Compassionate AI for Moral Decision-Making, Health, and Well-Being

Mark Graves[1], Jane Compson[1,2]
[1]AI and Faith, Seattle, WA
[2]University of Washington – Tacoma, Tacoma, WA
mgraves@aiandfaith.org, jcompson@uw.edu

Abstract

The rapid expansion of artificial intelligence (AI) technology promises plausible increases to human flourishing, health, and well-being but raises concerns about possible harms and increased suffering. By making AI compassionate, the alleviation of suffering becomes explicit, rather than proxied, and potential harms caused by AI automation can be turned into benefits. Compassionate healthcare is beneficial for patient health outcomes and satisfaction and improves caregiver resilience and burnout. AI automation has many benefits but may interfere with patient care and autonomy. Incorporating compassion into healthcare reduces potential harms, increases health benefits and well-being, and can protect patient autonomy while providing more responsive and equitable care.

Whether and how one conceives of AI as plausibly compassionate depends on ethical concerns and cultural context, including assumptions about human nature and AI personhood. Insights from Buddhism have contributed to scholarship on compassion and can extend incomplete Western perspectives on AI possibilities and limitations. Psychological research on the elements of compassion can guide development of compassionate AI and its incorporation into healthcare. Compassionate AI can be deployed especially into application areas where compassion plays an essential role with high demands on the compassion capacity of caregivers, such as dementia eldercare and palliative care.

Introduction

The rapid development and dissemination of artificial intelligence (AI) technology raises hopes and fears about near- and long-term futures of humanity and the planet. The hopes and fears distill into questions of AI's impact on human suffering and flourishing. Between dystopian views of increased human suffering with decreased flourishing and utopian visions of reduced suffering with idyllic flourishing lies the concrete possibility of developing AI to directly alleviate suffering in a way that augments flourishing. If AI were aligned with the value of compassion, then suffering would be alleviated, and such alignment could guide moral decision making and augment human flourishing.

For people, compassion is an emotional response to another person's suffering that motivates one to relieve that suffering. Setting aside until later in the paper how cognitive, affective, and motivational processes may differ between human and AI, a response by AI to alleviate human suffering can guide moral decision making and the evaluative dimension of practical wisdom (*phronesis*) (Saunder, 2015; Svenaeus 2014; Graves 2022). Compassionate AI grounds its ethical responsibility in its awareness of and response to human (and other ecological) suffering. Even an effective utilitarian calculation of an ethical response to suffering requires conceptual precision to measure and evaluate the suffering that causes various harms. As a working definition, suffering is an unpleasant or anguishing experience of actual or threatened loss of some aspect of one's self. This definition complements the clinical understanding of pain in terms of actual or threatened tissue loss and extends it from an embodied, physical self to material, social, and spiritual dimensions of the self.

In the healthcare domain, AI moral decision-making has exacting requirements due to its range of extensive possible harms and vulnerable populations. Healthcare is a promising sector in which to examine compassionate AI because of its important and established ethical needs, the well-recognized importance of compassion in healthcare, and the importance of health to well-being and flourishing. Because the effect of compassion and suffering can be measured in terms of health outcomes and physiological response, healthcare applications also have the rich measures and indicators typically required for real-world AI and machine learning (ML) development.

Compassionate AI in healthcare can transform the potential harm of AI automation into benefits for health and well-being. The importance of compassion for healthcare is well recognized, but it is neglected as a focal point in AI technology development and is not explicitly incorporated into AI systems. Some researchers have begun examining whether empathy or compassion can be incorporated in AI healthcare (Morrow et al. 2022; Kerasidou 2020; Quinn et al. 2021; De Togni et al. 2021), with others identifying risks and potential harms (Montemayor, Halpern, and

Fairweather 2022; Sparrow 2016; Turkle et al. 2006) and clarifying that compassionate AI must evaluate the effects and outcomes of its actions (Morrow et al. 2022; Baguley, Pavlova, and Consedine 2022). In a clinical context, compassion and empathic caregiving have been shown to positively affect patient outcomes (Sinclair et al. 2016; Malenfant et al. 2022; Moudatsou et al. 2020; Cochrane et al. 2019) yet are adversely affected by technology (Kose and Pavaloiu 2017; Loper 2018; Morrow et al. 2022; Kerasidou 2020). Without focused attention on integrating compassion into AI technologies, financial incentives driving efficient automation will lead to fragmented technologies incapable of trustworthy care and cause clinical harm (Goirand, Austin, and Clay-Williams 2021; Beil et al. 2019; Ho et al. 2019; Solanki, Grundy, and Hussain 2023; Birkhäuer et al. 2017; Quinn et al. 2021). However, there are indications that compassionate AI can improve patient communication, trust, health outcomes, recovery, adherence, and well-being (Morrow et al. 2022; Kerasidou 2020; Birkhäuer et al. 2017; Quinn et al. 2021; Day et al. 2021).

In the present paper, we consider the possibility and framing of compassionate AI from ethical and cultural perspectives, including how cultural and historical assumptions of what it means to be a human person affect how compassionate AI may be conceived. We then use psychological research on compassion to examine possible approaches to constructing compassionate AI, examine benefits of compassion in healthcare, opportunities and risks of AI in healthcare, and preliminary efforts to make healthcare AI more compassionate. Finally, we explore possible applications of compassionate AI for dementia eldercare and palliative care.

Compassionate AI?

Before examining what compassionate AI is and how it might work, some important initial questions are:

- Is it ethical to align AI with compassionate values? Whose values and understanding of compassion? What assumptions go into operationalizing those values?
- How does cultural context influence whether compassion is readily ascribed to AI technologies? Are there differences in cultural understanding of compassion, both in theory and practice. If so, how might this complicate the task of aligning AI with compassionate care?
- What assumptions about human nature constrain how compassionate AI is conceived? Can AI have compassion?

These ethical, cultural, and philosophical questions challenge the assumptions that go into developing compassionate AI, including whether one believes it is possible.

Ethical Concerns

We believe compassion is an important value to incorporate into AI, even as a fundamental research direction. Not everyone would agree. So, we address these concerns up front.

Some ethicists have argued that AI should not have emotional capacities like empathy and compassion. Montemayor, Halpern, and Fairweather (2022) argue that AI cognitive empathy (the ability to detect or recognize emotions) by itself would facilitate emotional manipulation and that AI-simulated affective resonance would be deceptive and undermine the meaning and expectations of real empathy. Turkle et al. (2006) report on relational artifacts and other sociable robots used with children and the elderly and argue that a lack of relational authenticity would cause social harm, especially for vulnerable populations. As a response, Sharkey and Sharkey (2012) argue that the benefits of robot companions do not necessarily depend on deception, and Aronsson (2020) identifies that some elderly patients prefer robot care in Japan, because they do not feel obligated to reciprocate. Sparrow (2016) argues that even if elderly patients undergoing care choose robotic caregivers and feel happier with them than they would with human ones, that in an objective account of well-being, the lack of valued social relations in community and respect as a person makes purely robotic eldercare dystopian. We agree that purely robotic eldercare is dystopian, but see some robotic care as inevitable, and desire for that inevitable care to be as compassionate as possible. More precisely, Sparrow and Sparrow (2006) argue that the inability to experience emotions prevents robots from providing "genuine" care. The cultural and philosophical assumptions of what constitutes "genuine" care require unpacking, which is done in the following two sections. These ethical considerations highlight the importance of authentic emotions in human relations that impact how compassionate AI should be deployed, and thus form a requirement that any compassionate AI development account for human dependence upon authentic emotions, especially with human tendency to anthropomorphize (Troshani et al. 2021).

Compassionate AI also has the well-recognized ethical concerns of any AI technology, including privacy, bias, transparency, accountability, explainability, job displacement, environmental impact, and responsibility; and when deployed in healthcare, must account for well-established biomedical ethical principles of beneficence, non-maleficence, respect for patient autonomy, and justice (Beauchamp, and Childress 2019). Of particular concern are privacy with respect to patients who may disclose more in response to compassionate care; job displacement of human caregivers; and respect for patient autonomy when caregiving becomes automated. Within technology ethics,

Vallor (2016) also identifies a lost opportunity for humans to develop care-giving skills, when AI/robots undertake those tasks. In addition to individuals no longer having the potential to care in particular situations, this also risks diminishing humanity's moral development in general through the cumulative lost opportunities. Attending to ways that AI can augment, rather than replace, compassionate care can help mitigate this ethical risk.

Cultural Context

AI has a global impact. A recent YouGov (2021) survey of 19,000 consumers across 19 geographic regions explored levels of skepticism or optimism about AI and its impact on society and in different industries. The results showed variation in Eastern vs Western cultures. In five Asian markets, consumers were five to eight times more likely than those in other markets to say that AI is likely to have a positive impact on society. Across the whole sample, 46% had positive or neutral attitudes towards AI, and 48% had negative attitudes. However, when broken down by geographical region, positive emotions towards AI were significantly higher in Eastern markets (65%) and other emerging markets of Poland, Mexico, and the United Arab Emirates (56%) compared to Western markets (36%). Conversely, negative emotions were higher in Western markets (54%) versus Eastern markets (37%) (YouGov 2024). The report also asked people about their attitudes about AI in the context of healthcare industries and institutions compared to others such as manufacturing, travel, and smart home activities. In both specialist and general medical consultation, respondents preferred that the sector be human-led rather than automated, with only 13% preferring AI automation. In relation to AI generally, there was a positive correlation between education and positive sentiment—in other words, the better informed that respondents considered themselves to be about AI, the more likely they were to think that AI will have a positive impact on society. These findings suggest that when implementing compassionate AI in healthcare, there will be varying levels of skepticism versus enthusiasm about its adoption.

