut 2023). Future studies may look at how trust in AI can be defined according to these three dimensions in different contexts.

To promote the responsible design, implementation, and use of AI-based solutions, it is important to consider the following four levers:

- **AI itself is not inherently ethical or responsible:** The ethical nature of AI depends entirely on the teams that design, train, implement, and use it (Floridi and Cowls 2022, Jobin et al. 2019). Therefore, there is a real interest in offering the internal teams concerned appropriate, long-term support that encourages responsible practices (Lu et al. 2023).
- **Individual and collective ethics and responsibility are mutually nurturing/reingorcing:** It is therefore virtuous to create internal spaces for reflective exchange and apply practices that contribute to organizational learning (Purvis and Zhang 2024), addressing the ethical issues of AI-based solutions in order to promote and exercise a culture of individual and collective responsibility, shared by all staff members (Taylor 2024).
- **Awareness of responsibility is essential:** No one can behave responsibly without being aware of the contours of their own responsibility. Awareness provides us with a context and gives meaning to our actions. In addition to employees, customers and external partners must understand their responsibility when using AI-based solutions and be aware that these algorithms do not guarantee ethically acceptable behaviour (Mittelstadt et al. 2016). It is the duty of every company creating, providing, and distributing AI-based services to equip all users with the knowledge to use AI consciously, make informed decisions, and assess the impact of their actions (Santoni de Sio and Mecacci 2021).
- **Clear and transparent communication is key:** Responsibility also includes the ability to respond in a clear and intelligible manner, without avoidance, to the questions put to the company by all its stakeholders. This presupposes that the company is able to clearly explain its AI-related choices in an honest, didactic, and educational way (Orr and Davis 2020), possibly through the use of social media platforms (Guechtouli and Purvis 2024).

4 | THREE THEORETICAL FRAMEWORKS ROOT THE APPROACH IN A SYSTEMIC PERSPECTIVE

The chosen theoretical frameworks are called upon regarding their relevance to the research and their complementarity. They root the entire approach in systemic theory, which is vital as both AI and responsibility function as organising concepts, and their combination forms a third system.

- **Edgar Morin's Complex Thinking** provides a comprehensive method for reading the whole picture. As a form of systemic knowledge, it encourages a holistic approach to understanding and embracing

reality's heterogeneity. Morin cautions that complexity should not be seen as "the master word that will explain everything" but "the awakening word that urges us to explore everything" (Morin 1982).

- **Jacques Ardoino's Multi-referential Approach** applies these systematic principles to the field of training. By examining the same research object through multiple disciplinary lenses, this approach allows for the production of a prismatic analysis that is better able to account for the issues at stake in a composite reality by handling "several disciplinary languages, without confusing them" (Ardoino 2000).
- **Yves Clot's Clinic of Activity** extends these concepts into the realm of professional transformation. Through participatory and dialogical observation of individual and collective actions, this approach empowers employees to enhance their capacity to act, both in their specific tasks and within the broader organisation (Clot 2005).

Both well-established and current, these three reference frameworks present a dual anchoring in human and systemic perspectives. They are also accompanied by proven ethnographic collection tools that allow for the rigorous articulation of theoretical insights and practical actions. Together they facilitate the development of concrete proposals aimed at improving both reflection and practice.

4.1 | Edgar Morin's Complex Thinking: A Method for Understanding the World

In his 1982 book *Science avec conscience*, Edgar Morin explained the concept of complexity and outlined a method for analysing it: "The method of complexity asks us to think without ever closing concepts, to break closed spheres, to reestablish the articulations between what is disjointed, to try to understand multidimensionality, to think with singularity, with locality, with temporality" (Morin 1982).

Morin uses "complexity" in its etymological sense, *complexus*, meaning "that which is woven together, that which is entangled, that which is made up of different interlocking elements" (Morin 1982). Postulating that reality cannot be reduced to an agglomeration of simple elements, he opposes what he considers to be the simplifying and reductive scientific logic of many theories, which explain complex realities on the basis of some of their elementary components (atomist, linguistic, genetic, psychoanalytic theories, etc.). For Morin, everything is both situated and global, which implies that no knowledge is possible on the basis of closed, separate elements, without apprehension of the whole in which they participate.

Complex thinking aims to understand the world, without seeking to reduce or globalise it. It seeks to connect phenomena that are usually analysed independently, even though together they make up an irreducible entity, since "complexity is insimplifiable" (Morin 1982). Complex thinking therefore produces a holistic reading of human systems and phenomena, as opposed to a fragmentary (atomistic) reading, in which the whole is both more and less than the sum of its parts. Morin thus specifies that "there are emergent qualities, that is, qualities that arise from the organisation of a whole, and that can retroact on the parts", but also that "the parts can have qualities that are inhibited by the organisation of the whole". For Morin, "the real problem (of thought reform) is that we have learned too well to separate. It is better to learn to connect. To connect, that is to say, not only to establish an end-to-end connection but to establish a connection that is made in a loop".

