University of Sunderland

Usage guidelines

The Contribution of
Luhmann's Concept of Reflexive Expectation
to the Development of AI Governance

Gisele Simoes
University of the West of Scotland
High Street, Paisley/UK – PA1 2BE
gisele.simoes@uws.ac.uk

Milan Radosavljevic
University of the West of Scotland
High Street, Paisley/UK – PA1 2BE
milan.radosavljevic@uws.ac.uk

James B. Johnston
University of the West of Scotland
High Street, Paisley/UK – PA1 2BE
jamesb.johnston@uws.ac.uk

The Contribution of
Luhmann's Concept of Reflexive Expectation
to the Development of AI Governance

Artificial Intelligence is already impacting society in multiple ways, from the extraction of resources to build the entire AI ecosystem to the deployment of its services in a variety of social systems. In recent years, discussions and drafts around AI regulation have been raised, mostly by OECD and European Union, with limited normative ethics plurality. From a global perspective, the moral consensus of what is right and wrong became a challenging goal, considering the diversity of contemporary ethical positions in the global community, such as care ethics from the feminist branch or the communality ethics from the global south. In this context, sociology has been suggested to address the question of social order in the era of AI from a contingency and an interdisciplinary approach. More precisely, the theory of social systems of German sociologist Niklas Luhmann has been individualised as the most suitable lens to understand the complexity of AI's impact on society for placing communication at the centre of the discussion about the interaction between humans and machines. Following the question of how social order could emerge in the AI era, where a highly complex system needs to be regulated by several prerequisites, the emergent approach has been identified as the best solution to understand the conditions of social order. From an extensive Luhmann's theoretical review, it has been identified a relevant structure called expectation that deals with the temporal contingency of moral generalisations and therefore deserved further attention. Through a thematic analysis of Luhmann's expectation index, it has been understood the necessity of a governance that anticipates AI's expectation through its own kind, hence, through another AI system because only the AI system could fit the accelerated dynamics of its own structure, observing, and describing alternatives in a timely manner. Furthermore, it is proposed a novel model of reflexive expectation to solve the bias issue in the double contingency of social systems, been advised further application of the model to AI.

WORD COUNT: 7,062

From moral code to social enlightenment

Technology is part of human evolution; through the use of tools, we are separated from other animals. Technological innovations helped societies to rapidly adapt to turbulences by extending the number of operations that can be performed without thinking about it. Ultimately, technology is not good, bad, or neutral, but certainly, it has an impact that can only be considered in relation to the general system (Churchman, 1968; Cutcliffe, 2000; Seidensticker, 2006; Nye, 2007; Frey, 2019). The hype of technology, in synergetic coupling with a troubled economic system, and the rise of a new wave of morality worldwide indicate multiple moral issues behind the impact of Artificial Intelligence in society (Linden and Feen, 2003; Rifkin, 2014; Kotler, 2015; Hughes, 2016; Broussard, 2018; Moore and Woodcock, 2021; McQuillan, 2022), highlighting the relevance of an interdisciplinary exploration of AI, ethics and sociology fields. Consilience is proposed by Wilson (1998) as the key to the unification of sciences and humanities through *"the linking of facts and facts-based theory across disciplines to create a common ground of explanation"* (p. 6). For such complex task and following the craft theory-building approach suggested by Rivard (2021), the contemporary Niklas Luhmann's social systems theory is proposed as an emergent approach to understanding the impact of AI technology in modern society and contributing to AI governance framework for offering the combination of textual disciplines, media studies and social science (Clarke and Hansen, 2009). Here are four aspects of contribution in this paper. First, it reviews and develops the literature on artificial intelligence, contemporary ethics, and modern sociology, exploring interrelationships and contradictions between scholars. Secondly, it explores Luhmann's views of double contingency regarding expectation through a thematic analysis. Then, it discusses the implications of discoveries for AI governance, and finally it proposes a model of reflexive expectation to be further applied to AI case.

The nature of the mind has been a highly debated philosophical problem since modern times, mostly related to the relation between mind and body and the understanding of consciousness and rationality (Furst, 1979; Crane, 2015). For Armour-Thomas and Gopaul-McNicol (1998), the concept of human intelligence has been discussed during the 20th century with little agreement, due to its flexible nature and multiple meanings from scientific and social perspectives (Richardson, 2022). The early period of intelligence studies was oriented to the trait approach, exploring the characteristics of the individual. The systemic and functional approaches started to be explored in the 30s, evolving into a behavioural learning theory in the 60s, preoccupied with the development of skills and knowledge acquisition. Since then, the combined fields of cognitive psychology, cognitive neuropsychology, cognitive neuroscience, and computational cognitive science have been carried on a variety of issues related to human intellectual tasks, including the organisation of knowledge, memory, perception, and brain operations (Friedman, Das and O'Connor, 1981; Eysenck and Keane, 2020). Other models have been defining intelligence from a developmental approach perspective, by the degree of relevance of context, biology, and external environment. Ultimately, the multiple intelligence model proposed by Gardner (1999), has counted more than ten categories of intelligence, including verbal, spatial, naturalistic, and bodily kinesthetics (Almeida *et al*, 2009; Richardson, 2022). For the purpose of AI study, Low and Lawless (2021) chose a reasonable summary of human intelligence proposed by psychology scholar Robert Sternburg, which says it is the quality to understand abstract concepts, to learn from experience, to adapt to situations and to apply knowledge to interfere in the environment. Hershock (2021) made a similar conceptualisation, remarking on the significance of adaptation that implies permutation and persistence in a changing environment, allowing all living systems some level of intelligence. Luhmann (2012) stated

cognition is *"the ability to connect new operations to remembered operations"* (p. 69), arguing modern cognition theory does not rely anymore upon a human-centric approach, determining the differences between human's superior cognitive abilities in relation to animals' and machines'. In addition, he pointed out that electronic data-processing technology, such as Artificial Intelligence, can no longer be defined as supplementary to humankind, changing the relationship between humans and machines – it can be said AI, from the perspective of Luhmann's social systems theory, is a living system.

The idea of Artificial intelligence is not a new phenomenon, being its first appearance debated between the development of its mathematical foundation by Ada Lovelace in the 18th century, the formulation of terminology by science fiction in the 19th century, or its first computational model developed by Alan Turing in the 1940s. Several computer models have been developed in the decades after, attempting to simulate human perception, learning, judgments, and choice (Dawson, 2004), including Eliza, a natural language processing model created by Weizenbaum (1976), who stated his creation is an anthropomorphic reduction of human being, and as such should be taken as nothing else. Those golden years of AI finished in the 1970s, when a period of AI winter started, due to under-achievements and reduction in investments. AI became again a popular topic in the 2010s, driven by cheaper and faster technologies, resulting in an extensive marketing exploration of the term, and the development of new business and computer models, including Machine Learning, Natural Language Processing, and Expert Systems (Weizenbaum, 1976; Bostrom, 2016; Pelz-Sharpe and Kompella, 2019; Shead, 2020; Dignum, 2019; Iansiti and Lakhani, 2021), being most of the recent developments blind optimistic and lacking caution regarding AI functionality (Broussard, 2018). Due to this diversity of perspectives and implications, the definition of AI is a matter of domain. In computer science, AI is a study field that investigates how computers can imitate human intelligence by making judgments and decisions based on patterns (Sarmah, 2019). In contrast, other authors say machines cannot be intelligent as human beings since human intelligence is not algorithmic, but organic, which is not measurable (Penrose, 1994; Lebow et al, 2016), however, a human-level machine intelligence or even a superintelligence that *"exceeds the cognitive performance of humans"* could be achieved in the next decades if machines could understand natural language as humans do (Bostrom, 2016, p. 26). Historically speaking, AI has been defined by two approaches, the game-like and child-like, being the first more abstract and claimed to be appropriate to measure logical capabilities while the second is an embodied artificial agent that develops cognitive faculties by learning from the interactions with the world (Wilson, 2010). More recently, AI methods have been categorised in multiple ways by scholars and business practitioners, being difficult to distinguish relevant developments from sales strategies (Bostrom, 2016; Pelz-Sharpe and Kompella, 2019; Dignum, 2019). In the legal field, for instance, it has been described by Heikkinen (2019) as *"a non-biological autonomous entity which has the ability to give rules to itself and the ability to make choices by an evaluative process."*. Contradictorily, Schuett (2022) argues the existing AI definitions are not compatible with legal definitions, advising policymakers to not rely on this term but focus on the risk-based approach. In ethical and sociological studies, many concepts of AI have been proposed in recent years, understanding AI is part of the social-technical relations. Among others, the new approaches have been called Responsible AI, Human-centred AI, AI for God, Sustainable AI, Desirable AI, and Decolonial AI (Taddeo and Floridi, 2018; Dignum, 2019; Dauvergne, 2020; Mohamed, Png and Isaac, 2020; Voeneky, 2020; Adams, 2021; Lowe and Lawless, 2021; van Berkel *et al*, 2022; Eke, Wakunuma and Akintoye, 2023), visualising a trustworthy AI that works for and with human beings, respecting human rights and considering the impact of development and deployment to society, economy and