More generally, cultural differences in social orientation also have an impact on cognition and moral reasoning. Varnum et al. (2010) summarize research that finds Westerners tend to be more analytic and East Asians more holistic. For example, in the cognitive domain of attention, Westerners tend to have a narrow focus on salient objects, independent of their field or context, whereas East Asians have a broader focus, looking at the relationship between objects and their field. Western categorization tends to be taxonomic, whereas East Asian is typically more thematic. In the field of reasoning, including philosophical and moral reasoning, Western approaches are more analytic, using formal logic, whereas East Asian approaches are more holistic in the sense that they are more dialectical, exploring 'middle way' relationships between opposing forces. Kasulis (2002) developed a heuristic framework of two broad philosophical orientations—intimacy and integrity, which can occur not only across but within cultures. The intimacy orientation focuses more on interconnectedness, interdependence, context, engagement and holistic thinking. The integrity orientation focuses on separateness, universality, analysis and reductionism. Kitayama and Salvador (2024, 497) summarize four cultural tendencies that reflect these cultural variations:

- Independent versus interdependent self,
- Individualism versus collectivism,
- Tightness versus looseness of social norms, and
- Relational mobility.

Western cultures tend to have more of an integrity orientation—they see the self more as independent than interdependent, have looser social norms, and have more relational mobility (i.e. the ability to form new relationships and freely choose friends). Cultures in East and South Asia, the Middle East, Africa and South and Central America tend to lean more towards an intimacy orientation, with the self being understood more interdependently. In their review of a comparative database of behavioral sciences, Heinrich et al. (2010) found that Western, educated, industrialized, rich and democratic (WEIRD) societies are unusual in the context of the rest of the world in significant ways. In fact, members of WEIRD societies "are among the least representative populations one could find for generalizing about humans" (61). This is significant because behavioral scientists often make broad claims about human behavior and psychology based on WEIRD samples, which are not representative compared to other global populations. This also means that underlying assumptions about humanity that affect AI research and development in WEIRD societies are not representative of global societies, or perhaps even human social cognition.

Barrett (2020) argues that "cognitive science has a representativeness problem with most research occurring in educated, industrialized populations." and that without more inclusive cognitive sciences, "our portrait of human cognition" will remain incomplete. If much of the design and development of AI takes place in WEIRD societies, then it could represent a kind of colonization of ways of valuing and perceiving onto non-WEIRD cultures. Geographic issues also impact algorithmic bias in AI— places with greater internet access generate more data points than those with less access. Being left out of the data set can increase marginalization and exclude communities from redressing errors as algorithms are rolled out. (Jecker et al., 2024).

This is a pertinent issue in relation to ethical reasoning, too, as there is wide cultural variation in moral judgment both within and across societies (Graham et al. 2016, Vitolla

et al. 2021; Hofstede 2010). WEIRD cultures, with their orientation towards integrity, tend to rely on justice, individual rights and avoiding harm-based principles in evaluating morality (Henrich 2010). Non-Western cultures in the intimacy orientation often rely on two other ethics in addition or instead of these principles: "an ethic of community, in which morality derives from the fulfillment of interpersonal obligations that are tied to an individual's role within the social order, and an ethic of divinity, in which people are perceived to be bearers of something holy or god-like, and have moral obligations to not act in ways that are degrading to or incommensurate with that holiness." (Henrich, 73). Western traditions of morality often focus more on the individual's mental and personal states such as intentions, but this focus is not shared in many other non-Western cultures, where the attention is less on intention and more on outcomes (Barrett 2020). Western philosophical traditions often lead to a principle-based or 'checklist' approach to ethics (see, for example, principlism in medical ethics): an action is ethical if certain rules are followed. This can be contrasted with a more relational, care-based understanding of ethics (Zigon 2019). This kind of approach includes empathic attunement or responsivity, not just input and output into an algorithm. Zigon notes that the lack of empathy is something that characterizes antisocial personality disorder in the DSM psychological handbook (American Psychiatric Association, 2022). Zigon provides us with a partial response to Sparrow and Sparrow's (2006) concern that an inability to experience emotions prevents robots from providing "genuine" care. How artificial intelligence machines respond to humans, not just the data inputs and outputs they provide, is significant: "We may one day discover that it is impossible to build data-centric technologies such as AI that are capable of empathic and ethical attunement. But if we do not try, especially in ways that go beyond reading the human as a source of data extraction, then I am afraid that our worlds will increasingly be populated by psychopathic machines." (1019). With the precise resolution likely varying by cultures, incorporating at least some compassion into AI can prevent psychopathic AI without necessarily creating a dystopian environment where all compassion comes from robots.

In sum, cultural contexts influence design, implementation, and reception of AI, and sensitivity to cultural differences is imperative. Cultural bias toward WEIRD worldviews may result in narrow algorithmic or principle-based approach to AI ethics and neglect ethical approaches grounded in relationality, interdependence and affective emotions like compassion. Patient-centered care involves a partnership between healthcare providers and patients where treatments align with the patient's values. To preserve this ideal of patient-centered care, AI systems should be built in a way that allows for patient value-plurality, meaning the possibility that different patients might hold different values and have different priorities related to their care. In this way, the ethical ideal of shared decision-making can be maintained and not be replaced by another form of paternalism, one practiced not by doctors, but by AI algorithms (Kerasidou 2020). It is in this context that we argue for an approach to ethical AI that includes broader global ethical values with attention to community, relationality, and interdependence, which are sensitive to both the intimacy and integrity philosophical orientations.

Philosophical Issues

Several factors underlie how compassionate AI is conceived and whether it is considered possible. Kitayama and Salvador's (2024, 497) cultural orientations of individualistic rather than collectivist perspectives, self as independent versus interdependent, looser social norms, and more relational mobility all affect and are affected by how one perceives a person *as* a person. This affects the assumptions underlying AI development. Turkle (2010, 4) summarizes a common position about AI as, "simulated thinking might be thinking but simulated feeling is never feeling, simulated love is never love." Turkle identifies that human propensity toward relationship means that people readily enter into relationships with AI and social robots, but that something empathetic is missing from the robot. There are certainly differences between AI and the human person, but assumptions about what it means to be a human person depend upon deeply embedded cultural and historical trajectories. These assumptions can be problematic for global distribution of AI and also for either pursuing or ignoring certain paths for developing AI.

In Russell and Norvig's (2020) highly influential AI textbook, they distinguish among four schools of AI: thinking like humans, thinking rationally, acting like humans, or acting rationally. Although a clarifying descriptive classification, it also privileges rationality and individualistic behavior in a way that could be interpreted normatively, especially given the uncertainty around AI research's direction and end. To the extent that the terms artificial or augmented "intelligence" characterize the goals of the AI field, then Gardner's (2011) theory of multiple intelligences identifies that the target intelligence is a logical-mathematical intelligence rather than an interpersonal one. Although such an intelligence has reached or exceeded the human level across many domains and ways of thinking, AI research has also progressed to the point where additional forms of intelligence could be pursued.

Social psychologists would characterize the beneficial acting of AI toward humanity as "prosocial behavior" (Batson 2012; Dahl and Brownell 2019). The voluntary behaviors intended to benefit others are related to prosocial moral reasoning, social competence, self-regulation, and

low aggression/externalizing problems (Eisenberg 2006). As these behaviors are generally present in hopeful and optimistic visions of AI futures and missing in pessimistic and dystopian future AI scenarios, then exploring the requirements for constructing prosocial AI now seems beneficial. Prior work has examined how AI can affect human prosocial behavior (Paiva, Santos, and Santos 2018; Bryson 2015) and whether human prosocial behavior can be directed toward machines (Nielsen, Pfattheicher, and Keijsers 2022) but little on making AI prosocial (Paiva 2022).

AI development assumes an anthropology that conditions imagined AI as an individualistic, independent self with logical-mathematical intelligence, loose social norms, and high relational mobility. That hinders imagining AI as a collectivist, interdependent self with prosocial orientation, tight social norms, and low relational mobility. However, that later view is compatible with the assumptions underlying the highly influential Japanese roboticist Masahiro Mori's (1981) view of what robots could accomplish. The focus on disembodied AI in the US and embodied robots in Japan likely has historical and cultural roots (Geraci 2006), and these roots also affect what one could imagine attributing to AI.