In the context of a responsible design, implementation and use of AIS, the value of complex thinking lies in its ability to grasp a composite ecosystem while energising relationships between stakeholders, based on three main principles (Morin 1988):

- The "dialogical relationship" enables the apparent contradictions that co-exist in a system to be brought into dialogue (e.g. AI contributes to the fight against climate disruption while increasing the climate debt). The result is an ability to understand and articulate antagonistic but radically inseparable dimensions that make up the same complex reality. The "dialogical relationship includes the idea that antagonisms can be stimulating and regulating" (Morin 1982). It is therefore based on the confident expression of each stakeholder, reciprocal active listening, mutual understanding, empathy and positive criticism, all effects sought by responsible design, one of the two reflexive methods of the experimental trunk described below (Part 5). It enables each stakeholder to grasp the point of view of the others, to understand the respective objectives, to energise cooperation, and to ensure that decision-making mechanisms are based on a variety of criteria. This fully overlaps with the intentions of the support proposal described at the end of the article (Part 6).
- The "principle of recursivity" extends the notion of the feedback loop developed by Wiener[‡], the founding father of cybernetics, and explains the loops of creation, organisation and destruction that form and operate in any system (nature, society, business, consciousness, etc.). Thus, "human individuals produce society in and through their interactions, but society, as an emerging whole, produces the humanity of these individuals by providing them with language and culture" (Le Moigne and Morin 1999). This recursiveness is particularly suited to the experimental trunk. Indeed, it strengthens the individual and collective capacity to assume responsibility: "Any explanation must be complemented by understanding, any understanding must be complemented by explanation. Finally, ethical recursion strengthens us immunologically against our tendency to blame others, making them scapegoats for our faults".
- The "hologrammatic principle"[§] corresponds to a form of reciprocal inclusion that Morin explains with the formula "the part is in the whole and the whole is in the part" (Morin 1990). This is the case of the genetic heritage of an individual, which is found in each and

[‡] In "The human use of human beings: Cybernetics and society", published in 1950, Wiener predicted that intelligent machines would put an end to human labour. He also announced that the use of cybernetics without a "post-industrial" evolution of the structures of society would lead to an unprecedented increase in unemployment and social exclusion, and could even gradually destroy democracy.
[§] Also known as the hologrammic or holoscopic principle.

every cell that makes it up, just as "we produce the society that produces us" (Morin 1982). This is also true "for the company, which has its own rules of functioning and within which the laws of society as a whole play out" (Morin 1990). This principle activates the lever according to which "individual and collective ethics and responsibility feed each other" (Chapuis and Guégan 2023). There can be no awareness of teams without awareness of the individuals who make them up, and vice versa.

4.2 | Jacques Ardoino's Multi-Referential Approach: A Comprehensive Method for Articulating the Plurality of Perspectives and Hybridising Knowledge

The multi-referential approach is a practical application of complex thinking, aimed at analysing situations and improving learning.

In the 1950s and 1960s, in the context of the development of applied human sciences, Jacques Ardoino conceived of a new scientific approach that could go beyond the classical analytical approach of the so-called "exact" sciences. Like Edgar Morin, Ardoino felt that the principle of disjunction-reduction on which these sciences are based ("the mutilating alternative", (Morin 1982)) is inadequate to account for the profusion of phenomena, situations, practices and interactions observed in the human and social sciences, of which Ardoino notes "the luxuriance, the abundance, the richness" (Ardoino 1993).

The multi-referential approach therefore seeks to make complex professional and social practices intelligible via an analysis articulating the reading grids of different disciplines, most often bordering. The same research object can thus be observed, understood, and described according to the respective perspectives of sociology, psychology, psychosociology, biology, ethnology, history, economics, semantics, pedagogy, organisation, institution, etc. "Fully assuming the hypothesis of the complexity, even the hyper-complexity, of the reality about which we are questioning ourselves, the multi-referential approach proposes a plural reading of its objects (practical or theoretical), from different angles, involving as many specific perspectives and languages, appropriate to the required descriptions, according to distinct reference systems, supposed, explicitly recognised as non-reducible to each other, that is to say heterogeneous" (Ardoino 2000).

The multi-referential approach does not hesitate to state and maintain the tension between several versions of the same research object, assuming that this is the only way to obtain a global vision. The aim is to enhance the skills of individuals and the collective, to improve inter-individual exchanges, and to highlight unsuspected levers for action.

