AIS Transactions on Human-Computer Interaction

Volume 12 | Issue 3 Article 1

9-30-2020

Human-Centered Artificial Intelligence: Three Fresh Ideas

Ben Shneiderman
University of Maryland, ben@cs.umd.edu

Recommended Citation

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction, 12*(3), 109-124. https://doi.org/10.17705/1thci.00131
DOI: 10.17705/1thci.00131

Transactions on Human-Computer Interaction

Volume 12 Issue 3

9-2020

Human-Centered Artificial Intelligence: Three Fresh Ideas

Ben Shneiderman

Department of Computer Science and Human-Computer Interaction Lab, University of Maryland, College Park, ben@cs.umd.edu

Recommended Citation

Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, *12*(3), pp. 109-124.

DOI: 10.17705/1thci.00131

Available at http://aisel.aisnet.org/thci/vol12/iss3/1

Transactions on Human-Computer Interaction

Research Commentary DOI: 10.17705/1thci.00131 ISSN: 1944-3900

Human-Centered Artificial Intelligence: Three Fresh Ideas

Ben Shneiderman

Department of Computer Science and Human-Computer Interaction Lab, University of Maryland, College Park

ben@cs.umd.edu

Abstract:

Human-Centered AI (HCAI) is a promising direction for designing AI systems that support human self-efficacy, promote creativity, clarify responsibility, and facilitate social participation. These human aspirations also encourage consideration of privacy, security, environmental protection, social justice, and human rights. This commentary reverses the current emphasis on algorithms and AI methods, by putting humans at the center of systems design thinking, in effect, a second Copernican Revolution. It offers three ideas: (1) a two-dimensional HCAI framework, which shows how it is possible to have both high levels of human control AND high levels of automation, (2) a shift from emulating humans to empowering people with a plea to shift language, imagery, and metaphors away from portrayals of intelligent autonomous teammates towards descriptions of powerful tool-like appliances and tele-operated devices, and (3) a three-level governance structure that describes how software engineering teams can develop more reliable systems, how managers can emphasize a safety culture across an organization, and how industry-wide certification can promote trustworthy HCAI systems. These ideas will be challenged by some, refined by others, extended to accommodate new technologies, and validated with quantitative and qualitative research. They offer a reframe -- a chance to restart design discussions for products and services -- which could bring greater benefits to individuals, families, communities, businesses, and society.

Keywords: Human-Centered Artificial Intelligence, Human-Computer Interaction, Artificial Intelligence, design, reliable, safe, trustworthy, Copernican Revolution

Fiona Nah was the accepting senior editor for this paper.

1 Introduction

Artificial Intelligence (AI) dreams and nightmares, represented in popular culture through books, games, and movies, evoke images of startling advances as well as terrifying possibilities. The contrast is often between a blissful place where intelligent machines meet all human needs and a dystopian future in which robots control and dominate humanity. In both cases, people are no longer in charge; the machines rule. However, there is a third possibility; an alternative future filled with computing devices that amplify human abilities a thousand-fold, empowering people in remarkable ways while ensuring human control. This compelling prospect, called Human-Centered AI (HCAI), enables people to see, think, create, and act in extraordinary ways, by combining potent user experiences with embedded AI methods to support services that users want (Li, 2018; Robert et al., 2020; Shneiderman, 2020a).

To counter the widespread belief that AI-driven human-like robots will take over, this paper describes how to make successful technologies that augment and enhance humans rather than replace them. This shift in thinking could lead to a safer, more understandable, and more manageable future. An HCAI approach will reduce the prospects for out-of-control technologies, calm fears of robot-driven unemployment, and give users better control of privacy and security.

This fresh vision is meant as a guide for researchers, educators, designers, programmers, managers, and policy makers in shifting toward language, imagery, and ideas that advance a human-centered approach. Putting people at the center will lead to the creation of powerful tools, convenient appliances, and well-designed products and traditional services that empower people, build their self-efficacy, clarify their responsibility, and support their creativity.

2 Related Work

Numerous books celebrate AI, and at the same time raise fears of out-of-control robots and computers taking over (Figure 1), including some by prominent writers (Marcus & Davis, 2019; O'Neil, 2016; Russell, 2019). Some authors raise doubts about overly optimistic promises, but nevertheless generally push for more AI and even Artificial General Intelligence (AGI), which includes the commonsense reasoning that indicates full human abilities. The HCAI approach advances the goals of AI, while ensuring human control.

Figure 1. A sample of the many popular AI-centered books

A central issue is how to define AI. I rely on a definition that suggests that AI systems and algorithms do what people do: perceive, think, decide, and act. A broadened definition would include the capacity to recognize and respond to emotions, to adapt to new circumstances, and recommend or rank alternatives. The products and services of AI systems include pattern recognition (of images, texts, spoken words, faces,

signals, etc.), generation (of images, texts, spoken words, faces, signals, etc.), natural language processing and translation, and game playing (checkers, chess, go, etc.). Other major AI topics are robot design -- especially in relation to social robots -- and autonomous systems of many kinds. AI research and development have become major topics for businesses and governments around the world. Business applications range from internal management to customer support, while government applications include policing and policy making.