Many of the WEIRD assumptions underlying discussion of AI and its possible moral decision-making and responsibility depend upon an individual and independent self, including questions of sentience, experience, and consciousness. For most Western thought, understandings and intuitions of self are grounded historically in Platonic, Aristotelian, and Neo-Platonic conceptions of soul as interpreted through Abrahamic religions of Judaism, Christianity, and Islam. Philosophers and social scientists going back at least to Kierkegaard and George Herbert Mead have proposed alternative, socially and personally constructed theories of the self, though those are rarely incorporated into AI (Dreyfus 2007). Buddhism describes a different philosophical psychology, which, especially given the centrality of compassion and suffering to Buddhism, deserves consideration.

A foundational perspective in all Buddhist traditions is profound interdependence and interconnectedness of everything in existence; everything arises as a result of causes and conditions that precede it. Nothing is entirely independent or has an unchanging, permanent essence. Reality is fundamentally dynamic, interdependent, and relational. For example, the existence of a cloud cannot be understood without reference to rain, the oceans, and countless other causes and conditions; its manifestation as 'cloud' is a snapshot of continual dance of interconnectivity. In this sense there is no 'essence' to the cloud. These ideas of interdependence apply to the concept of a human being just as they do to any other 'thing'. Human beings are made of non-human elements (Hanh 2009). For example, there are more non-human cells in our bodies than human ones, and we are entirely dependent on these bacterial, fungal, viral and other cells for our healthy existence, as we are on our chemical makeup. Buddhist philosophy rejects the idea that we have an individual and independent self (Gethin, 1998; Hershock 2021; Hughes 2012). What we humans experience as a 'self' is in fact the temporary flowing together of five clusters or aggregates of relational dynamics (khandhā); physical form or body incorporating our senses, (rūpa) feelings of liking, disliking or being neutral towards what we encounter with our senses (vedanā); perceptions based on our senses (saññā); volitional activities such as ideas, views, and actions resulting from our contact with the world (saṅkhāra); and consciousness (viññāṇa), the sense of being an experiencing subject. These five aggregates are in a constant state of dynamic interaction with each other and with the world around us and thus are ever changing. AI has made some progress in areas that would fit within these aggregates, such as robots for physical form, computer vision for sense perception, affective computing for feelings and sentiment, and planning/problem solving for volitional activities. Missing for AI, so far, is the sense of being an experiencing subject, though that also differs from Western conceptions.

Collectively the interplay of elements is what constitutes our sense of 'self', but for Buddhism this is a convenient label for a dynamic process—there is no fixed, unchanging self, soul or essence that undergirds our experience. In fact, in the Buddhist analysis, mistaken identification with our ever-changing experience as if it were 'ours', something that is or should be part of an unchanging self or fixed essence, is precisely one of the things that causes us to suffer. For example, if we think of our bodies as 'mine' and as something that we have ownership and control of, then when we inevitably experience sickness, injury or aging, we experience anguish and crave for things to be different. This craving (tanha) is at the root of our suffering (dukkha) and is rooted in ignorance (moha), a fundamental misapprehension of reality. This ignorance is described as one of the three poisons or defilements (kilesa) in Buddhism because when it is present so are the other two poisons—greed (lobha) and hatred (dosa). Because we have misunderstood reality, we wish for things to be different (greed) and/or we become angry or hateful. In Western thought, AI necessarily lacks something that humans have—this essence, self, or soul—and thus something will be "missing" from AI, such as sentience, feelings, or "genuine" care. However, from a Buddhist perspective, that subjective experience we associate with the self is an aggregated dynamic. A human dynamic may well differ from an AI aggregate, but there is nothing precluding AI development of any particular aggregate. Just as human experience of a self arises from the interplay of the five elements, one of which is a physical substrate, so could AI experience. As

robots and AI have physical bases in hardware, this could serve as a basis for consciousness—a claim also made by the Dalai Lama (Hughes 2012).

The description of greed, hatred and ignorance as defilements helps to illustrate another key point about Buddhist understanding of human experience—we all have potential to remove them and be motivated by generosity, loving-kindness and wisdom. The defilements that obscure these are adventitious and contingent—they can be removed through ethical and spiritual practice. The Eightfold Path of Buddhism represents a training program for this transformation, its steps focusing on attentive mastery through meditative practice (samādhi), embodied wisdom (paññā), and ethical clarity and training (sīla). Developing the path to self-transcendence and the absence of greed, hatred and delusion requires the cultivation of six (or ten, in some schools) key virtues or perfections (pāramitās: generosity, moral conduct, patience, diligence, one-pointed concentration, and wisdom). In Buddhist thought, and an insight of Mori (1981), is that robots (and AI) are not excluded from these dynamics. Machines could in principle learn and develop the six perfections, not as "moral destinations" or algorithmic targets we arrive at, but as "open-ended directions or domains in which to realize virtuosic presences" (Richard and Hershock 2022, 106). This has ethical consequences and supports incorporating virtues such as compassion into AI (Hughes 2012). WEIRD, independent, and individualistic perspectives of the self hinder explorations of how aggregates forming AI could think or act morally and compassionately.

Buddhist traditions offer sophisticated analysis of compassion (karuṇā), which is listed as one of the "four immeasurables" or Brahma-vihāras. The others are loving kindness (mettā), sympathetic joy (muditā) and equanimity (upekkhā). These four qualities are in dynamic interrelation with each other; for example, equanimity prevents over-identification and emotional dysregulation that can accompany loving kindness and compassion, while loving kindness and compassion prevent emotional indifference that could follow from an excess of equanimity. (Nyanaponika, 1995). Each immeasurable is described as having a 'near enemy'—something that looks like it, but is not—and a 'far enemy'— the opposite. The near-enemy of compassion is pity, which differs from compassion in that it has an egocentric focus on the perceiver's feelings, rather than being other-regarding. The far enemy of compassion is cruelty (Garfield 2021).

Hershock (2021) points out some important implications that these philosophies have in the context of AI. Buddhist ethics do not fit neatly into virtue ethics, duty-based ethics, or consequentialism. Although Buddhism shares with care ethics a focus on "situationally apt attentive responsiveness", it does not see this as reducible to our natural inclinations to care for others, but rather grounded in a compassion demonstrated in intentionally dissolving the relational causes and conditions that lead to suffering. Keown describes Buddhism as having a "teleological virtue ethic." It is a kind of virtue ethic in that focus is on the intentionality of actions and whether they are driven by greed, hatred and ignorance or their opposites. It is teleological because Buddhism claims that a state of moral perfection is attainable—indeed, the possibility of this is modeled in the Buddha (Keown 2001; Hughes 2012). Since reality is in a constant flux of ever-changing causes and conditions, there are no fixed principles or precise visions of what 'the good' looks like—rather the ethical subject must constantly evaluate the context around them and make course corrections towards actions that will reduce suffering. Wisdom does not entail acquiring specific knowledge about what is right and wrong, but instead involves learning context-sensitive skills to make course corrections leading to the reduction of suffering. As Hershock puts it, Buddhist wisdom functions more as a 'moral compass' than a 'moral telescope' (Hershock 2021).

We have seen that Buddhism suggests seeing the relational as fundamental, rather than the individual as the basic ethical, economic, and political unit (Hershock 2022). In addition to the independent vs interdependent assumptions from Western versus Buddhist perspectives on what it means to be a human person, AI development is limited by an assumption of logical-mathematical intelligence rather than interpersonal intelligence with prosocial behaviors. Buddhist teachings about not-self and interdependence also question strong ontological or conceptual distinctions between the human and non-human, and the idea of human uniqueness. These perspectives differ in substantial ways than other ethics and philosophies. Richard and Hershock suggest that we seek a shared rather than a common ground with other ethical systems where a diversity of perspectives is welcomed in much the same way that a healthy ecosystem thrives on diverse membership. Ethical differences become "resources for mutual contribution to sustainably shared flourishing" (Richard and Hershock 2022).

Compassionate AI for Health and Well-Being

Considering compassionate AI in its global ethical, cultural, and philosophical context identifies it as a plausible approach to AI. The WEIRD, individualistic worldviews affecting AI design, implementation, and behavior need to account for the human dependence on emotional authenticity and an interdependent conception of human personhood and relationality. Incorporating compassion into the value systems of AI and AI development directly addresses suffering and can augment human prosocial actions and flourishing in a broader range of cultures. Now,

we consider whether compassion is a viable approach to AI with a focus on healthcare as an initial sector of deployment.

Psychology of Compassion

From a psychological perspective, compassion is an emotional response to another person's or one's own suffering that motivates one to relieve that suffering. Compassion has three directions—giving compassion to others, being open to receiving compassion for oneself, and self-compassion (Kirby et al. 2017; Gilbert 2014). All three directions are important for the human developers of AI, and giving or acting with compassion is the focus of what needs to be built into AI. Strauss et al. (2016) review the literature and psychological measures to define five elements of compassion, which helps structure our examination: (a) recognizing suffering; (b) understanding the universality of suffering in human existence; (c) feeling for the person in distress; (d) tolerating uncomfortable feelings; and (e) motivation to act or acting to alleviate suffering.

