Human-centered AI: The role of Human-centered Design Research in the development of AI

Conference Paper · August 2020

DOI: 10.21606/drs.2020.282

Jan Auernhammer
Stanford University

DRS2020
BRISBANE, 11–14 AUG
SYNERGY

Human-centered AI: The role of Human-centered Design Research in the development of AI

Jan Auernhammer[a*],

[a] Stanford University
*Corresponding author e-mail: jan.auernhammer@stanford.edu
doi: https://doi.org/10.21606/drs.2020.282

Abstract: Artificial Intelligence has the tremendous potential to produce progress and innovation in society. Designing Artificial Intelligence for people has been expressed as essential for societal well-being and the common good. However, human-centered is often used generically without any commitment to a philosophy or overarching approach. This paper outlines different philosophical perspectives and several Human-centered Design approaches and discusses their contribution to the development of Artificial Intelligence. The paper argues that humanistic design research should play a vital role in the pan-disciplinary collaboration with technologists and policymakers to mitigate the impact of Artificial Intelligence. Ultimately, Human-centered Artificial Intelligence incorporates involving people and designing Artificial Intelligence systems for people through a genuine human-centered philosophy and approach.

Keywords: Human-Centered; Artificial Intelligence; Design Research; Ethics

1. Introduction

Artificial Intelligence (AI) has the potential to transform industries and societies. This transformative potential requires to consider implications for people such as ethics and well-being. For this reason, the European Commission (EC), Massachusetts Institute of Technology, Stanford University, and others committed to the design, development, and application of AI for the good of people (Jobin, Ienca, & Vayena, 2019; Xu, 2019). For example, a high-level Expert Group of the EC presented ethics guidelines for trustworthy AI (EC, 2019). The Massachusetts Institute of Technology announced a US$1 billion commitment to AI to address the global opportunities and the need for ethical application ("MIT reshapes itself to shape the future," 2018). Similarly, Stanford University established an Institute for Human-centered AI. This commitment illustrates both the potential as well as the need for designing AI systems human-centered, including the ethically and trustworthy development of AI. The EC Expert Group stated that AI is not an end in itself, but rather a promising means of enhancing individual and societal well-being and bringing progress and innovation (EC, 2019). These commitments indicate that human-centered approaches have a pivotal role in the development and use of AI technology for the well-being of people. However, Bannon (2011) pointed out that researchers often use the term human-centered in a generic way to encompass a range of distinct research themes without any commitment to a philosophy and overarching conceptual framework.

This article discusses two different philosophical perspectives in AI, the rationalistic and design perspective (Winograd, 1996). This discussion is followed by examining these different perspectives and their implications in guiding the development and application of human-centered AI. The article elaborates in more detail the humanistic design perspective and discusses several Human-centered Design (HCD) approaches and their role in the development of AI. HCD research is the overarching term encompassing various research approaches that center people in the design process. The article focuses on the humanistic design perspective because of the over-valuing of the rationalistic approach that grew up around the advances in pure science and technology of the mid 20th century (Winograd, 1996). The article concludes that a comprehensive or pan-disciplinary design approach is needed to design, develop, and advance human-centered AI.

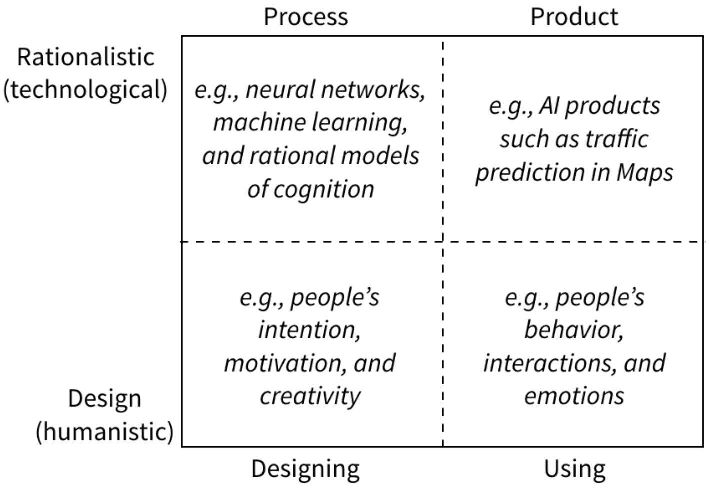
2. Background

Artificial Intelligence systems and technologies such as machine learning, natural language processing, expert systems, and robotics have evolved over the last 80 years since Alan Turing (1950) proposed the Turing-Test (Crevier, 1993; Grudin, 2009). In this evolution, two distinct philosophical perspectives have emerged in how humans and computers should interact (Grudin, 2009; Lieberman, 2009; Winograd, 2006). These are the “rationalistic” and “design” perspectives (Winograd, 1996). The discourse about these two philosophical perspectives or cultures of science and humanities is ongoing over decades beyond the development of AI (Snow, 1993). This philosophical divide between the two perspectives was present in the early developments of AI, represented by John McCarthy (“rationalistic”) and Douglas Engelbart (“design”) (Markoff, 2005; Winograd, 2006).

From the rationalistic perspective, the term AI encompasses the theory and development of computer systems that mimic human abilities and perform tasks that require human intelligence. AI research focuses on mathematical and technological advancement. People are seen as “cognitive machines” (Winograd, 2006; Winograd & Flores, 1986). The role of this perspective in the research of AI is the advancement of the understanding of statistical language, neural networks, and machine learning that form the adaptive mechanisms of AI (Winograd, 1996).

The design perspective views AI as a problem-solving tool to advance people’s capabilities and improve human conditions. The design perspective focuses on the interaction or involvement of the human with the computer (Winograd, 1996; Winograd & Flores, 1986).

This perspective sees human thought and human physical embodiment as inseparable (Dreyfus, 1992; McKim, 1972). The benefit of the design perspective is that it allows coping with the real-world complexity and messiness of the human situation (e.g., Rittel & Webber, 1973). The design perspective has its strength in the interactions of people with the AI system. Figure 1 shows the two perspectives in relation to two main areas of design research.



	Process	Product
Rationalistic (technological)	*e.g., neural networks, machine learning, and rational models of cognition*	*e.g., AI products such as traffic prediction in Maps*
Design (humanistic)	*e.g., people's intention, motivation, and creativity*	*e.g., people's behavior, interactions, and emotions*
	Designing	Using

Figure 1 illustrates the spectrum of the rationalistic view and design perspectives. The rationalistic perspective focuses on thought and people as a formal symbolic representation and focuses on process and product knowledge. The design perspective focuses on knowledge creation about the interactions between people and the enveloping environment, including technologies when designing and using artifacts.

