Human-Centered AI: A New Synthesis

Ben Shneiderman(✉)

University of Maryland, College Park, MD 20742, USA
ben@cs.umd.edu

Abstract. Researchers, developers, business leaders, policy makers and others are expanding the technology-centered scope of Artificial Intelligence (AI) to include Human-Centered AI (HCAI) ways of thinking. This expansion from an algorithm-focused view to embrace a human-centered perspective, can shape the future of technology so as to better serve human needs. Educators, designers, software engineers, product managers, evaluators, and government agency staffers can build on AI-driven technologies to design products and services that make life better for the users. These human-centered products and services will enable people to better care for each other, build sustainable communities, and restore the environment.

Keywords: Human-computer interaction · Artificial Intelligence · Design · Visualization · Control panels · User interfaces

1 Introduction: A New Synthesis

A new synthesis of disciplines is emerging, in which AI-based intelligent algorithms are combined with human-centered design thinking to make Human-Centered AI (HCAI). This synthesis of disciplines increases the chance that technology will empower rather than replace people. In the past, researchers and developers focused on building AI algorithms and systems, stressing machine autonomy and measuring algorithm performance. The new synthesis values AI and gives equal attention to human users and other stakeholders by raising the prominence of user experience design and by measuring human performance. Researchers and developers for HCAI systems value meaningful human control, putting people first by serving human values such as rights, justice, and dignity, thus supporting goals such as self-efficacy, creativity, responsibility, and social connections.

The higher level goal is to support the 17 United Nations Sustainable Development Goals (https://sdgs.un.org/goals), which were established in 2015 to set aspirations for 2030 (Fig. 1). These goals include elimination of poverty, zero hunger, quality education, and reduced inequalities. Other ambitions address environmental issues such as climate action, life on land, life below water, and sustainable cities and communities. Linking these ambitions to user interface design leads to potent technologies that help people in their relationships, healthcare, education, community efforts and more.

Fig. 1. The 17 United Nations' Sustainable Development Goals (SDGs) (https://sdgs.un.org/goals)

This new synthesis developed in [1–4], is presented in depth in [5] (https://hcil.umd.edu/human-centered-ai/). The supporting concepts come from three fresh ideas for changing technology design so as to bring about a human-centered orientation. These ideas are the:

1) **HCAI framework** that guides creative designers to ensure human centric thinking about highly automated systems [3]. The examples include familiar devices, such as thermostats, elevators, self-cleaning ovens, and cellphone cameras, as well as life critical applications, such as highly automated cars and patient controlled pain relief devices. The dated 1-dimensional model with ten levels of autonomy/automation assumes a zero-sum approach which means that more automation means less human control. However, thoughtful designers can deliver *high levels of human control and high levels of automation*, as we all do with digital cameras and many other devices. In short, the new way of thinking is based on a 2-dimensional model with human control and computer automation as separate axes (Fig. 2). Digital cameras show how high levels of human control (framing, zooming, decisive moment to click, etc.) can be integrated with high levels of automation (aperture, focus, jitter reduction, etc.).

2) **Design Metaphors** suggest how the two central goals of AI research, science and innovation are both valuable, but researchers, developers, business leaders, and policy makers will need to be creative in finding effective ways to combine them to benefit the users [4]. There are four design metaphors that can be used to combine the two goals of AI research:
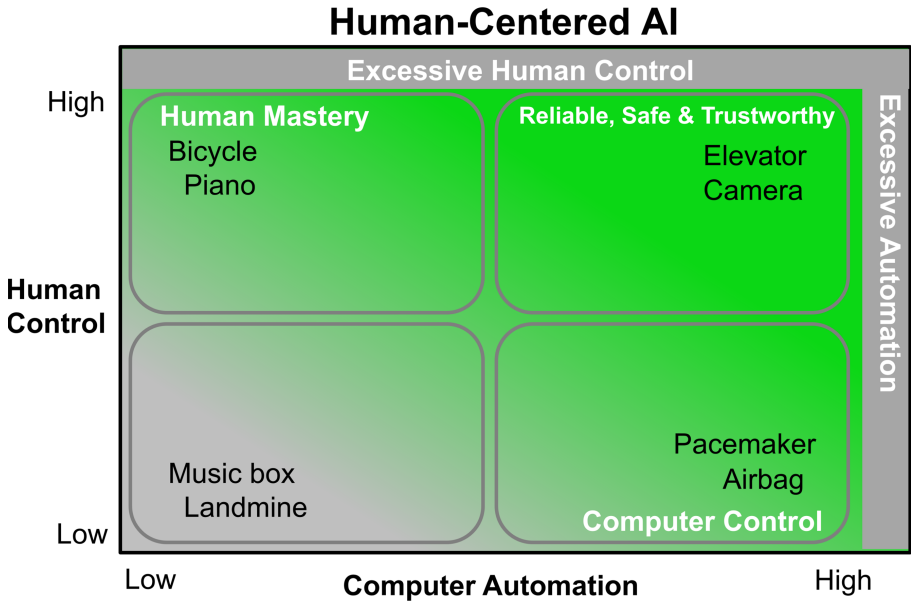
Human-Centered AI