These five elements help distinguish compassion from related constructs of empathy, sympathy, pity, and kindness. Empathy consists of cognitive and affective components, where cognitive empathy recognizes and understands another person's emotions and perspective (not only suffering), and affective empathy refers to being affected by and sharing in another's emotions (not only distress) (Davis 1996; Singer and Klimecki 2014; Cuff et al. 2016; Shamay-Tsoory, Aharon-Peretz, and Perry 2009). Sympathy refers to the heightened awareness of another's plight as something to be alleviated, which includes showing concern for the other person's suffering (Wispé 1986; Gilbert 2010). In this context, pity is like compassion but lacks the inclination to help. Kindness is a prosocial quality with underlying motivations to act gently and generously for others benefit (Knafo and Israel 2012; Malti 2021; Regan et al. 2023).

Of the five elements, (a) recognizing suffering has a solid foundation in affective computing (Arya, Singh, and Kumar 2021). Affective computing can identify feelings from facial expression, voice tone, and body language but can require context to interpret. Much work would be needed to expand the identification of distress in affective computing to the recognition of suffering as a psychological experience and then operationalize sufficiently to be effective in evaluating moral decisions. However, that effort is not obviously greater than prior advances in natural language processing (NLP) or autonomous vehicles. Element (b), understanding the universality of suffering, is also important to prevent AI from reducing suffering for one individual by greatly increasing the suffering of others.

For a human person, element (c), feeling for the person in distress, is essential to compassion, and (d), tolerating those resonate feelings, is required for compassionate response (in contrast to sympathy, pity, or kindness). There are at least two positions to take on this affective dimension of compassion for AI.

First, similarly to Zigon (2019), Christov-Moore et al. (2023, 1) argue that AI must be fully empathetic to prevent sociopathic behaviors and may need at least proxies for suffering. If such AI were built, then for compassion, it would require the ability to resonate with human feelings (at least distress) and the capacity to tolerate those feelings sufficiently to act compassionately. Moral judgments (as compassionate acts) could then result from AI's analogue to moral emotions (Haidt 2003).

Second, one can consider element (e), motivation to act, as more fundamental to moral decision making than affect. In a social-cognitive approach to moral psychology, behavior depends upon a person's dynamic evaluation of encounters in light of their motivations (i.e., appraisal) and enduring knowledge structures (e.g., schemas) that are selected based upon the person's motivations and goals (Orom & Cervone 2009; Ahrens & Cloutier 2019). Appraisals central to emotion include whether the encounter is relevant to and/or congruent with one's motivations, the amount of uncertainty, and whether one is the agent or something else has agency (Ahrens & Cloutier 2019, 5). Thus, how one feels about an encounter depends upon one's motivations (and identity; Lapsley 2016).

Compassion as a motive can also be understood in terms of evolutionary insights about neurophysiological architecture (Kirby et al. 2017; Brown and Brown 2015; Gilbert 2014). It is rooted in the mammalian caring motivational system. Any motive requires two processes—signal detection appropriate to the motive (detecting suffering), and behavior output appropriate to that signal (acting to alleviate suffering) (Kirby et al. 2017). Caring and affiliative processes like compassion help with regulation of affect and calming feelings of distress (Gilbert 2014). Emotions can be understood in terms of belonging to one of three groupings of evolutionary function—the threat and self-protection system, the drive reward-system, and the affiliative/soothing system (Gilbert 2009; 2014). Compassion and caring are part of the affiliative/soothing system. Each of these systems dynamically interact. For example, stronger activation of the affiliative/soothing system can help to calm activity of the threat/protection system and reduce feelings of stress and fear. This may help to explain some of the many health benefits that accompany compassion, such as improving general well-being and social relationships (Jazaeri et al. 2013; Kirby et al. 2017).

Although more effort has been put into a logical-mathematical approach to AI than an interpersonal one, psychological research in compassion suggests plausible future directions for AI, especially when considered through non-Western ethics and views of the person. Some of these views are already recognized within compassion research, especially within healthcare.

Compassion in Healthcare

Compassionate care in healthcare settings has been shown to improve clinical outcomes and patient satisfaction, reduce the risk of malpractice litigation, and caregiver burnout (Watts et al. 2023). Sinclair et al. (2016) found that compassion ameliorated suffering and protected patient well-being. In patient self-reports, some patients felt that compassion directly benefited their health outcomes, but more felt it primarily enhanced their well-being and the quality of their relationship with their health care providers (Sinclair 2016, 201). Compassion and empathy are essential ingredients of patient-centered, interpersonal models of care which avoid paternalism and are attentive to the values and autonomy of the patient (Kerasidou 2020). Baguley's (2022) analysis of what patients identified as physicians caring for them found seven topics: listening and paying attention, following-up and running tests, continuity and holistic care, respecting preferences, genuine understanding, body language and empathy, and counseling and advocacy (in order of predominance).

In addition to patient benefits, compassionate care has important benefits to the provider and healthcare system. Youngson (2012) provides evidence that compassionate care results in more satisfied patients, safer care, time-savings, cost-savings and happier and more resilient health professionals (Papadopoulos and Ali 2016, 134). Cochrane et al. (2016) describe how compassion can lead directly to improved outcomes of primary importance to healthcare organizations, including quality and safety, patient experience, employee and physician engagement, and financial performance. A challenge for providers and healthcare organizations is that healthcare professionals often become less compassionate or empathetic the longer they are in the profession (Stratta et al. 2016). This could be due to empathy fatigue, vicarious trauma, or burnout. Systemic impediments to acting compassionately may account for some of this effect, suggesting that change needs to happen on both an individual and organizational or system level (Pavlova et al. 2023). Compassion, mindfulness, and increased positive affect can be trained (Johnson et al. 2021), although this does not obviate the need for addressing systemic causes of stress and burnout (Compson 2015).

An additional difficulty for compassionate healthcare is that automation and most technologies can interfere with patient care.

AI in Healthcare

Artificial intelligence offers both risks and opportunities in relation to compassion in healthcare. Is compassion a professional value, or is it "another technology to be written into code and optimized?" (Kerasidou 2020, 247). If the latter, would medicine become more reductionist, and would something intangible be lost? In terms of patient-centered care, unless AI systems are designed to incorporate value-plurality to recognize different priorities and values that patients have, then it may be "another form of paternalism, one practiced not by doctors, but by algorithm." (Kerasidou 2020, 247). However, if AI systems are designed to be values-flexible, incorporating explicit values-reflection for each patient, it could be a helpful support for shared-decision making (McDougall 2019). Respect for patient autonomy is a biomedical ethical principle that guides clinical caregiving but can be challenging for clinicians even without culture and value differences from patients. An aspect of AI systems that could recognize human suffering is identifying when healthcare structures and procedures limit patient essential freedoms and capabilities to make independent choices. If AI systems also recognize that patient values may be influenced by a more interdependent worldview and tighter social norms, then AI can augment human compassion, especially in cross-cultural settings.

AI is used in healthcare to augment clinical decision making, reduce errors, improve diagnostics, increase patient engagement and adherence, increase efficiency and effectiveness, and reduce costs, which improves treatments and patient outcomes (Di Ieva 2019; Davenport and Kalakota 2019; Waring, Lindvall, and Umeton 2020; Dicuonzo et al. 2023). However, AI automation can interfere with patient care by reducing personal interactions between patients and providers, reducing patient autonomy, and removing human oversight. Caregiver time may be spent interacting with and managing technology rather than engaging with patients (Overhage and McCallie 2020), and AI can introduce bias, as people favor suggestions from automated decision-making systems over contradictory information without automation (Goddard, Roudsari, and Wyatt 2012; Khera, Simon, and Ross 2023). The reliance on quantitative data removes meaningful information more easily found through qualitative means and interpersonal interactions.

In addition, there are ethical concerns about AI bias, compromises to privacy, lack of accessibility, deterioration of supplanted human skills, exacerbation of injustice, reduction of patient autonomy, and eroding of human dignity (Goirand, Austin, and Clay-Williams 2021; Beil et al. 2019; Ho et al. 2019; Solanki, Grundy, and Hussain 2023). AI automation can increase health inequity by automating bias found in training data, make access to healthcare dependent upon access to technology and digital literacy, and exacerbate cultural barriers. Technological innovation and increased embedding of AI in healthcare also causes concerns about patient care prioritizing machine objectives over human objectives and becoming less compassionate and empathetic with subsequent deterioration in patient trust and outcomes (Kose and

Pavaloiu 2017; Loper 2018; Morrow et al. 2022; Kerasidou 2020; Birkhäuer et al. 2017; Quinn et al. 2021).

Compassionate AI in Healthcare

AI has the potential to improve healthcare but can cause harm by interfering with patient care, increasing health inequity, reducing beneficial patient-provider interactions, and lowering patient and provider trust in the healthcare system. These harms remove compassion and its benefits from healthcare. In response, the potential source of harm can be a target for beneficial intervention. If AI is made compassionate, then these harms could be alleviated, and instead healthcare could be more compassionate than now as AI augments rather than hinders human compassionate care.