2.1 Philosophical perspectives in guiding the principled development of AI

The rationalists and humanistic designers are approaching the question of ethics and the human impact of AI differently. The rationalistic perspective focuses on developing aggregated or normative models and principles. The design perspective focuses on examining the messiness of the human situation through "enlightened trial and error" (Winograd, 1996).

Based on the rationalistic perspective, government, non-profit, industry, and research organizations have outlined general principles and guidelines of ethical AI (e.g., Jobin et al., 2019). These guidelines provide advice on the development (process) and application (product) of trustworthy AI. The EC Expert group expressed guidelines such as lawful, ethical, and robust (EC, 2019). Firstly, lawful includes that all AI applications need to respect laws and regulations. Secondly, trustworthy AI applications need to fulfill ethical principles. Lastly, AI applications require to be robust. Robustness requires the safe, secure, and reliable performance of the AI, and safeguards should be foreseen to prevent any unintended adverse impacts (EC, 2019).

However, such general guidelines do not represent real-world complexity. Firstly, laws and policies often evolve slower than technological development. As stated in the EC (2019) report, laws are not always up to speed with technological developments, be out of date with ethical norms, or not well-suited addressing context-specific issues. Secondly, ethical principles and moral choices are not universal as surveyed and identified in the Moral Machine study (Awad et al., 2018). Different cultures have different ethical perspectives. General principles and guidelines may not provide the answer in guiding ethical questions in context-specific cases such as autonomous driving. Lastly, robustness does not represent real-world complexity as the social impact of AI is hard to predict or foresee. Katz (2017) expressed that the human species has been excellent in anticipating the consequences of failure, but not so good at preparing for the consequences of success. Similarly, Muehlhauser & Helm (2012) expressed the potential of unintended consequences in AI. As laws may not be up-to-date, universal principles cannot answer context-specific ethical questions, and robustness does not prevent unintended consequences, the humanistic design perspective may provide a more suitable approach to examine the societal impact of AI.

In contrast to the rationalistic perspective, the design perspective utilizes interventions through methods such as prototyping to examine the emerging ethical dilemmas in the interactions between people and AI systems. Methodologies such as Wizard of Oz prototyping are ideal for examining experiences of machine learning (Browne, 2019). Such prototyping methodologies allow examining human needs, behavior, interactions, experience, and cognition in situ (e.g., Martelaro & Ju, 2017a; Norman & Draper, 1986). Examining the emerging interaction, behavior, and needs allows assessing the impact on people. However, experiments require designers and design researchers to decide which ethics perspective is included and how each perspective is assessed and integrated. These decisions pose ethical dilemmas with consequences for AI behavior and its societal implication (Baum, 2020). The implication of these choices indicates that design researchers need to consider various aspects of human implication in the design experiment beyond merely paying close attention to human and social factors (Bannon, 2011). As a remedy, different HCD approaches provide specific viewpoints in researching human concerns in AI.

3. Human-centered Design Research in Artificial Intelligence

HCD is the design approach that centers people and their needs, motivations, emotions, behavior, and perspective in the development of a design. However, involving people in the design does not necessarily mean that they are “centered.” Language often reveals how people are viewed in different approaches. In engineering, people are often perceived as “human factors,” an influencing factor in the performance of the technology. In management, people are often considered as “human resources” or “human capital.” HCD requires viewing humans as people. People with different prior experiences, needs, desires, ambitions, interests, irrational decision making, and lifestyles embedded within specific cultural contexts. HCD is a shift of viewing humans not as a part of the system but central in every aspect of the design.

Several scholars developed design approaches that center human values in the design (e.g., Arnold, 1959; Dreyfuss, 1955; McKim, 1959; Papanek, 1973, 1983). Each of the different HCD approaches provides a valuable perspective when designing for people. The outlined approaches were selected to provide different perspectives such as societal, diversity, interaction, and human needs in the design of computer systems such as AI. Each HCD approach provides an opportunity to research and examine the impact of AI on people from a specific lens. The next sections discuss in detail the implication and value of several HCD approaches in researching the implications for people of AI systems.

3.1 Human-centered Systems and Artificial Intelligence research

The first HCD approach is *Human-centered Systems*. This HCD approach allows investigating the changes in large-scale social structures resulting from the design, implementation, and use of computer systems such as AI (Sawyer, 2005). This human-centered perspective views technology as interdependent of political, ideological, and cultural assumptions of a society that give rise to it (Cooley, 1980; Kling, 1977; Kling & Scacchi, 1980, 1982; Kling & Star, 1998). Figure 2 illustrates this interrelation.

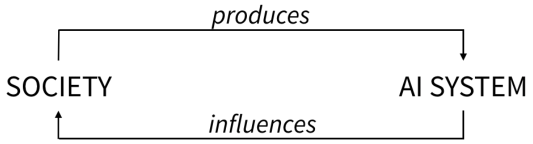


Figure 2 illustrates the relationship between society, including political, cultural, and ideological assumptions and AI technology. Examining technology within the social context allows identifying socio-economical processes that gave rise to the AI systems as well as the influences on society by AI systems.

This HCD research approach allows examining the impact of AI systems on social organization (e.g., Lisetti, Brown, Alvarez, & Marpaung, 2004; Serenko, Ruhi, & Cocosila, 2007). For example, Pee, Pan, & Cui (2019) investigated the collaborative knowledge work of the AI robotic system and people in the workplace of a hospital. They identified distinct forms of knowledge embodiment, as well as the effects of embodiment on social connections between people and AI robotic systems. Such studies allow identifying effects produced by the design and use of the AI system such as multiple and sometimes paradoxical effects, benefit of some groups more than others, moral and ethical consequences, and reciprocal relationships with the broader social context (Sawyer, 2005). Another critical aspect of Human-centered Systems is the designers' ideologies and cultural context in the design of the work system. Cooley (1980) expressed that ideologies such as Taylor's (1907) scientific management will determine how people are treated within the designed system. The next HCD approach addresses this concern.

3.2 Social Design and Artificial Intelligence research

Social Design addresses the designers' ideological issue, as expressed by Cooley (1980). It is an HCD design approach that focuses on the designer's role in society. This approach includes the designer's responsibilities in the design choices made that impact society, as emphasized by Papanek (1973). The designer is the translator of societal needs into an AI systems design, as illustrated in Figure 3.