Fig. 2. The HCAI Framework shows a 2-dimensional model with low and high human control and low and high computer automation. It makes clear that it is possible to have high human control and high computer automation. There are also situations in which human mastery or computer control are desired and situations in which excessive human control and excessive automation need to be prevented.

(1)   intelligent agents and supertools
(2)   teammates and tele-operated devices
(3)   assured autonomy and supervised autonomy
(4)   social robots and active appliances

Journalists, headline writers, graphic designers, and Hollywood producers are entranced by the possibilities of robots and AI, so it will take a generation to change attitudes and expectations towards a human-centered view. With fresh thinking, researchers, developers, business leaders, and policy makers can find combined designs that will accelerate HCAI thinking. A greater emphasis on HCAI will reduce unfounded fears of AI's existential threats and raise people's belief that they will be able to use technology for their daily needs and creative explorations. It will increase benefits for users and society in business, education, healthcare, environmental preservation, and community safety.

A key change in thinking is to design user interfaces and control panels to give users of supertools and active appliances greater understanding of the state of the machine and what it will do in the next 10, 60, or 600 seconds. Users are in control when they have visual previews (or alternate interfaces for uses with visual impairments) of what their computer could do, so they can select from alternatives and initiate actions, then follow it through during execution. This what users of digital cameras and navigations systems

(Figure 4) already have, but the guidelines need to be applied in other applications. Similar designs for industrial robots, drones, financial trading systems, ship navigation, and medical devices follow the *Human-Control Mantra*: Preview first, select and initiate, then view execution (Fig. 3).
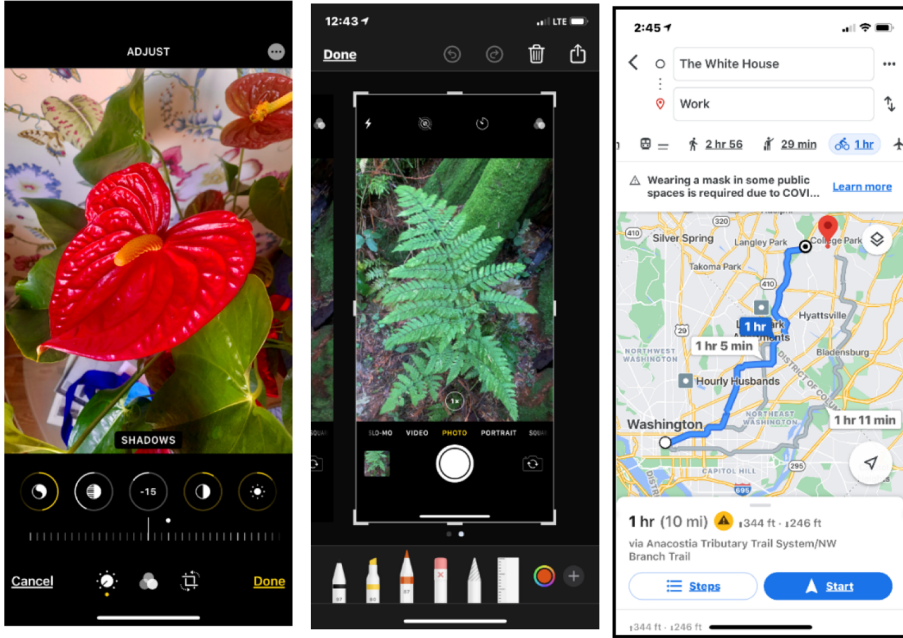

Fig. 3. Digital cameras show users previews of the image they would get and let users control features before, during and after taking photos. Navigation systems show users several possible routes, which they choose from and then they start the system when they are ready.