With a view to capacitating the exercise of responsibility in the field of AI, the interest of the multi-referential approach lies in its ability to articulate complementary perspectives and take a multi-disciplinary look. This contribution is valuable given the diversity of the direct or indirect stakeholders and the professions involved in the design, implementation and use of an AIS. Thus, the three types of multi-referentiality distinguished by Jacques Ardoino will be used:

- The multi-referentiality of understanding is mobilised during the questioning, listening and description phase in order to obtain "a 'vision' of things that is at once 'systemic', comprehensive and hermeneutic" (Ardoino 1991). It is particularly well-suited to the interviews that will be conducted with experts, as well as to the lexical clarification and illumination of the self-evidences. That is a necessary prerequisite for the creation of a glossary that can be understood by everyone, and for the use of a shared language.
- The interpretative multi-referentiality mobilises the work of the multi-referentiality of understanding during the analysis of practices and situations. Based on clinical listening (at the subject's bedside), it is inseparable from any in-depth reflective practice.
- The explanatory multi-referentiality is "oriented towards the production of knowledge" (Ardoino 1991), and of discourse. Articulating in a comprehensible way different disciplinary points of view will serve as a basis for compiling the glossary, for fostering a culture of collective responsibility, and for the educational actions planned for the future. Its aim is "to identify, sort out, distinguish, recognise and differentiate the very different meanings that the terms used can take on, depending on the psychology and social positions of the interlocutors and the various partners, as well as with regard to the broader circumstances in which the situations are situated" (Ardoino 2000).

Thus, "sometimes multi-referential analysis will apply to the intelligibility of concepts and notions, sometimes to that of situations" (Ardoino 1991).

4.3 | Yves Clot's Clinic of Activity: A Method for Understanding and Transforming Work from Within

The clinic of activity applies complex thinking and the multi-referential approach within a work collective. It brings into discussion the activity itself, that is to say the professional gestures, the criteria for a "well-done" piece of work, as well as the different possible resources for coping with complex situations. The aim is to collectively construct solutions that no individual would have thought of on their own.

To define the clinic of activity, conceptualised by the occupational psychologist Yves Clot (Clot 2005), it is first necessary to clarify the meaning of the following terms.

- The word clinical is derived from the Greek *klinikos*, meaning "concerning the bed". Clinical thus refers to being "at the bedside of". It is used by Yves Clot (Clot 2005) to describe the posture of listening and observing as close as possible to the activity of the person doing the work, that is to say "at the bedside of the work being done".

- The word "activity" does not only refer to an observable action. It covers all the manual, intellectual and cognitive operations that are actually put into play at every moment by the person acting in order to achieve prescribed objectives while taking into account the constraints and the context. The reality of activity therefore includes what is not done, what we try to do without success, what we forbid ourselves from doing, etc., because the suspended, prevented, set aside activities shape the real activity.

The clinic of activity thus seeks to access the logic of the actor or actress, which generally differs from that of the person observing. It is a question of going beyond what can be observed (the action) in order to uncover hidden knowledge and to reveal possible internal conflicts that are played out in the situation (the activity).

Yves Clot considers that the only experts on a given job are the people who do it, not those who prescribe it. His goal is therefore to go to the actors' and actresses' bedside, i.e. to intervene in work situations, in order to transform the work itself in the course of the intervention. As Clot explains, "the centre of gravity of psychological investigation in work is shifting. It moves from diagnosis to the invention of a framework and a mechanism where those who are concerned can begin to think collectively about work in order to reorganise it" (Clot 2005). This shift is key to understanding the professional context and actively transforming it.

This dynamic approach aligns with the evolving perspective on AI ethics, which has transitioned from a static to a dynamic understanding of ethical issues in AI, emphasising the importance of continuous engagement and adaptability (Sahut et al. 2023).

In the context of this research, the value of the clinic of activity lies particularly in the underlying idea that the transformations are achieved by the collectives themselves in their work environment. The dual analysis generated by the researcher's intervention enables the actors and actresses to become aware of their work and, if necessary, to plan ways of acting differently. In this sense, the clinic of activity promotes the development of the power to act of professionals and of the collective they form in the organisation of work, which is necessary to respond to the singular challenges posed by AI.

5 | TWO COMPLEMENTARY REFLECTIVE METHODS FORM THE TRUNK OF THE EXPERIMENT

5.1 | Responsible Design: A Design and Innovation Approach

Responsible design appeared in the 2000s and is derived from design thinking. Practised in multidisciplinary groups (therefore multi-referenced) with the aim of improving usage situations or experiences, design thinking promotes a culture of exploration through:

- "Clinical" research into the needs of users ("need finding") (Bason and Austin 2022);
- Rapid prototyping (Norman and Verganti 2014);
- Continuous iteration and adjustment (Bason and Austin 2022).

