

A Research Framework Focused on AI and Humans instead of AI versus Humans

Gerhard Fischer
Center for LifeLong Learning & Design (L3D)
University of Colorado, Boulder
gerhard@colorado.edu

1 Introduction

The arguments in this position paper are grounded in my professional career as a faculty member in Computer Science and Cognitive Science. For the last three decades, our research in the Center for Lifelong Learning & Design (L3D) has been centered on human-centered design, intelligence augmentation, and distributed cognition with a focus how to transcend the unaided individual human mind with socio-technical environments [Arias et al., 2016; Arias et al., 2001].

The theme of this workshop “*AI for Humans or Humans for AI*” does not have a simple answer [Markoff, 2016]. My arguments are focused to support the “AI for Humans” perspective [Fischer & Nakakoji, 1992; Shneiderman, 2022]. Our research activities [Fischer, 2021] and my contributions to previous CoPDA workshops explored problems beneficial to the needs of people, societies, and humanity by postulating “Quality of Life” as an overarching design objective [Fischer, 2018; Fogli et al., 2020], enriching the discourse about “AI for Humans” beyond a discussion of efficiency and productivity.

2 AI: What is it?

2.1 Differentiating AI Approaches

There is no generally accepted definition for AI and there is no defined boundary to separate “AI systems” from “non-AI systems”. Despite this shortcoming AI is currently being considered worldwide as a “deus ex machina” and it is credited with miraculous abilities to solve all problems and exploit all opportunities of the digital age. Figure 1 makes an attempt to unpack the meaning of AI into more specific research areas [Fischer, 2021] by differentiating between

- *Artificial General Intelligence (AGI)* is the envisioned objective to create intelligent agents that will match human capabilities for understanding and learning any intellectual task that a human being can. While some researchers consider AGI as the ultimate goal of AI, for others AGI remains speculative as no such system has been demonstrated yet. Opinions vary both on whether and when AGI will arrive, if at all.
- *AI for Specific Purposes (AISP)* is an engineering discipline that explores specific well-defined problems for which AI systems performs better than human beings. Many successful contributions have occurred in achieving these objectives providing the basis for the current hype surrounding AI. Human involvement is not a relevant design criteria in these approaches.
- *Human-Centered AI (HCAI)* (closely related to *intelligence augmentation* [Engelbart, 1995; Markoff, 2016]) is focused on improving the quality of life of humans by creating AI systems that amplify, augment, and enhance human performance with systems that are reliable, safe, and trustworthy [Shneiderman, 2022].

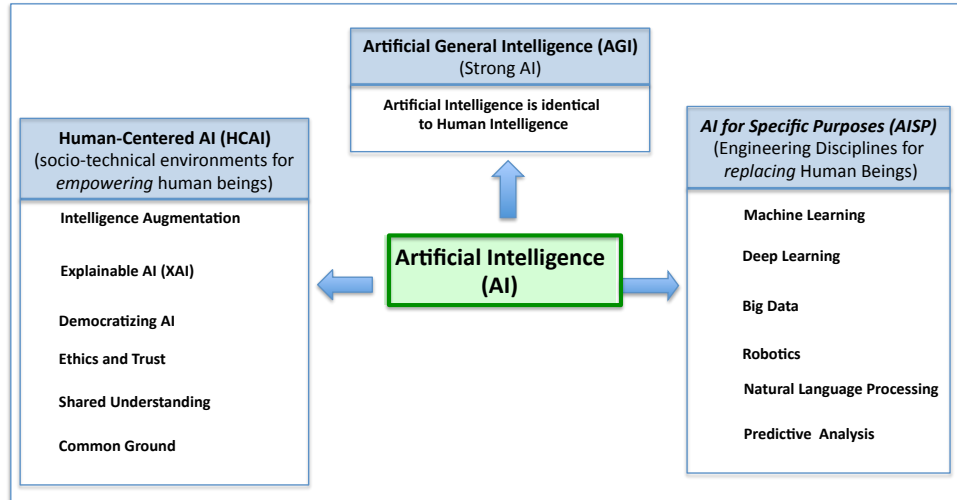


Figure 1: Differentiating AI Approaches

2.2 Contrasting “Humans for AI” versus “AI for Humans”

While the growth of technology is certain, the inevitability of any particular future is not. Contrasting “AI for Humans” versus “Humans for AI” represents an important objective to articulate design guidelines about the future of technological developments.

