

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

Abstract. Despite lacking a shared understanding and a generally accepted definition, Artificial intelligence (AI) is promoted and credited with miraculous abilities to solve all problems. To gain a more nuanced and deeper understanding of the design trade-offs associated with AI, this paper proposes a research framework that contrasts two competing frameworks: (1) AI *versus* Humans (characterized by strong AI and Artificial General Intelligence) focused on replacing human beings and (2) AI *and* Humans (characterized by intelligence augmentation and human-centered AI) focused on empowering human beings as individuals and communities.

The arguments in the paper are supported by research activities that explored conceptual frameworks and inspiring prototypes. These developments have resulted in gaining a deeper understanding of how AI-type systems can contribute to quality of life aspects with a specific focus on rethinking and reinventing learning, education, working, and collaboration in the digital age.

Keywords: AI, AI versus Humans, AI and Humans, Intelligence Augmentation, Quality of Life, Design Trade-offs, ChatGPT

1 Introduction

The theme of this special issue based on the Sixth International Workshop on Cultures of Participation in the Digital Age (CoPDA'2022) "*AI for Humans or Humans for AI*" [1] does not have a simple answer [2]. The frameworks and arguments of this paper are focused on supporting the "*AI for Humans*" perspective [3, 4]. Our research activities explored problems beneficial to the needs of people, societies, and humanity by postulating "quality of life" as an overarching design objective [5, 6], enriching the discourse about AI beyond a discussion of efficiency and productivity.

The paper does not provide prophecies or definite answers but by exploring spaces of different possibilities tries to articulate questions, challenges, design trade-offs, frameworks, and inspirational prototypes grounded in our research experience for contributing to the foundation of a scientific discourse about the future of AI with a focus on how to transcend the unaided individual human mind with socio-technical environments [7, 8].

Section 2 attempts to characterize AI by describing different AI approaches. The initial differentiation is further elaborated by the two major themes *AI versus Humans* in Section 3 and *AI and Humans* in Section 4. Section 5 illustrates the different approaches with some specific examples and Section 6 identifies the promises and pitfalls of using ChatGPT in educational settings. The final two sections describe some lessons learned and summarize the overall implications of the paper.

2 AI: What is it?

During the last few years, a lot has been written about AI and new AI research centers have been created around the world.

Most of the currently existing AI approaches and visions are over-inclusive, discuss AI in very general terms, and reflect the common tendency to use the “AI” label as a catchall marketing phrase. AI is a “*suitcase word*” to which people attribute multiple meanings [9] and it does not have a simple meaning and should not be used as if it does.

AI is too poorly defined, too interesting, too ubiquitous, to be left to the hype spread by utopists and politicians. Early attempts to create an understanding what AI might be with the development of prototypes (as documented in the “first AI book” [10]) received little credit for their innovative contributions and have subsequently been reduced to the status of being additional contributions to computer science. Many of these contributions were later understood well enough and became useful and common enough, and subsequently, they were not labeled AI anymore.

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 world-wide 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 attempts to unpack the meaning of AI into more specific research areas [11] 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 perform 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 criterion in these approaches.
- *Human-Centered AI (HCAI)* [4] (closely related to *intelligence augmentation* [2, 12]) 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.

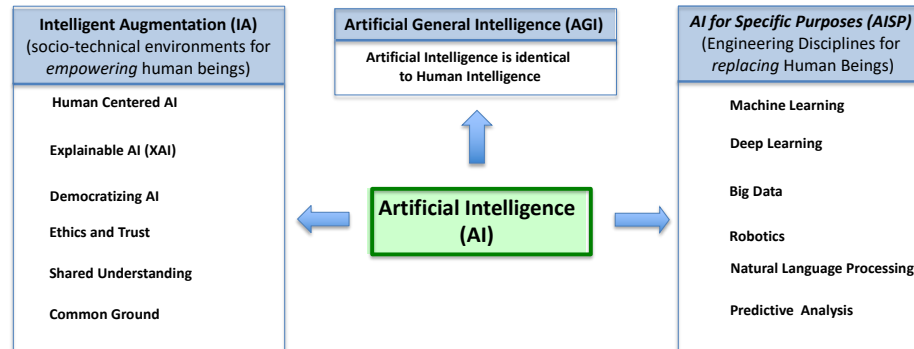


Figure 1: Differentiating AI Approaches

2.2 Characterizations and Visions of AI by Optimists, Pessimists, and Realists

The current views and the goals and objectives of future developments of AI can be differentiated into three major categories (the edited book by [13] contains over one hundred short opinion pieces of prominent AI researchers and critics that address these different perspectives):

- *AI Optimists* (often articulating utopians scenarios for future developments) [14] believe that AGI is a realistic and desirable goal for the not too distant future and advocating trans-humanism and singularity are desirable and inevitable objectives. In the past optimists have often spread hype about the potential of developments including: (1) AI based claims about expert systems [15], (2) the power of e-memories [16], and (3) the unique educational opportunities of MOOCs [17] which subsequently turned out to be unrealistic objectives.
- *AI Pessimists* argue that AI has failed and the objectives of AGI are dangerous for the future of the human race [18] and several public figures expressed concern that AGI could result in human extinction [19].
- *AI Realists* assert that there is on one hand substantial progress in pursuing and incorporating AI approaches addressing fundamental societal problems and on the other hand there are just as many unsolved problems for which human intelligence will be far superior to artificial ones for decades to come [4]. They address the reluctance in many segments of society to allow computers to take over tasks that simple models perform demonstrably better than humans. Their efforts are focused not on AGI but on the improvement of super-tools in the AISP domains and using HCAI for creating supportive socio-technical environments. They argue that we have to shift our view of progress, putting

the emphasis on social and personal flourishing rather than technological enhancements.