Design thinking offers a framework for balancing a company's economic ambitions with the possibilities of technology and the desires and needs of the users of the product or service. However, while effective, this approach nevertheless has its limitations. One significant challenge is the power dynamics involved in AI implementation strategies, which can result in an "illusion of transparency" where stakeholders may believe they have full control or insight into a system, while hidden biases and decisions remain opaque (Lissillour and Monod 2024). It also struggles to integrate all the facets of the "suitcase" term "user". In fact, the term covers various roles (customer, citizen, worker, parent, etc.), whose aspirations can be contradictory (Verbeek 2011). Additionally, design thinking does not really look beyond the users of the product or service observed, which tends to ignore the impact on other direct or indirect stakeholders. In this respect, the example of applications such as Airbnb or Waze is obvious: they perfectly meet the needs of their users, but their impact on the daily lives of the neighbors of the rented flats, as well as the residents of the roads used as traffic easing routes by motorists in a hurry, has not been anticipated, or at least not avoided.¶

This raises the question of what about the products and services (whether AI-based or not) designed and offered by companies. Will a service that is useful today and fully meets the needs of the average user still be relevant in five years' time? Furthermore, what are its actual and undesired impacts on other non-users, on society and social connections, on the environment, on future generations, etc.?

These are the values and interests of responsible design, also known as sustainable design: while relying on an iterative, collaborative approach and founded on the empathetic approach inspired by design thinking, responsible design enables:

- To consider the concepts of inclusion, diversity, and systemic thinking;
- To pay particular attention to users, with all their singularities;
- To take into account the context, the complexity of situations and cultures;
- To anticipate the positive and negative consequences of what we design on direct users, but also on other individuals or groups, on society, and on the world.

To achieve this, the responsible design approach offers a range of tools on which it is possible to capitalise. The aim is to constitute a panoply of references specifically adapted to the responsible design of an AIS, taking into account its unique features and the particular points of vigilance inherent in it.

¶ Val-d'Oise (France): how local residents bypassed Waze, fed up with car traffic - Le Parisien https://www.leparisien.fr/societe/val-doise-comment-des-riverains-ont-court-circuite-waze-excedes-par-le-trafic-automobile-03-05-2021-3PYSSTDVGNABVKYVQXRFA6AFEM.php

The tools and principles of responsible design make it possible to assess what is being designed, or even what has already been designed, by describing the risks and design errors to be avoided. By inviting the design teams to ask themselves questions that are sometimes disturbing, they are forced to delve deeper into aspects that have been forgotten, whether voluntarily or through ignorance. Lastly, they make it possible to precisely document the decisions to be taken when faced with a dilemma, provide a basis for argumentation with the project's actors and actresses, support the communication of doubts, and indicate a more virtuous path.

In the context of the design, implementation, and use of AIS, responsible design thus contributes to the awareness that each and every AIS raises ethical issues and that it is a question of dealing with them without evacuating them. It aims to accentuate the beneficial impacts, reduce the deleterious impacts, and even produce regenerative impacts.

5.2 | Reflective practice: a training approach that reinforces know-how through analytical capability

Reflective practice is rooted in the philosophy of learning through experience developed by the American philosopher John Dewey in the early 20th century. He highlighted that what is specifically human in human action is, on the one hand, the awareness of this action and, on the other hand, the regulation of this action. Dewey postulated that human beings learn by acting, as soon as their actions are the subject of a methodical reflective approach. For Dewey, reflective practice constitutes "the best way of thinking" (Dewey 1933). Consequently, it appears that the improvement in practitioner skills is proportional to their reflective awareness of lived experience.

In 1955, Donald Schön, a pedagogue specialising in reflective learning through practice, published his doctoral thesis on Dewey and discovered that the most effective professionals are those who methodically and rigorously 'reflect' 'during' their actions. This reflection in action is both a 'reflection on' the action and a 'reflection for' the action (a reflection for the improvement of the action). Schön thus established that each and every professional person is capable of improving his or her "reflection within" the action (and therefore of improving his or her actions), as long as he or she takes the time to reflect on it methodically afterwards. Taking this step back also allows us to uncover the hidden knowledge that we implicitly know about our actions without necessarily being able to explain them, and that resides in our actions. It was Donald Schön and his colleague Chris Argyris who introduced the concept of "reflective practice" to teaching, learning, and training (Argyris 1995).

From the 1990s onwards, reflective practice has been particularly encouraged for professionalisation purposes, at a time when institutions were increasingly demanding autonomy from their staff in terms of updating their skills. As the educational psychologist Philippe Perrenoud points out, "the more we move towards qualified professions, the more the organisation limits prescribed work and, willy-nilly, delegates to employees the concern of creating or adapting procedures to deal with the complexity of situations" (Perrenoud 1997).

