



The Emotional Risk Posed by AI (Artificial Intelligence) in the Workplace

Den emosjonelle risikoen som AI (kunstig intelligens) utgjør på arbeidsplassen

Maria Danielsen

Doktorgradsstipendiat, Institutt for filosofi og førstesemesterstudier, UiT, Norges arktiske universitet.

Danielsen arbeider med en doktoravhandling som bygger på og er en videreutvikling av Martha Nussbaums teori om emosjoner. Med grunnlag i denne emosjonsteorien undersøker avhandlingen tre temaer. Del 1. undersøker mening, del 2. undersøker emosjonell risiko forbundet med å overlate beslutningstaking på arbeidsplassen til kunstig intelligente systemer. Del 3. knytter sammen de to første delene og forklarer hvorfor kunsten er særskilt egnet til å formidle emosjonell risiko knyttet til innføringen av kunstig intelligens på arbeidsplasser og i andre institusjoner i samfunnet. maria.danielsen@uit.no

Sammendrag

Eksistensiell risiko i forbindelse med allestedsnærværende kunstig intelligens er hyppig diskutert og er oftest beskrevet som fare for masseødeleggelse, personvernmissbruk og singulariteten. Målet med denne artikkelen er å utforske et nytt felt når det gjelder eksistensiell risiko og peker på en annen type fare, nemlig emosjonell risiko. Verdier er en hovedkilde til emosjoner og ved å utøve påvirkning og press på våre verdier, kan kunstig intelligente systemer utsette oss for ulike former for emosjonell risiko, som for eksempel tap av omsorg og tap av mening. Artikkelen første del er en kort presentasjon av et studie på en av Tysklands største banker. Denne banken har innført kunstig intelligente systemer for å overta en av bankens viktigste roller som er å ta beslutninger på lånesøknader. Andre del gjør rede for hvordan mennesker aktivt bruker verdier for å fatte beslutninger. Tredje del forklarer sammenhengen mellom verdier og emosjoner. Fjerde del vender tilbake til bankstudien og knytter den opp til del to og tre ved å gi eksempler på hvordan de ansatte forholdt seg til sin egen rolle i banken, og hvordan forholdet til kundene deres endret seg emosjonelt etter det kunstig intelligente systemet ble innført for å ta alle beslutninger på lånesøknader.

Nøkkelord

Kunstig intelligens, beslutningstaking, instrumentelle verdier, myke verdier, emosjoner

Abstract

The existential risk posed by ubiquitous artificial intelligence (AI) is a subject of frequent discussion with descriptions of the prospect of misuse, the fear of mass destruction, and the singularity. In this paper I address an underexplored category of existential risk posed by AI, namely emotional risk. Values are a main source of emotions. By challenging some of our most essential values, AI systems are therefore likely to expose us to emotional risks such as loss of care and loss of meaning. Part one presents a study of a leading bank in Germany where an AI system was implemented to replace humans in decision-making processes. Part two explains why humans actively make use of values to make decisions. Part three shows the connection between values and emotions. Part four relates parts two and three to the bank study by giving concrete examples of how the employees saw their roles as workers, and how the relationships to their customers changed emotionally after the AI system was implemented to make decisions.

Keywords

Artificial intelligence,, decision-making, instrumental values, soft values, emotions

What does it mean to be human in the present day and age? For example, what is it that we really value about ourselves that makes us different from other life forms and machines? What do other people value about us that makes some of them willing to offer us jobs? Whatever our answers are to these questions at any one time, it's clear that the rise of technology must gradually change them. (Tegmark 2017, p. 82)

When considering the risks of AI systems, now becoming ubiquitous in society, we should not focus only on the technology and the capacities of a super intelligent AI, i.e., the prospects of misuse, mass destruction, surveillance, or the singularity. We should also inquire into how AI systems impact us existentially and emotionally by introducing an exclusive focus on instrumental values. Recently published studies of human interaction with AI systems and devices prepare the groundwork for taking on such investigations.¹ For example, how will it affect workers and their clients emotionally when soft values lose their impact in the workplace in favour of instrumental values?²

AI systems are increasingly used for decision-making in organisations and institutions in contemporary society (Berente et al. 2021, p. 1437). Managers, employees, customers, and clients—all those who relate to AI systems in the workplace—do not have insight into what the decisions are based on, and yet these systems have authority to make irreversible decisions that might severely impact our lives, for example by approving or rejecting our loan applications.