Frameworks centered on “**Humans for AI**” [Kurzweil, 2006] are grounded in objectives such as

- technological advances are more important than people;
- requiring people to work on technology’s terms;
- using people as stopgaps to do the parts of a task that machines can not yet do;
- restricting perspectives to “can we do it?” and ignoring challenges derived from the questions “should we do it?” by insufficiently considering potential drawbacks such as (a) the loss of meaningful work (b) the loss of personal control (if big data is watching us, how can we retain personal freedom?), and (c) an increase in the digital divide and inequality (those who own the data own the future).

In contrast frameworks centered on “**AI for Humans**” [Fischer & Nakakoji, 1992; Shneiderman, 2022] are grounded in objectives such as

- humans and computers are different therefore focusing on *complementing* rather than emulating and replacing human capabilities by computers;
- human-centered design, where the work starts with *understanding people’s needs* and capabilities;
- transcending the unaided individual human mind by exploring the potential of *distributed cognition*;
- *identifying* situations in which autonomous, intelligent technology should be deployed, often in areas characterized by the “three D’s”: dull, dirty, and dangerous; and
- sparking design efforts for *exploring a synthesis of humans and AI* by integrating their strengths and reducing their weaknesses as identified by a design trade-off analysis.

3 “AI and Humans” and “AI versus Humans”

Throughout history, there have always been two distinct forces at play: the *substituting* force, which replaced human workers and the *complementing* force which empowered human beings [Susskind, 2020].

3.1 Distributed Cognition: AI and Humans

A fundamental challenge for research in computer science, cognitive science, and the learning sciences is to understand thinking, learning, working, and collaborating by exploiting the power of omnipotent and omniscient technology. We need to understand what tasks should be reserved for educated human minds and the collaboration among different human minds, and what tasks

can and should be taken over or aided by cognitive artifacts. In an information-rich world, the true power comes not from more information, but from information that is personally meaningful, relevant to people's concerns, and relevant to the task at hand.

Distributed cognition [Hollan et al., 2001] is a fundamental framework by which to marry the intellectual power of the human mind with appropriate technologies. People think in conjunction and partnership with others and with the help of culturally provided tools [Salomon, 1993]. Distributed cognition complements our biological memory with external symbolic memory [Bruner, 1996] and extends the individual mind with the social mind. Distributed cognition transcends the individual, unaided human mind [Sloman & Fernbach, 2017] but it comes at a cost: external symbolic representations entail complex media that require extensive learning efforts by humans.

Many of our research efforts have addressed this challenge including:

- *domain-oriented design environments*, focused on supporting human problem-domain interaction and not only human-computer interaction [Fischer, 1994];
- the *Envisionment and Discovery Collaboratory*, supporting communities of interest in Renaissance communities with boundary objects [Arias et al., 2016]; and
- *context-aware systems* based on user and task models reducing information overload [Fischer, 2012].

"AI and Humans" as a research strategy is focused on complementing and augmenting human abilities with socio-technical systems for supporting more inclusive societies instead of increasing the digital divide [Fogli et al., 2020]. To be successful, *mutual understanding* represents an important challenge for the "AI and Humans" approach in order to overcome hurdles such as (1) the lack of self-knowledge (i.e., these systems are unaware what they know and not know) and (2) by being black boxes they are incapable of explaining how they reach their decisions in terms understandable to humans (e.g.: their reasoning is based on correlations derived from "Big Data" [Mayer-Schönberger & Cukier, 2013] whereas humans understand and argue based on causality).