The argumentation and focus of this paper is grounded in the basic assumption that the objectives of the “AI Realists” are the most promising way to pursue.

3 AI versus Humans

Throughout history, there have always been two distinct forces at play: the *substituting* force, which replaced human workers by automation and the *complementing* force which empowered human beings [20, 21]. 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), (4) freeing them for higher cognitive functions (e.g.: having cars with automatic transmissions) and (5) identifying situations in which autonomous, intelligent technology should be deployed, often in areas characterized by the “three D’s”: dull, dirty, and dangerous;
- 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 them.

An early attempt leading to great expectations for AI systems replacing human beings was the development of *expert systems* in the 1980s [15]. These developments provided the first phase of broad-based enthusiasm for automating high-level human activities that would lead to substantial economic advantages. The expectations did not materialize, and subsequently, researchers identified fundamental limitations of the expert systems approach [22] 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?

The issues, open questions, and ethical issues associated with an “AI versus Humans” framework can be illustrated with a few specific examples:

- **Automatic line calling systems in tennis:** these systems have become standard (and obligatory) use in major tennis tournaments eliminating human line judges. They use multiple cameras to track the ball's trajectory and, within milliseconds, determine whether the ball landed in or out. The general assumption (and partial verification) is that these systems are more accurate and surpass human abilities to make the right decision for this specific task. The decision reached by the automatic system is final and cannot be challenged by the players or the referee.
- **Self-driving cars:** In contrast to automatic line calling systems in tennis, self-driving cars still represent a design option to be fully realized in the future. While the AI optimists argue for numerous potential advantages (e.g.: increased safety, and accessibility for those unable to drive, such as the elderly, or visually impaired), the AI pessimists believe that this objective is neither desirable nor achievable in the near future, and many AI realists argue that intelligent driver support systems (grounded in an “AI and Human”

framework) represent the preferable strategy to pursue [23]. Another limitation of many efforts and articles focused on self-driving cars is that they are treated merely as an isolated technological solution to the problems for the fundamental problem of “Mobility for All in the Future” [5].

- **Using AI in grading student exams:** this has been done for a long time, but it poses numerous questions. It is best applicable to questions that are represented as multiple choice questions and for which right and wrong answers exist whereas questions that require critical thinking, creativity, or deep understanding, AI-based systems have severe limitations to provide accurate grading and meaningful feedback to students [24].
- **Automatic warfare:** from the four examples this goal is the most serious and questionable design objective for an “AI versus Human” approach which raises fundamental ethical questions [25]. In this context asking the technological question “can we do it” should definitely give priority to the moral and ethical questions “should we do it”?

Many of the current “AI versus Humans” systems have 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.: in many cases their reasoning is based on correlations derived from “Big Data” [26] whereas humans understand and argue based on causality).

Frameworks centered on “*AI versus Humans*” [14] 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).

4 AI and Humans

Frameworks centered on “**AI and Humans**” [3, 4] 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 starting with *understanding people’s needs* and capabilities;
- transcending the unaided individual human mind by exploring the potential of *distributed cognition*; 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.

4.1 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 (AI versus Humans) or aided (AI and Humans) 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. The big question for the years to come will be: "How will the line be drawn between humans and machines, and who will draw it?"

Distributed cognition [27] is a fundamental framework marrying 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 [28]. Distributed cognition complements our biological memory with an external symbolic memory [29] and extends the individual mind with the collective mind.

Figure 2 conceptualizes and illustrates a long-term historical perspective (with a broad qualitative representation) of how individual minds were complemented by collective human minds aided by technologies.