In the following I will discuss an in-depth study from one of Germany's largest banks where an AI system was implemented to replace human decision-making. Using concrete examples from this study, I will show how the deployment of AI systems intensifies instrumental values and does away with a whole set of values (soft values). Values can mean many things, but one way to understand values is by distinguishing between soft values and instrumental (hard) values. Examples of instrumental values are numbers, measures, things that can be quantified. I define soft values as relational, moral, and eudemonistic, such as trust, responsibility, and care. While humans can consider and apply both instrumental values *and* soft values when they practice decision-making, AI systems can only apply instrumental values. As scholars have pointed out, AI systems are supercarriers of instrumental rationality because they can only operate through logical and mathematical procedures, typically for purposes of optimization (Lindebaum et al. 2019, p. 248).

Replacing humans with AI systems in the workplace implies that we lose activities where we develop and sharpen soft values. Since AI systems can neither reason by nor apply soft values, soft values lose impact. I will base my argument on Martha Nussbaum's account of values as a main source of emotion to show why replacing humans with AI systems has emotional consequences. Beneath emotions such as "doubt", "guilt", and "love", are soft values (and sometimes also clusters of values) such as "loyalty", "trust", and "compassion".

It is important that we become aware of the relationship between values and emotions in the context of ubiquitous AI technology. If the implementation of AI systems entails that soft values lose impact or disappear, it could critically impact human emotions. For example, feelings of uncertainty and indifference may increase. The feeling of care and meaningfulness in relation to others and to our work may decrease or disappear alto-

-
1. In a special issue of *MIS Quarterly* in 2021, "Managing AI", a group of scholars raises concerns about the growing application of AI in organizations, particularly about the effect AI systems has on workers, employees, and clients. I will discuss questions from an article in this issue (Berente et al, 2021) where I will focus on value change in the transition from human to artificial. I will include an empirical study of human interaction with AI devices and systems. (The first is Mayer, A.-S., et al. 2020. See also Kudina, O. and P.-P. Verbeek 2019.)
 2. In the literature, instrumental values are also called "hard" values. When I cite other scholars, I might use "hard" instead of instrumental.

gether: something we should perceive of as radically uncertain and potentially risky on an existential level.

Structure of paper

I start this paper by considering a recent in-depth study of one of Germany's largest banks (Main Finance), where the AI system "CleverLoan" was deployed to replace humans in decision-making processes. I then include a section showing why values are essential to decision-making and provide a background for the claim that AI systems are supercarriers of instrumental rationality. To investigate why this is emotionally relevant, I will discuss Nussbaum's account of values and their connection to emotions. I then return to the study of Main Finance, which includes interviews with employees showing that while the bank owner's (instrumental) goal of increased profit and decreased default rates succeeded with the implementation of CleverLoan, the employees lost knowledge and banking skills they had acquired in the practice of decision-making. Moreover, they could no longer practice values such as responsibility, trust, transparency, accountability, and care towards their clients. I will follow up with examples of how these changes were expressed emotionally.

Cleverloan #1: An introduction to the case

AI systems are increasingly being deployed in organizations and universities.³ According to Berente et al., more than half of businesses were implementing some form of AI technology in the year 2020, and its application continue to grow at an astonishing speed (Berente et al. 2021 p. 1433–37). One example is the AI system CleverLoan, which was introduced as a tool for decision-making in one of Germany's largest banks.⁴ Before the introduction of CleverLoan, the bank's service desk employees (now reassigned to *loan consultants* or *front desk workers*) were responsible for advising customers and deciding whether to grant a loan, and on what terms. They conceived of decision-making as their main task in the banking business and were trained in the industry through education and experience, skills appreciated by their customers and by themselves. Once CleverLoan was implemented, employees were no longer involved in decision-making processes.