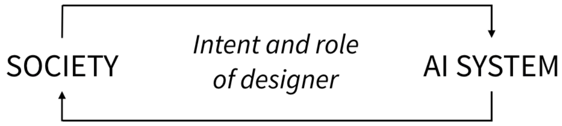


Figure 3 illustrates the intent and role of designers in the translation of societal needs to an AI system. Examining the socio-economical system as well as the role and intent of designers allows identifying underlying values and motivation of specific AI solutions.

Researching AI through this HCD approach allows investigating the intention, decision making, and responsibilities of designers in the development and implementation of the AI system. This investigation includes the socio-economical processes underpin designers' choices. For example, Facebook's business model of targeted advertisement drives the design and implementation of the AI that generates user engagement by filtering content, which impacted democratic elections (Eisenstat, 2019). Such AI systems employed on social media platforms have ethical consequences that need to be considered and resolved (Kane, 2019). Examining if the AI system and business models aim to replace or enhance people provide insights about the values and role of the designer in society (Norman, 2017). Examining the socio-economic and political processes in which the designer is embedded, as well as the intention choices made, can make underlying motivation and values explicit. Designers need to recognize their role, ideology, and socio-economical processes in which they are embedded for the design of AI systems beneficial for society. An HCD approach that can overcome the isolated viewpoint of the designer is Participatory Design.

3.3 Participatory Design and Artificial Intelligence research

Participatory Design focuses on the democratization of participation in systems design (Bodker, 1996; Ehn & Kyng, 1987). This HCD approach engages in questions of human impact such as democracy, power, and control (Ehn, 2017). It enables co-creating with different stakeholders (e.g., users, employees, and customers), as illustrated in Figure 4.

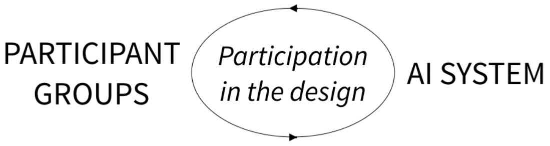


Figure 4 illustrates the participatory design approach in which the user and other stakeholders are involved in the design process. People co-create the design solution.

Participatory Design is essential in the design of AI systems (Neuhauser et al., 2013). This HCD research approach contributes by informing AI developers and cyberneticians about subtle distinctions among sub-groups in society who could benefit from better functional access to computing systems and their adoption (Neuhauser & Kreps, 2011). Participatory Design can help create ideas for the AI System through diverse methods. However, it only represents the design space for the period of the project, and users need to have a basic understanding of what AI can do and not do (Bratteteig & Verne, 2018). Participatory Design is often limited in real-life participation as participants are often only involved while the project is running, power issues are not addressed as management makes final decisions, and participation is resource-intensive and often never compensated (Bodker, 1996). Participatory Design addresses the perspectives of people involved in the project. Another HCD approach that takes a similar view is Inclusive Design.

3.4 Inclusive Design and Artificial Intelligence research

Inclusive Design is an approach that includes and considers the needs and behavior of diverse groups in the design to make mainstream products, services, and systems accessible, usable, and useful for as many people as possible (e.g., Waller, Bradley, Hosking, & Clarkson, 2015). Figure 5 illustrates the inclusion of diverse groups in the design of AI systems.

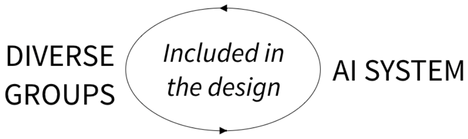


Figure 5 illustrates the interrelation of including and considering diverse groups in the design and development of AI systems.

This HCD research approach examines the usability, accessibility, and usefulness for specific groups of people in the AI design. Inclusive computer system designs enable accessibility and usability and have a positive impact on ethical issues (Abascal & Nicolle, 2005). An important aspect is the categorization of individual differences such as physiological, psychological, and sociological aspects to accommodate differences in the design (Benyon, Crerer, & Wilkinson, 2000). The choices of whom to include has ethical implications in AI systems. For example, excluding specific groups of people when training AI systems result in race and gender bias (Garcia, 2016; Leavy, 2018). Developing diverse design teams can prevent machine biases in the design of AI systems. Building AI systems that overcome biases is not only a matter of having more diverse, and diversity-minded design teams as AI systems themselves can help identify gender and racial biases (Daugherty, Wilson, & Chowdhury, 2019). HCD experiments that include and consider diverse groups enable identifying and making biases explicit early in the design process. Such experiments can include interactions between people and AI systems.

3.5 Interaction Design and Artificial Intelligence research

Interaction design was first proposed by Bill Moggridge and Bill Verplank (Moggridge, 2007). It focuses on understanding and designing human-machine interactions by examining people's behavior, actions, and cognitive processes within the interactions (Norman, 1986, 2013; Norman & Draper, 1986). Figure 6 visualizes the human-AI system interaction.

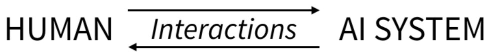


Figure 6 illustrates the human-design interaction in which AI system designs are altered to create a more useful, usable, and meaningful interaction and experience for people.

This HCD research allows examining human-AI systems interactions through prototyping (Browne, 2019; Houde & Hill, 1997). In particular, a reverse Turing Test like prototype termed Wizard of Oz is a common research tool in Human-AI interaction research. Such prototyping was used in AI interaction research such as autonomous driving interactions (e.g., Fu et al., 2019; Rothenbücher, Li, Sirkin, Mok, & Ju, 2015) or human-robot interaction (e.g., Martelaro, Nneji, Ju, & Hinds, 2016; Shibata, Tashima, & Tanie, 1999). Human behavior, activities, and emotions are directly observed in the interaction with the AI system. These observations allow examining the immediate influence of an AI system on people. Similarly, Xu (2019) argues that HCI researchers can contribute to ethical AI design and technological enhancement. This HCD approach enables identifying potentially harmful interactions and creating valuable experiences for people. It can nudge people into an economically healthier behavior, as outlined by Thaler & Sunstein (2009). The same approach is utilized to persuade people into intended behavior through technology such as AI.

3.6 Persuasive Technology and Artificial Intelligence research

Persuasive Technology is the approach that attempts to intentionally change human attitudes, behavior, or both through technology (Fogg, 2003). Figure 7 illustrates the underlying persuasive ability within human-computer interactions.

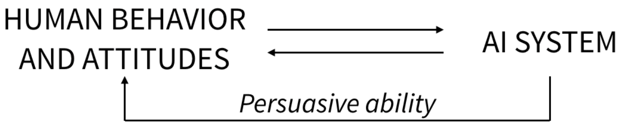


Figure 7 illustrates the relationship between the persuasive ability of technology to change human behavior and attitudes within the interactions of technology.