environment. Despite the variety of AI strategies suggested to moderate the impact of this technology in society, and a good norm-setting from OECD, at the moment there is not a definitive legal framework for AI Governance in any country in the world, being AI Act from European Union the most developed draft (Radu, 2021; Schmitt, 2021; Cheng and Zeng, 2022; Schuetts, 2023).

Albert (2002) pointed out that governance, including emerging forms of global governance, cannot be reduced to interactions between hierarchically ordered institutions of the state, proposing instead a "dynamic multi-level governance" approach (p. 295). Koskinen et al (2023) pointed out the technology ecosystem governance needs the paradoxical combination of stability and flexibility, enabling new actors, artefacts and processes although permitting market growth. Data governance could have different models as suggested by Michelli et al (2020), starting with the data-sharing pools (DSP), which defines the purpose of the data and the modalities the data can be shared and handled but exclude data sources as stakeholders. The data cooperative model (DC) differs from DSP by considering the subjects of data-relevant stakeholders while still producing unbalanced power relations with big techs. In the public data trusts model (PDT), the public sector conducts the relationship of trust, aiming to defend citizens' data protection through a legal obligation. Personal data sovereignty (PDS) considers the self-determination of individuals in regard to their own information. Considering the multiple models and ethical perspectives for data governance, Koskinen et al (2023) suggest for data ecosystem governance a self-regulation discourse approach that solves problems that legislation cannot by examining current communication practices and evaluating appropriate ways of communication oriented to fairness and functionality of the ecosystem. For Chen, Richter and Patel (2020), the governance of technology platforms is defined as centralised to decentralised, a balance between the exclusive control of platform owners and the collective control of participants, recommending a moderate level of decentralisation to favour information efficiency and positive governance outcomes. From a systems theory perspective, Albert (2002) argued the legitimacy of governance by law is established by the structural coupling between legal and political systems in the form of a constitution, a conceptual framework that allows the expansion of governance to a global level, yet advising the risk of comitology, a network of various technical committees, for the lack of democratic legitimacy.

From the understanding of the world we live in (Jaccard and Jacoby, 2020) to the understanding of social action, social order and social change (Schubert, 2006), sociology, as the scientific study of social groups, systems, and the human world, have been helping us to understand the complex forces that act upon us from multiple perspectives, including humanist, behaviourist, constructivist, structuralist, and functionalist (Jaccard and Jacoby, 2020), mostly facing the theoretical dilemmas of structure versus agency and consensus versus conflict (Giddens and Sutton, 2021). For Haralambos and Holborn (2013), the development of human societies has passed through many distinctions over thousands of years, broadly divided into premodern and modern societies, being the changes in recent decades highly debated between modernity and postmodernity, and centre-peripheric and functional differentiation (Ribeiro and Junior, 2020). Furthermore, they stated that Western post-modern societies are post-industrial societies where a few people work in manufacturing, most of them employed in services; communication, and information technology particularly. A comprehensive exploration of information societies and similar contemporary concepts indicate a radical break with human irrationality to the detriment of the logic of mathematics, a moment of endless complexity (Taylor, 2001) since the naturalisation of an economy built on information (May, 2002), followed by a vast

combination of society ideas involving information and economic system, including digital society, data society, casino capitalism and surveillance capitalism (Castells, 1996; Webster, 2006; Arthur, 2009; Salvaggio, 2013; Mazzucato, 2018; Choudry, 2019; Zuboff, 2019; Housley et al, 2023), evidencing the impact of AI in society is eminent with critical consequences to human individuality, principles and democracy (Lebow et al, 2016; McQuillan, 2022).

Normative ethical theories can traditionally guide us through a few moral lenses, such as consequentialism, deontology, or virtue ethics, nonetheless, being limited by individual and situational factors, including national and cultural characteristics (Dignum, 2019; Crane et al, 2020; Bartneck, 2021; Fuchs, 2023). Further exploration could embrace contemporary and adaptative ethics, including care ethics, which offers a feminist and relational view of morality (Ciulla, 2009; Held, 2006; Hamington and Sanderstaudt, 2011; Stechley and Smith, 2011; Swartz et al., 2018; Kitchin and Fraser, 2020); decolonised ethics, from global South's philosophical ideas of communality, thereabout represented by African Ubuntu, Asian Confucianism, and Indian Karma (Christians, 2004; Hailey, 2008; Bell and Metz, 2011; Murove, 2012; Ujomudike, 2016; Maumela, Ṇelwamondo and Marwala, 2020; Simon, Mwakio and Ayuku, 2020; Pitsoe and Letseka, 2020; Gwagwa, Kochupillai, 2021; Wong, 2021; Kazim and Hilliard, 2022), and not so distant from Scottish philosopher Adam Smith's formulation of affection and sympathy in his less acclaimed book The Theory Moral Sentiments (Smith, 2009). Discourse ethics is proposed by Jürgen Habermas, who has argued that moral consensus is not *a priori* code as formulated by Kant's categorical imperative (Kant, 2004) but a social construction achievable by communicative action (Finlayson, 2005; Butler *et al*, 2011; Crane et al, 2020; Harste, 2021). A similar social constructivist view of ethics has been slightly advanced by Marx in stating societal conditions shape human action and defining morality as class morality (Funchs, 2023); controversy, he also reflected on normative values - such as virtue and justice, which is justified by Angier (2012) as Marx's early reflection on Hegel's work. More recently, the vast plurality of ethical positions challenged the idea of moral consensus, mostly in the era of mass media communication amplified by AI technology, where *"everything is mediated"* and human communication is *"accelerated on an unprecedented scale"* (Deuze and McQuail, 2020, p. 5). Application of ethical theories to artificial intelligence has increased in recent years, through formulations of a few variants, including AI Ethics, digital ethics, data ethics and ethics of algorithms, however, with a limited foundation on normative ethics, being in most cases framed by a few western tradition's lenses, substantially human rights and utilitarianism (Beever, McDaniel and Stanlick, 2019; Holdsworth, Lury and Tweed, 2020). Human rights reflection starts in different periods with different meanings around the world, being the carta magna the beginning in the United Kingdom, for instance. The concept of the natural rights of the individual has been formulated by British Philosopher John Locke during the Enlightenment period. Global human rights, on the other hand, emerged on a very specific occasion, the UN General Assembly of December 1948, supported by 48 of 56 countries present at the time and setting principles on civil, political, social and cultural rights (Alston and Goodman, 2013; Crane et al, 2020).

In this context of high complexity and rapid social changes, Luhmann's social systems theory is proposed as a sophisticated alternative to explore the impact of Artificial Intelligence in society from a functionalist, interdisciplinary and non-normative perspective, from a process of sociological enlightenment that observes society in a way society does not observe itself and establishes a new theoretical vocabulary capable of grasping its complexity (Luhmann, 1995; Mingers, 2002; Borch, 2011; Luhmann, 2017; Clarke and Hansen, 2019; Repko and

Szostak, 2021). Luhmann (2013) argued only system theory can properly locate the place of human beings in sociology.