According to the study, managers in the banking industry sought to deploy AI systems because they experienced increased competition from the overall growth of digital platforms. Digitalization enables efficiency and flexibility, which in turn lead to a shift in customer demands. Additionally, the managers saw that implementing AI would decrease the default rates by eliminating human vulnerability and error. After CleverLoan was implemented, the system autonomously decided whether to approve a loan and provided the consultant with a final and irreversible decision. The employees at Main Finance were no longer responsible for deciding which loan applications should go through and which

3. See for example Khrono.no: <https://khrono.no/kraftsenter-for-kunstig-intelligens-har-flere-svakheter/734588> and Lindebaum et al., 2019.

4. The article "Unintended Consequences of Introducing AI systems for Decision Making" displays results from an in-depth study of one of Germany's largest banks, referred to as Main Finance, a bank consisting of 900 institutions with over 9000 branches and 135,000 employees, and a pioneer in applying AI systems (CleverLoan) to substitute human decision-making in the consultation process for granting loans to private customers (Mayer et al, 2020, p. 240). The scholars behind this article refer to Main Finance as a pioneer when it comes to implementing AI systems to replace human judgment, and the study is thus a great opportunity to examine the risks and consequences of substituting human decision-making processes with artificial systems.

should not. They were reassigned to the role of messengers, communicating CleverLoan's decision to the customers.

I will return to this study later in the text and show that when employees lost the practice of decision-making, they experienced their work as less meaningful. To show the connection between decision-making, soft values and emotions, I will now investigate why values are essential in decision-making processes; moreover, I will discuss what soft values are compared to instrumental values, to get an idea of the influence AI systems have on contemporary society.⁵

Decision-making

Decision-making is a complex process of reasoning (rationality), and, as scholars have pointed out, decision-making is the most central task in organisations (Berente et al., 2021). For example, at its core, management is about making decisions, ranging from routine-based to complex and value-laden choices among different alternatives, often involving a great deal of uncertainty. To exercise decision-making entails evaluating situations differently, i.e., applying the right values in each context. If we want to solve a math problem, instrumental values such as numbers and calculation fit better than a soft value such as intimacy. If we must sacrifice something to help someone that needs help, the soft value of generosity fits better than the instrumental values of numbers and calculation. To decide, we have to deliberate between outcomes; we might succeed or fail. We consider different values and evaluate which ones to employ in our decision-making. Sometimes we make errors and mistakes, in which case we can be held responsible, or asked to revise our decisions or revise the values themselves.

Supercarriers of instrumental rationality

AI is fundamentally about machines making decisions autonomously. Current approaches to AI involve predictive models that outperform humans in certain domains. Scholars speak of AI as a process rather than a phenomenon in itself: a frontier of computational advancement that mirrors human intelligence, and whose main task is to address ever more complex decision-making problems (Berente et al. 2021, p. 1435).

According to Berente et al., AI systems may outperform humans by achieving a certain type of high performance and rational action that humans are not capable of. Technological devices are more accurate, predictable, and more capable than humans when it comes to handling data and computation, a form for rationality the authors identify as instrumental, and in contrast to value-based rationality (Berente et al. 2021, p. 1436). Value-based rationality amounts to the notion of substantive rationality, and instrumental rationality amounts to formal rationality, terms I will clarify with the help of Kalberg.

Instrumental values are ascribed to instrumental reasoning; soft values to value-based reasoning/substantive rationality (see footnote for clarification of terms).⁶ Kalberg's

5. If AI systems become able to develop soft values in the future, for example if they acquire "a needy nature" and their own agency (which, I will argue, are conditions for developing soft values), it raises other issues and discussions than those I focus on in this paper.

6. Kalberg's discussion of rationality is based on Max Weber's. Weber and Kalberg use the terms formal (*zweckrational*) and substantive (*wertrational*) rationality. Berente et al. (2021) translate formal to *instrumental* rationality, and substantive to *value-based* rationality. To avoid confusion, I will consistently use the terminology *instrumental* and *substantive* rationality.

account of values and rationality is crucial in a discussion of decision-making, for the question of what values we decide by—the same as asking what form of rationality we reason by—concerns the main differences between humans and machines.