This HCD research approach has two directions in the development of AI systems. Companies utilize the persuasive ability of AI systems to "hook" users to make highly profitable products (Eyal & Hoover, 2014). The other is that this HCD research provides increased knowledge about persuasive computers allowing people to adopt such technologies to enhance their own lives and recognize when technologies are using tactics to persuade them (Fogg, 1998). Researchers investigated AI systems such as robotic agents and ambient intelligence for the persuasive ability (e.g., Midden & Ham, 2008; Verbeek, 2009). Berdichevsky & Neuenschwander (1999) outlined a framework to consider and minimize ethical abuse of persuasive technologies. However, to be able to identify actual motivational strategies employed in AI technology, studying how the strategies work is required (Orji & Moffatt, 2016). Researching AI systems through this HCD research approach incorporates the examination of the AI system design and people's behavioral changes within everyday life.

3.7 Human-Centered Computing and Artificial Intelligence research

A move away from human-machine interactions to the design of "interspaces," as described by Winograd (1997), is *Human-Centered Computing*. Interspaces incorporate people's lifestyles and the system design as present in everyday life of people (Hallnäs & Redström, 2002). In this approach, intelligence is viewed as an attribute of the combination of human-machine-context (Ford, Hayes, Glymour, & Allen, 2015; Hoffman, Hayes, & Ford, 2001). Figure 8 illustrates this combination.

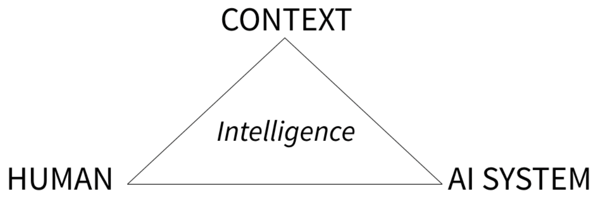


Figure 8 illustrates the interrelation of human-computer-context. Artificial intelligence exists only within this relationship and not only in the computer or the interactions.

This HCD research approach requires integrating diverse views including computer science, psychology, and design to understand the context and situated cognition and action of people using AI systems (Brezillon, 2003; Ford et al., 2015; Hoffman, Roesler, & Moon, 2004; Jaimes, Sebe, & Gatica-Perez, 2006). This approach produces better social outcomes than traditional conceptions of AI as it complements rather than substitutes human intelligence in systems (Hoffman, Hayes, Ford, & Bradshaw, 2012). Research through a Human-centered Computing lens contributes by understanding the influences of an AI system in people's everyday life, including their replacement and constraint of their capabilities and freedoms.

3.8 Need-Design Response and Artificial Intelligence research

The final HCD approach is the *Need-Design Response*, which is based on the work of McKim (1959, 1980). McKim (1959) expressed that every design is a response to a human need, which is often caused by the natural and cultural environment or context. From this perspective, every meaningful AI system design needs to fulfill a need within a specific context. This relationship between needs and an AI system in context is illustrated in Figure 9.

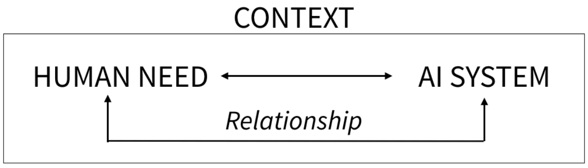


Figure 9 illustrates the "human need design"-relationship. Examining this relationship allows understanding the underlying need the AI system addresses.

This HCD research approach allows finding needs and responding through a meaningful design (Faste, 1987; McKim, 1959). The approach allows identifying the underlying drivers of people's motivation as described by Maslow (1987) in the use of designed systems such as AI. Martelaro & Ju (2017a, 2017b) used a Wizard of Oz prototype to find needs when interacting with the AI system in real-world contexts. The need-design relationship reveals moral implications in AI. This relationship is assessed as follows: Is the "AI solution," which fulfills the "human need" in the "context" potentially helpful or harmful? For example, should a financial *AI system* fulfill the *need for entertainment* by providing a game-based bank account? Such an AI system can easily persuade unhealthy economic behavior. In particular, addressing emotional needs can be highly profitable and requires high morality, as not all human needs are good needs (McKim, 1959). This HCD research approach enables making underlying needs explicit that the AI system design addresses. This approach becomes of particular relevance when AI systems are able to model socio-cultural specific expectations and behavior of people to predict human needs (Riedl, 2019).

4. HCD research for human-centered Artificial Intelligence

Each of the outlined HCD research approaches provides a specific perspective and contribution to the examination of ethical AI, as outlined in Table 1.

Human-centered Systems and *Social Design* provide a societal level of analysis. The first examines the changes in the social organization resulting from the AI implementation and use. Social Design examines the processes that give rise to AI systems, including the role of the designer. Both approaches provide an understanding of the political, socio-economical, and ideological dynamics in the design, development, and use of specific AI systems. Moral considerations include changes in social organization that disadvantage people and business models that drive the use of AI out of self-interest. Making these concerns and implications explicit and transparent is one aspect of Human-centered AI.

Participatory Design and *Inclusive Design* are approaches that address some of the concerns identified in Human-centered Systems and Social Design. The participation of stakeholders in the design and implementation of the AI system can provide essential perspectives in the re-design of the social organization such as work systems. By including many diverse groups throughout the design and development process allows designing the AI system to be accessible and meaningful for many diverse groups. This approach can overcome the one-sided view or interests of the designer or corporate. However, to examine and understand the direct or unanticipated impact of the AI system requires investigating the human-AI system interactions.

Table 1 Human-centered Design practices.