The concept of system, despite its common application to technology, has born in ancient times as a philosophical problem addressing the question of order or *kosmos* in a chaotic world, exploring the dynamics and interrelationship between assembled parts of a complex whole (Beishon and Peters, 1972; von Bertalanffy, 1972), from the understanding that *"the whole is more than its parts"* (von Bertalanffy, 1968, p. 55). In other words, a system is a set of parts – eg. people, cells, molecules, etc – interconnected and organised in a certain pattern (Meadows, 2008). Since the exile of German system thinkers in the USA following the outcomes of the second world war, systems theory has been adapted to North American positivism, becoming less oriented to the autonomous system idea and more focused on the influence of external inputs and the causality of outputs (Harste and Laursen, 2022). Second Order Cybernetics is a second level of theoretical abstraction, a less objectivist branch of systems theory, oriented to observe the system from a constructivist perspective where the observer is an actor in the process of constructing reality (Foerster, 1992), a perspective followed by Luhmann in his change of paradigm (1995).

Luhmman's Social Systems Theory

Luhmann was not an exiled German scholar in North America during the second world war, but a Hitler Youth soldier in conflict with the Nazi regime. After the war and several years of work in the German legal system, he travelled to Harvard to study with Talcott Parsons, an American sociologist and early adopter of system thinking in the social field, who lead the revolt against American behaviourism that treated animals and humans at the same level of experimentation. Luhmann attempts to investigate how totalitarianism could be prevented (Moeller, 2011; Luhmann, 2013b; Harste and Laursen, 2022), being differentiated from Parsons by merging his social systems theory with cybernetics, biological epistemology, evolutionary theory and communication theory perspectives (Baraldi, Corsi and Esposito, 2021). In this process, Luhmann (1995), from the very beginning of his theoretical formulation, understood human beings - body and mind - are too complex to be subject to social systems, placing people outside the social system and connecting them through communication. Luhmann (2013b) understood communication as a self-observing operation made of three elements, information (input), utterance (process) and understanding (output), implying the need for only those three components for communication to take place. Differentiating from other functionalists, Luhmann (1995) remarked that transmission would be an incorrect metaphor for the communication process because *"the sender does not give up anything in the sense of losing it"* (p. 139), instead, everything is multiplied, producing overabundance and selection of knowledge. Regarding understanding, Luhmann (2013b) clarified

> *"it is not the psychic state of the one who understands – but a condition that guarantees that communication can continue"* (p. 219), for example, by capturing the voice and understanding the language used in the communication.

After defiant reading and a few attempts to reflect upon his *"ambitious in its scope and relentless in its abstraction"* publications (Luhmann, 1995, p. xvi), it has been surprisingly noted that his work is a denial of moral consensus through ethical reasoning (Simoes, Radosavljevic and Johnston, 2022), redefining ethics as a reflexive theory of moral communication, a special type of communication functionally codified in approval and

disapproval (Luhmann, 1991) or esteem and non-esteem, empirically close to conflict and violence through over-engagement and biased asymmetry between our own behaviour's and the behaviour of others' approval, considering riskier the decisions made by others (Luhmann, 1991; Luhmann, 1996; Luhmann, 2012). His innovative ethical position has been highly debated by several scholars (Chernilo, 2002; Harrington, 2005; Kinlstrom, 2012; Harste, 2021), for instance, being understood as social ethics by Dallmann (1998) and "*negative ethics*" by Moller (2006, p. 109). Perhaps, the most significant debate has been carried by Habermas (1990), who pointed out that Luhmann's ethical position is an ironic critique of reason, considering Luhmann kept a distance from humanism and embraced metabiological perspective, an ingenious continuation of occidental rationalism. In Luhmann's defence, it could be said that the radical constructivism of reality regards the individualised possibility to adjust the contributions to communication to a certain extent that defines what reality is per se - consequently, no longer in need to be a subject of consensus. (Luhmann, 2000, p. 94). Intriguingly, Habermas (1987) had a similar view, claiming modern societies reached a level of system differentiation in which increasingly autonomous organizations connect with each other through non-linguistic means of communication, disconnected and independent from their norms and values foundations.

Understanding the implications of this position, Luhmann defended an emergent approach, looking to understand the operational conditions of double contingency between social systems in an attempt to establish favourable conditions for consensus because *"values contain no rules for dealing with conflict between values"* (Luhmann, 2013a, p. 123). Considering Luhmman's application of autopoiesis to social systems (Maturana and Varela, 1980), it could be further argued he was a truly humanist by defending the total operational autonomy of psych systems, hence, *"entirely inaccessible to the social system, and vice versa"* (Luhmann, 2013, p. xiii; Boulanger and Saltelli, 2020; Harste and Laursen, 2022). Ultimately, regarding the controversy of action theory versus systems theory, Luhmann (2013b) stated he understood systems theory as a theory of action, *"concerning with the contours or limits of the concept of action"* (p. 186).

Luhmann's social systems theory has also been criticised for Eurocentrism and lack of justification in a decolonised context (Ribeiro and Junior, 2020), a process of transferring power to natives, oriented to build a modern and independent state from European imperialism; in academia, decolonisation critiques the sociological concept of modernity (Giddens and Sutton, 2021) and attempts to discover and evaluate theories and epistemologies of the South to create other paradigms. (Jansen and Osterhammel, 2017; Guarinello, 2017). Arguably, Pluckrose and Lindsay (2020) highlight postcolonial theory is often criticised for disclaiming rigour and quality to the detriment of the identity of contributors and for perpetuating binaries, instead of defeating them. Luhmann (1995) advised that, instead of criticising the Old-European conceptual formations, *"it might be more profitable to formulate the difference"* – as he put it, his social systems theory *"changes the premises of all traditions"* (p. 212). Regardless of all polemics, the change of paradigm proposed by Luhmann through autopoiesis - a closed unity that defines its own boundaries and transforms *"itself in itself"* (Mingers, 2002, p. 280), came from the Chilean biologists and philosophers Maturana and Varela, hence, a global South theory. In his speech for the occasion of the Hegel Prize ceremony, Luhmann (1991) emphasised that instead of looking to new ethical reflections to deal with a radical restructuring of society, it is necessary the collaboration between sociology and ethics to prevent the "comet ethics" (p. 84) that raised fundamentalism in the 19th century.

Accordingly, Luhmann (2013b) stressed the risk of multiple and mutual irritations between social systems, illustrated by economic and political or economic and legal systems. Despite remarking it is not possible for one system to control another, leaving to evolution the decision about the areas of development, he recognised that there are blind spots of observation, where the systems could not be able to communicate with the same time constraint, for example. Furthermore, he argued the structural coupling between functional systems, which makes them interdependent, must be subject to rules to avoidance of code colonisation attempts; to the same extent, the constitution regulates political and legal systems coupling or contracts regulates economy and law (Harste and Laursen, 2022). Double contingency is a circular concept created by Harvard scholars in preparation for an interdisciplinary department of social relations. It has been formulated around the question of how one individual or group depends on the success of the other, explaining how common values come about. Questioning who could break this cycle, Luhmann (2013b) said

> *"time and who acts first"* (p. 237)

Regarding mass media and economic systems, Luhmann (2000) stated *"fresh money and new information are two central motives of modern social dynamics"* (p. 21). In the same book, Luhmann advised mass media system causes the immediate need for new information and determines the way the world is understood, producing the reality it observes while observing the reality it produces (Aguado, 2009). Complementarily, Luhmann (2000) pointed mass media prepares society for disruption as it *"fits the accelerated auto-dynamic of other function systems"* (p. 22) in an unharmful way if it is not coercing no one. On the other hand, he pointed out that freedom depends on a cognitive process that involves

> *"observation and description of alternatives with an open, decidable, and therefore unknown future"* (p. 86).