3.2 Automation: AI versus Humans

Automation can be a two-edged sword:

- at one extreme, it is a servant, relieving humans of (1) carrying out personally irrelevant tasks (such as checking the results of simple calculations or spelling corrections), (2) wasting time with low-level operations (e.g.: programming in machine languages), (3) protecting them from dangerous activities (e.g.: using robots to find hidden bombs), and (4) freeing them for higher cognitive functions (e.g.: having cars with automatic transmissions);
- at the other extreme, automation can reduce the status of humans to that of 'button pushers', and can strip their work of its meaning and satisfaction. In personal meaningful activities, humans enjoy the process and not just the final product, and they want to take part in something [Fischer et al., 2000].

An early attempt leading to great expectations for AI systems replacing human beings was the development of *expert systems* in the 1980s [Buchanan & Shortliffe, 1984]. These developments provided the first phase of broad-based enthusiasm for automating of high-level human activities that would lead to substantial economic advantages. The expectations did not materialize, and subsequently researchers identified fundamental limitation of the expert systems approach [Winograd & Flores, 1986] that led to the "AI-Winter" in the following decade. An interesting question to be asked today in a new phase of AI enthusiasm is whether we will see another "AI-Winter" in the years to come?

4 Examples for Illustrating the Different Approaches

4.1 Adaptive versus Adaptable Systems

Adaptive systems are grounded in the "AI versus Humans" approach: they change their behavior by themselves driven by context-aware mechanisms including models of their users and specific task contexts, whereas *adaptable systems* are examples for the "AI and Humans" approach allowing

users to adjust, modify, and extend systems in order to capture unforeseen and missing aspects of problems.

Many research efforts have not clearly differentiated between adaptable and adaptive systems. Table 1 represents an initial effort to compare and differentiate the two approaches. Such a differentiation will be important and useful by identifying the *design trade-offs* between them, demonstrating the possibility for a *successful integration*, and analyzing the *impact* of these developments.

Table 1: A Comparison and Differentiation between Adaptive and Adaptable Systems

	Adaptive Systems	Adaptable Systems
Definition	modifications and suggestions generate by the systems for specific tasks and users	users actively change the functionality of the system
Knowledge	contained in the system; projected in different ways	knowledge is curated, modified, and extended by users
Strengths	little (or no) effort by users; no special user knowledge is required; work <i>for</i> people	users are in control; users know their tasks best; work <i>with</i> people
Weaknesses	users lack control; common understanding is reduced resulting in filters bubbles; lack of explainability	users must do substantial work; require a learning effort; create a tool mastery burden; systems may become incompatible
Mechanisms required	models of users, tasks, and dialogs; big data resources; intelligent agents	meta-design environments supporting modifiability, tailorability, and evolution
Application domains	active help systems, critiquing systems, recommender systems	open systems, co-designed systems, end-user development
Primary Techniques	automation grounded in Artificial Intelligence (AI) approaches	human involvement grounded in Intelligence Augmentation (IA) approaches

4.2 Learning Environments

Making learning part of life is a necessity rather than a possibility or a luxury to be considered for addressing the complex, systemic problems occurring in a world undergoing constant change.

Different kinds of problems require different kinds of learning approaches and different socio-technical environments supporting these approaches. Outside the classroom, much learning and problem solving takes place as individuals explore personally meaningful problems, engage with each other in collaborative activities while making extensive use of media and technologies.

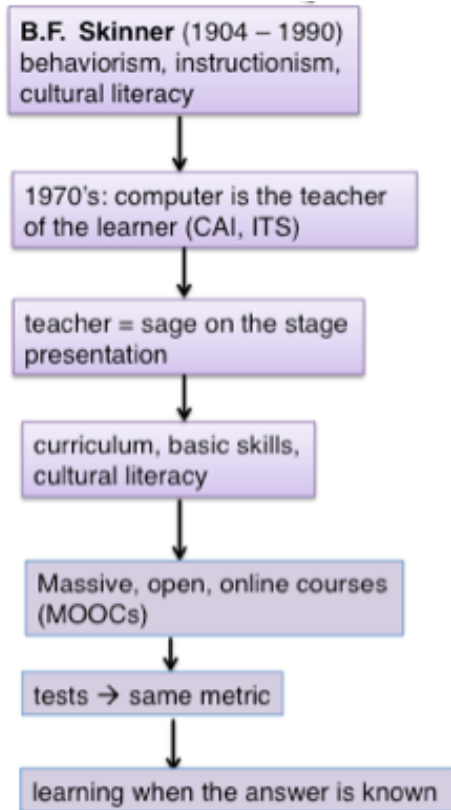
In classroom environments instructionist approaches dominate and learning is conceptualized as an isolated process of information transmission and absorption whereas outside of schools learning is a much more complex activity. Computational environments from the early beginnings have been conceptualized and employed to support human learning in these two different settings and two fundamentally different approaches have emerged:

- *intelligent tutoring systems* [Anderson et al., 1995], in which the problem is given by the teacher or the system, and
- *interactive learning environments* [Papert, 1980], in which tools are provided that allow learners to explore problems of their own choice.

Intelligent tutoring systems can provide substantial more support because the designers of the environments know (at design time) the types of problems the learners will work on (at use time). To support learners in interest-driven, self-directed activities, interactive learning environments need to be augmented with mechanisms (such as domain-oriented design environments, critiquing

systems, and context-awareness) that can offer help and support for learners who get stuck or who do not know how to proceed.

Instructionism: Automating the Teacher with Intelligent Tutoring Systems



Constructionism: Empowering the Learner with Interactive Learning Environments

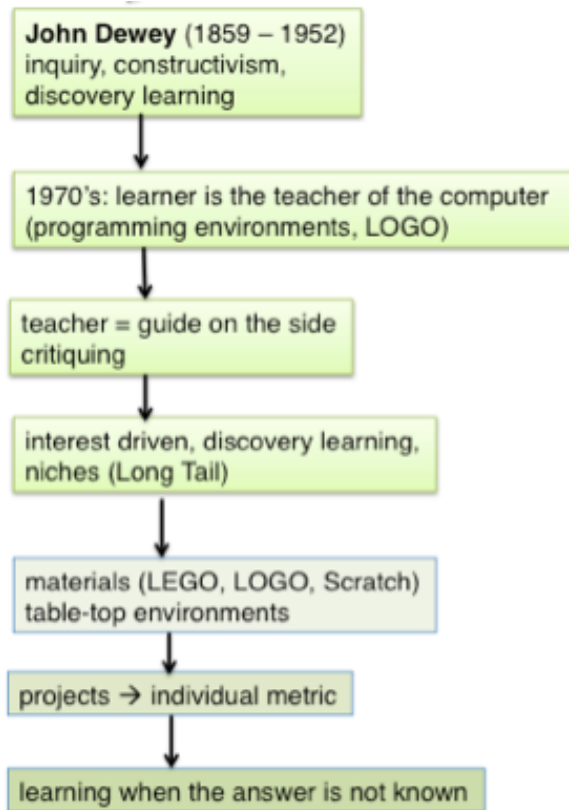


Figure 2: A Comparison of Intelligent Tutoring Systems and Interactive Learning Environments

5 Research Challenges Associated with the “AI and Humans” Framework

Arguing for the strong preference in our own research for a framework grounded in the objective “AI and Humans”, it should not be overlooked that this framework presents several important pitfalls [Fischer, 2018] that require careful attention and further exploration including:

- **overreliance:** despite all the technological support for humans in a distributed cognition framework, which capabilities do humans need to learn to avoid overreliance on external tools? How can “tools for living” and “tools for learning” be differentiated in specific contexts?
- **deskilling:** will humans lose (1) basic mathematical capabilities by using hand-held calculators; (2) the ability to spell by using spelling correctors; (3) important geographical knowledge by using navigation systems; and (4) the motivation learning a foreign language by using automated translation systems?
- **learning demands associated with powerful and complex tools:** will AI technologies that empower human beings in distributed cognition approaches require reasonable learning efforts for humans to understand the possibilities and the limitations of these tools?
- **establishing different discourses:** will discourses and investigations facilitated and supported by “AI and Humans” technologies provide opportunities for exploring motivation, control, ownership, autonomy, and quality of life?

- **quality of life:** will “AI and Humans” approaches provide us with more time, less stress, and more control or will they lead to participatory overload problems by requiring the engagement in problems that individuals consider irrelevant for them;

For all these research issues that are no simple answers, only *design trade-offs* [Fischer, 2018]. And because *there are no decontextualized sweet spots* for analyzing these design trade-offs, the investigations must be situated and explored in specific contexts.