Approach	HCD research in AI	Ethical contribution	Authors
Human-centered System (HCS)	Implementation and use of AI systems and the impact on social systems such as organizations	Examination of moral and ethical consequences for all participants of the AI system design	(Cooley, 1980; Kling, 1973; Kling & Star, 1998)
Social Design (SD)	Socio-economical systems and designers' ideologies and responsibilities that give rise to specific AI systems	Identification of socio-economical dynamics and ideologies that drive the design of unethical AI systems	(Margolin & Margolin, 2002; Papanek, 1973)
Participatory Design (PD)	The democratization of the design and development of AI systems by including different user groups	Discovery of different ethical perspectives in the participant groups in the design of the AI system	(Bodker, 1996; Ehn & Kyng, 1987; Neuhauser & Kreps, 2011)
Inclusive Design (ID)	Inclusion of typically excluded communities to explore alternatives in the design of AI systems	Identification of, e.g., machine biases that disadvantages or discriminates a specific group of people	(Abascal & Nicolle, 2005; Benyon et al., 2000; Spencer, Poggi, & Gheerawo, 2018)
Interaction Design (IxD)	Interface and interaction design to provide a useful and usable AI system	Examination of usability, accessibility of the AI system and potentially harmful people's behavior and experience in the interaction with the AI system	(Norman, 2013; Norman & Draper, 1986; Winograd, 1996, 2006; Xu, 2019)
Persuasive Technology (PT)	AI systems that persuade people towards an intended behavior	Identification of ethical questionable behavioral triggers of the AI system	(Berdichevsky & Neuenschwander, 1999; Fogg, 1998, 2003)
Human-centered Computing (HCC)	Adaptation and organization of everyday life around AI systems that produces a capability between human and computer system	Examination of interspaces enabled by the AI system that replaces or constrains people's capability and lifestyle	(Brezillon, 2003; Ford et al., 2015; Hoffman et al., 2004; Jaimes et al., 2006)
Need-Design Response (NDR)	Design and development of AI systems for physical, intellectual and emotional needs of people	Identification of unethical AI systems design based on needs such as addiction	(Faste, 1987; Gilmore et al., 1999; McKim, 1959, 1980)

Interaction Design, *Persuasive Technologies*, and *Human-centered Computing* enable the examination of the interaction and behavior of people with the AI system. Examining and iteratively developing the human-AI system interactions allows identifying potential harm and enables creating safe and secure experiences and solutions for people. For example, in the communication between pedestrians and autonomous vehicles, purposefully designed signals can improve the interactions (e.g., Moore, Currano, Strack, & Sirkin, 2019). The Persuasive Technologies perspective allows examining the motivational strategies that are embedded in the system by identifying the resulting attitude and behavior of people when using and interacting with the AI system. By incorporating the Human-centered Computing philosophy in the design of AI systems, the focus of the design challenges moves away from purely technical problems to the enhancement and support of human capabilities through the AI system.

The *Need-Design Response* approach by McKim (1959, 1980) that has been popularized by the term Design Thinking allows identifying human needs in the early phases of the design project through practices such as need-finding. Understanding and having a sensitivity for which human needs the design of the AI system addresses, can overcome ethical implications at the beginning. Need-finding and rapid prototyping techniques allow the identification of, e.g., the potential addiction resulting from meeting unhealthy people's needs through the AI system. For example, AI designs that utilize the need for social acceptance in an unhealthy way to generate constant user engagement can lead to addictive behavior on social media platforms. This HCD research approach allows identifying the relationship between needs and AI designs.

Each of the HCD approaches allows examining specific societal and human implications of an AI system design. The ability to examine these implications allows HCD research to play a vital role in the design and development of AI systems.

5. The role of HCD research for Human-centered AI

The challenge of building AI systems and technologies that are ethical and beneficial for society is two challenges in one, the technical (rationalistic) and beneficial technology (design) challenge (Baum, 2017). Interestingly, scholars identified that there is little discourse between the fields of user experience (design) and machine learning (rationalistic) (Yang, Banovic, & Zimmerman, 2018). However, as illustrated in Figures 2 to 9, each HCD approach requires the technical side in the design, development, and use of AI systems. For the resolution of this disparity, several scholars have proposed a comprehensive design approach that incorporates psychology, art, business management, and science to bridge the gap between humanism and science (Arnold, 1959; McKim, 1959, 1980; Winograd, 2006). There is a great opportunity for the fields of AI and HCD to collaborate to make progress on growing concerns around fairness, accountability, and transparency of AI systems (Abdul, Vermeulen, Wang, Lim, & Kankanhalli, 2018; Loi, Wolf, Blomberg, Arar, & Brereton, 2019). This pan-disciplinary approach requires to bring the strength of the rationalistic approach for developing new AI technologies and techniques and the advantage of the humanistic design for developing useful, meaningful, and ethical AI systems together. The role of HCD research is to examine the human impact of AI systems through various HCD approaches by collaborating with and informing AI technologists. Figure 10 illustrates the pan-disciplinary approach in comparison to only a rationalistic (technology) and only a design (humanistic) approach.

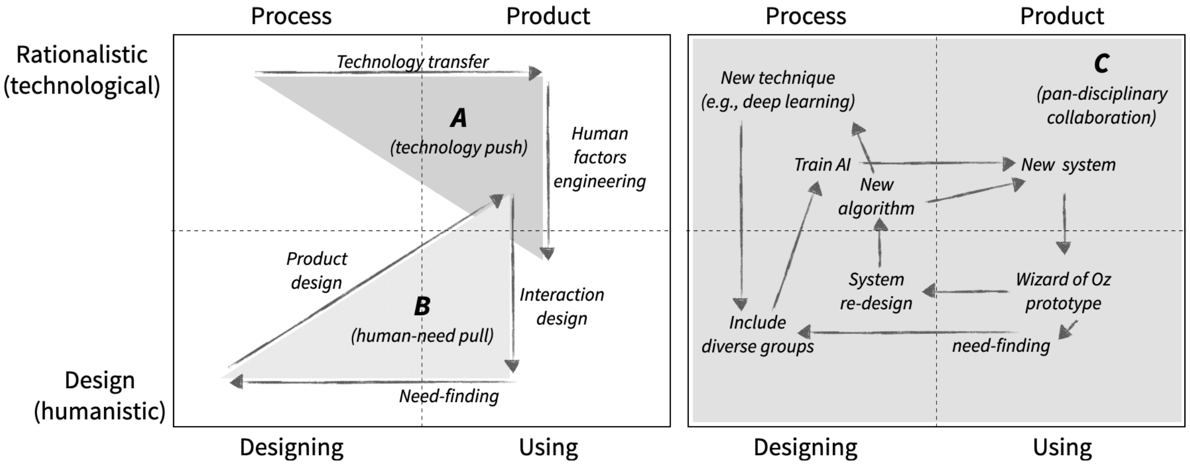


Figure 10 exemplifies three different approaches to designing and developing AI systems. A – Develops new technological solutions, translated into a product, and pushed to people. B – Utilizes existing technologies, and re-design them for specific human-needs. C – Exemplifies the pan-disciplinary collaboration of rationalistic and humanistic approaches.