According to Kalberg, substantive rationality directly orders action into patterns. However, it does not do this based on a pure means/end calculation of solutions (i.e., instrumental), but in relation to past, present, or potential value postulates. Value postulates are understood as *entire clusters of values* that vary in comprehensiveness, internal consistency, and content. In brief terms, according to Kalberg, substantive rationality exists as a manifestation of an individual's inherent capacity for value-rational action (Kalberg 1980, p. 1155). Substantive rationality is understood as the form reason should take in relational and moral issues, because it includes the consideration of moral, relational and eudemonistic values (soft values). According to Raz, we do not only make use of substantive reasoning for moral purposes; it is also essential for “understanding” and for “connected knowledge”.

In contrast, instrumental rationality is reasoning according to abstract rules, as something that relates to spheres of life that are rule-governed, and more specifically to universally applied rules, laws and regulations, and decisions arrived at *without regard to persons* (Kalberg 1980, p. 1158). Instrumental rationality is understood as logically and mathematically engineering optimal decisions for specific objectives, as a means/end approach to problem solving (Berente et al. 2021, p. 1436).

Information technology (IT) has long been associated with instrumentality, and in that sense, we can conceive of the development of AI systems as a continuation of a tendency in society. But, as scholars have stated, AI involves a sort of instrumental control that is new, because AI can be more comprehensive and interactive than previous generations of IT (Berente et al. 2021, p. 1436). AI systems perform better than humans for purposes involving calculation and computation, in tasks associated with instrumental reason. Less accounted for are the problems that arise when instrumental rationality interferes with—or comes at the cost of—substantive rationality or reasoning according to soft values.

Georg Ritzer has stated that the main purpose of modern technology deployed in organisations is to illuminate human error and vulnerability for the purposes of “rationality” and control. (Ritzer 2007, p. 52). However, eliminating human error and vulnerability entails that soft values—for example responsibility, trust, care—disappear. As Ritzer pointed out, modern Western technology is a product of instrumental reasoning and is opposed to soft values. Moreover, according to him, the great sources of uncertainty and unpredictability in any rationalizing system are *people*, and the replacement of human technology with nonhuman technology is to minimize uncertainty and vulnerability with the goal of increased control (Ritzer 2007, p. 52).

Instrumental (hard) values, e.g., numbers, quantification, abstract rules, and principles, can successfully be computed into a system, but soft values such as trust, responsibility, and care cannot be transferred to machines. Additionally, as Ritzer states, the notion of “rationality” in the West is culturally shaped to consider reasoning according to instrumental values as more rational than reasoning according to soft values (Ritzer 2007, p. 43). On a societal level, such a conception of rationality can help to justify the neglect or disappearance of soft values through the replacement of humans with machines.

Why is this something to be concerned about? Why should we protect soft values like responsibility, trust, accountability, and care in decision-making processes in the workplace, in large societal institutions and organisations like banks, immigration offices and universities? Why is it a problem that soft values fall away in decision-making processes in our institutions? I will use the following sections to investigate the role of values according

Martha Nussbaum. Her account of values is eudemonistic, which means that humans apply values because we are *concerned with flourishing*.

Values according to Nussbaum

According to Nussbaum in *Upheavals of Thought: The Intelligence of Emotions*, “value” is understood as an appeal to the “actual sense of value of human beings”. She asks whether a life without such and such value would be so impoverished that we would be unwilling to think of it as a human life at all? (Nussbaum 2009, p. 30). The set of values she operates with are based on her list of “human capabilities” through which she articulates our basic needs for survival and for flourishing.⁷ Examples are life, bodily health, bodily integrity, senses, imagination, thought, and emotions.⁸ An underlying premise of human values is that they originate from “needs” that we cannot rid ourselves from without losing something essential to us.⁹ For example, if my chances of providing basic needs such as food and shelter were blocked, I would not be able to sustain life, I would be sick, feel desperate, and in the end, die. And if I could not provide for my psychological needs and need for self-fulfilment, as friendship, trust, ability to learn and to master something, I would be lonely, depressed, lack self-esteem, and I could start thinking that life is not worth living. Nussbaum conceives of values (needs) in relation to freedom and opportunity, and to our ability to fashion a life in accordance with our own view of the deepest and most important part of us. She bases her account of needs and values on Aristotle’s view, where “needs” can be conceived of as where virtue begins (Reader 2017, p. 113).

