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

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 firmly grounded in "AI for Humans". Our research activities and my contributions to previous CoPDA workshops explored problems more beneficial to the needs of people, societies, and humanity by postulating "Quality of Life" as an overarching design objective [Fischer, 2018], enriching the discourse about "AI for Humans" beyond a discussion of efficiency and productivity

2 Contrasting AI for Humans versus Humans for AI

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 considering seriously 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] are grounded in objectives such as

- humans and computers are different 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, unsafe;
- sparking design efforts for *exploring a synthesis of humans and AI by* integrating their strengths and premises rather than their weaknesses and perils as identified by a design trade-off analysis.

Throughout history, there have always been two distinct forces at play: the substituting force, which harmed workers, but also the helpful complementing force, which did the opposite [Susskind, 2020].

3 Distributed Cognition — Humans and AI

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 such 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. Printed media serve only as representational media in this context whereas computational media have the power to serve as interpretive media. 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 [Fischer, 1994];
- the Envisionment and Discovery Collaboratory, supporting communities of interest with boundary objects and supporting not only Renaissance scholars but Renaissance communities [Arias et al., 2016];
- context-aware systems to reduce information overload [Fischer, 2012].

We need new ways of thinking and new approaches in which we address the basic question associated with distributed intelligence and the design of sociotechnical systems: Which tasks or components of tasks are or should be reserved for educated human minds, and which can and should be taken over or aided by cognitive artifacts?

4 Learning Environments: An Example for Illustrating the Different Approaches

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

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 places as individuals explore personally meaningful problems, engage with each other in collaborative activities while making extensive use of media and technologies. Many past educational systems have been built on the assumption

- that teaching is necessary for learning to occur;
- that teaching and learning is inherently linked; and
- that a curriculum can and should be developed to create a cultural literacy.

In such a culture, teachers taught learners about the world and learning was conceptualized as an isolated process of information transmission and absorption. It ignored the fact that in today's world, more and more knowledge, especially advanced knowledge, is acquired well past the age of formal schooling, and in many situations through educational processes that do not center on the traditional school.

From the very early beginnings that computational environments have been employed to support human learning, 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 (such as LOGO [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). In interactive learning environments, little support is given when a learner is stuck since it supports autonomous learning. In order to support self-directed learning, they 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 when the information needs to be contextualized to the task at hand and to the learner's needs and interests.

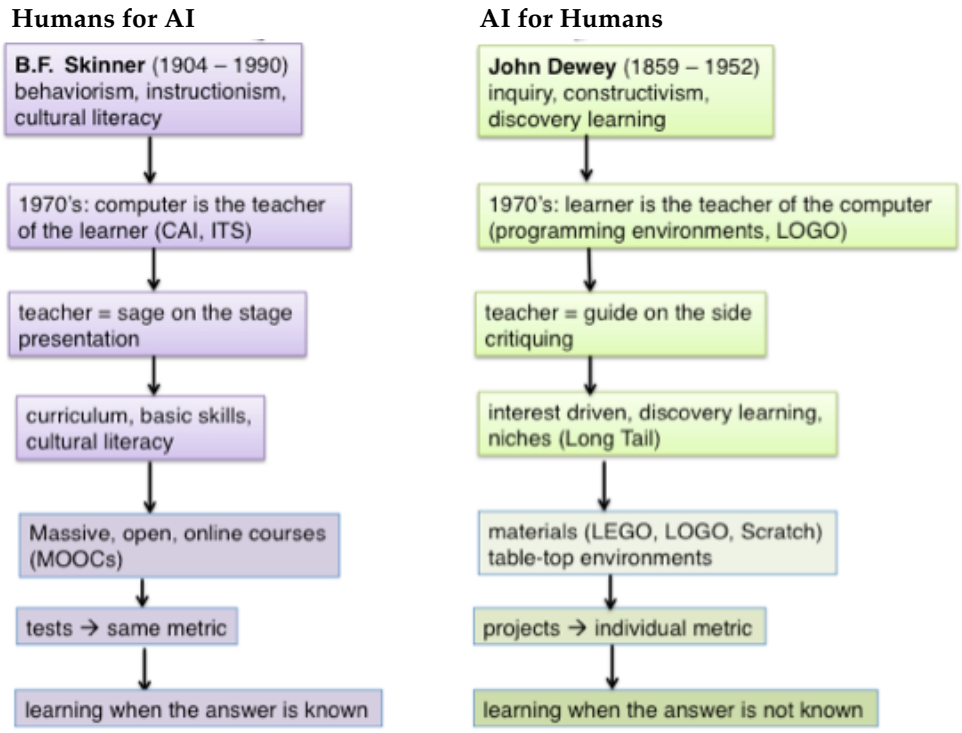

Figure 1: **Instructionism (Intelligent Tutoring Systems)** **Constructionism (Interactive Learning Environments)**

5 Research Challenges Associated with the Framework "Humans and AI"

Arguing for the strong preference in our own research for a framework grounded in the basic objective "Humans and AI", it should not be overlooked that this framework presents a number of important design trade-offs that require careful attention and further exploration including:

- **overreliance:** despite all the technological support for humans in a distributed cognition framework, which capabilities should humans learn to avoid overreliance on external tools? How should we differentiate between "tools for living" (the … and "tools for learning" (…) in specific contexts?
- **deskilling**: by using (1) hand-held calculators will humans loose basic mathematical capabilities; (2) spelling correctors will humans loose the ability to spell; (3) navigation systems will humans loose important geographical knowledge; and (4) translation systems will humans avoid the effort to learn a foreign language
- **participation overload:** in the context of meta-design will the support for active engagement lead to participation overload (particularly in personally irrelevant activities?
- **learning demands associated with powerful and complex tools:** to exploit the benefits of AI technologies that empower humans beings in distributed cognition approaches (rather them replace them) requires often substantial learning efforts for humans to understand the possibilities and the limitations of these tools
- **quality of life:** will the AI technologies provide us with more time, less stress, more control or will they cause a shift in authority from from humans to algorithms

(especially in case of tools that we do not understand and that cannot provide us with explanations anout their actions);

- **establishing different discourses:** to deeply understand the potential transformation of human lives enriched rather than limited by AI technologies, our discourses and investigations must not only be focused around technological issues but explore motivation, control, ownership, autonomy, and quality because changes in complex environments are not primarily dictated by technology but they are the result of an incremental shift in human behavior and social organization; they require a co-design of social and technical systems, and use models and concepts that focus not only on the artifact but exploit the social context in which the systems will be used.

For all these research issues that are no simple answers, only design trade-offs. And because *there are no decontextualized sweet spots* to analyze 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'2020 workshop is the 6th CoPDA workshop (see Figure 2 for an overview). 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. My claim: all of the workshops have identified basic research challenges derived from real problems. Such an effort could lead to the articulation of a coherent and important theme(s), an edited book, or a EU research project.

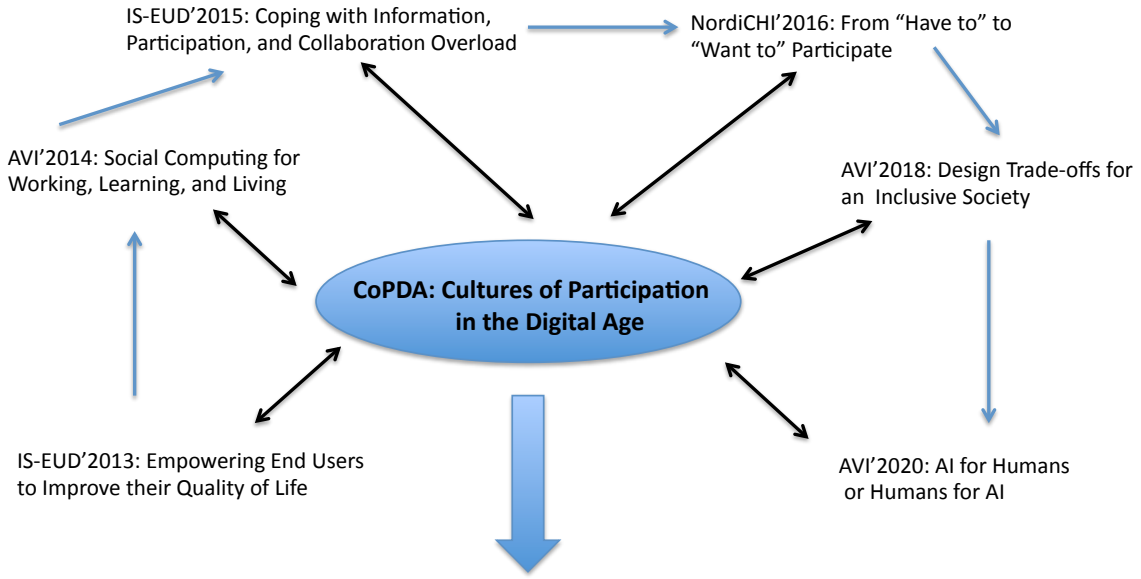

Figure 2: An Overview of the CoPDA Workshops

7 Conclusions

We are in a period of major changes in technology, impacting almost all areas of human lives. The world-wide euphoria about artificial intelligence 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*. Although the positive impacts are celebrated, the negative impacts are often treated as unfortunate but

unavoidable side effects. Suppose instead we adopt the view that these negative side effects are so severe that we need different frameworks for designing the future of digitalization.

Today, much of our technology is designed through a technology-centered approach. Basically, technologists and technology companies invent and design what they can but then leave many tasks that cannot be done by machines (yet??) to people, thereby forcing people to work on the technology's terms.

Technological developments facilitate activities that could not be done before — e.g.: (1) the Internet making MOOCs a reality to reach ten thousands or more people; (2) the development of sophisticated driver-assistance systems and/or self-driving cars (currently limited to very restricted environments) increasing traffic safety.

With exciting new possibilities on the horizon, it is critically important not only to ask and critically assess *"can we do something"* but *"should we do something"* (e.g.: to delegate decision about life and deaths to algorithm in automated warfare) should be considered as the central question, requiring that issues derived from ethics, values, impact, control, and autonomy are taken into account. The future depends on ourselves, and we should not depend on technological developments.

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.

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. Tortora, S. Levialdi, & M. Tucci (Eds.), *Proceedings of the Conference on Advanced Visual Interfaces (AVI 2012),* ACM, Capri, Italy (May), pp. 287-294.

Fischer, G. (2018) "Design Trade-Offs for Quality of Life " *ACM Interactions* XXV.1 (January + February 2018), pp. 26-33.

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.

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

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.

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.