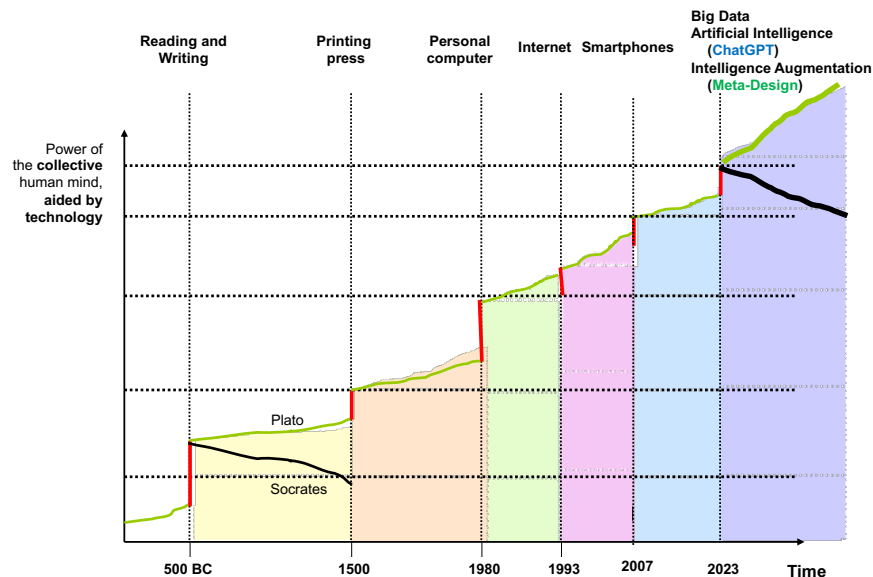


Figure 2 Transformative Technologies and Cultural Changes Supporting Distributed Cognition

While these transformative technologies were revolutionary inventions that transformed society in many positive ways, they also brought challenges (e.g.: external symbolic representations entail complex media that require extensive learning efforts by humans [30]) and unintended consequences illustrate the complex dynamics of

technological change. Table 1 briefly summarizes by example the opportunities and unintended consequences for “Reading and Writing” and “The Printing Press”. The changes in Figure 2 can be analyzed because they have taken place in the past. Section 6 of the article explores the design challenges and trade-offs of AI that we face today for increasing “quality of life”.

Table 1: Opportunities and Unintended Consequences of Technological Changes

Technology	Cultural Changes	Opportunities	Unintended Consequences
Reading and Writing	Oral Societies → Literal Societies	writing supports the preservation of knowledge (Plato)	writing will lead to a weakening of the human mind (Socrates)
Printing Press	democratization of knowledge	share and scrutinize work more broadly and effectively	spread of misinformation; cultural homogenization

4.2 Making AI more helpful for everyone—embedding AI features in applications

“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 [6].

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 [31];
- the *Envisionment and Discovery Collaboratory*, supporting communities of interest and Renaissance communities with boundary objects [8]; and
- *context-aware systems* based on user and task models reducing information overload [32].
- *socio-technical* environments creating more inclusive societies thereby allowing people of all ages and all abilities to exploit information technologies for personally meaningful activities [6].

Our research activities have always considered the question “*Will AI replace Humans?*” (i.e. teachers, decision-makers, radiologists, ...) as the less relevant question compared to the question “*Will Humans who use AI be better informed, and more creative than humans who don’t use AI?*”

On a much larger scale (but illustrating the same objective), Google in its 2023 I/O keynote session (<https://io.google/2023/program/396cd2d5-9fe1-4725-a3dc-c01bb2e2f38a/>) illustrated its ideas and plans to embed AI features in numerous tools that millions of people use every day to “make AI more helpful for everyone” by “boosting creativity and productivity”.

4.3 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 [5] that require careful attention and further exploration.

Over-reliance. Despite all the technological support for humans in a distributed cognition framework, which capabilities do humans need to learn to avoid over-reliance on external tools? How can “tools for living” and “tools for learning” be differentiated in specific contexts [33]?

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; (4) the motivation to learn a foreign language by using automated translation systems? How will motivation, control, ownership, and autonomy be impacted by deskilling [21, 34]

Learning demands are 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 super-tools [4]? What should students learn about AI and what can they learn being supported by AI.

5 Examples and Frameworks for Illustrating the Different Approaches

Our framework juxtaposing “AI and Humans” and “AI versus Humans” is illustrated in this section by three specific themes: (1) design trade-offs between the two approaches, (2) the merits, challenges, and pitfalls associated with adaptive and adaptable systems, and (3) the choosing learning and education as a paradigmatic example to exemplify the arguments describes in the previous sections.

5.1 No Best Solutions—Only Design trade-offs

Design is choice: it is an argumentative process with no optimal solutions [35].

There are no simple answers only *design trade-offs* to explore the merits of the different approaches “AI versus Humans” in comparison to “AI and Humans”. And because *there are no decontextualized sweet spots* for analyzing these design trade-offs, the investigations must be situated and explored in specific contexts.

Some of the major design trade-offs that we have explored are [5]:

- will *distributed cognition* support the unaided human mind or will it lead to an over-reliance on external tools thereby suppressing the acquisition of basic knowledge and skills (e.g.: navigation systems may have a negative impact on geographical knowledge)?
- will *interest-driven learning* lead to insufficient exposure to basic skills?