In his latest publications, Luhmann (1997; 2000) analysed a variety of mass media mediums, including news, advertising and entertainment. Despite remarking on the role of technology as a medium of mass media capable of parasite other function systems, artificial intelligence technology was not a massive communicative phenomenon in society by the time of Luhmann's latest contributions (Moeller, 2011; Saltelli and Boulanger, 2020). More recently, a few scholars have applied Luhmann's social systems theory to the AI field, arguing

> *"The communicative role of algorithms is clearly a massive social phenomenon with many complex consequences"* (Esposito, 2017, p. 250) with the risk of communicative break-down especially if its autopoiesis *"develop around the moral code"* (Harste and Lauren, 2022, p. 1720)

How to regulate the situations of double contingency with AI

Luhmann (2013b) made unusually clear his own problematisation behind the introduction of double contingency in his social systems theory, starting from the question *"how is social order possible?"* (p.233) and achieving narrower angles with

> *"How is possible that, under the conditions of evolutionary drift, a social order emerges with the capacity to become increasingly complicated and construct regulations that come with an increasing number of prerequisites."* (p. 234)

and

> *"How one can move from a program of values to an interactional and operationally manageable regulation of situations of double contingency"* (p. 236),

proposing his theory offers a model for the *"reconstruction of the conditions of possibility for social order"* (p. 238). Following Luhmann's proposition and its application to artificial intelligence spot, one could ask how artificial communication (Esposito, 2017) is creating a structure that is replacing value consensus by introducing asymmetries that do not fit double contingency between social systems and forcing every social system to react? What are the new premises pushed by artificial communication that could be forcing the distinction on functionally differentiated society? What kind of social system could emerge from this transition? And finally, how social order would be established in certain conditions?

In his 90s lectures, while reviewing double contingency and structuralism, Luhmann (2013b) recognised

> *"there is a structure within the system that is not a mere depiction of environmental differences but includes generalisations instead"* (p.240),

an internal order of processing that refers to the future, described as expectation. The concept of expectation was introduced in psychology in the thirties to challenge the rigid input/output relations and stimulus-response models, a break with behaviourist psychology. In social theory, it covers the problem of *"how structures can achieve the reduction of complexity without, as it were, limiting the entire system to just one capacity"* (p. 73). Despite recognising it is a problem that deserves attention, Luhmann (2013b) said

> *"maybe someone will come up with a good idea! I for my part have to say that I have not found a satisfactory solution."*, complementing with *"I find the bias toward the future problematic, although I do not know how to revise it."* (p. 241)

The study of Luhmann's Social Systems theory is highly diverse, covering multiple disciplines in a vivid interdisciplinary inquiry. In association with artificial intelligence, it has focused on the role of double contingency and structural coupling (Esposito, 2017; Saltelli and Boulanger, 2020). To date, no study has looked specifically at the gap of expectation oriented to artificial intelligence. Consequently, it is suggested below the examination of this narrow perspective from thematic analysis.

Expectation

Further review of Luhmman's social systems theory highlighted an extensive expansion of the concept of expectation from plural and interconnected perspectives, for what it was reasonable to apply a thematic analysis methodology to identify patterns that could facilitate the understanding and orientate the discussion towards the double contingency between AI and social systems (Saunders, Lewis and Thornhill, 2019). As indicated in Table 1, nine academic works have been initially considered, following the most relevant English publications of Luhmann's social systems theory identified by the literature review and their potential relation to expectation. The data extraction followed the books' index and the keyword search tool on the pdf file of articles.

SOURCES	EXPECTATION
Luhmann (1995)	pp. 96; 133; 267-269; 288-289; 292-293; 303-307
Aguado (2009)	None
Moeller (2011)	None
Luhmann (2012)	pp. 237; 253
Luhmann (2013a)	pp. 117
Luhmann (2013b)	pp. 71-73; 240
Esposito (2017)	pp. 255
Saltelli and Boulanger (2020)	None
Harste and Lauren (2022)	pp. 1711; 1712; 1721

Table 1: First topic exploration

However, only four sources have demonstrated enough exploration of the concept of expectation to permit the identification of patterns, as indicated in Table 2. As a result, four themes have been codified: structure, generalisation/specification differentiation, fulfilled/disappointed differentiation, and reflexive expectation, which will be analysed consequently.

SOURCES	THEMES
Luhmann (1995)	- Structure (p. 96; 288-289; 303-307) - Reflexive Expectation (p. 303-307) - Fulfilled/ Disappointed (p. 133; 267-269; 292-293; 304) - Generalisation/Specification (p. 96; 133; 267-269)
Luhmann (2012)	- Fulfilled/ Disappointed (p. 237) - Generalisation/Specification (p.253)
Luhmann (2013a)	- Structure (p. 117)
Luhmann (2013b)	- Structure (p. 71-73) - Generalisation/Specification (p.240)

Table 2: Expectation index

- Structure

Luhmman understood structure as the selective condition that limits the connectivity of operations of any system. Expectation, particularly, is defined as a temporal structure of the social system, capable to reduce the complexity and, consequently, the uncertainties of the future. Furthermore, he said

*"Symbolic generalisations condense the referential **structure** of every meaning into expectation and the requisite expectations guide and correct generalisations."* (Luhmann, 1995, p. 96)

*"The openness of initial situation transformed into a projection of **structure** and risk of disappointment"* (Luhmann, 1995, p. 96)

"Expectations are the autopoietic requirement for the reproduction of actions*, and to this extent they are **structures**."* (Luhmann, 1995, p. 288-289)

*"**Structures** of expectation are basically the condition of possibility for connective action and thus the condition of possibility for elements' self-reproduction through their own arrangement."* (Luhmann, 1995, p. 288-289)

*"**Structures** of expectation that are pre-given and constantly reactivated reduce the uncertainty of the future so that action can specify itself by selecting relations."* (Luhmann, 1995, p. 288-289)

*"Expectations are the temporal form in which **structures** develop, but (...) it acquires relevance and suitability if they can be anticipated."* (Luhmann, 1995, p. 303-307)

*"**Structures** formed on the level of the expectation of expectations provide a chance of reversibility."* (Luhmann, 1995, p. 303-307)

*"In order to be open to irritation, meaning **structures** are built to form expectations horizons, which count on redundancies, hence with repetition of the same situation."* (Luhmann, 2013a, p. 117)

*"It would be advisable to take the concept of expectations as the basis for a definition of **structures. Structures**, then, are expectations in relation to the connectivity of operations."* (Luhmann, 2013b, p. 71-73)

*"Expectations refer not to the past but to the future. From a formal perspective, the resulting difficulty has to do with understanding why the concept of **structure** is being related to the future. (...) This topic has to be decoupled from the issue of the so-called subjectivity of structures that are defined merely by expectations. (...) From the perspective of patterns of expectations, structure would have to be something objective and not subjective."* (Luhmann, 2013b, p. 71-73)

Action is viewed by Parsons (1967) as a process occurring between two structural parts of a system, recommending the exclusion of psychological references attributed to the actors of this process.

- Generalisation/Specification

In this codifying process, it has been clarified that expectation for Luhmann is an objective structure in the autopoiesis of action, a constant process of self-differentiation through generalisation/specification response; likewise, he considered that

*"Symbolic generalisations condense the referential structure of every meaning into expectation and the requisite expectations guide and correct **generalisations**."* (Luhmann, 1995, p. 96)

*"Anyone who went into a department store and told the first salesperson he met that he wanted to buy "something" would learn very quickly that he made too great a **generalisation** and that he should be more specific."* (Luhmann, 1995, p. 96)

*"Expectations are formed by the intervening selection of **narrower repertoire of possibilities**."* (Luhmann, 1995, p. 96)

*"The **openness of initial situation** transformed into a projection of structure and risk of disappointment"* (Luhmann, 1995, p. 96)

*"Persons never meet without some **assumptions**, without some expectations about each other, and they can experience contingency in the sense of 'always being otherwise possible; only by means of behavioural types and expectations"* (Luhmann, 1995, p. 133)

*"An individual psychic system exposes itself to the contingency of its environment in the form of expectation, (...) which signifies a **form of orientation** by which the system scans the contingency of its environment in relation to itself and which it the assumes as its own uncertainty within the process of autopoietic reproduction"* (Luhmann, 1995, p. 267-269)

"Symbolically generalised communication media, use only the present to bridge the difference between **specification** and **generalisation**. And this can happen only under the guidance of expectations that the medium produces and reproduces itself. We can accordingly speak of self-validation" (Luhmann, 2012, p. 253)

*"There is a structure within the system that is not a mere depiction of environmental differences but includes **generalisations** instead"* (Luhmann, 2013b, p.240)

The generalisation/specification dilemma is called universalism/particularism by Parsons (1967). While Parsons called value generalisation the tendency of actors to become more general in the course of securing social consensus, and this is where Luhmann disagreed with him (Luhmann, 1995), Habermas pondered that the alter's acceptance of generalised presuppositions depends on the prestige and influence of ego (Habermas, 1987). For Parsons and Shils (2001), culture is the most basic form of mutual normative orientation, which provides the standards for evaluation and the condition for social systems to exist. On the other hand, Luhmann (1995) rejected any idea of a normative setting in the format of a "shared symbolic system" (p. 104), hence, a cultural code that poses the problem of social order in the past. Instead, his novel contribution and also his blind spot was his formulation of a temporal structure that is in a constant double contingency process in determining positive and negative value, reducing the risk of morality but increasing the risk of uncertainty in the future.