This marks a difference between human reasoning processes (processes which we actively apply values to) and reasoning processes conducted by machines. A lot more is at stake for us personally when we deliberate between and make use of values than what is the case with machines. In the following I will explain the connection between values and emotion. I will argue that when we lose activities that we need for the shaping and defining of soft values, when this whole set of values falls away in favour of AI systems and an exclusive focus on instrumental values, we are likely to lose important emotions.

Values and the relation to emotions

According to Nussbaum, values are a main source of emotions. She conceives of emotions as thoughts, i.e., *beliefs*, but of a peculiar kind, hence the title of her seminal work: *Upheavals of Thought: The Intelligence of the Emotions*. The distinct thing about the thoughts or beliefs embedded in emotion is that they involve appraisal or evaluation, or put differently, belief about values. We can conceive of an emotion as “a thought of an object combined with thought of the object’s salience or importance” (Nussbaum 2001, p. 23). The *upheaval* (of thought) is the evaluative aspect, something that has a physical, or felt expression. For example, what do I feel if I believe that I am being lied to by a trusted friend? Or what do I

7. According to Nussbaum, the presence or absence of human values are typically understood to be a mark of the presence or absence of human life. Realizing and fulfilling values in a truly human way includes developing and exercising one’s human powers such as social and rational skills, and not merely operating at an animal level (Nussbaum 1999, p. 234).

8. For an extensive account, see Nussbaum (1999).

9. I will argue that her account of human values is built on Aristotle’s account of “necessity”, as four types. The first sense is of being required for life or existence. The second is of being required to achieve a good or avoid an evil. The third is of being coerced against will or nature. And the fourth is of being logically compelled, as in demonstration (Reader 2017).

feel if I am being rejected by a partner, a boss, or friends without getting any explanation as to why this is happening, in other words if am being ghosted? I am likely to experience emotions like anxiety, as being less likable, and a feeling of having been deceived. My emotions express the valence of the values of trust, loyalty, and friendship, self-respect, just to mention a few. In a similar way, I feel *safe* and *comfortable, at ease* if I believe that the values of trust, loyalty, and honesty (transparency) are upheld in my relationships with a partner, friends and professionally at work.

The structure of emotions, as values that on a deeper level are needs combined with evaluative belief, is illustrated in this passage from *Upheavals of Thought*: “(T)he peculiar depth and the potentially terrifying character of the human emotions derives from the especially complicated thoughts that humans are likely to form about their own need for objects, and about their imperfect control over them” (Nussbaum 2001, p. 16). Notice that needs and lack of control go hand in hand. This is the reason why emotions, according to Nussbaum, render us ultimately vulnerable. In sum, for Nussbaum emotions are value judgments: thoughts containing things we value, on a personal level, things which we are not entirely in control of realizing or fulfilling. This is not to say that emotions are concerned exclusively with the individual who has the emotion. Values and emotions such as compassion, empathy, and the appreciation of the emotions of others are part of what flourishing means in her account (Nussbaum 2001, p. 307).

I have highlighted the structure of values and emotions because it marks an important distinction when we compare humans to machines. An AI system does not need anything to survive, it does not care if it sustains life and is not concerned about its own or other peoples’ opportunities, for example about flourishing. For that reason, an AI system does not and cannot apply certain values (soft values, relational values) in reasoning processes and decision-making as we do, either. AI systems do not take risks and are not vulnerable to the extent we are.

We should be more aware of the distinction between soft and instrumental values in the context of AI systems, and consider when we should apply soft values, when we should we apply instrumental values, and how we should relate them. We should question whether AI systems should handle decisions that involve moral and relational issues such as decision-making, given that soft values, i.e., responsibility, accountability, trust, and care, are not transferrable to machines.

I now return to the study entitled “Unintended Consequences of Introducing AI systems for Decision Making” and the discussion of CleverLoan, and I will highlight questions regarding values in the transition from human to artificial. If human values fundamentally cannot be transferred to the algorithms by which AI systems make decisions, it means that these values are lost in the process. I will discuss the changes and consequences of replacing humans with the AI system, as seen from the employee’s perspective, and back this up with concrete examples of the mechanism between values and emotions. The study shows that the conception of values such as responsibility and transparency changes when they are outsourced to a machine, and that soft values fall away in the transition between human and artificial. I will focus on the values of responsibility, transparency, and reliance/trust.