A second central issue is the boundary between automation and autonomy. Some researchers and developers believe that autonomous systems based on “machine learning”, “deep learning”, and “neural nets” enable greater capacity for AI systems to be adaptable, resilient, and “intelligent” as compared to traditional automated systems. Other researchers and developers believe that there is no clear boundary between automation and autonomy, and that the unpredictability of machine autonomy has dangers. They also believe in the importance of human autonomy, while preferring the language of interdependence. Since traditional automated systems are likely to be redesigned to include more autonomous features based on AI algorithms, the boundary between automated and autonomous is likely to become even fuzzier. This commentary addresses automated and autonomous designs as if they are in the same category.

3 Fresh Ideas to Reframe Old Beliefs

Reframing old beliefs with a fresh vision is among the most powerful tools for change. It can liberate researchers and designers from old beliefs, allowing them to adopt a different perspective. This paper suggests that HCAI can liberate researchers and designers to realize that there are other ways of thinking about future technology. Ancient astronomers, with a few exceptions, saw the earth as the center of the solar system with the sun and planets revolving around it. In the 16th century, Copernicus reframed astronomy by making a convincing case for a sun-centered model, which he showed to be more accurate, thus enabling better predictions and further discoveries. Similarly, HCAI reframes AI by replacing algorithms and AI systems with humans at the center. That is why I have termed HCAI a Second Copernican Revolution (Figure 2). Traditional discussions suggest that humans are “in-the-loop” around AI, while the HCAI reframing suggests that AI is “in-the loop” around humans, who are now the center of attention.

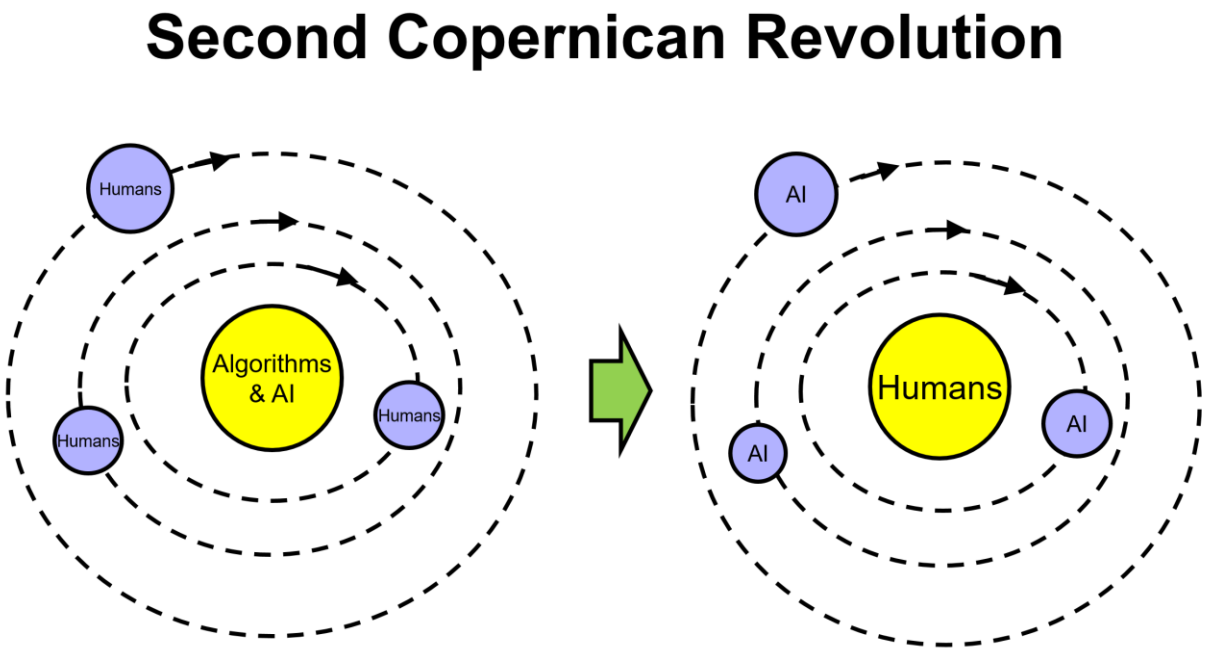

Figure 2. A second Copernican Revolution puts humans at the center of attention

The reframing to HCAI has deeper implications. In the past, researchers and developers focused on building AI algorithms and systems, stressing machine autonomy, measuring algorithm performance, and celebrating what AI systems could do. In contrast, HCAI’s design thinking approach puts the human users

at the center, emphasizing user experience design, measuring human performance, and celebrating the new powers that people have. Researchers and developers for HCAI systems focus on user needs, explainable systems, and meaningful human control. People come first; serving human needs is the goal. Some may see this as an extreme metaphor, but it emphasizes the profound importance of providing an appropriate concept for future technologies that will promote human dignity.