This comprehensive and pan-disciplinary collaboration design can play the role of mitigating the AI impact on society through collaborating with policymakers and decision-makers. New AI technologies and techniques can be prototyped and experimented through specific designs in a specific context. The pan-disciplinary experiments allow informing decision-makers such as investors and policymakers of the potential human impact of designed AI systems at different stages of the design. This pan-disciplinary collaboration is required to design and develop human-centered AI systems as well as inform and develop policies and guidelines. Figure 11 illustrates this comprehensive design approach.

In this comprehensive approach, the role of Design Research is to examine the potential societal impact of AI technology and systems through “enlightened trial and error.” Design Research can help to overcome the problem of the out-of-date law, allows examining ethical questions in context-specific cases and can help in identifying unintended consequences by learning faster the impacts on people. To be able to contribute to human-centered AI in this pan-disciplinary collaboration, design researchers are required to develop design experiments and prototypes that can examine various implications of AI systems on the well-being of individuals and society.

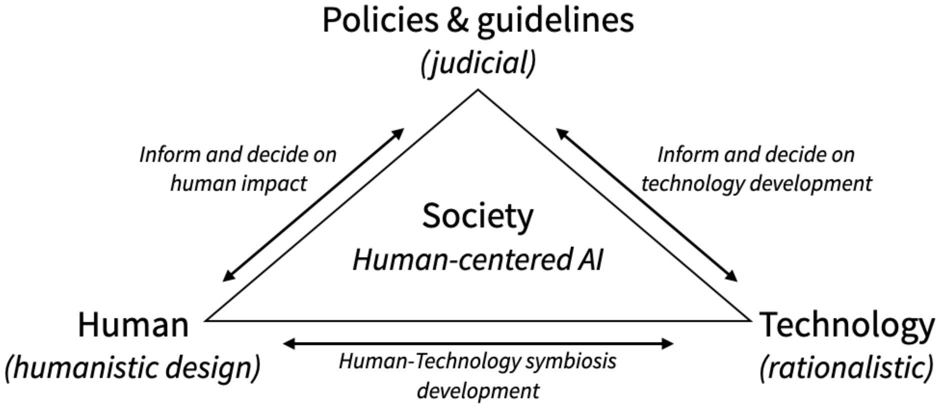

Policies & guidelines
(judicial)
Inform and decide on human impact
Inform and decide on technology development
Society
Human-centered AI
Human
(humanistic design)
Technology
(rationalistic)
Human-Technology symbiosis development

Figure 11 illustrates a comprehensive design approach that incorporates rationalistic AI technology development, humanistic AI design, and legal guidelines for new emergent technologies based on the examination of human impact through HCD research.

6. Conclusion and outlook

This article provided an overview of different *Human-centered Design* approaches and their contribution to the development of Human-centered AI. The discussed HCD approaches illustrate that human-centered AI has essential to do with people, including the designer, user, and other stakeholders and their ideologies, practices, activities, interactions, and needs. Pan-disciplinary research that combines fields such as psychology, cognitive science, computer science, engineering, business management, law, and design is required to develop a genuinely human-centered AI approach.

Educators need to develop pan-disciplinary programs that enable students to understand both the humanistic and rationalistic perspective and are able to collaborate by understanding the technological, societal, and political and policy perspectives and challenges incorporated in the design, development, and use of AI. This educational approach has the promise of designing AI systems and technology that create delightful experiences for people in their everyday life by developing empathetic and creative designers and technologists that combine these viewpoints.

Pan-disciplinary research holds the opportunity to examine the impact of AI on society as well as spark new meaningful innovation for people. For design researchers, an essential task is the development of fruitful experiments for each societal concern in the design, development, and use of AI systems and technology.

Vitruvius Pollio (1914), in approximately 20-30 BC, believed that an architect should focus on three central themes when preparing a design for a building: firmitas (strength), utilitas (functionality), and venustas (beauty). Today, Human-centered AI needs to focus on three integrated perspectives when designing AI systems: rationalistic (technology), humanistic (people), and judicial (policies).

Acknowledgments: The author would like to thank Larry Leifer and the wider designX-lab community for the countless invaluable discussions and insights.

5. References

Abascal, J., & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers, 17*(5), 484-505. doi:10.1016/j.intcom.2005.03.002

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). *Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda*. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. https://doi-org.stanford.idm.oclc.org/10.1145/3173574.3174156

Arnold, J. E. (1959). *Creative engineering seminar, 1959*. Stanford, CA: Stanford, University.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., . . . Rahwan, I. (2018). The Moral Machine experiment. *Nature, 563*(7729), 59-64. doi:10.1038/s41586-018-0637-6

Bannon, L. (2011). Reimagining HCI: toward a more human-centered perspective. *interactions, 18*(4), 50-57. doi:10.1145/1978822.1978833

Baum, S. D. (2017). On the promotion of safe and socially beneficial artificial intelligence. *AI & SOCIETY, 32*(4), 543-551. doi:10.1007/s00146-016-0677-0

Baum, S. D. (2020). Social choice ethics in artificial intelligence. *AI & SOCIETY, 35*(1), 165-176. doi:10.1007/s00146-017-0760-1

Benyon, D., Crerer, A., & Wilkinson, S. (2000). Individual Differences and Inclusive Design. In C. Stephanidis (Ed.), *User Interfaces for All: Concepts, Methods, and Tools.* Mahwah: CRC Press.

Berdichevsky, D., & Neuenschwander, E. (1999). Toward an ethics of persuasive technology. *Commun. ACM, 42*(5), 51–58. doi:10.1145/301353.301410

Bodker, S. (1996). Creating Conditions for Participation: Conflicts and Resources in Systems Development. *Human–Computer Interaction, 11*(3), 215-236. doi:10.1207/s15327051hci1103_2

Bratteteig, T., & Verne, G. (2018). *Does AI make PD obsolete?: exploring challenges from artificial intelligence to participatory design*. Paper presented at the Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2, Hasselt and Genk, Belgium.

Brezillon, P. (2003). Focusing on context in human-centered computing. *IEEE Intelligent Systems, 18*(3), 62-66. doi:10.1109/MIS.2003.1200731

Browne, J. T. (2019). *Wizard of Oz Prototyping for Machine Learning Experiences*. Paper presented at the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk. https://doi-org.stanford.idm.oclc.org/10.1145/3290607.3312877

Cooley, M. (1980). *Architect or Bee? The Human / Technology Relationship*. Slough: Langley Technical Services.

Crevier, D. (1993). *AI: the tumultuous history of the search for artificial intelligence.* New York: Basic Books.