CleverLoan #2: The transition from human to artificial

The study accounts for intended, unintended, positive and negative consequences. As intended, the instrumental value of increasing the bank’s profits succeeded after CleverLoan was implemented (Mayer et al. 2020, p. 246). This result is not the object of our dis-

cussion but demonstrates the success of AI systems and why business owners may want to implement them. However, I shall investigate the positive and negative *unintended* consequences, beginning with a general overview of the employees' experiences.

Unintended Consequences

One employee reports that transferring responsibility to the AI system has made the work easier for her. She will now confidently and easily communicate decisions on whether to grant a loan to the customers. As I mentioned in the beginning of this paper, the employers cannot interfere with the decision made by CleverLoan, neither do they know the underlying algorithms by which CleverLoan makes its decisions. Customers are aware of this and, strange as it sounds, customers now accept rejections more easily. Employers and employees perceive these consequences as positive (Mayer et al. 2020, p. 246–47).

As for the negative consequences, the employees state that the system systematically excludes certain customer groups, and that the AI system is inconsistent regarding interest rates. The front-desk workers do not understand why two seemingly identical clients end up getting rates that may differ, something they report as an increased moral burden. Moreover, employees say that they experience loss of competence and loss of critical thinking. Prior to the introduction of CleverLoan, they valued their ability to reflect on their work, but since the AI system was implemented they no longer understand the banking business (Mayer et al. 2020, p. 247–48). This has led to knowledge outsourcing, and moreover to losses in the relationships between customers and employees.

Responsibility: One result of replacing the service desk worker with CleverLoan is that the employee is no longer involved in decision-making, but must nevertheless communicate and justify the outcome to the customer. In the article, employees state that they do not feel responsibility towards the clients anymore because they can neither intervene in the decision nor alter the outcome. As a result, one employee states that she no longer cares what the decision is—whether a loan is approved or declined—as illustrated in this quote: “In the end, because I . . . and the bank do not take any risk with (CleverLoan), I more or less don't care which loan goes through and which doesn't” (Mayer et al. 2020, p. 249).

The replacement of responsibility was conceived of as one of the positive unintended consequences. Employees describe the communication as easier and quicker after they transferred the responsibility to CleverLoan. For example, in this quote, a staff member says: “For me it is easy because I just have this tool and I don't have to decide on my own. I can delegate the responsibility more or less” (Mayer et al. 2020, p. 246). As the manager of Main Finance put it:

Before you had to be familiar with the loan business, and there weren't that many people who could make reliable decisions on allocating loans. So, with CleverLoan it suddenly became relatively easy. You could put the required data on a query procedure and then the system says yes or no. (Mayer et al. 2020, p. 245)

The value of responsibility contains a cluster of things we value, like the ability to reflect critically on our work and our decisions, which in turn requires other values, as knowledge and expertise, not to mention the social value of fairness towards one's clients. All values mentioned above (responsibility, critical reflection, knowledge, skill, fairness) are lost in the transition. After CleverLoan was deployed, consultants were not required to have any understanding of banking other than pushing buttons. In the interview, one of the managers at Main Finance states that employees have become dependent on the AI system.

[Employees] would now make nonsense [loan decisions] if they had to decide themselves. Another disadvantage is that reference and knowledge are now lost because employees no longer understand the loan business. (Mayer et al. 2020, p. 249)

If no one holds the skills to do follow-up tests to check that the AI system is making reasonable decisions, we become utterly vulnerable towards the machine and can only hope that it serves our interest.

Transparency: In the context of AI, “transparency” takes on different meanings than those we are familiar with, for example openness, and giving reasons and explanations for one’s opinions and decisions. The CleverLoan system is supposed to be transparent, as legally required by European law. If its decision is to reject a loan application, the system automatically generates an explanation for the rejection. However, the explanation does not disclose all reasons for the decision, not even when they are inconsistent. As mentioned, two seemingly identical applicants may be given different interest rates. According to Mayer et al., only a few AI provider managers, i.e., developers and programmers, know the underlying algorithms and the overall statistical model, which they keep strictly confidential and do not share with anyone (Mayer et al. 2020, p. 243–48).