- will **communication tools** increase and widen the collaboration with other learners, or will it increase social isolation because individuals will be sitting at home in front of a computer and interacting less with other people?
- will **personalization** reduce the information overload problem, or will it create filter bubbles preventing people from understanding each other [36]?
- will digital technologies widen the “**digital divide**” or reduce it by making educational opportunities available to many more learners of all economic classes and in all parts of the world (e.g. with Massively Open Online Courses (MOOCs))?
- will excessive reliance on technology lead to “**over-reliance**”, where individuals become dependent on technology and begin to rely less on their own cognitive abilities leading to a decline in cognitive abilities over time if individuals do not use and exercise their cognitive abilities regularly?
- will “**foundation models**” offer significant benefits in terms of augmentation, efficiency, and innovation, or will their application have negative impacts in terms of human skill development, misuse of data, and insufficient attention to ethical concerns?

A careful analysis of design trade-offs can provide the foundation for identifying the best possible mixes between different approaches.

5.2 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 of the “AI and Humans” approach allowing users to adjust, modify, and extend systems in order to capture unforeseen and missing aspects of problems [37].

Many research efforts have not clearly differentiated between adaptable and adaptive systems. Table 1 represents an 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 2: 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

	Adaptive Systems	Adaptable Systems
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

5.3 Learning Environments

AI in education initiatives and computational environments to support human learning have been developed from the early beginnings grounded in the two different frameworks discussed in this paper.

Instructionist approaches and *intelligent tutoring systems* [38] have been grounded primarily in the “AI versus Humans” approach whereas *constructionist approaches* and *interactive learning environments* [39, 40] have been grounded primarily in the “AI and Humans” approach. Table 2 characterizes and contrasts the major dimensions of the two approaches.

Table 3: Instructionism versus Constructionism

	Instructionism	Constructionism
focus	basic skills, cultural literacy; presentation of an organized body of knowledge (basic skills) computer is the teacher of human learners	problem is based on the learner’s needs and interests; learner is the teacher of the computer (e.g.: with programming)
Inventor contributor	B.F. Skinner (1904 – 1990)	John Dewey (1859 -1952) Seymour Papert (1928-2016)

	Instructionism	Constructionism
characteristics	problem is given by the teacher or the systems; learning supported from the supply side; adult-run education; prescriptive	learning supported from the demand side; child-run education; permissive
strengths	organized body of knowledge; pedagogically and cognitively structured delivering instruction efficiently and effectively	real interests, personally meaningful tasks, high motivation providing students with opportunities to set their own goals, build on their own interests
weaknesses	limited relevancy to the interests of the learner or the task at hand	coverage of important concepts may be missing; unstructured learning episodes; lack of coherence
primary role of teachers	sage on the stage—presents what they know and are prepared for	guide on the side – confronted with unfamiliar problems
planning versus situated responses	anticipating and planning of the learning goals and content	learning needs arise from the situational context
distribution over lifetime	decreasing in importance from school to university to lifelong learning	increasing in importance from school to university to lifelong learning
assessment	“standard” assessment instruments (same metric for all students) are applicable	“innovative” assessment instruments (individual metrics) are needed

AI tutors work best on problems that are highly structured and well-defined (an example being a new AI tutor called Khanmigo from Khan Academy [41]). Instructionist AI tutors are best suited for problems that can be precisely formulated, have a well-defined answer and strategies for getting to the answer.

Constructivist approaches are more suited for “wicked, ill-defined, and open” problems [42] for which framing the problem is as important as solving it. Learners should be exposed to these kinds of problems that have no final answers but require the exploration of design trade-offs (see Section 5.1). To foster and support these transitions, new models of learning and teaching are needed, not just technology overlays on old modes of teaching and learning [43, 44]. Going to school is more than getting exposed to new information but it is participating in communities. “Learning about” needs to be complemented by “learning to be”[45, 46].

6 ChatGPT

The newest AI development is *Generative AI Models* which have taken the world by storm during the last few months. While generative AI models have a broad scope including a variety of media, *Large Language Models (LLMs)* are specifically focused

on understanding and generating human language. Earlier related models such as Latent Semantic Analysis [47] were limited by insufficient computational power and small training corpora. Today LLM models are less constrained by these limitations, are continually changed, further developed, and additional models and platforms are appearing rapidly [48].

ChatGPT is currently the most prominent example of a LLM and it provides a specific example and a contemporary context for new interpretations of the theme and the specific topics of this article.

ChatGPT responds to the **question** explored in this article:

"Is ChatGPT more in line with an "AI versus Humans" or an "AI and Humans" framework"

with the **answer**:

OpenAI's approach to developing models like ChatGPT is grounded in the philosophy of "AI and Humans" rather than "AI versus Humans." The goal is to develop AI as a tool that can complement, augment, and assist human abilities rather than replace or compete with them. A few key points to consider:

1. **Collaborative Tool:** ChatGPT and similar models are designed to work alongside humans. For example, they can be used as research assistants, writing aids, or tutors, helping users to obtain information, craft content, or learn new topics.
2. **Augmentation Over Replacement:** While AI can automate certain tasks, the emphasis is on augmenting human capabilities. For instance, while ChatGPT can help draft content, human creativity, judgment, and context-awareness remain crucial.
3. **Ethical Design:** OpenAI's mission is to ensure that artificial general intelligence benefits all of humanity. The development and deployment of AI models are undertaken with careful consideration of their societal impact, emphasizing cooperation and shared benefits.
4. **Limitations of AI:** Current AI models, including ChatGPT, lack true understanding, consciousness, emotions, or general intelligence. They work based on patterns in the data they were trained on. Humans bring context, ethics, emotions, and deeper understanding to the table, making the collaboration between AI and humans essential for many tasks.
5. **Feedback Loop:** OpenAI often relies on the user community and broader public to provide feedback on its models, pointing out strengths, weaknesses, and potential areas of concern. This collaborative feedback loop is integral to refining and improving AI technologies.