Putting humans at the center of design thinking does not mean that designers should mimic human form and behavior. The alternative is to serve human needs by way of comprehensible, predictable, and controllable tools, appliances, and user experiences. Lewis Mumford's (1934) book *Technics and Civilization* provides a helpful guide to the evolution of new technologies. He describes "The Obstacle of Animism", which is the tendency of new technology designers to use humans or animals as a guide to design. A key example of the limitation of animism is the shift from using two human legs to four wheels to transport heavy loads. Mumford wrote that "the most ineffective kind of machine is the *realistic* mechanical imitation of a man or other animal." He leaves us with an important history lesson: "for thousands of years animism has stood in the way of … development."

Mumford's description of technology evolution accurately describes the move away from human forms in early smiling simulated bank teller machines to our current form fill-in Automated Transaction Machines and the shift from intelligent tutoring machines with human faces or avatars to Massive Open Online Courses (MOOCs) with well-designed user interfaces that give students a clear sense of what is happening and what their choices are. This reframing avoids the deception of a simulated human. It puts humans at the center by increasing human control, even though there are high levels of computer automation and AI algorithms. While bio-inspired designs can be helpful starting points, designers should keep in mind other possibilities and evaluate performance of multiple alternatives.

Numerous psychological studies by Clifford Nass and others (Reeves and Nass, 1996; Nass and Moon, 2000) showed that when computers are designed to be like humans, users respond and engage in socially appropriate ways. However, other designs might lead to superior performance. Designing systems to be like people reduces the chance that designers will take advantage of unique computer features that have no human analog, including sophisticated algorithms, advanced sensors, information abundant displays, and powerful effectors.

This controversy might be summarized as:

> **Nass's Fallacy:** Since many people are willing to respond socially to robots, it is appropriate and desirable to design robots to be social or human-like.
>
> **Shneiderman's Conjecture:** Successful robots utilize the distinctive features of machines. Robots will become more tool-like, tele-operated, and under human supervisory control through well-designed user interfaces that avoid human-like features.

Another fallacy lies in the belief that computers should become our teammates, partners, and collaborators. Psychologists point out the difficulty in accomplishing this goal because human teammates have such distinctive ways of coordinating with each other (Klein et al., 2004):

> **Teammate Fallacy**: Humans work in teams, so computers should be designed to have the same behaviors as humans.
>
> **Computers-in-the-Loop Reality**: Humans work with others in teams, crews, and groups, with computers best designed as helpful tools that continuously provide information and carry out tasks, but do so under human control.

While a majority of researchers believe that social robots and robot teammates are inevitable (Wang & Siau, 2019), I think Mumford's historical perspective and current design successes show that these social and teammate views will give way to functional designs.

I believe it is helpful to remember that computers are not people and people are not computers. Boden et al. (2017) convey a similar idea: "Robots are simply not people." Humans have legal and moral responsibilities over the design of machines, including robots. Yes, there are lessons to be learned from bio-inspired designs, but the value of such designs should be studied in comparison with interactive visual user interfaces, such as with navigation systems, and physical interfaces, such as with drone and game controllers. Human-like speech communication with applications like Siri or Alexa has a role when hands and eyes are busy, while high information capacity of visual user interfaces has many advantages, such as with digital cameras or e-commerce shopping.

This bumper sticker (Figure 3) emphasizes centering on the social nature of human collaboration with computers in the loop, available to support human efforts.

Figure 3. Bumper sticker "Humans in the Group; Computers in the Loop"

Breaking free from the old belief that computers should be like human teammates can liberate designers to more readily take advantage of the distinctive capabilities of algorithms, databases, sensors, effectors, etc. The U.S. Air Force emphasized the distinction by using the term Remotely Piloted Vehicles to convey that a human pilot was responsible. The Mars Rovers have a whole room of human controllers at NASA's Jet Propulsion Laboratory to operate the Rovers, as they perform well-designed activities on Mars. The da Vinci Surgical Systems website says they "don't perform surgery. Your surgeon performs surgery with da Vinci by using instruments that he or she guides via a console" (www.davincisurgery.com). Remember Mumford's (1934) message that successful designs are not based on human forms.

The challenge for designers is to understand what human-like features are useful, such as a human-like voice for virtual assistants like Siri, Alexa, and Cortana, which allow hands-free access to information. However, designers must also come to understand that visual displays of long textual lists, a map, or diagrams, are sometimes better than a spoken response. For instance, the information abundance of visual displays has many advantages over an ephemeral spoken response that could be drowned out by ambient noise. In addition to limiting design choices, humanizing robots can lead to three problems: mistaken usage based on emotional attachment to the systems, false expectations of robot responsibility, and incorrect beliefs about appropriate use of robots (Robert, 2017).

A frequently successful strategy is to provide interactive visual interfaces that present abundant information in compact, spatially stable, tiled layouts so users can maintain situational awareness. They can interpret the status, formulate plans, and carry out tasks while monitoring performance. Modest highlights can draw attention to new information or time sensitive features. Machine-generated recommendations, alerts, and alarms can be placed in stable positions so users can turn to them when they wish. User-controlled adaptation of the window layouts, list sort orders, or featured highlights enables users to tailor displays to their needs and pass their layouts on to colleagues.