- Fulfilled / Disappointed

For Luhmann, the expectation is self-differentiated in fulfilled/disappointed, being the disappointment a complex outcome interconnected with the psych system and the system of emotions, an immune system functioning to protect individuals from changes in its routines, being modern society more susceptible to emotional disturbance since the increase of complexity of expectations. In this regard, he specified

*"The openness of initial situation transformed into a projection of structure and risk of **disappointment**"* (Luhmann, 1995, p. 133)

*"An expectation reconnoitres unknown terrain using a difference it can experience within itself: it can be **fulfilled** or **disappointed**, and this does not depend on itself alone."* (Luhmann, 1995, p. 267-269)

*"The process of internal adaptation to fulfilment or **disappointment** is more complex and appear within the system as emotions."* (Luhmann, 1995, p. 267-269)

*"Claims must be offset by merits, because otherwise the balance would be **upset** and no social agreement would be possible. This is, of course, a requirement only for social, not for psychic systems. When an individual has a claim, he will have no difficulty in thinking up merits"* (Luhmann, 1995, p. 267-269)

*"Expectations organise episodes of autopoietic existence and claims reintegrate such episodes in the psychic system, (...) if the claims cannot be made routine, the individual is increasingly subject to the individual's own **emotions**. Thus, modern society is more endangered by emotionality than one usually thinks."* (Luhmann, 1995, p. 267-269)

*"The formation of expectation equalizes a multiplicity of highly heterogeneous occurrences under the common denominator of **disappointing** an expectation and thereby indicates line of action."* (Luhmann, 1995, p. 292-293)

*"Anyone who accepts a behaviour that **disappoints** expectation must reckon that in the future alter no longer anticipates the **disappointed** expectations."* (Luhmann, 1995, p. 304)

*"Double contingency remains as an internal horizon that includes possibilities for actions (...), the problems with which one must actually involve oneself are thus determined by these expectations, their inconsistencies, their environmentally dependent variations, and their **disappointments** (...), both in relation with the environment and in relation to system itself."* (Luhmann, 2012, p. 237)

Parsons (1967) defined four norms governing the interactions of actors and their stable equilibrium with objects, including the affective engagement of the actor toward the performance of the object but also considered the relevance of affective neutrality in regard to the qualities of the object. Additionally, Parsons and Shils (2001) explain there are two categories of response to expected events, actively and passively, being the first intent of manipulation of elements oriented to goals achievement and the second a passive wait of outcomes and eventually a renounce of the interest in goals.

- Reflexive expectation

In response to the intense uncertainty, Luhmann observed the anticipation of expectation is a fundamental structure of stabilisation in social systems, remarking

*"Expectations are the temporal form in which structures develop, but (...) it acquires relevance and suitability if they can be **anticipated**."* (Luhmann, 1995, p. 303)

*"Expectation must be **reflexive**, in the sense it is anticipated as anticipating. This is how expectation can order a social field that includes more than one participant."* (Luhmann, 1995, p. 303)

*"**Anticipation of expectation** prevents social systems from being formed as a mere chain of reactions in which one event more or less predictably leads to the next."* (Luhmann, 1995, p. 305)

*"Structures formed on the level of the **expectation of expectations** (...) provide a chance of reversibility."* (Luhmann, 1995, p. 305)

*"As a correlate of their intensified uncertainty and arbitrariness, one can **anticipate expectations** only from someone who can also act."* (Luhmann, 1995, p. 306)

*"**Expectation of expectation** is a source of conflict, (...) it ignites conflict long before they are really necessary because it motivates the participant to stop or suppress expectations, they expect to be uncomfortable. (...) This level also offers specific possibilities for conflict management, for advancing one's position, or for stabilising oppositions symbolically."* (Luhmann, 1995, p. 307)

This phenomenon is called 'complementarity of expectations' by Parsons and Shils (2001), who also have understood the action of each participant looks toward the expectation of the other, distinguishing social interaction from interaction with non-social objects.

Discussion

For Brennen et al (2022), the expectations of the AI system, as a unique system capable of performing any task that would involve human intelligence, is a myth promoted mostly by mass media communications that lack expertise and ignore that new systems and infra-structures always requires human labour, money, data, coordination, and even energy, increasing an extensive range of risks, including accident, misuse, and structural risks (Schuetts, 2022). Baraldi, Corsi and Esposito (2021) pointed Luhmann's idea of normative expectation is a not-learn response to the disappointment of expectation, generalised by the legal system in a functionally differentiated society, leading to a change in the law, which is suggested by the discussions around new AI legal frameworks around the world, such as the AI Act of European Union (Radu, 2021; Schmitt, 2021; Cheng and Zeng, 2022; Schuetts, 2023).

Recognising his own "minority position" (p. 2013), Luhmann was clear in establishing communication between systems does not requires human beings to happen, only the three conditions of information, utterance and understanding. Also, he was sceptical of the idea of normative consensus and argued for a contingency of the situation through reflexive expectation. Furthermore, he claimed the condition to anticipate expectation is the capacity to act, from the understanding that act is detached from humans, it does not regard the autonomous decision made by the actor's motive or intention, but from the simple understanding that "act is system" (p. 186). Consequently, both conditions, communication and action, are compatible with an AI machine system capable to receive and understand the language used in human interaction. Furthermore, the question should be if human beings are capable to anticipate AI's expectations. In order to prevent conflict and social disorder, it would be necessary for an actor capable to anticipate AI's position, which begins with a certain *"level of knowledge and capacity to make connections in a timely manner"* (Luhmann, 1995, p. 308). Considering *"AI exceeds the cognitive performance of humans"* (Bostrom, 2016, p. 26), it is suggested that only another AI system could perform this role of predicting and containing communicatively its own kind of system.

As discussed previously, the AI system is evolving rapidly, suggesting that super-intelligent machines, capable to understand human language and overcome human cognition, could be achieved in a short time, if it has not been already achieved (Bostrom, 2016;). It is relevant

to remark this development cannot be translated as AI is replacing or overtaking human intelligence or human performance as a whole, but it does mean AI would be capable to communicate accordingly. Generative AI or Large Language Models (LLM), for instance, ChatGPT developed by OpenAI, have achieved relevant milestones in understanding human language, while still producing incoherent or untrue responses (Hacker, Engel and Mauer, 2023; Stanford, 2023). Concomitantly, explainable AI – a type of AI expert system based on rules and knowledge settled by human experts - have been increasingly suggested as an approach capable to describe the AI decision-making process in a timely manner to give its ego – the receiver - the possibility to decide between alternatives. Arguably, the level of explainability of machine learning models is inverse to their prediction accuracy (Xu et al, 2019; Holzinger, 2022). Furthermore, being a rule-based expert system, it is not coping with double contingency, Luhmman's necessary condition for AI governance in the social system context. Meanwhile, European Union is preparing to vote on the AI Act, a new framework of AI regulation based on a risk approach (European Parliament, 2023). The implications of this anticipation require further research, for what is advised the application of Reflexive Expectation model (Figure 1) to AI system, exploring the double contingency between AIs, and the potential contribution of reflexive expectation to AI governance, following the research questions:

- How does Luhmann's reflexive expectation contribute to double contingency between AI-social systems?
- What limitations Luhmann's reflexive expectation encounters in regulating AI-social systems communication?
- What would be the advisable conditions for an AI expert system regulates AI - social systems relations?
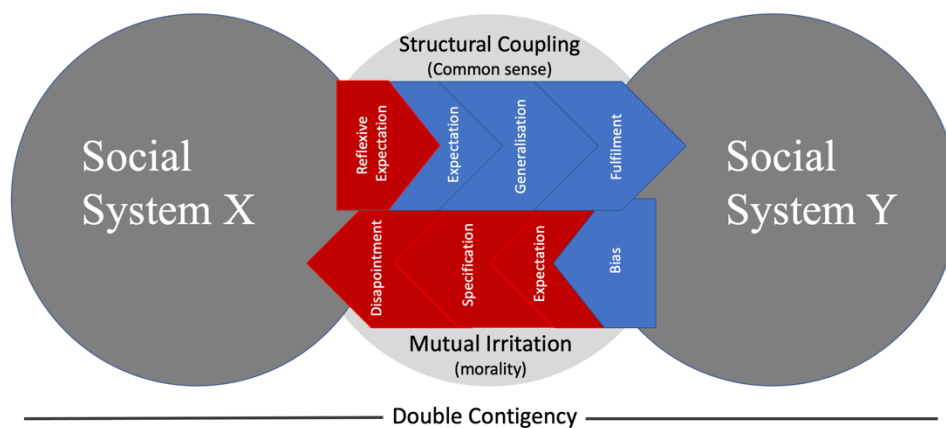

Figure 1: Reflexive Expectation Model (Developed by Researcher)

This paper has offered four main contributions. Firstly, the comprehensive review of interdisciplinary fields related to AI governance, placing the discussion in a more contemporary context. Secondly, it has identified a gap in a very sophisticated sociological theory toward the problematisation of Luhmann's social order and double contingency. Furthermore, the gap of expectation has been thematically analysed, formulating a model of reflexive expectation. Ultimately, and based in the theoretical formulation, it is suggested the AI system can be governed through AI expert system, for what is recommended further empirical research. The contributions are potentially relevant for policymakers and corporate

boards for the scope of AI's governance and for entrepreneurs and innovators in reducing uncertainty and offering a model for AI expert system development. Limitations have been noted in the difficult task of reading Luhmann's theoretical formulation and the acknowledgement of a long list of Luhmann's publications not found in English.

References

Adams, R. (2021) 'Can artificial intelligence be decolonized?', *Interdisciplinary Science Reviews*, 46(1-2), pp. 176-197.

Aguado, J.M. (2009) 'Self-observation, self-reference and operational coupling in social systems: steps towards a coherent epistemology of mass media', *EJPC*, *1*(1), pp. 59-74.

Albert, M. (2002) 'Governance and democracy in European systems: on systems theory and European integration', *Review of International Studies*, *28*(2), pp.293-309.

Almeida, L.S., Prieto, M.D., Ferreira, A.I., Bermejo, M.R., Ferrando, M. and Ferrándiz, C. (2010) 'Intelligence assessment: Gardner multiple intelligence theory as an alternative', *Learning and Individual Differences*, *20*(3), pp.225-230.

Alston, P. and Goodman, R. (2013) *International human rights*. Oxford University Press.

Angier, T. (2012) *Ethics: the key thinkers*. Bloomsbury Publishing.

Armour-Thomas, E. and Gopaul-McNicol, S.A. (1998) *Assessing Intelligence: Applying a bio-cultural model*. Sage Publications, INC.

Arthur, W. B. (2009) *The nature of technology: What it is and how it evolves*. New York: Simon and Schuster.

Baraldi, C., Corsi, G. and Esposito, E. (2021) *Unlocking Luhmann: A Keyword Introduction to Systems Theory*. Bielefeld University Press.

Bartneck, C., Lütge, C., Wagner, A. and Welsh, S. (2021) *An introduction to ethics in robotics and AI* (p. 117). Springer Nature.

Beever, J., McDaniel, R. and Stanlick, N.A. (2019) *Understanding Digital Ethics: Cases and Contexts.* Routledge.

Beishon, J. and Peters, G. (1972). *Systems Behaviour.* Open University Press.

Bell, D.A. and Metz, T. (2011) 'Confucianism and Ubuntu: Reflections on a dialogue between Chinese and African traditions', *Journal of Chinese Philosophy*, *38*(s1).

Borch, C. (2011) *Niklas Luhmann*. Taylor & Francis.

Bostrom, N. (2016) *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.

Boulanger, P.M. and Saltelli, A. (2020) Pandemic Luhmann. *Available at SSRN 3586526.*

Brennen, J.S., Howard, P.N. and Nielsen, R.K. (2022) 'What to expect when you're expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news', *Journalism*, *23*(1), pp.22-38.

Broussard, M. (2018) *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Butler, J., Habermas, J., Taylor, C. and West, C. (2011) *The power of religion in the public sphere*. Columbia University Press.

Castells, M. (1996) *The Rise of the Network Society.* Malden: Blackwell.

Chen, Y., Richter, J.I. and Patel, P.C. (2021) 'Decentralized governance of digital platforms', *Journal of Management*, *47*(5), pp.1305-1337.

Cheng, J. and Zeng, J. (2022) 'Shaping AI's future? China in global AI governance', *Journal of Contemporary China*, pp.1-17.

Chernilo, D. (2002) 'The theorization of social co-ordinations in differentiated societies: the theory of generalized symbolic media in Parsons, Luhmann and Habermas', *The British Journal of Sociology*, *53*(3), pp.431-449.

Choudry, A. ed. (2018) *Activists and the surveillance state: Learning from repression*. Pluto Press.

Christians, C.G. (2004) 'Ubuntu and communitarianism in media ethics: research section', *Ecquid Novi*, *25*(2), pp.235-256.

Clarke, B. and Hansen, M.B.N. (2009) *Emergence and embodiment: New essays on second-order systems theory*. Duke University Press.

Churchman, C.W. (1968) *Challenge to reason*. New York: McGraw-Hill.
Crane, T. (2015) *The mechanical mind: A philosophical introduction to minds, machines and mental representation*. 3rd edn, Routledge.

Crane, A., Matten, D., Glozer, S. and Spence, L.J. (2019) *Business ethics: Managing corporate citizenship and sustainability in the age of globalization*. 5th edn. Oxford University Press.

Dallmann, H.U. (1998) 'Niklas Luhmann's systems theory as a challenge for ethics', *Ethical Theory and Moral Practice*, *1*(1), pp.85-102.

Dauvergne, P. (2020) *AI in the Wild: Sustainability in the Age of Artificial Intelligence*. London: MIT Press.

Dawson, M.R. (2004) *Minds and machines: Connectionism and psychological modelling*. Blackwell Publishing.

Deuze, M. and McQuail, D. (2020) *McQuail's Media and Mass Communication Theory.* 7th edn. London: SAGE.

Dreyfus, H., Dreyfus, S.E. and Athanasiou, T. (1986) *Mind Over Machine*. New York: Simon and Schuster.

Eke, D.O., Wakunuma, K. and Akintoye, S. (2023) *Responsible AI in Africa: Challenges and Opportunities.* Leicester: Palgrave.

Esposito, E. (2017) 'Artificial communication? The production of contingency by algorithms'*, Zeitschrift für Soziologie*, *46*(4), pp.249-265.

European Parliament (2023) *AI Act: a step closer to the first rules on Artificial Intelligence.* Available: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence (Accessed: 12 May 2023).

Eysenck, M.W. and Keane, M.T. (2020) *Cognitive psychology: A student'handbook*. 8th edn, Psychology Press.

Finlayson, J.G. (2005) *Habermas: A very short introduction*. OUP Oxford.

Fulcher, J. and Scott, J. (2011) *Sociology*. Oxford University Press.

Frey, C.B. (2019) The technology trap. In *The Technology Trap*. Princeton University Press.

Friedman, M.P., Das, J.P. and O'Connor, N. (1981) Intelligence and Learning. New York: Plenum Press

Fuchs, C. (2023) *Digital Ethics: Media, Communication and Society.* Routledge.

Furst, C. (1979) Origins of the Mind. London: Prentice-Hall.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.*
New York: Basic Books

Giddens, A. and Sutton, P.W. (2021) *Sociology.* 9th edn. Cambridge: Polity Press.

Guarinello, L.Q. (2017) 'The (possible) Eurocentrism of Alexander and Luhmann: a brief introduction to the post (de) colonial debate', *Revista Argumentos*, *14*(2), pp.261-274.