Daugherty, P. R., Wilson, H. J., & Chowdhury, R. (2019). Using artificial intelligence to promote diversity. *MIT Sloan Management Review, 60*(2).

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason.* New York: MIT Press.

Dreyfuss, H. (1955). *Designing for People.* New York: Simon and Schuster.

EC. (2019). *ETHICS GUIDELINES FOR TRUSTWORTHY AI*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai:

Ehn, P. (2017). Scandinavian Design: On Participation and Skill. In D. Schuler & A. Namioka (Eds.), *Participatory Design: Principles and Practices* (pp. 41-77): CRC Press.

Ehn, P., & Kyng, M. (1987). The Collective Resource Approach to Systems Design. In G. Bjerknes, P. Ehn, & M. Kyn (Eds.), *Computers and Democracy - a Scandinavian Challenge* (pp. 17-58). Aarhus: Gower Publishing.

Eisenstat, Y. (2019). I worked on political ads at Facebook. They profit by manipulating us. The company can't avoid damaging democracy. *Perspective.*

Eyal, N., & Hoover, R. (2014). *Hooked: How to Build Habit-Forming Products.* New York: Penguin Publishing Group.

Faste, R. A. (1987). Perceiving Needs. *SAE Transactions, 96*, 419-423. doi:10.2307/44472796

Fogg, B. J. (1998). *Persuasive computers: perspectives and research directions*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, California, USA.

Fogg, B. J. (2003). *Persuasive Technology Using Computers to Change What We Think and Do.* San Francisco: Elsevier Inc.

Ford, K. M., Hayes, P. J., Glymour, C., & Allen, J. (2015). Cognitive Orthoses: Toward Human-Centered AI. *AI Magazine, 36*(4). doi:10.1609/aimag.v36i4.2629

Fu, E., Sibi, S., Miller, D., Johns, M., Mok, B., Fischer, M., & Sirkin, D. (2019, 9-12 June 2019). *The Car That Cried Wolf: Driver Responses to Missing, Perfectly Performing, and Oversensitive Collision Avoidance Systems.* Paper presented at the 2019 IEEE Intelligent Vehicles Symposium (IV).

Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal, 33*(4), 111-117. doi:10.1215/07402775-3813015

Gilmore, D., Trump, R., Velazquez, V., Coughlan, P., Fulton Suri, J., Prokopoff, I., . . . Saperstein, E. (1999). *User-Centered Design in Practice.* Paper presented at the User Centered Design in Practice - Problems and Possibilities, Stockholm, Sweden.

Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *AI Magazine, 30*(4). doi:10.1609/aimag.v30i4.2271

Hallnäs, L., & Redström, J. (2002). From use to presence: on the expressions and aesthetics of everyday computational things. *ACM Trans. Comput.-Hum. Interact., 9*(2), 106-124. doi:10.1145/513665.513668

Hoffman, R. R., Hayes, P., Ford, K. M., & Bradshaw, J. M. (2012). *Collected Essays on Human-Centered Computing, 2001-2011*: IEEE Computer Society Press.

Hoffman, R. R., Hayes, P. J., & Ford, K. M. (2001). Human-centered computing: thinking in and out of the box. *IEEE Intelligent Systems, 16*(5), 76-78. doi:10.1109/MIS.2001.956085

Hoffman, R. R., Roesler, A., & Moon, B. M. (2004). What is design in the context of human-centered computing? *IEEE Intelligent Systems, 19*(4), 89-95. doi:10.1109/MIS.2004.36

Houde, S., & Hill, C. (1997). Chapter 16 - What do Prototypes Prototype? In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction (Second Edition)* (pp. 367-381). Amsterdam: North-Holland.

Jaimes, A., Sebe, N., & Gatica-Perez, D. (2006). *Human-centered computing: a multimedia perspective*. Paper presented at the Proceedings of the 14th ACM international conference on Multimedia, Santa Barbara, CA, USA. https://doi.org/10.1145/1180639.1180829

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389-399. doi:10.1038/s42256-019-0088-2

Kane, T. B. (2019). Artificial Intelligence in Politics: Establishing Ethics. *IEEE Technology and Society Magazine, 38*(1), 72-80. doi:10.1109/MTS.2019.2894474

Katz, B. M. (2017). THE GOOD PLACE, OR NO PLACE AT ALL Silicon Valley and the Paradox of Place. In J. McGuirk & B. McGetrick (Eds.), *California Designing Freedom*: Phaidon Press.

Kling, R. (1973). *Toward a person-centered computer technology.* Paper presented at the ACM National Conference, Atlanta, Ga.

Kling, R. (1977). The Organizational Context of User-Centered Software Designs. *MIS Quarterly, 1*(4), 41-52. doi:10.2307/249021

Kling, R., & Scacchi, W. (1980). Computing as Social Action: The Social Dynamics of Computing in Complex Organizations. In M. C. Yovits (Ed.), *Advances in Computers* (Vol. 19, pp. 249-327): Elsevier.

Kling, R., & Scacchi, W. (1982). The Web of Computing: Computer Technology as Social Organization. In M. C. Yovits (Ed.), *Advances in Computers* (Vol. 21, pp. 1-90): Elsevier.

Kling, R., & Star, S. L. (1998). Human centered systems in the perspective of organizational and social informatics. *SIGCAS Comput. Soc., 28*(1), 22-29. doi:10.1145/277351.277356

Leavy, S. (2018). *Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning*. Paper presented at the Proceedings of the 1st International Workshop on Gender Equality in Software Engineering, Gothenburg, Sweden. https://doi.org/10.1145/3195570.3195580

Lieberman, H. (2009). User Interface Goals, AI Opportunities. *AI Magazine, 30*(4). doi:10.1609/aimag.v30i4.2266

Lisetti, C. L., Brown, S. M., Alvarez, K., & Marpaung, A. H. (2004). A social informatics approach to human-robot interaction with a service social robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34*(2), 195-209. doi:10.1109/TSMCC.2004.826278

Loi, D., Wolf, C. T., Blomberg, J. L., Arar, R., & Brereton, M. (2019). *Co-designing AI Futures: Integrating AI Ethics, Social Computing, and Design*. Paper presented at the Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion, San Diego, CA, USA. https://doi.org/10.1145/3301019.3320000

Margolin, V., & Margolin, S. (2002). A "Social Model" of Design: Issues of Practice and Research. *Design Issues, 18*(4), 24-30. doi:10.1162/074793602320827406

Markoff, J. (2005). *What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry*: Penguin Publishing Group.