3) **Governance Structures** bridge the gap between widely discussed ethical principles and the practical steps needed to realize them [3]. Software team leaders, business managers, organization leaders, and government policy makers (Fig. 4) will have to adapt proven technical practices, management strategies, and independent oversight methods [1], so they can achieve the desired goals of:

(1) **Reliable systems** based on proven software engineering practices,
(2) **Safety culture** through business management strategies, and
(3) **Trustworthy certification** by independent oversight and government regulation

These ideas will need to be refined in practice, tuned to the needs of each industry, and adjusted as innovations emerge. They are gaining acceptance, but there is still resistance from those who believe in established ways of working.
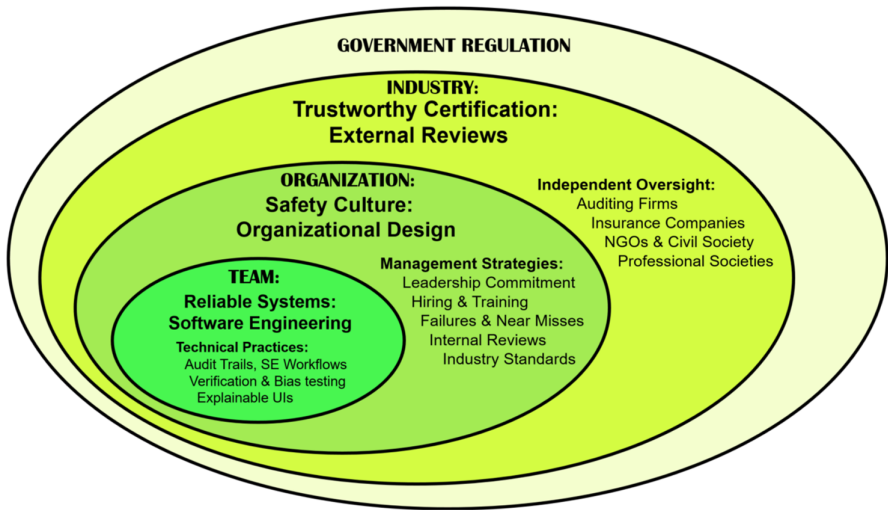
Governance Structures for Human-Centered AI

Fig. 4. Governance structures for Human-Centered AI with four levels: reliable systems based on software engineering (SE) practices, a well-developed safety culture based on sound management strategies, trustworthy certification by external review, and government regulation.

I am well aware that my vision for the future is still a minority position, so there is much work to be done to steer researchers, developers, managers, and policy makers to this new synthesis. Other challenges come from the numerous threats such as misinformation, cyber-crime, political oppression, and online bias, hatred and bullying.

However, I feel confident that the future is human-centered -- filled with supertools and active appliances that amplify, augment, and enhance human abilities, empowering people in remarkable ways while ensuring human control. This compelling HCAI prospect enables people to see, think, create, and act in extraordinary ways, by combining engaging user experiences with embedded AI algorithms to support services that users want. The HCAI prospect contributes to hope-filled agendas for healthcare, community safety, economic development, racial justice, and environmental sustainability.

References

1. Shneiderman, B.: Opinion: the dangers of faulty, biased, or malicious algorithms requires independent oversight. Proc. Nat. Acad. Sci. **113**(48), 13538–13540 (2016). http://www.pnas.org/content/113/48/13538.full
2. Shneiderman, B.: Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. ACM Trans. Interact. Intell. Syst. **10**(4), 31 (2020). https://doi.org/10.1145/3419764. Article 26
3. Shneiderman, B.: Human-centered artificial intelligence: reliable, safe, & trustworthy. Int. J. Hum.-Comput. Interact. **36**(6), 495–504 (2020). https://doi.org/10.1080/10447318.2020.1741118

4. Shneiderman, B.: Design lessons from AI's two grand goals: human emulation and useful applications. IEEE Trans. Technol. Soc. **1**, 2 (2020). https://ieeexplore.ieee.org/document/9088114

5. Shneiderman, B.: Human-Centered AI. Oxford University Press (2022, forthcoming)