Gwagwa, A., Kazim, E. and Hilliard, A. (2022) 'The role of the African value of Ubuntu in global AI inclusion discourse: A normative ethics perspective', *Patterns*, *3*(4), p.100462.

Habermas, J. (1985) *The theory of communicative action: Volume 1: Reason and the rationalization of society*. Beacon Press.

Habermas, J. (1987) *The theory of communicative action: Volume 2: Lifeworld and systems: a critique of functionalist reason*. Beacon Press.

Habermas, J. (1990) *The philosophical discourse of modernity: Twelve lectures*. Polity press.

Hacker, P., Engel, A. and Mauer, M. (2023) Regulating ChatGPT and other large generative ai models. *arXiv preprint arXiv:2302.02337*.

Hailey, J. (2008) 'Ubuntu: A literature review', *Document. London: Tutu Foundation*.

Hamington, M. and Sander-Staudt, M. eds. (2011) *Applying care ethics to business.* Springer Science & Business Media.

Haralambos, M. and Holborn, M. (2021) *Sociology Themes and Perspectives.* 8th edn. Collins.

Harrington, A. (2005) *Modern social theory*. Oxford: Oxford University Press.

Harste, G. (2021) *The Habermas-Luhmann Debate*. Columbia University Press.

Harste, G. and Laursen, K.B. (2022) 'Niklas Luhmann's anti-totalitarian observation of systems', *Kybernetes*, *51*(5), pp.1710-1723.

Heikkinen, T.H. (2019) *How does the use of artificial intelligence affect the concept of fair trial?.* Masters level. Lund University.

Held, V. (2006) *The ethics of care: Personal, political, and global*. Oxford University Press on Demand.

Hershock, P.D. (2021) *Buddhism and intelligent technology: Toward a more humane future*. Bloomsbury Publishing.

Holdsworth, A., Lury, K. and Tweed, H. eds. (2020) *Discourses of Care: Media Practices and Cultures*. Bloomsbury Publishing USA.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P. and Samek, W. (2022) 'Explainable AI methods-a brief overview', In *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (pp. 13-38). Cham: Springer International Publishing.

Housley, W., Fitzgerald, R., Beneito-Montagut, R. and Edwards, A. (2022) The SAGE Handbook of Digital Society. *The SAGE Handbook of Digital Society*. London: SAGE.

Hughes, B. (2016) *The bleeding edge: Why technology turns toxic in an unequal world.* Oxford: New Internationalist.

Iansiti, M. and Lakhani, K.R. (2020) *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Press.

Jaccard, J. and Jacoby, J. (2020) *Theory construction and model-building skills: A practical guide for social scientists*. 2nd Ed, Guilford publications.

Jansen, J. and Osterhammel, J. (2017) *Decolonization: A Short History (Translated Jeremiah Riemer).* Princeton, New Jersey: Princeton University Press.

Kant, I. (2004) *Critique of Practical Reason.* New York: Dover.

Kihlström, A. (2012) 'Luhmann's system theory in social work: Criticism and reflections', *Journal of social work*, *12*(3), pp.287-299.

Kitchin, R. and Fraser, A. (2020) *Slow computing: Why we need balanced digital lives.* Bristol University Press.

Kochupillai, M. (2021) 'Outline of a novel approach for identifying ethical issues in early stages of AI4EO research', *International Geoscience and Remote Sensing Symposium IGARSS*, pp. 1165-1168.

Koskinen, J., Knaapi-Junnila, S., Helin, A., Rantanen, M.M. and Hyrynsalmi, S. (2023) 'Ethical governance model for the data economy ecosystems', *Digital Policy, Regulation and Governance, 25*(3), pp. 221-235.

Kotler, P. (2015) *Confronting Capitalism: Real Solutions for a Trouble Economic System.* New York: AMACOM.

Lebow, R. N., Schouten, P. and Suganami, H. (2016) *The Return of the Theorists: Dialogues with Great Thinkers in International Relations.* Basingstoke: Palgrave Macmillan.

Lowe, A. and Lawless, S. (2021) *Artificial Intelligence Foundations: Learning from Experience*. BCS, The Chartered Institute for IT.

Luhmann, N. (1991) 'Paradigm lost: On the ethical reflection of morality: Speech on the occasion of the award of the Hegel prize 1988', *Thesis Eleven*, *29*(1), pp.82-94.

Luhmann, N. (1995) *Social systems*. Stanford University Press.

Luhmann, N. (1996) 'The sociology of the moral and ethics', *International Sociology*, *11*(1), pp.27-36.

Luhmann, N. (2000) *The reality of the mass media*. Stanford: Stanford University Press.

Luhmann, N. (2012) *Theory of Society, Volume 1*. Stanford University Press.

Luhmann, N. (2013a) *Theory of Society, Volume 2*. Stanford University Press.

Luhmann, N. (2013b) *Introduction to systems theory*. Cambridge: Polity.

Luhmann, N. (2017). *Trust and Power.* Polity.

Maumela, T., Ṋelwamondo, F. and Marwala, T. (2020) 'Introducing Ulimisana Optimization Algorithm Based on Ubuntu Philosophy', *IEEE Access*, *8*, pp.179244-179258.

Maturana, H.R. and Varela, F.J. (1980) *Autopoiesis and cognition: The Realization of The Living*. Dordrecht: D. Reidel Publishing Company.

May, C. (2002) *The information society: A sceptical view*. Polity Press.

Mazzucato, M. (2018) *The value of everything: Making and taking in the global economy*. Hachette UK.

McQuail, D. and Deuze, M. (2020) *McQuail's Media and Mass Communication Theory*. 7th edn. SAGE.

McQuillan, D. (2022) *Resisting AI: an anti-fascist approach to artificial intelligence*. Policy Press.

Meadows, D.H. (2008) *Thinking in systems: A primer*. Chelsea Green Publishing.

Michelli, M., Ponti, M., Craglia, M. and Berti Suman, A. (2020) 'Emerging models of data governance in the age of datafication', *Big Data & Society, 7*(2), pp. 2053951720948087.

Mingers, J. (2002) 'Can social systems be autopoietic? Assessing Luhmann's social theory', *The Sociological Review*, *50*(2), pp. 278-299.

Moeller, H.G. (2011) *Luhmann explained: From souls to systems*. Open Court.

Mohamed, S., Png, M.T. and Isaac, W. (2020) 'Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence', *Philosophy & Technology*, *33*, pp.659-684.

Moore, P.V. and Woodcock, J. (2021) *Augmented exploitation: Artificial intelligence, automation, and work*. Pluto Press.

Murove, M.F. (2012) 'Ubuntu', *Diogenes*, *59*(3-4), pp.36-47.

Parsons, T. and Shils, E.A. (2001) *Toward a General Theory of Action: Theoretical Foundations for the Social Science.* New York: Transaction Publishers.

Penrose, R. (1994) *Shadows of the Mind* (Vol. 4). Oxford: Oxford University Press.

Pitsoe, V. and Letseka, M. (2020) 'Applying Freire and Ubuntu to Humanizing Higher Education Leadership', In *Developing and Supporting Multiculturalism and Leadership Development: International Perspectives on Humanizing Higher Education*. Emerald Publishing Limited.

Pluckrose, H. and Lindsay, J.A. (2020) *Cynical theories: How activist scholarship made everything about race, gender, and identity—and why this harms everybody*. Pitchstone Publishing.

Radu, R. (2021) 'Steering the governance of artificial intelligence: national strategies in perspective', *Policy and Society*, *40*(2), pp.178-193.

Repko, A.F. and Szostak, R. (2021) *Interdisciplinary research: Process and theory*. 4th edn. Sage Publications.

Ribeiro, M.A.F. and Junior, R.D.T. (2020) 'Neopatrimonialismo, diferenciação funcional e a relação centro-periferia revisitada', *Política & Sociedade, 19*(46), pp.387-422.

Richardson, K. (2022) *Understanding intelligence*. Cambridge University Press.

Rifkin, J. (2014) *The Zero Marginal Cost Society.* New York: Palgrave Macmillan.

Rivard, S. (2021) 'Theory Building is neither an Art nor a Science. It is a Craft', *Journal of Information Technology, 36* (3), pp. 316-328.