Reflective practice is particularly used in medicine, social work, child protection, teaching, training, professional sports, etc. These professional fields are characterised by strong interpersonal dynamics, long-term care, a significant impact on the physical and/or social lives of the people being cared for, as well as a perpetual adjustment of knowledge and skills. In the flow of events, reflective practice sharpens practitioners' ability to 'know how to analyse' (Paquay et al. 2012), but also to strive to consciously know their knowledge.

In the context of designing, implementing, and using SIAs, the usefulness of reflective practice lies in its ability to maintain lucidity over time and flexibility in action, to enhance skills on an ongoing basis, and to make hidden knowledge visible so that it can be shared. "This process of reflection in the course of action and on action is at the heart of the art that enables practitioners to play their cards right in situations of uncertainty, instability, singularity, and conflict of values" (Schön 1994). These are precisely the kinds of situations that arise for teams working on or with AI, and the reflexive process appears to be an invaluable tool to help them deal with them. The whole thing is reminiscent of the ability of 'antifragile' systems and societies to take advantage of unexpected events, according to Nassim Taleb[#] (Taleb 2014).

As Edgar Morin reminds us: "We are at the dawn of a long-term and in-depth effort (...) to provide scientific activity with the means for reflexivity, that is to say, for self-questioning" (Morin 1982).

6 | A PROPOSAL FOR PRACTICAL SUPPORT FOR TEAMS TO DEPLOY THE CANOPY OF A GENUINE SHARED CULTURE OF RESPONSIBILITY

The experimental ambition is therefore to support project teams and individuals involved in the design and/or implementation and/or use of AIS, by articulating two complementary toolboxes:

1. Responsible design, to make employees aware of ethical risks, confront societal and environmental issues, respond to them in real-life situations, and ensure informed and committed design and innovation;
2. Reflective practice, in order to cultivate employees' capacity for self-analysis, improve professional practices, exercise vigilance, and legitimise doubt, through an ongoing training approach.

The aim of the approach is to create and maintain a shared culture of individual and collective responsibility in order to encourage the commitment of each and every person within the company and to make

[#] While the fragile fear unexpected events, the robust are indifferent to them.

the very exercise of responsibility a professional gesture (Chapuis and Guégan 2023).

6.1 | More collective, proactive and reactive moments to exercise responsibility

The support envisaged proposes to punctuate the projects with "responsible collective moments," whether proactive (i.e., planned as part of the project follow-up) or reactive (i.e., convened in response to an acute ethical questioning). The proactive sessions will be scheduled to support the people concerned at all stages of a project:

- Framing, construction and preparation of data;
- Model selection, development and training;
- Validation and deployment;
- Operation;
- End-of-life management for the product, service, or technological component.

During these sessions, the participants – representing all stakeholders – will collaborate to review both the technical and ethical systems in place and determine the necessary actions, such as adjustments, system correction, or stress tests (e.g., to check robustness, safety, and guard against the risk of cyber-attacks or embezzlement) (Binns 2018, ?). This review is crucial to understanding how AI shifts power dynamics within organisations, particularly in cases where AI systems are perceived as controlling or supportive roles, as seen in customer relationship management (Monod et al. 2023).

In the event of an unforeseen issue relating to ethics or responsibility, "à la carte" sessions can be organised to enable design teams, those in charge of operations, users, those affected or representatives of affected stakeholders, to work together in order to find effective and rapid solutions. Here again, the approach will enable participants to analyse the problem, study what may have caused it, decide collegially on the actions to be taken to remedy it, avoid a repetition of the incident either on the project itself or on related projects, and capitalise on the entire incident. In such cases, the clinic of activity and the multi-referential approach will be mobilised (Clot 2005, Morin 1982). The details of how these sessions will be convened and run will be specified during the experimental phase.

As the proposed support is aimed at all the players involved in the project, it requires scrupulous identification of all stakeholders and clarification of their roles in the design, implementation, and/or use of the AIS.

The proposed support method will ensure that all these stakeholders, direct and indirect, can participate in the responsible design of AI, either by being direct "makers" of the said design or by influencing it by sharing their points of view, needs, expectations, and/or fears. The dialogical relationship and the multi-referential approach will be mobilised here, and the principle of "always designing for everyone" of responsible design applied.

As well as "connecting" the project actors and actresses, as Morin encourages, these spaces for dialogue will make it possible to both:

- Examine social, societal and environmental issues;
- Analyse current professional practices to better take into account these issues;
- Improve the way we design in order to respond to them in a convincing manner;
- Deliberate on possible directions and their implications, or even on the temporary declutching of the AIS in the event of deadlock or persistent dilemma;
- Increase the skills and knowledge of all those taking part.

All this will make it possible to assert and defend enlightened collective choices while at the same time educating people, which is part of corporate social responsibility.