In summary, the design and intent behind ChatGPT and similar models by OpenAI are very much in line with a cooperative "AI and Humans" framework. The aim is to harness the strengths of AI while recognizing and valuing the unique capabilities of humans.

For all the major technological innovations illustrated by Figure 2, the fundamental transformations are not caused by the technology per se but its use (by individual

learners, by workers, and by decision-makers) and the associated cultural transformations [44]. All the developments and their assessments of Figure 2 have taken place in the past—whereas the ongoing development and use of ChatGPT is in the future. Figure 3 displays different future development paths for ChatGPT—unique opportunities to positively contribute to quality of life by further enriching the power of the aided, collective human mind (the green curve) or having negative impacts (the black curve).

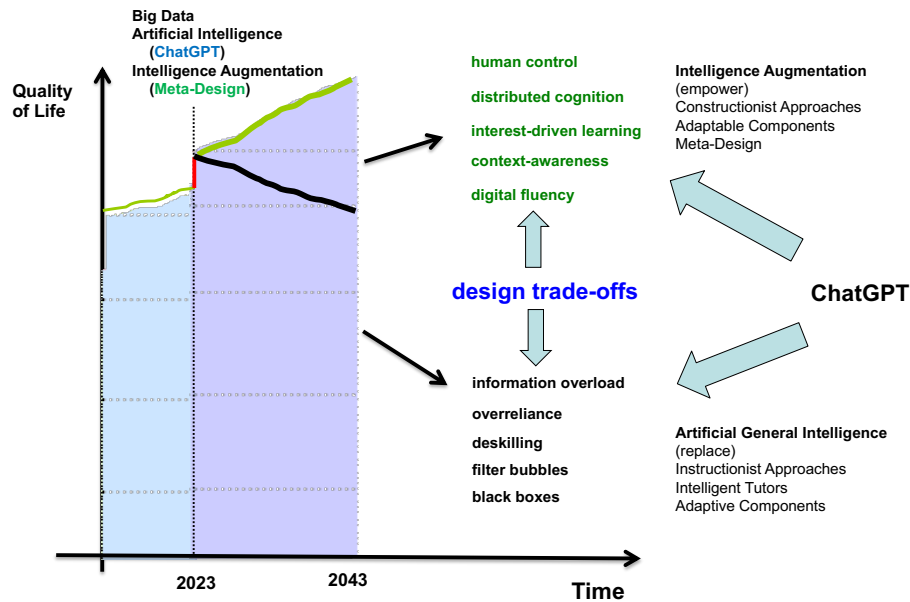


Figure 3: Different Futures for ChatGPT to Impact Learning and Quality of Life

The following paragraph provides one possible narrative for using ChatGPT:

Personal Speechwriters—A Narrative for Using ChatGPT. High-level politicians, industry leaders, and other decision makers are using speechwriters to assist them in giving presentations for a broad range of different topics. If one or more speechwriters provide a text and the “important persons” deliver the speech by reading the text literally given to them—they engage in the least demanding activity: all they need to be able to do is read.

What the listeners of the speech hope for and expect is that the “important person” superimposed her/his own idea based on the ideas provided by the speech writers.

Most people writing an essay, a job application, or an admission letter for a university can not afford to have another human as a speech writer. They can use ChatGPT as their personal speech writer in the following two ways:

- *Behavior-1*: use the text created by ChatGPT as their final product (governed by a “AI versus Human” perspective)

or

- *Behavior-2*: they can use ChatGPT to generate a first draft and then rewrite the texts to reflect their own voices and experiences (governed by a “AI and

Human” perspective). In this case ChatGPT can be a powerful tool for assistance by blending its computational power with human intuition, expertise, and ethical judgment.

The major discussion issues that have developed over the last few months in the context of ChatGPT include:

New Forms of Plagiarism versus Inspirations for Creative Reflections. Behavior-1 has led to broad-based concerns of educators and decision-makers that learners and job applicants submit the texts produced by ChatGPT as their own work creating a new form of plagiarism representing a modern form of cheating. But the major damage may be experienced by the persons engaging in such activities: it denies them the learning experiences and creative insights that occur with the process of refining initial ideas into a refined document.

Behavior-2 represents an interesting and thought-provoking use of ChatGPT similar to exploring a topic with literature references, exploring Wikipedia articles, and consulting with friends and colleagues. Norman in his essay “In Defense of Cheating” [49] argues convincingly that in schools the use of isolated work is encouraged and cooperative work is punished whereas in the real world and lifelong learning, this behavior is desired and critically important [50]. In collaborative brainstorming and decision-making, ChatGPT can serve as an “additional member,” offering insights or alternative perspectives that may not have been considered, thereby enriching the creative process.