Users may also sit near other operators so they can mutually monitor what others are doing, ask for assistance, and work together easily when they need to respond to fast moving events. NASA spacecraft control rooms, counter-terrorism centers, police command centers, medical workstations, air traffic control centers, and stock market trading rooms are common examples (Figure 4). Many variations on smaller interactive dashboards are increasingly used on laptops, such as the currently common COVID dashboards, while mobile device displays can also provide compact information abundant displays such as in navigation apps.

Figure 4. Bloomberg terminal for financial analysts and traders

4 Three Fresh Ideas

The three ideas, summarized in this commentary, may enable researchers, educators, designers, programmers, managers, and policy makers to create new possibilities for future products and services.

High levels of human control AND high levels of automation are possible: The first idea is a fresh way to think of technology design, and breaks out of the traditional one-dimensional view that led readers to believe that more automation meant less human control. This one-dimensional view was first described by Sheridan and Verplank (1978), even though Sheridan (1992, 2000) continued to have strong views about the need for supervisory control.

I accepted this one-dimensional framework as a way to understand what seemed like a necessary tradeoff. I described it in the first edition of *Designing the User Interface: Strategies for Human-Computer Interaction* (Shneiderman, 1987), with a section titled “Balancing Automation and Human Control”. However, in recent years, I became unsettled by this view as I came to increasingly value human control even with high levels of automation. That section is now titled: “Ensuring Human Control While Increasing Automation” (Shneiderman et al., 2016), which at first may seem perplexing -- but give it a chance.

The two-dimensional HCAI framework (Shneiderman, 2020a) shows how creative designers can imagine highly automated systems that keep people in control. It separates human control (y-axis) from computer automation (x-axis) (Figure 5). The examples of **high levels of human control AND high levels of automation** include familiar devices, such as thermostats, elevators, self-cleaning ovens, and dishwashers, as well as life critical applications like highly automated cars and patient controlled pain relief devices. Smartphone cameras exemplify this new possibility of human control and automation, in which AI algorithms are used for setting focus and aperture, while compensating for hand shaking movement. However, the user is in control of where to point the camera, what zoom level is best, and when to take the photo. They have control over filters before they take their photos and can edit, crop, and adjust lighting afterward. Users can adjust the aperture setting and exposure level by touching the image location they are interested in, but a further helpful user control would be to select the shutter speed to allow longer exposures for creative images of waterfalls or intentionally blurred images of swirling dancers.

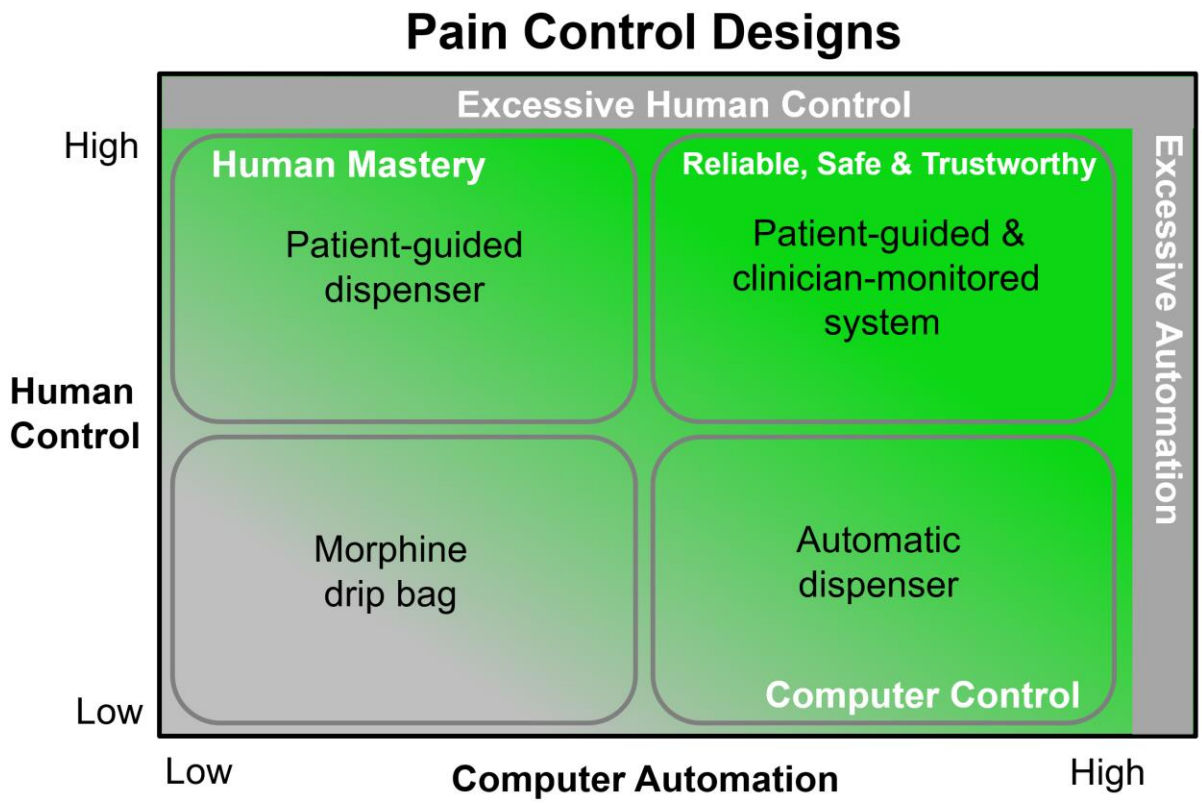

Figure 5. The HCAI two-dimensional framework for thinking of new designs, applied to pain control device designs