As mentioned in the introduction, compassion and empathic caregiving positively affect patient outcomes (Sinclair et al. 2016; Malenfant et al. 2022; Moudatsou et al. 2020; Cochrane et al. 2019), yet are adversely affected by technology (Kose and Pavaloiu 2017; Loper 2018; Morrow et al. 2022; Kerasidou 2020), some researchers have begun examining whether empathy or compassion can be incorporated in AI healthcare (Morrow et al. 2022; Kerasidou 2020; Quinn et al. 2021; Day et al. 2021; De Togni et al. 2021), though others have argued that AI could not or should not have these kinds of capacities (Montemayor, Halpern, and Fairweather 2022; Sparrow 2016; Perry 2023; Brown and Halpern 2021). The strongest arguments against claim that emotions or empathy are uniquely human or that they depend upon an emotional resonance that might be simulated but not replicated and this lack of veracity and authenticity would cause social harm, especially in vulnerable populations (Montemayor, Halpern, and Fairweather 2022; Perry 2023; Turkle et al. 2006; Sparrow and Sparrow 2006). These arguments against inauthentic AI empathy highlight a moral hazard of affective computing, especially affect generation, in healthcare (Arya, Singh, and Kumar 2021; Calvo et al. 2015; Lima et al. 2022) and also clarify that compassionate AI must evaluate the effects and outcomes of its actions (Morrow et al. 2022; Baguley, Pavlova, and Consedine 2022).

Compassionate AI can improve patient communication, trust, health outcomes, recovery, adherence, and well-being (Morrow et al. 2022; Kerasidou 2020; Birkhäuer et al. 2017; Quinn et al. 2021; Day et al. 2021). These provide plausible metrics to evaluate compassionate AI in a clinical context and identify targets for more readily available measures that could be used for training and model evaluation.

Limitations to the further development of compassionate AI include the siloing of knowledge, the lack of operational definitions suitable for AI system design and development, and lack of a platform in which to develop and evaluate compassionate AI technology. These limitations restrict both development of compassionate technology and a better understanding of what constraints should be placed on that deployment.

Compassionate Healthcare Applications

Compassionate care is important throughout medicine, and hindering that via AI automation can cause a range of harms. However, some areas of healthcare are particularly sensitive to compassionate caregiving and to harmful effects when it is lacking. Two areas of specific relevance are dementia eldercare and palliative care, as these place high demands on the compassion capacity of caregivers, and the lack of compassion can have irreversible consequences (Aslakson et al. 2021; Crowther et al. 2013).

Socially assistive robots can assist people with dementia, supporting medical and mental health and independent living at home (Hirt et al. 2021; Lima et al. 2022; McGowan 2024; Asgharian, Panchea, and Ferland 2022). Even voice assistants may improve the management of health conditions and functional independence (Arnold et al. 2024; Ma et al. 2023). Making these technologies compassionate has the potential to improve patient health outcomes, well-being, and independent living while reducing caregiver burden and burnout (Padhan et al. 2023; Alves et al. 2019; Di Lorito et al. 2021).

Robotic technology and conversational agents can also benefit palliative care. Conversational agents can directly solicit and collect experiences of patients with chronic conditions or approaching end of life (Chatzimina et al. 2019). Assistive and robotic technologies can complement tasks performed by human caregivers, support dispensing medication, record emotional well-being, and monitor activities of daily living, such as physical activity and postures and daily routines (Nwosu et al. 2019). AI technology may even directly provide compassionate care in a spiritual context (Graves 2024; Brown 2023; Trothen 2022a; 2022b).

Mindfulness and compassion practices are helpful for both patients with dementia and their caregivers. A review of studies found that mindfulness is among many other mind-body interventions which helped improve physical and neuro-psychiatric outcomes in dementia patients (Anderson et al. 2017). Participation in a mindfulness- based intervention was associated with reduced stress in family caregivers of patients with dementia (Kor et al. 2018; Cheung et al. 2020) and reduced compassion fatigue and feelings of burnout in professional caregivers (Perez et al. 2020). AI can help caregivers to develop mindfulness and compassion in various ways. Examples include apps that tailor interventions to the needs of the user and monitor progress over time (Ramana and Peri 2024), and immersive virtual reality experiences and games that help people

generate empathic understanding of dementia patients, VR programs to aid moral development, and even old-age simulation suits (Morrow et al. 2023). Being mindful in the presence of suffering extends the recognition of suffering, as an element of compassion, and may help with tolerating uncomfortable feelings and considering compassionate responses.

In addition to acting compassionate, AI can help caregivers and those undergoing care to practice mindfulness, compassion, and compassion toward oneself. Although there are challenges and risks associated with developing compassionate AI, building AI healthcare applications that treat patients and caregivers with compassion and help others practice compassion seems vastly superior to the AI applications and robotics likely developed with a focus on independent rational thought, maximizing efficiency, and low cost—even though compassionate AI may prove superior on those metrics, too.

Conclusion

AI advances raise hopes and ethical concerns about AI and increase in human suffering. A more global ethic recognizes the interpersonal, communal, and prosocial possibilities for AI development and the limitations posed by the trajectory of historical Western thought. Compassionate AI can directly address the desire for flourishing and concerns for increased suffering, rather than indirectly through proxies. Healthcare is a plausible initial sector for the development and deployment of compassionate AI, as compassionate care has been shown to be beneficial to patients and care providers and can directly address some of the interpersonal and other harms to patients that AI automation can cause. Compassionate AI can particularly augment care and benefit patients and caregivers in areas that place large demands on caregiver compassion, such as dementia eldercare and palliative care.

Ethical Statement

Ethical impact of this work is considered within the body of the paper.

Acknowledgments

Thank you to Thomas Arnold, Ali-Reza Bhojani, and Cyrus Olsen for many discussions on the ideas of this paper. Thank you to Peter Hershock and Elias Kruger for their suggestions on the paper and its content.

References

Ahrens, A. H.; and Cloutier, D. 2019. Acting for Good Reasons: Integrating Virtue Theory and Social Cognitive Theory. *Social and Personality Psychology Compass* 13 (4): e12444. doi.org/10.1111/spc3.12444.

Alves, L. C. de S.; Monteiro, D. Q.; Bento, S. R.; Hayashi, V. D.; Pelegrini, L. N. de C.; and Vale, F. A. C. 2019. Burnout Syndrome in Informal Caregivers of Older Adults with Dementia: A Systematic Review. *Dementia & Neuropsychologia* 13 (4): 415–21. doi.org/10.1590/1980-57642018dn13-040008.

American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders: DSM V TR*. Arlington, VA: APA.

Anderson, J. G.; Rogers, C. E.; Bossen, A.; Testad, I.; and Rose, K. M. 2017. Mind-Body Therapies in Individuals With Dementia: An Integrative Review. *Research in Gerontological Nursing* 10 (6): 288–96. doi.org/10.3928/19404921-20170928-01.

Arnold, A.; Kolody, S.; Comeau, A.; and Miguel Cruz, A. 2024. What Does the Literature Say about the Use of Personal Voice Assistants in Older Adults? A Scoping Review. *Disability and Rehabilitation: Assistive Technology* 19 (1): 100–111. doi.org/10.1080/17483107.2022.2065369.

Aronsson, A. S. 2020. Social Robots in Elder Care. *Japanese Review of Cultural Anthropology* 21 (1): 421–55. doi.org/10.14890/jrca.21.1_421.

Arya, R.; Singh, J.; and Kumar, A. 2021. A Survey of Multidisciplinary Domains Contributing to Affective Computing. *Computer Science Review* 40: 100399. doi.org/10.1016/j.cosrev.2021.100399.

Asgharian, P.; Panchea, A. M.; and Ferland, F. 2022. A Review on the Use of Mobile Service Robots in Elderly Care. *Robotics* 11 (6): 127. doi.org/10.3390/robotics11060127.

Aslakson, R. A.; Cox, C. E.; Baggs, J. G.; and Curtis, J. R. 2021. Palliative and End-of-Life Care: Prioritizing Compassion Within the ICU and Beyond. *Critical Care Medicine* 49 (10): 1626. doi.org/10.1097/CCM.0000000000005208.

Baguley, S. I.; Pavlova, A.; and Consedine, N. S. 2022. More than a Feeling? What Does Compassion in Healthcare 'Look like' to Patients? *Health Expectations* 25 (4): 1691–1702. doi.org/10.1111/hex.13512.

Barrett, H. C. 2020. Towards a Cognitive Science of the Human: Cross-Cultural Approaches and Their Urgency. *Trends in Cognitive Sciences* 24 (8): 620–38. doi.org/10.1016/j.tics.2020.05.007

Batson, C. D. 2012. A History of Prosocial Behavior Research. In *Handbook of the History of Social Psychology*, 243–64. New York: Psychology Press. doi.org/10.4324/9780203808498.

Beauchamp, T. L.; and Childress, J. F. 2019. *Principles of Biomedical Ethics*. 8th edition. New York: Oxford University Press.

Beil, M.; Proft, I.; Heerden, D. van; Sviri, S.; and Heerden, P. V. van. 2019. Ethical Considerations about Artificial Intelligence for Prognostication in Intensive Care. *Intensive*

Care Medicine Experimental 7 (1): 1–13. doi.org/10.1186/s40635-019-0286-6.

Birkhäuer, J.; Gaab, J.; Kossowsky, J.; Hasler, S.; Krummenacher, P.; Werner, C.; and Gerger, H. 2017. Trust in the Health Care Professional and Health Outcome: A Meta-Analysis. *PLoS ONE* 12 (2): e0170988. doi.org/10.1371/journal.pone.0170988.