ChatGPT hallucinates. While ChatGPT is a powerful language model with impressive capabilities, it is not infallible and can produce hallucinated content. Recognizing this limitation is crucial for its responsible and effective use. One of the major concerns with hallucinations is the potential spread of misinformation. If users are not aware of this limitation or do not cross-check information, they might take inaccurate statements as fact. But misinformation is not unique to ChatGPT. Misinformation, inaccurate information, and one-sided information may also be found in news media, and Wikipedia or may be spread by professors giving lectures to students. The appropriate response to cope with misinformation is for users and learners to engage in *critical thinking* and attempt to verify all critical information using other sources.

Banning ChatGPT. Caught in the excitement and extensive use of ChatGPT, school boards and whole countries banned (at least temporarily) the use of ChatGPT. As indicated in Figure 2, previous transformational technological innovations have led to cultural transformations emphasizing in an “AI and Human” framework that human beings were freed from (1) learning things by heart (a necessity in oral societies by writing them down (in literal societies); (2) learning phone numbers by having address books on their phones); (3) finding places with maps by having navigation systems; (4) conducting large scale numerical computations by having calculators; and (5) learning huge number of facts by consulting the internet and Wikipedia. Society needs to help students and teachers to understand ChatGPT’s strengths and risks, rather than just forbidding them to use the technology.

7 Implications for the Future

Albert Einstein "My pencil and I are smarter than I am."

Einstein's remark about his pencil underscores the extended capabilities, collaboration, and complexities enhancing the unaided, individual human mind with new additional and transformative support for distributed cognition (see Figure 2). The assertion is also supported by Bruner's concept of scaffolding and his recognition that human intellect is sustained by socio-technical environments [29]): “*a human’s use of mind is dependent upon her/his ability to develop and use tools or instruments or technologies that make it possible for her/him to express and amplify her/his powers*”. Learning from the past by being aware that all of the changes (indicated in Figure 2) provided opportunities as well as unintended consequences (see Table 1), the challenges associated with using AI in the years to come to benefit humans and improve their quality of life need to be explored for specific problems and use situations (see Figure 3). This article tries to transcend global statements and assumptions by contrasting the promises and pitfalls of focusing AI on replacing human beings (with “AI versus Humans”) or empowering human beings (with “AI and Humans”).

Quality of Life does not have a simple meaning (it is another “suitcase word”; see Section 3) and should not be used as if it does. Quality of life (QOL) is the general well-being of individuals and societies and should not be confused with the concept of standard of living, which is based primarily on income. In the context of this article, its focus is on assessing its relevance for the digital age [5]. Its focus should not be on strong or weak AI (transcending the distinction indicated in Figure 1) but explore a post-AI attitude, where the goal is to prioritize human well-being [51-53]. As indicated in Figure 3, examples of quality of life dimensions related to ChatGPT considered *positive* are human control in personally meaningful activities, distributed cognition, support for interest-driven learning and context-awareness, and digital fluency and potential *pitfalls* are information overload, over-reliance, deskilling, filter bubbles, and systems being black boxes. The overall effect of ChatGPT on quality of life will depend on how it is integrated into society, how individuals choose to utilize it, and an understanding and safeguards to mitigate its risks.

AI is not new and it has made contributions in different areas (see Figure 1). *AI systems for specific purposes* are closely aligned with the “replacement” paradigm of “AI versus Humans” (see Section 3) whereas *Human-Centered AI* explores and supports the “empowerment” paradigm of “AI and Humans” (see Section 4).

The current worldwide excitement, envisioned opportunities, and fears based on foundation models have reached levels never seen before. Governments (at the national US level [54] and international level [55]) are exploring *regulation frameworks* for the protection of human rights, transparency, explainability, fairness, accountability, safety, appropriate human oversight, ethics, bias mitigation, privacy, and data protection needs to be addressed. *The results are often stuck with global statements* (e.g.: “huge opportunities come with risks that could threaten global stability and undermine our values.”). The most *interesting questions* are often not asked. The question “Will AI replace teachers, radiologist, journalists?” should be replaced by the assertion to be investigated “Teachers, radiologist, journalist who use AI replace those who don’t”.

Academic communities have contributed very little to these discussions. One particular contribution could and should be to clarify to role of AI in addressing *specific*

challenges as this paper has attempted to differentiate between “Humans versus AI” and “Humans and AI”.

8 Conclusions

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

The transformative changes that will take place in the future are not out there to be discovered but they will be *designed* (taking ideas from AI optimists, pessimists, and realists into account). Technological change is not inevitable. The impact of changes upon people and society can be 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 that address the *basic question* associated with the themes “AI and Humans” and “AI versus Humans” by exploring: (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)?