The two-dimensional HCAI framework (Figure 5) clarifies that there are situations that require high levels of computer automation, such as airbag deployment and embedded pacemakers (lower right quadrant), and other situations that require high levels of human control, such as piano playing or bicycling (upper left quadrant). In the case of pain control designs, the non-automated morphine drip bag (lower left) was improved by an automated dispenser (lower right). Further improvements give patients a trigger to request limited additional morphine (upper left), while advanced designs combine a patient-guided system with a clinician-monitored system (upper right).The two-dimensional HCAI framework also recognizes that there can be problems with excessive automation (right side), such as the Boeing 737 MAX crashes, stock market flash crashes, and parole or hiring decisions based on machine learning with biased datasets. There can also be problems with excessive human control (top), such as drunk drivers and suicidal pilots. Improved systems with well-designed interlocks to prevent excesses must be part of every design process.

Shift from emulating humans to empowering people: The second idea is to show how the two central goals of AI research -- emulating human behavior (AI science) and developing useful applications (AI engineering) -- are both valuable, but that designers go astray when the lessons of the first goal are put to work on the second goal. Often the emulation goal encouraged beliefs that machines should be designed to be like people, when the application goal might be better served by providing comprehensible, predictable, and controllable designs. While there is an understandable attraction for some researchers and designers to make computers that are intelligent, autonomous, and human-like, that desire should be balanced by appreciating that many users want to be in control of technologies that support their abilities, raise their self-efficacy, respect their responsibility, and enable their creativity.

Shneiderman (2020b) describes four such design tradeoffs that challenge designers of HCAI applications, and offers combined designs that bring the best of both (Figure 6):

1. Intelligent Agent and Powerful Tool
2. Simulated Teammate and Tele-Operated Device
3. Autonomous System and Supervisory Control
4. Humanoid Robot and Mechanical-like Appliance

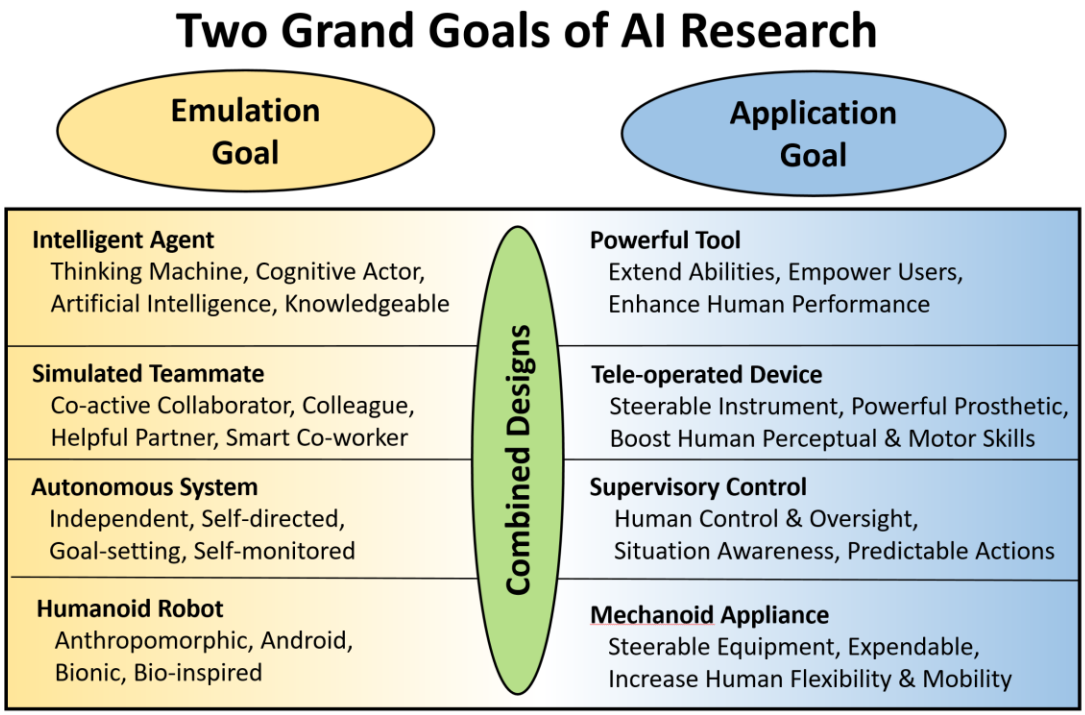

Figure 6. Four issues raised by the Emulation and Application Goals, which lead to combined designs that balance intelligent, autonomous, and human-like qualities with an appreciation that many users want to be in control of technologies that support their self-efficacy, responsibility, and creativity

For example, when thinking of how to communicate with a simulated teammate, a common approach is natural language speech generation to report status; however, a rescue robot would be more effective if it simply sent a video image with abundant continuous visual data about location, temperature, air quality, obstacles, and possible future routes. Similarly, humanoid rescue robots with two legs gave way to more reliable four wheeled or treaded devices and aerial drones that were tele-operated appliances (Murphy, 2014).