At Main Finance, the underlying algorithm of the AI system is neither transparent nor comprehensible to the loan consultants. Furthermore, the algorithm changes regularly, which makes the decisions difficult to understand. (...) This lack of understanding creates a feeling of uncertainty in their interaction with customers. (Mayer et al. 2020, p. 254)

The relationship between workers and their clients was altered, as the quote illustrates. While it is possible to hold human beings accountable, to ask them questions or to provide reasons, such efforts were futile in the face of CleverLoan. The loan consultants (or front-line workers) could not always explain or defend decisions they did not understand themselves, and experienced a loss of confidence and a feeling of uncertainty. For clients, being rejected without explanation and with no way to negotiate or find out the reasons for the decision, resembles the experience of being “ghosted”.¹⁰

Reliance (trust): Oddly enough, frontline workers perceive that customers more easily accept the system’s decision than they would if the decision were made by the loan consultant, as this quote from one of the employees demonstrates:

It helps arguing, I’d say, in difficult ... loan applications, decision communications, decision-making, as strange as it sounds. You don’t have deep insight into how the tool works, but interestingly, the acceptance of a ‘No’ is often easier for the customer when a machine has decided it for him/her (Mayer et al. 2020, p. 247).

The service desk employees who have been reassigned to a loan consultant role say that they experience less pressure (i.e., less accountability) now that they completely rely on the AI system. “The autonomy of the AI system gives these employees the security and confidence to step into the new role as loan consultants” (Mayer et al. 2020, p. 247). I will not dwell on the difference between trust and reliance, but it is not clear whether customers and employees rely on the system because it is better than humans at computation, or if they trust that

10. “Ghosting—the act of ending a relationship by ceasing communication without explanation—is a type of ostracism that threatens a person’s basic psychological needs for belonging, self-esteem, meaningful existence, and control. The experience of ghosting creates uncertainty within the relationship and may vary based on individual differences in the need for closure, which is the desire to avoid ambiguity” (Leckfor et al. 2023, p. 1).

the system will make the best decision. If they can neither trust nor rely on the AI system, why do they accept its decision? Is it not more likely that employees and customers feel powerless when confronted with an autonomous black box?

The increasing digitalization in all industries has led to a shift in customer demands in the areas of flexibility and efficiency. At Main Finance, a main incentive behind the implementation of the AI system was to increase profits by meeting these demands. Berente et al. point out that these systems are increasingly being implemented across all industries and represent the biggest societal change in modern times, and that the role of managers in the burgeoning societal transformation involving AI can therefore hardly be overstated. As they state it, it is the managers that make all the key decisions about AI, and they must account for challenges of moral and ethical characters when replacing humans with AI systems. These include issues regarding values as privacy, fairness, justice, bias, and a host of other thorny issues (Berente et al. 2021, p. 1434). Managers should not be concerned with instrumental goals such as profit alone; they should also be concerned with the flourishing of their employees and clients. In this sense, the issue confronting managers sums up the main issue with AI systems, namely giving priority to instrumental values over soft values.

The interviews with employees at Main Finance testify that they lost responsibility and that this led them to no longer care about what decisions were made. Furthermore, the algorithms the AI system employs are not transparent, and whether the employees agreed with the system's decisions or not, they could not interfere with them. This is something the employees experienced as an increased moral burden. They lost confidence and experienced feelings of uncertainty in relation to their clients. We also saw that clients easily accepted the system's decision even when it worked against their own interest. This, I proposed, is not really a sign of trust towards the AI system's decisions, that these decisions were correct or treated them fairly, but instead a sign of psychological powerlessness before an authoritative black box acting like "a ghoster". The emotional consequences that I have given examples of here are loss of care, an experience of moral burden—for example being involved in processes where clients are not treated fairly—loss of confidence, uncertainty, and powerlessness.

Concluding considerations

My interpretation of soft values and the relation between values and emotions demonstrates that the most striking difference between humans and machines is that humans are vulnerable towards those things we value and need the most. In this sense, our vulnerability is a main reason we value others, and as I mentioned, vulnerability is a source of virtue.¹¹ AI systems do not need anything: they do not care whether they sustain life, or how they are perceived, and they cannot take any risks. They are invulnerable compared to humans. As I have shown in this paper, the nature of AI systems entails that while they perform better than humans in reasoning according to instrumental values, for example in promoting efficiency and increasing profit, their nature is not susceptible to considering or practicing soft values.