The decision to come down on either side of the debate is difficult because there are no obvious right answers. This paper tried to provide some arguments that investigating the possibility space of “AI and Humans” is one of the most important and fascinating research themes of the future. This should include exploring a post-AI attitude, where the goal is to prioritize human well-being (e.g.: quality of life, meaningful work, and learning), and a detailed, scientifically grounded discussion about the design trade-offs associated with technological developments where the challenge “*can we do it*” is complemented by the ethical question “*should we do it*”. As we have fewer constraints on what we can do, it is inevitable that our values will matter more than ever.

Acknowledgments

The author would like to thank to organizers and the contributors of the COPDA 2022 “AI for Humans or Humans for AI?” workshop, and Ben Shneiderman, Eileen Kintsch, Clayton Lewis, Johan Lundin, Anders Morch, Don Norman, and Mitchel Resnick, who inspired me with their ideas and who helped me to clarify my own understanding of the topic of this paper.

The reviewers of an earlier submitted version provided arguments that led to numerous improvements of the paper.

My thanks also go to the “Hanse-Wissenschaftskolleg Institute for Advanced Study” in Delmenhorst, Germany (HWK; <http://www.h-w-k.de/en/hwk-overview.html>) which supported me with a fellowship for 6 months during the years 2021 and 2022 to explore the themes discussed in the paper.

References

1. Barricelli, B.R., et al. *CoPDA 2022: "AI for Humans or Humans for AI?" Proceedings of the Sixth International Workshop on Cultures of Participation in the Digital Age*. 2022; Available from: <https://ceur-ws.org/Vol-3136/>.
2. Markoff, J., *Machines of Loving Grace (The Quest for Common Ground Between Humans and Robots)*. 2016: Harpercollins
3. Fischer, G. and K. Nakakoji, *Beyond the Macho Approach of Artificial Intelligence: Empower Human Designers - Do Not Replace Them*. *Knowledge-Based Systems Journal*, Special Issue on AI in Design, 1992. **5**(1): p. 15-30.
4. Shneiderman, B., *Human-Centered AI 2022*: Oxford University Press.
5. Fischer, G., *Design Trade-Offs for Quality of Life* *ACM Interactions* 2018. **25** (1): p. 26-33.
6. Fogli, D., et al., *Exploring design trade-offs for achieving social inclusion in multi-tiered design problems*. *Behaviour & Information Technology*, 2020. **39**(1): p. 27-46.
7. Arias, E.G., et al., *Transcending the Individual Human Mind—Creating Shared Understanding through Collaborative Design*, in *Human-Computer Interaction in the New Millennium*, J.M. Carroll, Editor. 2001, ACM Press: New York. p. 347-372.
8. Arias, E.G., H. Eden, and G. Fischer, *The Envisionment and Discovery Collaboratory (EDC): Explorations in Human-Centered Informatics*. 2016, San Rafael, California: Morgan & Claypool Publishers.
9. Minsky, M., *The Emotion Machine*. 2007: Simon & Schuster.
10. Feigenbaum, E.A. and J. Feldman, *Computers and Thought*. 1963, New York: McGraw-Hill Book Company.
11. Fischer, G., *End-User Development: Empowering Stakeholders with Artificial Intelligence, Meta-Design, and Cultures of Participation*, in *IS-EUD 2021 Proceedings*, D. Fogli, Tetteroo, D., Barricelli, B.R., Borsci, S., Markopoulos, P., Papadopoulos, G.A., Editor. 2021, Springer, LNCS 12724. p. 3–16.
12. Engelbart, D.C., *Toward Augmenting the Human Intellect and Boosting Our Collective IQ*. *Communications of the ACM*, 1995. **38**(8): p. 30-33.
13. Brockman, J., ed. *What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence*. 2015, Harper Perennial.
14. Kurzweil, R., *The Singularity is Near*. 2006: Penguin Books.
15. Buchanan, B.G. and E.H. Shortliffe, eds. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. 1984, Addison-Wesley Publishing Company: Reading, MA.

16. Bell, G. and J. Gemmell, *Total Recall: How the E-memory Revolution Will Change Everything*. 2009, New York, NY: Dutton.
17. Friedman, T.L. *The Professors' Big Stage* (*New York Times*, March 5). 2013; Available from: http://www.nytimes.com/2013/03/06/opinion/friedman-the-professors-big-stage.html?_r=0.
18. Hawking, S., et al., *Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?* Independent (October 23), 2017.
19. Isaacson, W., *Inside Elon Musk's Struggle for the Future of AI*. Time Magazin, 2023. **September**.
20. Susskind, D., *A world without work: technology, automation, and how we should respond*. 2020, New York, N.Y.: Metropolitan Books/Henry Holt & Company.
21. Carr, N., *The Glass Cage — Automation and US*. 2014, New York: Norton.
22. Winograd, T. and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*. 1986, Norwood, NJ: Ablex Publishing Corporation.
23. Norman, D.A., *The human side of automation*, in *Road Vehicle Automation 2*, G. Meyer and S. Beiker, Editors. 2015, Springer International Publishing. p. 73-79.
24. Muller, J.Z., *The Tyranny of Metrics*. 2018, Princeton: Princeton University Press.
25. Center for International Governance Innovation. *The Ethics of Automated Warfare and Artificial Intelligence*. 2022; Available from: <https://www.cigionline.org/the-ethics-of-automated-warfare-and-artificial-intelligence/>.
26. Mayer-Schönberger, V. and K. Cukier, *Big Data*. 2013, New York, NY: Houghton Mifflin Harcourt.
27. Hollan, J., E. Hutchins, and D. Kirsch, *Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research*, in *Human-Computer Interaction in the New Millennium*, J.M. Carroll, Editor. 2001, ACM Press: New York. p. 75-94.
28. Salomon, G., ed. *Distributed Cognitions: Psychological and Educational Considerations*. 1993, Cambridge University Press: Cambridge, United Kingdom.
29. Bruner, J., *The Culture of Education*. 1996, Cambridge, MA: Harvard University Press.
30. Sloman, S. and P. Fernbach, *The Knowledge Illusion — Why We Never Think Alone*. 2017, New York: Riverhead Books.
31. Fischer, G., *Domain-Oriented Design Environments*. Automated Software Engineering, 1994. **1**(2): p. 177-203.