Journalists, headline writers, and Hollywood producers have encouraged misleading notions about robots and AI, so the already active reframing process could take a generation to change attitudes and expectations. Using the concept of a second Copernican Revolution could help designers find combined approaches that accelerate creation of reliable, safe, and trustworthy applications. A greater emphasis on HCAI could reduce 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. The terminology is important, but equally important is the imagery used. To change beliefs, it will be valuable to shift imagery away from the aging cliché of human hands connecting with robot hands or social robots (Figure 7).

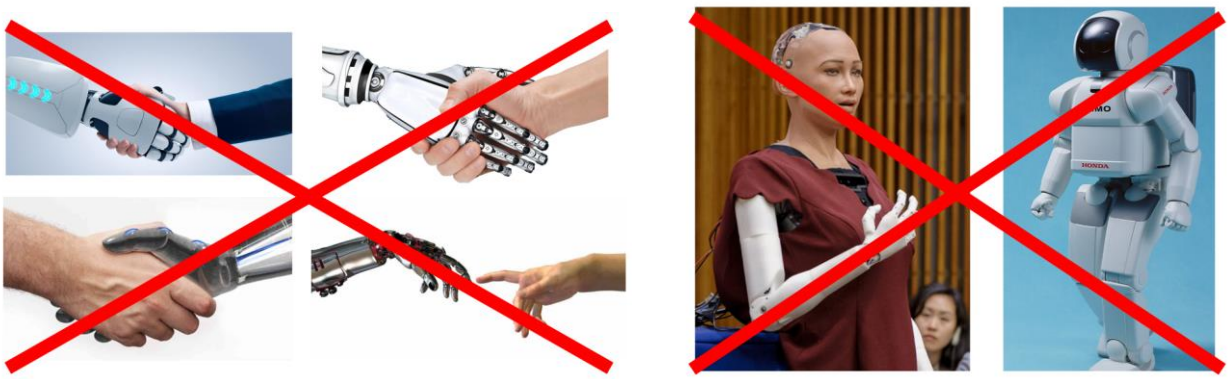

Figure 7. Cliché-ridden images of humanoid robot hands and social robots

Images of user interfaces on appliances and tele-operated devices are aligned with the idea of humans being in control and exercising creative judgment, while emphasizing human responsibility (Figure 8).

Figure 8. Appliances and tele-operated devices are the more likely future for human-centered technologies

Many applications involving machine and deep learning algorithms provide post-hoc explanations of why a decision refused mortgage or parole requests. However, exploratory user interfaces using interactive visual designs offer a more likely path to successful customer adoption and acceptance (Chatzimparmpas et al., 2020; Hohman et al., 2018; Nourashrafeddin et al., 2018; Yang et al., 2020). Well-designed interactive visual interfaces will improve the work of machine learning algorithm developers and facilitate comprehension by various stakeholders.

Governance structures for HCAI: The third idea bridges the gap between widely discussed ethical principles of HCAI and the practical steps needed to realize them. The 15 recommendations are based on the HCAI framework and the combined designs from emulation and application research (Shneiderman, 2020c). These recommendations suggest how to: (1) adapt proven software engineering team practices,(2) implement organization-wide management strategies to build a safety culture, and (3) establish independent oversight methods (Shneiderman, 2016) that can be applied industry-wide to improve performance (Figure 9).