Each community of experts also has its own tools in the service of ethics, used in the context of its own activity (cf. the data testing tools and possible sources of bias used by data scientists). In a multi-referential approach, the "responsible collective moments" will therefore also be an opportunity for the various project players to share the lessons they have learned from using these tools, compare their strategy with that of their colleagues whose objectives may be antagonistic (dialogical relationship), take part in the collegial decision as to the strategy to choose, adjust their own tools, etc. The culture of individual and collective responsibility will thus be nurtured.

6.2 | Access to an overview: Understanding the role of the individuals and the collective, and capitalising on their contributions

The experiment will apply the principles of reflective practice by questioning and improving the articulation between the use of these various expert tools and "responsible collective moments," ensuring these moments are as effective as possible.

For example, it will be important to carefully schedule collective deliberative questioning sessions on the topics covered in the AIS risk assessment checklists. The goal is to enable the product manager to complete these checklists with well-supported and informed decisions by the end of these collective moments (Argyris and Schön 1992, Schön 2008). The approach could go as far as contributing to the improvement of the checklists themselves, enhancing their relevance and practicality for real-world projects (Perrenoud 2012).

To make this process accessible and comprehensible to all project stakeholders, a tool in digital format, which can be consulted and amended by the project stakeholders, could be set up, like a "travel guide to responsible AI." This would make it possible to identify the key proactive stages and milestones necessary for responsible AI design. It would allow each and every person involved in the design process ("design maker") to track their own "responsible design journey," as well as

the parallel journeys of the other contributors to the project, highlighting intersections between their individual journeys and those of others. Specifically, stakeholders would have opportunities to:

- Share insights from personal tools and draw inspiration from what others have shared;
- Address certain key issues collectively;
- Analyse and resolve ethical dilemmas;
- Extract knowledge applicable to their own work.

Such a tool can also be used as a collective logbook, archiving:

- Significant information discovered along the way (including moments of reactive exchange responding to acute ethical questioning);
- Strategic directions agreed upon by the team;
- Specific data that could be of use to the collective (such as archetypal representations of indirect stakeholders or people who would benefit from using the technology but who are currently excluded from it).

This tool, its form and its content will be tested and refined in real-world projects and their teams to ensure it effectively supports the responsible design of AIS.

7 | THE LEVER OF THE CULTURE OF INDIVIDUAL AND COLLECTIVE RESPONSIBILITY

The introduction of reflective practice within teams developing and using AI should enable all the players concerned to become more personally involved in the implementation of responsible design and use of AI.

This approach mirrors successful strategies used in the medical field, particularly in the support given to young doctors' apprenticeships (Vierset 2016) and young doctors' ethical development through reflective practice (Bleakley 2006). Stakeholders in AI development, like medical professionals, can progress through five major thresholds of personal commitment, transitioning from passive to active involvement:

1. No perception of a problem, where individuals distance themselves from ethical considerations ("there is no problem");
2. Indirect involvement, with a feeling of detachment from responsibility ("I don't feel concerned; it's not up to me to take care of this, it's up to the designers, or the data scientists, or the project manager…");
3. Interest in the ethical question but inaction ("I feel concerned but I don't see what is in my control");
4. Personal action, where individuals feel a responsibility to take action ("I act because I am concerned");
5. Collective involvement, which is characterised by both personal and group action towards shared ethical goals ("I do this, and we do it together").

Cultivating a sense of individual and collective responsibility aims to move people up this ladder of involvement. The benefits are twofold:

- Fostering active engagement: Awakening an individual's drive to act ensures that ethics and responsibility are meaningfully embodied in AI-based products and services (Verbeek 2011);
- Enhancing motivation and fulfilment: By increasing the perceived meaning of their work, individuals feel more connected to the broader impact of their contributions (Perrenoud 2012).

Group sessions involving a wide range of stakeholders will further nurture this culture of individual and collective responsibility. These sessions will provide an opportunity for each participant to recognise the significance of their role and the importance of their actions within the collective effort, increasing their understanding of others' roles and strengthening the connection between their efforts and those of the group. In doing so, the group sessions will enhance the overall meaning and value of these multiple coordinated individual actions.

The hypothesis is that such group sessions will promote awareness of responsibilities and a progression of personal commitment, creating an environment where responsibility is shared and embraced collectively.

8 | CONCLUSION

The proposed reflective support aims to facilitate a clear and honest assessment of past, present and future designs. It encourages awareness of the positive and negative externalities and legitimises a commitment to product and service sustainability, counteracting the pressures of technical and social obsolescence. This reflective process is integral in shaping choices related to usage, offered functionalities, technical solutions, marketing strategies, and economic considerations.