Decision-making is conceived of as the main task in institutions and organisations and is a process of complex reasoning according to values. Humans realize, shape, and define soft values (responsibility, loyalty, trust, care, openness) in relation to others. For humans, such values are seen in relation to flourishing (freedom and opportunity), and as I showed

11. See for example, Reader (2015).

in this paper, such values are a main source of our emotions. By being replaced by AI systems, we lose activities where we shape and define those values: we lose values that promote life and flourishing, and thus we lose the emotions that follow from practicing and fulfilling those values. The biggest issue, however, is that we are letting AI systems take over and handle those values *for* us. The results, as the examples showed, were that staff and their clients started to experience emotions such as insecurity, carelessness/indifference, powerlessness, and loss of meaning. The speed and trajectory of AI systems deployed in institutions and organisations to take over essential tasks indicate that these, and similar emotional experiences, will be widespread in the future. With this, it is my hope that scholars continue to investigate the issue of the emotional risks posed by AI systems and ways to handle them.

Literature

- Aristotle. (1984). Nichomachean Ethics, in *The Complete Works of Aristotle*. Princeton University Press.
- Berente, N., et al. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450.
- Brock, Gillian and David Miller, “Needs in Moral and Political Philosophy”, The Stanford Encyclopedia of Philosophy. (Summer 2019 edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2019/entries/needs/>
- Flanagan, M., Howe, D. C., & Nissenbaum, H. (2008). In M. J. van den Joven & J. Weckert (eds.) *Information Technology and Moral Philosophy* (322–353). Cambridge University Press. pp. 322–353
- Hoffmann, Anna, I. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7): 900–915, <https://doi.org/10.1080/1369118X.2019.1573912>
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., ... & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, 7, 2378023121999581.
- Kalberg, S. (1980). Max Weber's types of rationality: Cornerstones for the analysis of rationalization processes in history. *American Journal of Sociology*, 85(5), 1145–1179.
- Kudina, O. and P.-P. Verbeek. (2019). Ethics from within: Google Glass, the Collingridge dilemma, and the mediated value of privacy. *Science, Technology, & Human Values* 44(2): 291–314.
- Leckfor, C. M., Wood, N. R., Slatcher, R. B., & Hales, A. H. (2023). From close to ghost: Examining the relationship between the need for closure, intentions to ghost, and reactions to being ghosted. *Journal of Social and Personal Relationships*.
- Lindebaum, D., et al. (2019). Insights from “the machine stops” to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, 45(1): 247–263.
- Matzner, T, & Ochs, C. (2019). Privacy. *Internet Policy Review*. 8(4).
- Mayer, A. S., Strich, F., & Fiedler, M. (2020). Unintended Consequences of Introducing AI Systems for Decision Making. *MIS Quarterly Executive*, 19(4). 239–257.
- Miller, T. (2019). But why? Understanding explainable artificial intelligence. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3), 20–25.
- Myrstad, F., & Tjøstheim, I. (2021). Out of Control. How consumers are exploited by the online advertising industry. *Forbrukerrådet*.
- Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press.
- Nussbaum, M. C. (2009). *Therapy of Desire*. Princeton University Press.
- Nussbaum, M.C. (1999). The Capabilities Approach. *International Labour Review*. 139.(3): 227–245.
- Raz, J. (2003). *The Practice of Value*. Clarendon Press.
- Reader, S. (2005). Aristotle on Necessities and Needs. *Royal Institute of Philosophy Supplements*, 57, 113–135. <https://doi.org/10.1017/S1358246100009176>
- Ritzer, G. (1998). The Weberian theory of rationalization and the McDonaldization of contemporary society. *Illuminating Social Life: Classical and Contemporary Theory Revisited*: 37–61.
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Penguin Books.

Toulmin, S. (2001). *Return to Reason*. Harvard University Press

Wolf, S. (2010). *Meaning in Life and Why It Matters*. Princeton University Press.