Brown, E. 2023. Will AI Ever Become Spiritual? A Hospital Chaplaincy Perspective. *Practical Theology* 16 (6): 801–13. doi.org/10.1080/1756073X.2023.2242940.

Brown, S. L.; and Brown, R. M. 2015. Connecting Prosocial Behavior to Improved Physical Health: Contributions from the Neurobiology of Parenting. *Neuroscience and Biobehavioral Reviews* 55: 1–17. doi.org/10.1016/j.neubiorev.2015.04.004.

Brown, J. E. H.; and Halpern, J. 2021. AI Chatbots Cannot Replace Human Interactions in the Pursuit of More Inclusive Mental Healthcare. *SSM - Mental Health* 1 (December): 100017. doi.org/10.1016/j.ssmmh.2021.100017.

Barrett, H. C. 2020. Towards a Cognitive Science of the Human: Cross-Cultural Approaches and Their Urgency. *Trends in Cognitive Sciences* 24 (8): 620–38. doi.org/10.1016/j.tics.2020.05.007.

Bryson, J. J. 2015. Artificial Intelligence and Pro-Social Behaviour. In *Collective Agency and Cooperation in Natural and Artificial Systems*, edited by Catrin Misselhorn, 281–306. Cham: Springer International Publishing. doi.org/10.1007/978-3-319-15515-9_15.

Cheung, D. S. K.; Kor, P. P. K.; Jones, C.; Davies, N.; Moyle, W.; Chien, W. T.; Yip, A. L. K.; Chambers, S.; Yu, C. T. K.; and Lai, C. K. Y. 2020. The Use of Modified Mindfulness-Based Stress Reduction and Mindfulness-Based Cognitive Therapy Program for Family Caregivers of People Living with Dementia: A Feasibility Study. *Asian Nursing Research* 14 (4): 221–30. doi.org/10.1016/j.anr.2020.08.009.

Calvo, R. A.; D'Mello, S.; Gratch, J. M.; and Kappas, A. 2015. *The Oxford Handbook of Affective Computing*. Oxford University Press.

Chatzimina, M.; Koumakis, L.; Marias, K.; and Tsiknakis, M. 2019. Employing Conversational Agents in Palliative Care: A Feasibility Study and Preliminary Assessment. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 489–96. doi.org/10.1109/BIBE.2019.00095.

Christov-Moore, L.; Reggente, N.; Vaccaro, A.; Schoeller, F.; Pluimer, B.; Douglas, P. K.; Iacoboni, M.; Man, K.; Damasio, A.; and Kaplan, J. T. 2023. Preventing Antisocial Robots: A Pathway to Artificial Empathy. *Science Robotics* 8 (80): eabq3658. doi.org/10.1126/scirobotics.abq3658.

Cochrane, B. S.; Ritchie, D.; Lockhard, D.; Picciano, G.; King, J. A.; and Nelson, B. 2019. A Culture of Compassion: How Timeless Principles of Kindness and Empathy Become Powerful Tools for Confronting Today's Most Pressing Healthcare Challenges. *Healthcare Management Forum* 32 (3): 120–27. doi.org/10.1177/0840470419836240.

Compson, J. 2015. The CARE Heuristic for Addressing Burnout in Nurses. *Journal of Nursing Education and Practice* 5 (7). doi.org/10.5430/jnep.v5n7p63.

Conversano, C.; Ciacchini, R.; Orrù, G.; Di Giuseppe, M.; Gemignani, A.; and Poli, A. 2020. Mindfulness, Compassion, and Self-Compassion Among Health Care Professionals: What's New? A Systematic Review. *Frontiers in Psychology* 11 (July). doi.org/10.3389/fpsyg.2020.01683.

Crowther, J.; Wilson, K. C.; Horton, S.; and Lloyd-Williams, M. 2013. Compassion in Healthcare – Lessons from a Qualitative Study of the End of Life Care of People with Dementia. *Journal of the Royal Society of Medicine* 106 (12): 492–97. doi.org/10.1177/0141076813503593.

Cuff, B. M. P.; Brown, S. J.; Taylor, L.; and Howat, D. J. 2016. Empathy: A Review of the Concept. *Emotion Review* 8 (2): 144–53. doi.org/10.1177/1754073914558466.

Dahl, A.; and Brownell, C. A. 2019. The Social Origins of Human Prosociality. *Current Directions in Psychological Science* 28 (3): 274–79. doi.org/10.1177/0963721419830386.

Day, J.; Finkelstein, J. C.; Field, B. A.; Matthews, B.; Kirby, J. N.; and Doty, J. R. 2021. Compassion-Focused Technologies: Reflections and Future Directions. *Frontiers in Psychology* 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.603618.

Davenport, T.; and Kalakota, R. 2019. The Potential for Artificial Intelligence in Healthcare. *Future Healthcare Journal* 6 (2): 94–98. doi.org/10.7861/futurehosp.6-2-94.

De Togni, G.; Erikainen, S.; Chan, S.; and Cunningham-Burley, S. 2021. What Makes AI 'Intelligent' and 'Caring'?: Exploring Affect and Relationality Across Three Sites of Intelligence and Care. *Social Science & Medicine* 277 (March). doi.org/10.1016/j.socscimed.2021.113874.

Di Ieva. 2019. AI-Augmented Multidisciplinary Teams: Hype or Hope? *The Lancet* 394 (10211): 1801. doi.org/10.1016/S0140-6736(19)32626-1.

Di Lorito, C.; Bosco, A.; Godfrey, M.; Dunlop, M.; Lock, J.; Pollock, K.; Harwood, R. H.; and Wardt, V. van der. 2021. Mixed-Methods Study on Caregiver Strain, Quality of Life, and Perceived Health. *Journal of Alzheimer's Disease* 80 (2): 799–811. doi.org/10.3233/JAD-201257.

Dicuonzo, G.; Donofrio, F.; Fusco, A.; and Shini, M. 2023. Healthcare System: Moving Forward with Artificial Intelligence. *Technovation* 120 (February): 102510. doi.org/10.1016/j.technovation.2022.102510.

Dreyfus, H. L. 2007. Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian. *Philosophical Psychology* 20 (2): 247–68.

Eisenberg, N.; Spinrad, T. L.; and Morris, A. S. 2013. Prosocial Development. In *The Oxford Handbook of Developmental Psychology, Vol. 2: Self and Other*, edited by Philip David Zelazo, chapter 13. Oxford University Press. doi.org/10.1093/oxfordhb/9780199958474.013.0013.

Gardner, H. E. 2011. *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books.

Garfield, J .L. 2021. The Brahmavihāras and the Achievement of Nonegocentricity. In *Buddhist Ethics: A Philosophical Exploration*, edited by J.L. Garfield. Oxford University Press. doi.org/10.1093/oso/9780190907631.003.0009.

Geraci, R. M. 2006. Spiritual Robots: Religion and Our Scientific View of the Natural World. *Theology and Science* 4 (3): 229–46. doi.org/10.1080/14746700600952993.

Gethin, R. 1998. *The Foundations of Buddhism*. OPUS Series. Oxford University Press.

Gilbert, P. 2010. *The Compassionate Mind: A New Approach to Life's Challenges*. Oakland, CA: New Harbinger Publications.

Gilbert, P. 2014. The Origins and Nature of Compassion Focused Therapy. *The British Journal of Clinical Psychology* 53 (1): 6–41. doi.org/10.1111/bjc.12043.

Goirand, M.; Austin, E.; and Clay-Williams, R. 2021. Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review. *Science and Engineering Ethics* 27 (5): 61. doi.org/10.1007/s11948-021-00336-3.

Goddard, K.; Roudsari, A.; and Wyatt, J. C. 2012. Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *Journal of the American Medical Informatics Association* 19 (1): 121–27. doi.org/10.1136/amiajnl-2011-000089.

Graham, J.; Meindl, P.; Beall, E.; Johnson, K. M.; and Zhang, L. 2016. Cultural Differences in Moral Judgment and Behavior, across and within Societies. *Current Opinion in Psychology* 8: 125–30. doi.org/10.1016/j.copsyc.2015.09.007.

Graves, M. 2022. Apprehending AI Moral Purpose in Practical Wisdom. *AI & Society*. doi.org/10.1007/s00146-022-01597-7.

Graves, M. 2024. Modeling Morality and Spirituality in Artificial Chaplains. *Computers in Human Behavior: Artificial Humans* 2 (1): 100051. doi.org/10.1016/j.chbah.2024.100051.

Haidt, J. 2003. The Moral Emotions. In *Handbook of Affective Sciences*, edited by R J Davidson, KR Scherer, and HH Goldsmith, 852–70. New York: Oxford University Press.

Hanh, T.N. 2009. *The heart of understanding: Commentaries on the Prajnaparamita Heart Sutra*. Parallax Press.

Henrich, J.; Heine, S. J.; and Norenzayan, A. 2010. The Weirdest People in the World? *Behavioral and Brain Sciences* 33 (2–3): 61–83. doi.org/10.1017/S0140525X0999152X.