The overall objective of this approach is to support project teams throughout the lifecycle of AI systems, marking a significant paradigm shift in:

- The time allocated to and by teams for reflection and ethical decision-making;
- The identification, consideration, and potential inclusion of all direct and indirect stakeholders;
- Long-term monitoring of the AIS impacts on these stakeholders;
- Ensuring individual and collective responsibility by maintaining ethical operating conditions for the AIS.

However, as Edgar Morin states, "a paradigm shift is a long, difficult, chaotic process that comes up against enormous resistance from established structures and mentalities" (Morin 2020). Indeed, the responsible

design, implementation, and use of AI are only in the early stages, and much more progress remains to be made.

Crucially, this necessary paradigm shift requires the approval and backing of management and must address employee motivation. One particularly powerful motivator is the opportunity to work for a company that values and prioritises ethics and human rights, and is committed to the ecological transition[‖].

The proposed reflective support will soon be tested in real-world projects (use cases), offering an opportunity to test the hypotheses and the systems selected against the endogenous and exogenous difficulties that can hamper the smooth running of a project. Along the way, monitoring tools and methods will also be developed to ensure the achievement of objectives and to assess the relevance of the hypotheses. The results of these evaluations will be discussed in future publications.

AUTHOR CONTRIBUTIONS

The authors contributed to conceptualization, writing, reviewing, editing and addressing reviewer comments.

ACKNOWLEDGMENTS

The authors thanks the editor, the authors, and anonymous reviewers for their contributions to this issue.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

Ardoino, J. (1991) L'analyse multiréférentielle. In: *Sciences de l'éducation, sciences majeures, actes des journées d'études tenues à l'occasion des*p. 21.

Ardoino, J. (1993) L'approche multiréférentielle (plurielle) des situations éducatives et formatives. *Pratiques de formation*, 25(26), 15–34.

Ardoino, J. (2000) Les postures (ou impostures) respectives du chercheur, de l'expert et du consultant. In: *Éducation et formation*pp. 70–91.

Ardoino, J. & Mialaret, G. (1995) L'intelligence de la complexité—pour une recherche en éducation soucieuse des pratiques. *Cahiers de la recherche en éducation*, 2(1), 203–219.

Argyris, C. (1995) *Savoir pour agir: surmonter les obstacles à l'apprentissage organisationnel*. : InterÉditions.

Argyris, C. & Schön, D.A. (1992) *Theory in practice: Increasing professional effectiveness*. : John Wiley Sons.

Bason, C. & Austin, R.D. (2022) Design in the public sector: Toward a human centred model of public governance. *Public Management Review*, 24(11), 1727–1757.

Binns, R. Fairness in machine learning: Lessons from political philosophy. In: *Conference on fairness, accountability and transparency, 2018*, pp. 149–159.

Bleakley, A. (2006) Broadening conceptions of learning in medical education: the message from teamworking. *Medical Education*, 40(2), 150–157.

Brynjolfsson, E. & Mcafee, A. (2017) Artificial intelligence, for real. *Harvard Business Review*, 1, 1–31.

Chapuis, V. & Guégan, D. Pour une conception et une utilisation responsables de l'intelligence artificielle. In: *ERGO'IA 2023, 2023*.

Clarke, R. (2019) Principles and business processes for responsible ai. *Computer Law Security Review*, 35(4), 410–422.

Clot, Y. Pourquoi et comment s' occuper du développement en clinique de l'activité. In: *Colloque ARTCO, Symposium International, 2005*, pp. 4–5.

Davenport, T.H. & Ronanki, R. (2018) Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.

Dewey, J. (1933) *How we think: A restatement of reflective thinking to the educative process*. : Heath (DC).

Floridi, L. & Cowls, J. (2022) A unified framework of five principles for ai in society. In: *Machine learning and the city: Applications in architecture and urban design*pp. 535–545.

Guechtouli, M. & Purvis, B. (2024) Social media for information sharing in an industrial setting: Evidence from the chinese automotive industry. *Management Research Quarterly*, 1(1), 4–13.

Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Jonas, H. (1984) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. : University of Chicago.

Le Moigne, J.L. & Morin, E. (1999) *Intelligence de la complexité*. : l'Harmattan.

Lissillour, R. & Monod, E. (2024) The illusion of transparency: Examining power dynamics in ai implementation strategies. *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, 30(80), 79–114. doi:10.3917/rips1.080.0079.

Lissillour, R. & Sahut, J.M. (2023) Uses of information systems to develop trust in family firms. *Business & Information Systems Engineering*, 65(2), 127–141.

Lu, Q., Zhu, L., Whittle, J. & Xu, X. (2023) *Responsible AI: Best Practices for Creating Trustworthy AI Systems*. : Addison-Wesley Professional.

Manzini, E. (2015) *Design, when everybody designs: An introduction to design for social innovation*. : .