6 The Past, the Present, and the Future of the CoPDA Workshops

The AVI'2022 workshop is the 6th CoPDA workshop (see Figure 3). An important challenge for the researchers getting together in the workshop this year may be to explore the foundational idea(s) that these workshops have pursued and how they are related to each other. A particular objective of all previous CoPDA workshop has been to *collectively identify important and interesting themes for future workshops* and my hope is that this will happen again this year by exploring *post-AI attitudes* prioritizing human well-being and quality of life as primary objectives.

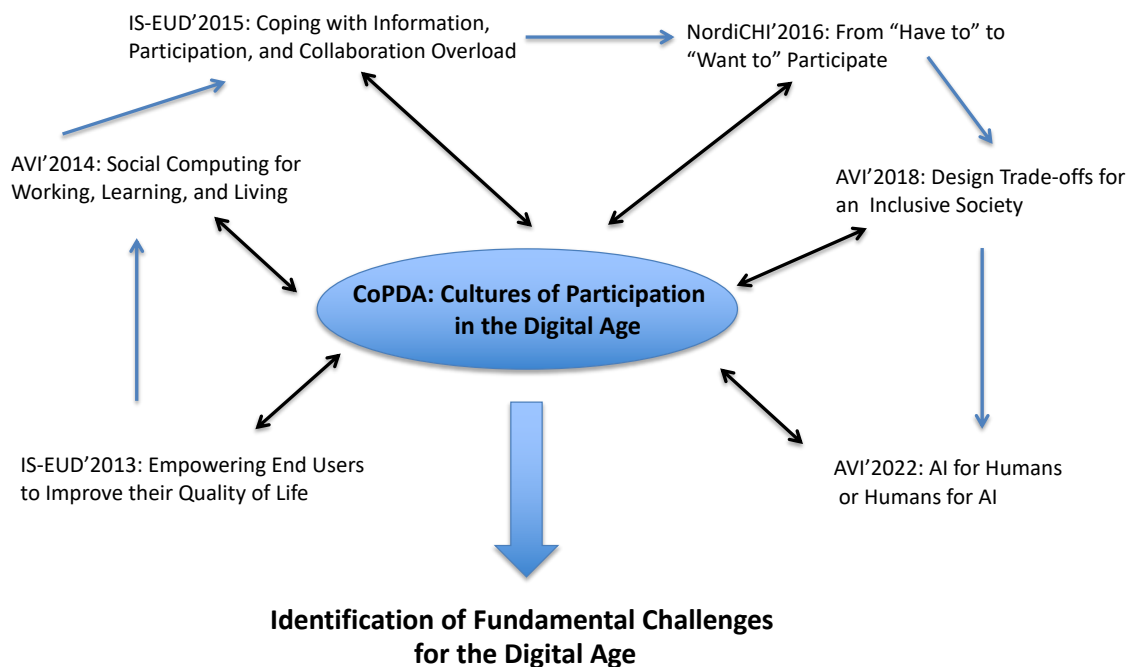


Figure 3: An Overview of the CoPDA Workshops

7 Conclusions

We are in a period of major changes in technology, impacting almost all areas of human lives. The world-wide euphoria about AI based on increases in computational and communication power, the advent of ubiquitous sensors supporting the Internet of Things, and powerful new software tools are changing education, work, healthcare, transportation, industry, manufacturing, and entertainment.

The impact of these changes upon people and society is both *positive* and *negative*. The *positive* impacts should be celebrated, and the *negative* impacts should be avoided rather than treated as unfortunate but unavoidable side effects. Future research needs to identify the positive and negative effects and provide evidence for the success and failure of specific developments.

We need new ways of thinking and new approaches in which we address the *basic question* associated with the themes “AI and Humans” and “AI versus Humans”: (1) which tasks or components of tasks are or should be reserved for educated human minds aided by cognitive artifacts (distributed cognition), and (2) which tasks can and should be taken over by AI systems acting independently (automation)?