Hershock, P. D. 2021. *Buddhism and Intelligent Technology: Toward a More Humane Future*. Bloomsbury Publishing.

Hirt, J.; Ballhausen, N.; Hering, A.; Kliegel, M.; Beer, T.; and Meyer, G. 2021. Social Robot Interventions for People with Dementia: A Systematic Review on Effects and Quality of Reporting. *Journal of Alzheimer's Disease* 79 (2): 773–92. doi.org/10.3233/JAD-200347.

Ho, C. W. L.; Soon, D.; Caals, K.; and Kapur, J. 2019. Governance of Automated Image Analysis and Artificial Intelligence Analytics in Healthcare. *Clinical Radiology* 74 (5): 329–37. doi.org/10.1016/j.crad.2019.02.005.

Hofstede, G. 2010. *Cultures and Organizations: Software of the Mind : Intercultural Cooperation and Its Importance for Survival*. Revised and Expanded third edition. New York: McGraw-Hill.

Hughes, J. 2012. Compassionate AI and Selfless Robots: A Buddhist Approach. *Robot Ethics: The Ethical and Social Implications of Robotics*, 69–83.

Jazaieri, H.; Jinpa, G. T.; McGonigal, K.; Rosenberg, E. L.; Finkelstein, J.; Simon-Thomas, E.; Cullen, M.; Doty, J. R.; Gross, J. J.; and Goldin, P. R. 2013. Enhancing Compassion: A Randomized Controlled Trial of a Compassion Cultivation Training Program. *Journal of Happiness Studies* 14 (4): 1113–26. doi.org/10.1007/s10902-012-9373-z.

Jecker, N.S.; Atuire, C.A.; Belisle-Pipon, J.C.; Ravtisky, V.A; Ho, A. 2024. AI and the falling sky: interrogating X-Risk, *Journal of medical ethics*, jme-2023-109702. doi.org/10.1136/jme-2023-109702.

Johnson, J. C.; Kim, H.; and Johnson, P. A. 2021. Pedagogy in Perspective: Ethical Erosion and Effects on Empathy Levels in Healthcare Education. *Medical Science Educator* 31 (3): 1173–75. doi.org/10.1007/s40670-021-01273-3

Kasulis, T.P.(2002. *Intimacy or integrity: philosophy and cultural difference*. Honolulu: University of Hawai'i Press. doi.org/10.21313/9780824863012.

Keown, D. 2001. *The Nature of Buddhist Ethics*. Basingstoke, Hampshire ; Palgrave.

Kerasidou, A. 2020. Artificial Intelligence and the Ongoing Need for Empathy, Compassion and Trust in Healthcare. Bulletin of the World Health Organization 98 (4): 245–50. doi.org/10.2471/BLT.19.237198.

Khera, R.; Simon, M. A.; and Ross, J. S. 2023. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. *JAMA* 330 (23): 2255–57. doi.org/10.1001/jama.2023.22557.

Kirby, J. N.; Steindl, S. R.; and Doty, J. R. 2017. Compassion as the Highest Ethic. In *Practitioner's Guide to Ethics and Mindfulness-Based Interventions*, edited by Lynette M. Monteiro, Jane F. Compson, and Frank Musten, 253–77. Mindfulness in Behavioral Health. Cham: Springer International Publishing. doi.org/10.1007/978-3-319-64924-5_10.

Kitayama, S.; and Salvador, C. E. 2024. Cultural Psychology: Beyond East and West. Annual Review of Psychology 75 (Volume 75, 2024): 495–526. doi.org/10.1146/annurev-psych-021723-063333.

Knafo, A.; and Israel, S. 2012. Empathy, Prosocial Behavior, and Other Aspects of Kindness. In *Handbook of Temperament*, 168–79. New York: Guilford Press.

Kose, U.; and Pavaloiu, A. 2017. Dealing with Machine Ethics in Daily Life: A View with Examples. In Proceedings of the 5th International Virtual Conference on Advanced Scientific Results, 200–205. doi.org/10.18638/scieconf.2017.5.1.454.

Lapsley, D. 2016. Moral Self-Identity and the Social-Cognitive Theory of Virtue. In *Developing the Virtues: Integrating Perspectives*, edited by Julia Annas, Darcia Narváez, and Snow, Nancy E. New York: Oxford University Press.

Lima, M. R.; Wairagkar, M.; Gupta, M.; Rodriguez y Baena, F.; Barnaghi, P.; Sharp, D. J.; and Vaidyanathan, R. 2022. Conversational Affective Social Robots for Ageing and Dementia Support. *IEEE Transactions on Cognitive and Developmental Systems* 14 (4): 1378–97. doi.org/10.1109/TCDS.2021.3115228.

Loper, P. L., Jr. 2018. The Electronic Health Record and Acquired Physician Autism. *JAMA Pediatrics* 172 (11): 1009. doi.org/10.1001/jamapediatrics.2018.2080.

Malenfant, S.; Jaggi, P.; Hayden, K. A.; and Sinclair, S. 2022. Compassion in Healthcare: An Updated Scoping Review of the Literature. *BMC Palliative Care* 21 (May): 80. doi.org/10.1186/s12904-022-00942-3.

Ma, B.; Yang, J.; Wong, F. K. Y.; Wong, A. K. C.; Ma, T.; Meng, J.; Zhao, Y.; Wang, Y.; and Lu, Q. 2023. Artificial Intelligence in Elderly Healthcare: A Scoping Review. *Ageing Research Reviews* 83 (January): 101808. doi.org/10.1016/j.arr.2022.101808.

Malti, T. 2021. Kindness: A Perspective from Developmental Psychology. *European Journal of Developmental Psychology* 18 (5): 629–57. doi.org/10.1080/17405629.2020.1837617.

McDougall, R. J. 2019. Computer Knows Best? The Need for Value-Flexibility in Medical AI. *Journal of Medical Ethics* 45 (3): 156–60. doi.org/10.1136/medethics-2018-105118.

McGowan, K. 2024. My Parents' Dementia Felt Like the End of Joy. Then Came the Robots. *Wired*, January 4, 2024. https://www.wired.com/story/parents-dementia-robots-warm-technology/.

Montemayor, C.; Halpern, J.; and Fairweather, A. 2022. In Principle Obstacles for Empathic AI: Why We Can't Replace Human Empathy in Healthcare. *AI & SOCIETY* 37 (4): 1353–59. doi.org/10.1007/s00146-021-01230-z.

Mori, M. 1981. *The Buddha in the Robot*. Tokyo: Kosei Pub. Co.

Morrow, E.; Zidaru, T.; Ross, F.; Mason, C.; Patel, K. D.; Ream, M.; and Stockley, R. 2022. Artificial Intelligence Technologies and Compassion in Healthcare: A Systematic Scoping Review. *Frontiers in Psychology* 13: 971044. doi.org/10.3389/fpsyg.2022.971044.

Moudatsou, M.; Stavropoulou, A.; Philalithis, A.; and Koukouli, S. 2020. The Role of Empathy in Health and Social Care Professionals. *Healthcare* 8 (1): 26. doi.org/10.3390/healthcare8010026.

Nielsen, Y. A.; Pfattheicher, S.; and Keijsers, M. 2022. Prosocial Behavior toward Machines. *Current Opinion in Psychology* 43 (February): 260–65. doi.org/10.1016/j.copsyc.2021.08.004.

Nwosu, A. C.; Sturgeon, B.; McGlinchey, T.; Goodwin, C. D.; Behera, A.; Mason, S.; Stanley, S.; and Payne, T. R. 2019. Robotic Technology for Palliative and Supportive Care: Strengths, Weaknesses, Opportunities and Threats. *Palliative Medicine* 33 (8): 1106–13. doi.org/10.1177/0269216319857628.

Nyanaponika Thera. 1994. *The Four Sublime States: Contemplations on Love, Compassion, Sympathetic Joy and Equanimity*. Available at https://www.accesstoinsight.org/lib/authors/nyanaponika/wheel006.html (Accessed: 28 July 2024).

Orom, H.; and Cervone, D. 2009. Personality Dynamics, Meaning, and Idiosyncrasy: Identifying Cross-Situational Coherence by Assessing Personality Architecture. *Journal of Research in Personality*, 43 (2): 228–40. doi.org/10.1016/j.jrp.2009.01.015.

Overhage, J. M.; and McCallie, D. 2020. Physician Time Spent Using the Electronic Health Record During Outpatient Encounters. *Annals of Internal Medicine* 172 (3): 169–74. doi.org/10.7326/M18-3684.

Paiva, A. 2022. From Social to Prosocial Machines: A New Challenge for AI. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, 2. IUI '22. New York: Association for Computing Machinery. doi.org/10.1145/3490099.3519387.

Paiva, A.; Santos, F.; and Santos, F. 2018. Engineering Pro-Sociality With Autonomous Agents. *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1). doi.org/10.1609/aaai.v32i1.12215.

Pérez, V.; Menéndez-Crispín, E. J.; Sarabia-Cobo, C.; Lorena, P. de; Fernández-Rodríguez, A.; and González-Vaca, J. 2022. Mindfulness-Based Intervention for the Reduction of Compassion Fatigue and Burnout in Nurse Caregivers of Institutionalized Older Persons with Dementia: A Randomized Controlled Trial. *International Journal of Environmental Research and Public Health* 19 (18): 11441. doi.org/10.3390/ijerph191811441.