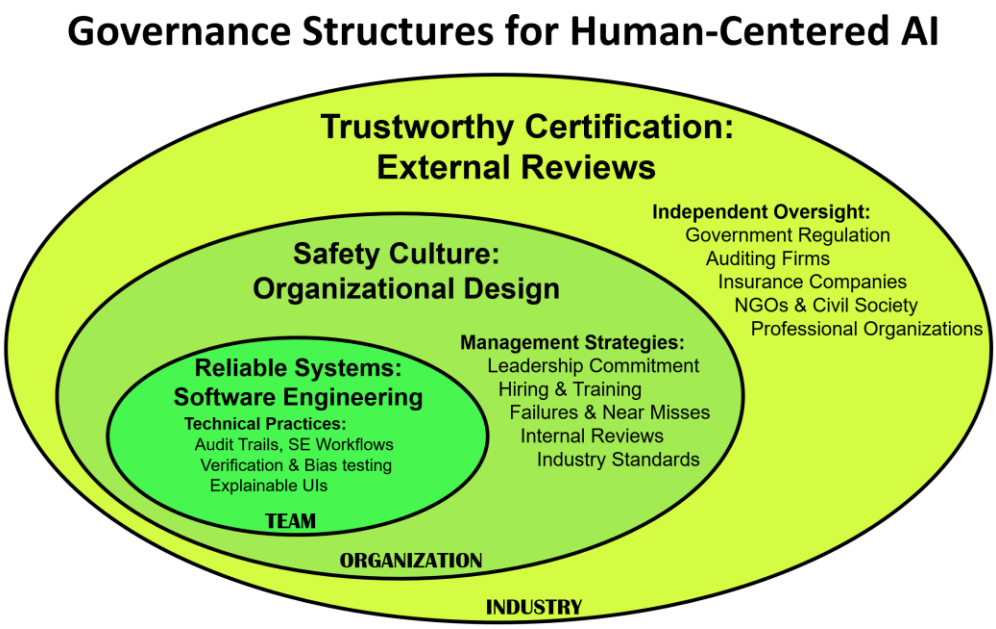

Figure 9. Governance structures to guide teams, organizations, and industry leaders

These new strategies guide software team leaders, business managers, and organization leaders in the methods to develop HCAI products and services that are driven by three goals:

1. **Reliable** systems based on proven software engineering practices with a team,
2. **Safety** culture through business management strategies within an organization, and
3. **Trustworthy** certification by independent oversight across an industry

These three goals apply to most commercial developments, but the 15 recommendations in Shneiderman (2020c) are tied to HCAI products and services (Table 1). These recommendations build on the HCAI framework by limiting the dangers of excessive automation and excessive human control, while steering practice to support the goals of reliable, safe and trustworthy products and services.

GOVERNANCE STRUCTURES FOR HUMAN-CENTERED AI
Reliable Systems Based on Sound Software Engineering Practices for a Team
1. Audit trails and analysis tools
2. Software Engineering Workflows
3. Verification and validation testing
4. Bias testing to enhance fairness
5. Explainable user interfaces
Safety Culture through Business Management Strategies within an Organization
6. Leadership commitment to safety
7. Hiring and training oriented to safety
8. Extensive reporting of failures and near misses
9. Internal review boards for problems and future plans
10. Alignment with industry standard practices
Trustworthy Certification by Independent Oversight for an Industry
11. Government interventions and regulation
12. Accounting firms conduct external audits
13. Insurance companies compensate for AI failures
14. Non-governmental and civil society organizations
15. Professional organizations and research institutes

Table 1: Recommendations for Human-Centered AI Systems for teams, organizations, and industry leaders

Technical practices for teams of programmers, designers, and software engineers include audit trails to enable analysis of failures, just like the flight data recorders (aviation black boxes, which are really orange boxes) that have helped make civil aviation a successful industry. Recommended practices include software engineering workflows, verification and validation testing of algorithms, bias testing to enhance fairness for training data and outcomes, and explainable HCAI user interfaces to enable inquiry and redress of grievances.

Management strategies for organizations begin by creating a safety culture with leadership commitment to safety that leads to better hiring practices and training oriented to safety. Other organizational strategies are extensive reporting of failures and near misses, internal review boards for problems and future plans, and alignment with industry standard practices. These strategies are promoted by internal committees, such as

Microsoft's Office of Responsible AI (https://www.microsoft.com/en-ca/ai/responsible-ai), or guidelines such as Google's Responsible AI Practices (https://ai.google/responsibilities/responsible-ai-practices/).

Trustworthy certification by industry, though subject to government interventions and regulation, can be done in ways that increase innovation. Other methods to increase trustworthiness include accounting firms to conduct independent audits, insurance companies to compensate for failures, non-governmental and civil society organizations to advance design principles, and professional organizations to develop voluntary standards and prudent policies.

5 Conclusions

The expansion of interest and application of AI research has triggered widespread public and government scrutiny of the ethical and responsible principles to guide designers, managers, and policy makers (Fjeld et al., 2020; IEEE, 2019). These principles are a useful starting point, but bridging the gap between ethics and practice requires realistic management processes.

This commentary endorses a Human-Centered AI approach for designing and developing systems that supports human self-efficacy, encourage creativity, clarify responsibility, and facilitate social participation. These foundational principles can help to guide designers towards vital technical goals, including privacy, security, fairness, reliability, safety, and trustworthiness.

The fresh ways of thinking in this commentary are based on a second Copernican Revolution that puts humans at the center of systems design thinking. It offers three ideas:

1. **High levels of human control AND high levels of automation are possible:** a two-dimensional HCAI framework that illustrates how it is possible to have high levels of human control AND high levels of automation.
2. **Shift from emulating humans to empowering people:** a plea to shift language, imagery, and metaphors away from portrayals of intelligent autonomous teammates towards descriptions of powerful tool-like appliances and tele-operated devices.
3. **Governance structures for HCAI:** a three-level governance structure that describes how software engineering teams can develop more reliable systems, how managers can emphasize a safety culture across an organization, and how industry-wide certification can promote trustworthy HCAI systems.

Promoting these three ideas is essential, but will not be easy. The entrenched beliefs about the human-like abilities of AI systems, which are held by some researchers, developers, journalists, and policy makers, slow progress towards a human-centered approach to new technologies.