32. Fischer, G., *Context-Aware Systems: The 'Right' Information, at the 'Right' Time, in the 'Right' Place, in the 'Right' Way, to the 'Right' Person*, in *Proceedings of the Conference on Advanced Visual Interfaces (AVI)*, G.T.e. al, Editor. 2012, ACM: Capri, Italy (May). p. 287-294.
33. Carmien, S. and G. Fischer, *Tools for Living and Tools for Learning*, in *Proceedings of the HCI International Conference (HCII), Las Vegas, July 2005*. 2005. p. (published on a CD).
34. Brynjolfsson, E. and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. 2014: W. W. Norton & Company.
35. Simon, H.A., *The Sciences of the Artificial*. third ed. 1996, Cambridge, MA: The MIT Press.
36. Pariser, E., *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. 2021: Penguin Books.
37. Fischer, G., *Adaptive and Adaptable Systems: Differentiating and Integrating AI and EUD in Proceedings of 9th International Symposium on End-User Development (IS-EUD 2023)*, D. Spano, Editor. 2023, Springer, LNCS, volume 13917: Cagliari, Italy. p. 3-18.
38. Anderson, J.R., et al., *Cognitive Tutors: Lessons Learned*. *The Journal of the Learning Sciences*, 1995. 4(2): p. 167-207.
39. Papert, S., *Mindstorms: Children, Computers and Powerful Ideas*. 1980, New York: Basic Books.
40. Resnick, M., *Lifelong Kindergarten — Cultivating Creativity through Projects, Passion, Peers, and Play*. 2017, Cambridge, MA: MIT Press.
41. Singer, N. *New A.I. Chatbot Tutors Could Upend Student Learning*. 2023; Available from: <https://www.nytimes.com/2023/06/08/business/khan-ai-gpt-tutoring-bot.html>.
42. Rittel, H. and M.M. Webber, *Planning Problems are Wicked Problems*, in *Developments in Design Methodology*, N. Cross, Editor. 1984, John Wiley & Sons: New York. p. 135-144.
43. Collins, A. and R. Halverson, *Rethinking Education in the Age of Technology — The Digital Revolution and Schooling in America*. 2009, New York, NY: Teachers College Press
44. Fischer, G., J. Lundin, and J.O.J. Lindberg, *Rethinking and reinventing learning, education, and collaboration in the digital age — from creating technologies to transforming cultures*. *International Journal of Information and Learning Technology* (Publisher: Emerald), 2020. <https://doi.org/10.1108/IJILT-04-2020-0051>.
45. Brown, J.S. and P. Duguid, *The Social Life of Information*. 2000, Boston, MA: Harvard Business School Press.

46. Giovannella, C., “*Learning by being*”: *integrated thinking and competencies to mark the difference from AIs*. *IxD&A Journal* 2023. **57**: p. 8-26.
47. Landauer, T.K., et al., eds. *Latent Semantic Analysis*. 2007, Erlbaum: Mahwah, NJ.
48. Sharma, A. *11 Best Generative AI Tools and Platforms*. 2023; Available from: <https://www.turing.com/resources/generative-ai-tools>.
49. Norman, D. *In Defense of Cheating*. 2005; Available from: <https://ubiquity.acm.org/article.cfm?id=1066347>.
50. Resnick, L.B., *Learning In School and Out*. Educational Researcher, 1987. **16**(9): p. 13-20.
51. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (Version 2)*. 2017; Available from: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.
52. Calvo, R.A. and D. Peters, *Positive Computing —Technology for Wellbeing and Human Potential*. 2014, Cambridge, MA: MIT Press.
53. Friedman, B. and D.G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*. 2019, Cambridge, MA: MIT Press.
54. The White House. *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*. 2023; Available from: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/?utm_source=substack&utm_medium=email.
55. UK Government. *The Bletchley Declaration by Countries Attending the AI Safety Summit*. 2023; Available from: <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.