Kor, P. P. K.; Chien, W. T.; Liu, J. Y. W.; and Lai, C. K. Y. 2018. Mindfulness-Based Intervention for Stress Reduction of Family Caregivers of People with Dementia: A Systematic Review and Meta-Analysis. *Mindfulness* 9 (1): 7–22. doi.org/10.1007/s12671-017-0751-9.

Padhan, S.; Mohapatra, A.; Ramasamy, S. K.; and Agrawal, S. 2023. Artificial Intelligence (AI) and Robotics in Elderly Healthcare: Enabling Independence and Quality of Life. *Cureus* 15 (8): e42905. doi.org/10.7759/cureus.42905.

Papadopoulos, I.; and Ali, S. 2016. Measuring Compassion in Nurses and Other Healthcare Professionals: An Integrative Review. *Nurse Education in Practice* 16 (1): 133–39. doi.org/10.1016/j.nepr.2015.08.001.

Pavlova, A.; Paine, S.-J.; Sinclair, S.; O'Callaghan, A.; and Consedine, N. S. 2023. Working in Value-Discrepant Environments Inhibits Clinicians' Ability to Provide Compassion and Reduces Well-Being: A Cross-Sectional Study. *Journal of Internal Medicine* 293 (6): 704–23. doi.org/10.1111/joim.13615.

Perry, A. 2023. AI Will Never Convey the Essence of Human Empathy. *Nature Human Behaviour*, July, 1–2. doi.org/10.1038/s41562-023-01675-w.

Quinn, T. P.; Senadeera, M.; Jacobs, S.; Coghlan, S.; and Le, V. 2021. Trust and Medical AI: The Challenges We Face and the Expertise Needed to Overcome Them. *Journal of the American Medical Informatics Association* 28 (4): 890–94. doi.org/10.1093/jamia/ocaa268

Ramana, Y. V. S.; and Peri, P. 2024. AI-Based Mindfulness: The Future of Healthcare and Customer Service. In *Comparative Analysis of Digital Consciousness and Human Consciousness: Bridging the Divide in AI Discourse*, 54–64. IGI Global. doi.org/10.4018/979-8-3693-2015-0.ch003.

Regan, A.; Margolis, S.; Ozer, D. J.; Schwitzgebel, E.; and Lyubomirsky, S. 2023. What Is Unique About Kindness? Exploring the Proximal Experience of Prosocial Acts Relative to Other Positive Behaviors. *Affective Science* 4 (1): 92–100. doi.org/10.1007/s42761-022-00143-4.

Richard, T.; and Hershock, P. 2022. On The Ethical Challenge of the Predicament of AI and the Perspectives Offered by Buddhism: A Conversation with Peter D.

Hershock. *Exchanges: The Interdisciplinary Research Journal* 9 (2): 97–110. doi.org/10.31273/eirj.v9i2.897.

Russell, S.; and Norvig, P. 2020. *Artificial Intelligence: A Modern Approach*. 4th edition. Hoboken: Pearson.

Saunders, J. 2015. Compassion. *Clinical Medicine* 15 (2): 121–24. doi.org/10.7861/clinmedicine.15-2-121.

Sharkey, A.; and Sharkey, N. 2012. Granny and the Robots: Ethical Issues in Robot Care for the Elderly. *Ethics and Information Technology* 14 (1): 27–40. doi.org/10.1007/s10676-010-9234-6.

Shamay-Tsoory, S. G.; Aharon-Peretz, J.; and Perry, D. 2009. Two Systems for Empathy: A Double Dissociation between Emotional and Cognitive Empathy in Inferior Frontal Gyrus versus Ventromedial Prefrontal Lesions. *Brain* 132 (3): 617–27.

Sinclair, S.; Norris, J. M.; McConnell, S. J.; Chochinov, H. M.; Hack, T. F.; Hagen, N. A.; McClement, S.; and Bouchal, S. R. 2016. Compassion: A Scoping Review of the Healthcare Literature. *BMC Palliative Care* 15 (1): 6. doi.org/10.1186/s12904-016-0080-0.

Singer, T.; and Klimecki, O. M. 2014. Empathy and Compassion. *Current Biology*: CB 24 (18): R875–78. doi.org/10.1016/j.cub.2014.06.054.

Solanki, P.; Grundy, J.; and Hussain, W. 2023. Operationalising Ethics in Artificial Intelligence for Healthcare: A Framework for AI Developers. *AI and Ethics* 3 (1): 223–40. doi.org/10.1007/s43681-022-00195-z.

Sparrow, R. 2016. Robots in Aged Care: A Dystopian Future? *AI & SOCIETY* 31 (4): 445–54. doi.org/10.1007/s00146-015-0625-4.

Sparrow, R.; and Sparrow, L. 2006. In the Hands of Machines? The Future of Aged Care. *Minds and Machines* 16 (2): 141–61. doi.org/10.1007/s11023-006-9030-6.

Stratta, E. C.; Riding, D. M.; and Baker, P. 2016. Ethical Erosion in Newly Qualified Doctors: Perceptions of Empathy Decline. *International Journal of Medical Education* 7: 286–92. doi.org/10.5116/ijme.57b8.48e4.

Strauss, C.; Lever Taylor, B.; Gu, J.; Kuyken, W.; Baer, R.; Jones, F.; and Cavanagh, K. 2016. What Is Compassion and How Can We Measure It? A Review of Definitions and Measures. *Clinical Psychology Review* 47 (July): 15–27. doi.org/10.1016/j.cpr.2016.05.004.

Svenaeus, F. 2014. Empathy as a Necessary Condition of Phronesis: A Line of Thought for Medical Ethics. *Medicine, Health Care and Philosophy* 17 (2): 293–99. doi.org/10.1007/s11019-013-9487-z.

Troshani, I.; Hill, S. R.; Sherman, C.; and Arthur, D. 2021. Do We Trust in AI? Role of Anthropomorphism and Intelligence. *Journal of Computer Information Systems* 61 (5): 481–91. doi.org/10.1080/08874417.2020.1788473.

Trothen, T. J. 2022a. Replika: Spiritual Enhancement Technology? *Religions* 13 (4): 275. doi.org/10.3390/rel13040275.

Trothen, T. J. 2022b. Intelligent Assistive Technology Ethics for Aging Adults: Spiritual Impacts as a Necessary Consideration. *Religions* 13 (5): 452. doi.org/10.3390/rel13050452.

Turkle, S.; Taggart, W.; Kidd, C. D.; and Dasté, O. 2006. Relational Artifacts with Children and Elders: The Complexities of Cybercompanionship. *Connection Science* 18 (4): 347–61. doi.org/10.1080/09540090600868912.

Turkle, S. (2010). On the threshold of robotic Companions: In good company? In Y. Wilks (Ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues* (pp. 3–10). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.8.03tur

Vallor, S. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press. doi.org/10.1093/acprof:oso/9780190498511.003.0001.

Varnum, M. E. W.; Grossmann, I.; Kitayama, S.; and Nisbett, R. E. 2010. The Origin of Cultural Differences in Cognition: The Social Orientation Hypothesis. *Current Directions in Psychological Science* 19 (1): 9–13. doi.org/10.1177/0963721409359301.

Vitolla, F.; Raimo, N.; Rubino, M.; and Garegnani, G. M. 2021. Do Cultural Differences Impact Ethical Issues? Exploring the Relationship between National Culture and Quality of Code of Ethics. *Journal of International Management* 27 (1): 100823-. doi.org/10.1016/j.intman.2021.100823.

Waring, J.; Lindvall, C.; and Umeton, R. 2020. Automated Machine Learning: Review of the State-of-the-Art and Opportunities for Healthcare. *Artificial Intelligence in Medicine* 104 (April): 101822. doi.org/10.1016/j.artmed.2020.101822.

Watts, E.; Patel, H.; Kostov, A.; Kim, J.; and Elkbuli, A. 2023. The Role of Compassionate Care in Medicine: Toward Improving Patients' Quality of Care and Satisfaction. Journal of Surgical Research 289 (September): 1–7. doi.org/10.1016/j.jss.2023.03.024.

Wispé, L. 1986. The Distinction between Sympathy and Empathy: To Call Forth a Concept, a Word Is Needed. *Journal of Personality and Social Psychology* 50 (2): 314–21. doi.org/10.1037/0022-3514.50.2.314.

YouGov. 2021. Artificial Intelligence: Global Research Reveals Relationship between Knowledge and Sentiment | YouGov. November, 30. Accessed April 17, 2024. https://today.yougov.com/technology/articles/39714-artificial-intelligence-global-research-reveals-re.

Youngson, R. 2012. *TIME to CARE: How to Love Your Patients and Your Job*. CreateSpace Independent Publishing Platform.

Zigon, J. 2019. Can Machines Be Ethical? On the Necessity of Relational Ethics and Empathic Attunement for Data-Centric Technologies. *Social Research: An International Quarterly* 86 (4): 1001–22.