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016) The ethics of algorithms: Mapping the debate. *Big Data Society*, 3(2), 2053951716679679.

Monod, E., Lissillour, R., Köster, A. & Jiayin, Q. (2023) Does ai control or support? power shifts after ai system implementation in customer relationship management. *Journal of Decision Systems*, 32(3), 542–565. doi:10.1080/12460125.2022.2066051.

Montuori, A. (2008) *Edgar Morin's path of complexity*. : Hampton Press.
URL https://www.academia.edu/213724/Edgar_Morins_Path_of_Complexity

Morin, E. (1982) *Science avec conscience*. : Fayard, Paris.

Morin, E. (1988) Le défi de la complexité. *Chimères. Revue des schizoanalyses*, 5(1), 1–18.

Morin, E. (1990) *Introduction à la pensée complexe*. : Editions du Seuil, Paris.

Morin, E. (2020) *Festival de incertidumbres. Tracts de crise*. No. 54. : .

Norman, D.A. & Verganti, R. (2014) Incremental and radical innovation: Design research vs. technology and meaning change. *Design Issues*, 30(1), 78–96.

Orr, W. & Davis, J.L. (2020) Attributions of ethical responsibility by artificial intelligence practitioners. *Information, Communication Society*, 23(5), 719–735.

Paquay, L., Altet, M., Charlier, E. & Perrenoud, P. (2012) *Former des enseignants-professionnels. Quelles stratégies? Quelles compétences?* : De Boeck Université, Bruxelles.

Perrenoud, P. (1997) *Formation continue et obligation de compétences dans le métier d'enseignant*. : Site de la faculté des Sciences de l'éducation de l'Université de Genève.

[‖] CSA Institute study for LinkedIn and ADEME (juin 2021) https://presse.ademe.fr/2021/06/etude-de-linstitut-csa-pour-linkedin-et-lademe-78-des-salaries-choisiraient-a-offres-equivalentes-de-rejoindre-une-entreprise-engagee-pour-la-transition-ecologique.html

Perrenoud, P. (2012) *Développer la pratique réflexive: Dans le métier d'enseignant*. : ESF.
Purvis, B. & Zhang, M. (2024) The contribution of social networks to organisational learning in an industrial family business. *Management Research Quarterly*, 1(2), 54–64.
Sahut, J.M., Braune, . & Lissillour, R. (2023) Développement de l'ia et questions éthiques: passage d'une perspective statique à une perspective dynamique. *Management Avenir*, 137(5), 137–158. doi:10.3917/mav.137.0137.
Santoni de Sio, F. & Mecacci, G. (2021) Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy Technology*, 34(4), 1057–1084.
Schön, D.A. (1994) *Le praticien réflexif: à la recherche du savoir caché dans l'agir professionnel*. : Logiques. Formation des Maîtres, Montréal.
Schön, D.A. (2008) *The reflective practitioner: How professionals think in action*. : Basic Books.
Taleb, N.N. (2014) *Antifragile: Things that gain from disorder*. Vol. 3. : Random House Trade Paperbacks.
Taylor, I. (2024) Collective responsibility and artificial intelligence. *Philosophy Technology*, 37(1), 27.
Verbeek, P.P. (2011) *Moralizing Technology: Understanding and Designing the Morality of Things*. : University of Chicago Press.
Vierset, V. (2016) Vers un modèle d'apprentissage réflexif. recueil de traces d'apprentissage formulées dans les log books des stagiaires en médecine. approches inductives. *Approches inductives*, 3(1), 157–188.
Wei, Y.S. (2024) An exploratory text analysis of the emerging intellectual structure of artificial intelligence-titled marketing publications. *Management Research Quarterly*, 1(2), 25–33.
Wright, R.W. (2003) *The Grounds and Extent of Legal Responsibility*. : .

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

AUTHORS BIOGRAPHY

Valérie Chapuis works as a researcher on responsible Artificial Intelligence (AI) at Orange Innovation. She is in charge of a research project focusing on human responsibilities in AI, which aims to enable, encourage and support the responsible design, implementation and use of it. Previously, she worked for a long time on the themes of inclusive and responsible customer service, innovative organisations enabling responsible innovation, and creativity. She has also lectured in education and training at Montpellier 3 University on the issue of gender and equal chances and rights. She is a graduate of Grenoble Ecole de Management (GEM) and of the Master 2 recherche "Analyse et Conception en Éducation et Formation" (ACEF) at Université Paul Valéry - Montpellier 3.

Delphine Guégan works as a UX Designer at Orange Innovation, focusing on responsible products and services. Previously, she worked for a long time as a research engineer in the fields of user experience, innovation by design approach, and creativity. She is a graduate of Telecom SudParis and of the École de Design Nantes Atlantique..