Martelaro, N., & Ju, W. (2017a). *The Needfinding Machine*. Paper presented at the Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria.

Martelaro, N., & Ju, W. (2017b). *WoZ Way: Enabling Real-time Remote Interaction Prototyping & Observation in On-road Vehicles*. Paper presented at the Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, Oregon, USA.

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). *Tell Me More: Designing HRI to Encourage More Trust, Disclosure, and Companionship*. Paper presented at the The Eleventh ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zealand.

Maslow, A. H. (1987). *Motivation and Personality.* Uttar Pradesh: Pearson.

McKim, R. H. (1959). Designing for the Whole Man. In J. E. Arnold (Ed.), *Creative engineering seminar, 1959*. Stanford, CA: Stanford, University.

McKim, R. H. (1972). *Experiences in Visual Thinking*. Belmont: Brooks/Cole Publishing Company Inc.

McKim, R. H. (1980). *Experiences in Visual Thinking.* Belmont: Brooks/Cole Publishing Company.

Midden, C. J. H., & Ham, J. R. C. (2008). *The persuasive effects of positive and negative social feedback from an embodied agent on energy conservation behavior*. Paper presented at the *(British) Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB 2008), ,* Aberdeen, UK.

MIT reshapes itself to shape the future. (2018). Retrieved from http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015

Moggridge, B. (2007). *Designing Interactions*: MIT Press.

Moore, D., Currano, R., Strack, G. E., & Sirkin, D. (2019). *The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles*. Paper presented at the Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, Netherlands. https://doi.org/10.1145/3342197.3345320

Muehlhauser, L., & Helm, L. (2012). Intelligence Explosion and Machine Ethics. In A. Eden, J. Søraker, J. H. Moor, & E. Steinhart (Eds.), *In Singularity Hypotheses: A Scientific and Philosophical Assessment*. Berlin: Springer.

Neuhauser, L., & Kreps, G. L. (2011). *Participatory Design and Artificial Intelligence: Strategies to Improve Health Communication for Diverse Audiences*.

Neuhauser, L., Kreps, G. L., Morrison, K., Athanasoulis, M., Kirienko, N., & Van Brunt, D. (2013). Using design science and artificial intelligence to improve health communication: ChronologyMD case example. *Patient Education and Counseling, 92*(2), 211-217. doi:https://doi.org/10.1016/j.pec.2013.04.006

Norman, D. A. (1986). Cognitive Engineering. In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design: New Perspectives on Human-computer Interaction* (pp. 31-61): Taylor & Francis.

Norman, D. A. (2013). *The Design of Everyday Things.* NewYork*: Revised and Expanded Edition*. New York: Basic Books.

Norman, D. A. (2017). Design, Business Models, and Human-Technology Teamwork. *Research-Technology Management, 60*(1), 26-30. doi:10.1080/08956308.2017.1255051

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design: New Perspectives on Human-computer Interaction.* Hillsdale: L. Erlbaum Associates.

Orji, R., & Moffatt, K. (2016). Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Informatics Journal, 24*(1), 66-91. doi:10.1177/1460458216650979

Papanek, V. (1973). *Design for the Real World*. London: Bantam Books.

Papanek, V. (1983). *Design for Human Scale*. New York: Van Nostrand Reinhold Company Ltd.

Pee, L. G., Pan, S. L., & Cui, L. (2019). Artificial intelligence in healthcare robots: A social informatics study of knowledge embodiment. *Journal of the Association for Information Science and Technology, 70*(4), 351-369. doi:10.1002/asi.24145

Pollio, V. (1914). *The Ten Books on Architecture.* London: Harvard University Press.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies, 1*(1), 33-36. doi:10.1002/hbe2.117

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169. doi:10.1007/BF01405730

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2015). *Ghost driver: a platform for investigating interactions between pedestrians and driverless vehicles*. Paper presented at the Adjunct

Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, United Kingdom.

Sawyer, S. (2005). Social informatics: Overview, principles and opportunities. *Bulletin of the American Society for Information Science and Technology, 31*(5), 9-12. doi:10.1002/bult.2005.1720310504

Serenko, A., Ruhi, U., & Cocosila, M. (2007). Unplanned effects of intelligent agents on Internet use: a social informatics approach. *AI & SOCIETY, 21*(1), 141-166. doi:10.1007/s00146-006-0051-8

Shibata, T., Tashima, T., & Tanie, K. (1999, 10-15 May 1999). *Emergence of emotional behavior through physical interaction between human and robot.* Paper presented at the Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C).

Snow, C. P. (1993). *The Two Cultures.* New York: Cambridge University Press.

Spencer, J., Poggi, J., & Gheerawo, R. (2018). *Designing Out Stereotypes in Artificial Intelligence: Involving users in the personality design of a digital assistant*. Paper presented at the Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good, Bologna, Italy. https://doi.org/10.1145/3284869.3284897

Taylor, F. W. (1907). On the art of cutting metals. In. New York: The American Society of Mechanical Engineers.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness*: Penguin Publishing Group.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind, 59*(236), 433-460.

Verbeek, P.-P. (2009). Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology. *NanoEthics, 3*(3), 231. doi:10.1007/s11569-009-0077-8

Waller, S., Bradley, M., Hosking, I., & Clarkson, P. J. (2015). Making the case for inclusive design. *Applied Ergonomics, 46*, 297-303. doi:https://doi.org/10.1016/j.apergo.2013.03.012

Winograd, T. (1996). *Bringing Design to Software.* New York: ACM Press.

Winograd, T. (1997). The Design of Interaction. In P. J. Denning & R. M. Metcalfe (Eds.), *Beyond Calculation: The Next Fifty Years of Computing* (pp. 149-161). New York, NY: Springer New York.

Winograd, T. (2006). Shifting viewpoints: Artificial intelligence and human–computer interaction. *Artificial Intelligence, 170*(18), 1256-1258. doi:https://doi.org/10.1016/j.artint.2006.10.011

Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design.* Indianapolis: Ablex Publishing Corporation.

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *Interactions, 26*(4), 42-46. doi:10.1145/3328485

Yang, Q., Banovic, N., & Zimmerman, J. (2018). *Mapping Machine Learning Advances from HCI Research to Reveal Starting Places for Design Innovation*. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. https://doi.org/10.1145/3173574.3173704

About the Authors:

Jan Auernhammer is a Research Engineer and Executive Director of the Human-centered Business Design Research Group and the Leifer NeuroDesignScience Program, Center for Design Research, Stanford University. His research interest is in the intersection of design, psychology, and management.