8 References

- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995) "Cognitive Tutors: Lessons Learned," *The Journal of the Learning Sciences*, 4(2), pp. 167-207.
- Arias, E. G., Eden, H., & Fischer, G. (2016) *The Envisionment and Discovery Collaboratory (EDC): Explorations in Human-Centered Informatics*, Morgan & Claypool Publishers, San Rafael, California.
- Arias, E. G., Eden, H., Fischer, G., Gorman, A., & Scharff, E. (2001) "Transcending the Individual Human Mind—Creating Shared Understanding through Collaborative Design" in J. M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium*, ACM Press, New York, pp. 347-372.
- Bruner, J. (1996) *The Culture of Education*, Harvard University Press, Cambridge, MA.
- Buchanan, B. G. & Shortliffe, E. H. (Eds.) (1984) *Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley Publishing Company, Reading, MA.
- Engelbart, D. C. (1995) "Toward Augmenting the Human Intellect and Boosting Our Collective IQ," *Communications of the ACM*, 38(8), pp. 30-33.
- Fischer, G. (1994) "Domain-Oriented Design Environments," *Automated Software Engineering*, 1(2), pp. 177-203.
- Fischer, G. (2012) "Context-Aware Systems: The 'Right' Information, at the 'Right' Time, in the 'Right' Place, in the 'Right' Way, to the 'Right' Person" in G. T. e. al (Ed.), *Proceedings of the Conference on Advanced Visual Interfaces (AVI)*, ACM, Capri, Italy (May), pp. 287-294.
- Fischer, G. (2018) "Design Trade-Offs for Quality of Life " *ACM Interactions* 25 (1), pp. 26-33.
- Fischer, G. (2021) "End-User Development: Empowering Stakeholders with Artificial Intelligence, Meta-Design, and Cultures of Participation" in D. Fogli, Tetteroo, D., Barricelli, B.R., Borsci, S., Markopoulos, P., Papadopoulos, G.A. (Ed.), *Is-EUID 2021 Proceedings*, Springer, LNCS 12724, pp. 3–16.
- Fischer, G., Greenbaum, J., & Nake, F. (2000) "Return to the Garden of Eden? Learning, Working, and Living," *The Journal of the Learning Sciences*, 9(4), pp. 505-513.
- Fischer, G. & Nakakoji, K. (1992) "Beyond the Macho Approach of Artificial Intelligence: Empower Human Designers - Do Not Replace Them," *Knowledge-Based Systems Journal, Special Issue on AI in Design*, 5(1), pp. 15-30.
- Fogli, D., Piccinno, A., Carmien, S., & Fischer, G. (2020) "Exploring Design Trade-Offs for Achieving Social Inclusion in Multi-Tiered Design Problems," *Behaviour & Information Technology*, 39(1), pp. 27-46.
- Hollan, J., Hutchins, E., & Kirsch, D. (2001) "Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research" in J. M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium*, ACM Press, New York, pp. 75-94.
- Kurzweil, R. (2006) *The Singularity Is Near*, Penguin Books.
- Markoff, J. (2016) *Machines of Loving Grace (the Quest for Common Ground between Humans and Robots)*, Harpercollins
- Mayer-Schönberger, V. & Cukier, K. (2013) *Big Data*, Houghton Mifflin Harcourt, New York, NY.
- Papert, S. (1980) *Mindstorms: Children, Computers and Powerful Ideas*, Basic Books, New York.
- Salomon, G. (Ed.) (1993) *Distributed Cognitions: Psychological and Educational Considerations*, Cambridge University Press, Cambridge, United Kingdom.
- Shneiderman, B. (2022) *Human-Centered Ai* Oxford University Press.
- Sloman, S. & Fernbach, P. (2017) *The Knowledge Illusion — Why We Never Think Alone*, Riverhead Books, New York.
- Susskind, D. (2020) *A World without Work: Technology, Automation, and How We Should Respond*, Metropolitan Books/Henry Holt & Company, New York, N.Y.
- Winograd, T. & Flores, F. (1986) *Understanding Computers and Cognition: A New Foundation for Design*, Ablex Publishing Corporation, Norwood, NJ.