The ideas in this commentary are meant to launch discussions that will advance global efforts such as the 17 United Nations Sustainable Development Goals (https://sustainabledevelopment.un.org/). Attaining these goals will require designers to combine technology developments with behavioral changes in order to improve healthcare and wellness, end poverty and hunger, and protect the environment. Well-designed technologies can support other UN Sustainable Development Goals such as to advance quality education, gender equality, global peace, and social justice.

Acknowledgments

Thanks to Editor in Chief Fiona Nah for guiding the review process and offering many helpful suggestions on writing and content. Thanks also to Gaurav Bansal, Dennis Galletta, Harry Hochheiser, Jonathan Lazar, Guru Madhavan, Avi Parush, Catherine Plaisant, Jennifer Preece, Lionel P. Robert, Steven M. Rosen, Daniel Shank, Keng Siau, Mark S. Smith, Wei Xu, and the anonymous reviewers for their supportive comments and thoughtful suggestions that greatly improved the draft versions. Thanks to Chris Carroll for thoughtful edits.

References

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B., & Winfield, A. F. (2017). Principles of robotics: Regulating robots in the real world. *Connection Science 29*(2), 124–129.

Chatzimparmpas, A., Martins, R. M., Jusufi, I., & Kerren, A. (2020). A survey of surveys on the use of visualization for interpreting machine learning models. *Information Visualization, 19*(3), 207-233.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1). https://cyber.harvard.edu/publication/2020/principled-ai

Hohman, F., Park, H., Robinson, C., & Chau, D. H. P. (2019). SUMMIT: Scaling deep learning interpretability by visualizing activation and attribution summarizations. *IEEE Transactions on Visualization and Computer Graphics, 26*(1), 1096-1106.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html and https://ethicsinaction.ieee.org/

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems 19*(6), 91–95.

Li, F.-F. (2018). How to make A.I. that's good for people. The *New York Times* (March 7, 2018). https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. New York, NY: Pantheon.

Mumford, L. (1934). *Technics and civilization*. Chicago: IL: University of Chicago Press.

Murphy, R. R. (2014). *Disaster robotics.* Cambridge, MA: MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.

Nourashrafeddin, S., Sherkat, E., Minghim, R., & Milios, E. E. (2018). A visual approach for interactive keyterm-based clustering. *ACM Transactions on Interactive Intelligent Systems), 8*(1), 1-35.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown Publishers.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. Cambridge, MA: MIT Press.

Robert, L. (2017). The growing problem of humanizing robots, *International Robotics & Automation Journal, 3*(1), 247-248.

Robert, L. P., Bansal, G., & Lütge, C. (2020). ICIS 2019 SIGHCI Workshop Panel Report: Human–Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence. *AIS Transactions on Human-Computer Interaction, 12*(2), 96-108.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*, New York: NY: Viking.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Sheridan, T. B. (2000). Function allocation: Algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies, 52*(2), 203–216.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Massachusetts Institute of Technology Cambridge Man-Machine Systems Lab.

Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.

Shneiderman, B. (2016). Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences, 113*(48), 13538-13540.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S. & Elmqvist, N. (2016). *Designing the user interface: Strategies for effective human-computer interaction*, Sixth Edition. Boston, MA: Pearson.

Shneiderman, B., (March 2020a). Human-centered artificial intelligence: Reliable, safe & trustworthy, *International Journal of Human-Computer Interaction, 36(*6), 495-504.

Shneiderman, B. (June 2020b). Design lessons from AI's two grand goals: Human emulation and useful applications, *IEEE Transactions on Technology and Society 1(*2), 73-82.

Shneiderman, B. (October 2020c). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems, *ACM Transactions on Interactive Intelligent Systems* (to appear).

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda, *Journal of Database Management, 30*(1), 61-79.

Yang, F., Huang, Z., Scholtz, J., & Arendt, D. L. (2020, March). How do visual explanations foster end users' appropriate trust in machine learning? In *Proceedings of the 25th ACM International Conference on Intelligent User Interfaces* (pp. 189-201).

About the Author

Ben Shneiderman is an Emeritus Distinguished University Professor in the Department of Computer Science, Founding Director (1983-2000) of the Human-Computer Interaction Laboratory (http://hcil.umd.edu), and a Member of the UM Institute for Advanced Computer Studies (UMIACS) at the University of Maryland. He is a Fellow of the AAAS, ACM, IEEE, and NAI, and a Member of the National Academy of Engineering, in recognition of his pioneering contributions to human-computer interaction and information visualization. His widely-used contributions include the clickable highlighted web-links, high-precision touchscreen keyboards for mobile devices, and tagging for photos. Shneiderman's information visualization innovations include dynamic query sliders for Spotfire, development of treemaps for viewing hierarchical data, novel network visualizations for NodeXL, and event sequence analysis for electronic health records. He is the lead author of *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed., 2016).

Photo at: http://www.cs.umd.edu/~ben/Photos/ben_6_10/5.jpg

Transactions on Human – Computer Interaction