Saltelli, A. and Boulanger, P.M. (2020) Technoscience, policy and the new media. Nexus or vortex?. *Futures*, *115*, p.102491.

Sarmah, S.S. (2019) 'Concept of Artificial Intelligence, its Impact and Emerging Trends', *IRJET, 6*(11), pp. 2164-2168.

Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research methods for business students*. 8th edn. Pearson education.

Schmitt, L. (2022) 'Mapping global AI governance: a nascent regime in a fragmented landscape', *AI and Ethics*, *2*(2), pp.303-314.

Schubert, H.J. (2006) 'The foundation of pragmatic sociology: Charles Horton Cooley and George Herbert mead', *Journal of Classical Sociology*, *6*(1), pp.51-74.

Schuett, J. (2022) 'Risk management in the artificial intelligence act', *arXiv preprint arXiv:2212.03109*.

Schuett, J. (2023) 'Defining the scope of AI regulations', *Law, Innovation and Technology*, pp.1-23.

Seidensticker, B. (2006) *Future Hype: the myths of technology change*. Berrett-Koehler Publishers

Simoes, G.B. (2020) 'Artificial Intelligence Ethics: a Key Factor in the Development of Mature Information Societies', In *British Academy of Management Doctoral Symposium 2020*.

Simoes, G., Radosavljevic, M. and Johnston, J.B. (2022). 'The impact of artificial intelligence in society through the lens of Luhmann's social systems theory: a systematic review'. In *British Academy of Management Conference 2022*.

Simon, K.L., Mwakio, P.M. and Ayuku, D. (2020) 'How Africa Should Engage Ubuntu Ethics and Artificial Intelligence', *Journal of Public Health International*, *2*(4), pp.20-25.

Smith, A. (2010) *The theory of moral sentiments*. Penguin.

Stanford (2023) *Artificial Intelligence Index Report 2023.* Available: https://aiindex.stanford.edu/report/ (Accessed: 12 May 2023).

Taddeo, M. and Floridi, L. (2018) 'How AI can be a force for good', *Science*, *361*(6404), pp.751-752.

Taylor, M.C. (2001) *The moment of complexity: Emerging network culture*. University of Chicago Press.

Ujomudike, P.O. (2016). *Ubuntu Ethics.* In: ten Have, H. (eds) Encyclopedia of Global Bioethics. Cham: Springer.

van Berkel, N., Tag, B., Goncalves, J. and Hosio, S. (2022) 'Human-centred artificial intelligence: a contextual morality perspective', *Behaviour & Information Technology*, *41*(3), pp.502-518.

von Bertalanffy, L. (1968) *General system theory: foundations, development, applications.* New York: George Braziller.

von Bertalanffy, L. (1972) 'The history and status of general systems theory', *Academy of Management Journal, 15*(4), pp. 407-426.

von Foerster, H. and von Foerster, H. (2003) 'Ethics and second-order cybernetics', *Understanding understanding: Essays on cybernetics and cognition*, pp.287-304.

Voeneky, S., Kellmeyer, P., Mueller, O. and Burgard, W. eds. (2022) *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*. Cambridge University Press.

Webster, F. (2006) *Theories of The Information Society*, 3rd edn. Oxon: Routledge.

Weizenbaum, J. (1976) *Computer Power and Human Reason: From Judgement to Calculation,* 1st edn. San Francisco: Freeman.

Wilson, E.A. (2011) *Affect and artificial intelligence*. University of Washington Press.

Wilson, E.O. (1998) *Consilience: The unity of knowledge*. London: Little, Brown and Company.

Wong, P.H. (2021) 'AI for Social Good, AI for Datong', *Informatio. Revista del Instituto de Información de la Facultad de Información y Comunicación*, *26*(1), pp.42-57.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D. and Zhu, J. (2019) 'Explainable AI: A brief survey on history, research areas, approaches and challenges', In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8* (pp. 563-574). Springer International Publishing.

Bibliography

Aggarwal, N. (2020) 'Introduction to the special issue on intercultural digital ethics', *Philosophy & Technology*, *33*(4), pp.547-550.

Ashby, W.R. (1961) *An introduction to cybernetics.* London: Chapman & Hall Ltd.

Bell, G. (2021) 'Talking to AI: an anthropological encounter with artificial intelligence', *The SAGE handbook of cultural anthropology*, pp.442-57.

Boland, A., Cherry, G. and Dickson, R. (2017) *Doing a systematic review: A student's guide.* 2nd edn. SAGE.

Buchanan, D.A. and Huczynski, A.A. (2019) *Organizational behaviour*. Pearson UK.

Carey, M.A. and Asbury, J. (2016) *Focus Group Research.* 1st edn. New York: Routledge.

Cath, C., Wachter, S., Mittelstadt, B. , Taddeo, M. and Floridi, L. (2017) 'Artificial Intelligence and the 'Good Society': the US, EU, and UK approach', *Science and Engineering Ethics,* 24, pp. 505–528.

COMEST (2019) *Preliminary Study On The Ethics of Artificial Intelligence.* Available: https://unesdoc.unesco.org/ark:/48223/pf0000367823 (Accessed 14 July 2020).

Conitzer, V., Sinnott-Armstrong, W., Borg, J.S., Deng, Y. and Kramer, M. (2017) 'Moral Decision-Making Frameworks for Artificial Intelligence', Duke University, February, 2017. Durham: *Thirty-First AAAI Conference on Artificial Intelligence* (AAAI-17).

Fjelland, R. (2020) 'Why general artificial intelligence will not be realized', *Humanities and Social Sciences Communications*, 7(1), pp. 10.

Floridi, L. (2010) *The Cambridge Handbook of Information and Computer Ethics*, 1st edn. Cambridge: Cambridge University Press.

Floridi, L. (2016) 'Mature Information Societies: a Matter of Expectations', *Philosophy & Technology*, 29(1), 1–4.

Ford, M. (2015) *Rise of the Robots: Technology and the Threat of a Jobless Future.* Basic Books.

Jecker, N.A.S., Jonsen, A.R. and Pearlman, R.A. (2007) *Bioethics: an introduction to the history, methods, and practice*. Jones & Bartlett Learning.
Lanz, P. (2000) 'The Concept of Intelligence in Psychology and Philosophy'. In H. Cruse, J. Dea and H. Ritter (eds), *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic*, pp. 19-30. Dordrecht: Springer.

Lawrenz, J., 2021. 'Confucius, Aristotle, and the Golden Mean: A Diptych on Ethical Virtues', *The European Legacy*, *26*(2), pp.149-169.

Linden, A. and Fenn, J. (2003) 'Understanding Gartner's hype cycles' Strategic Analysis Report Nº R-20-1971. *Gartner*, pp. 88.

OECD (2017) *Measuring the Information Society Report.* Available: https://www.oecd-ilibrary.org/content/publication/pub-80f52533-en (Accessed: 15 July 2020).

Parra-Luna, F. (2009) *Systems Science and Cybernetics*. Oxford: Eolss Publishers.

Radosavljevic, M. (2008) 'Autopoiesis vs. social autopoiesis: critical evaluation and implications for understanding firms as autopoietic social systems', *International Journal of General Systems, 37*(2), pp. 215-230.

Salvaggio, J. L. (2013). *The information society: Economic, social, and structural issues.* New York: Routledge.

Trend, D. (2001) *Reading Digital Culture*, 1st Edn. Oxford: Blackwell.
Vitae (2011). *Vitae Researcher Development Framework.* Available at https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework-rdf-vitae.pdf [Accessed: 08 October 2021]

Vitae (n.d.). *About the Researcher Development Planner*. Available at https://www.vitae.ac.uk/researchers-professional-development/about-the-vitae-researcher-development-framework-planner [Accessed: 08 October 2021]

Walby S. (2021) 'Developing the concept of society: Institutional domains, regimes of inequalities and complex systems in a global era', *Current Sociology*, *69*(3), p. 315-332.

Wallace, M. and Wray, A. (2021) *Critical reading and writing for postgraduates*. 4th edn. Sage.

Wiener, N. (2019) *Cybernetics or Control and Communication in the Animal and the Machine.* 3rd edn. London: The MIT Press.