Proceedings of the 56th Hawaii International Conference on System Sciences | 2023

How Can You Verify that I Am Using AI? Complementary Frameworks for Describing and Evaluating AI-Based Digital Agents in their Usage Contexts

Steven Alter
University of San Francisco
alter@usfca.edu

Abstract

This essay explains complementary frameworks for understanding and managing AI in usage contexts. In contrast with broad generalizations about the nature and impact of AI, those frameworks focus on specific AI-based digital agents used by people and/or machines performing purposeful activities in business, home, or societal environments. The agent responsibility (AR) framework helps in describing roles and responsibilities of specific AI-based digital agents in their usage contexts. The agent evaluation (AE) framework identifies six criteria that different stakeholders might use for evaluating AI-based digital agents.

1. Are AI-Enabled Systems Fundamentally Different from Other IT-Enabled Systems?

AI has become an umbrella term covering disparate IT applications that may involve machine learning (ML), text dictation, facial recognition, language translation, chatbots, expert systems, cognitive computing, robotics, and many other IT uses that might seem intelligent or smart. The diversity of across that range of applications leads to questioning common beliefs that AI uses are somehow unique. In contrast, this paper focuses on similarities between uses of AI and uses of other IT-intensive systems that use relatively new techniques:

1) Applications of AI and of IT (more generally) occur through algorithms embedded in digital agents serving work systems that may be sociotechnical (with human participants) or totally automated.

2) Types of roles and responsibilities that AI-based digital agents play in operational work systems overlap substantially with types of roles and responsibilities played by digital agents not associated with AI.

3) Applications of AI and other types of IT may be directed toward decision making and/or other facets of work such as communicating, coordinating, providing value, maintaining security, and so on.

4) The challenges of "managing AI" are similar to the challenges of managing most IT-intensive systems involving new or relatively unfamiliar techniques.

This paper applies the following definitions as a basis for explaining two frameworks for describing and evaluating AI applications even though there is no agreed upon definition of AI:

Algorithm. A specification for achieving specified goals within stated or unstated constraints by applying specific resources such as data inputs.

Agent. An entity that performs task(s) delegated by another entity. This definition assumes that actors, (entities that perform actions) may or may not be agents.

Algorithmic agent. A physical or digital agent that operates by executing algorithms.

Digital agent. An algorithmic agent that operates by executing algorithms encoded in software and that has no persistent physical existence.

AI-based digital agent. A digital agent whose main activities are guided by algorithms created using techniques associated with AI.

Activity. A purposeful action that is significant enough to identify when trying to understand a system's development or operation.

Work system. A system in which human participants *and/or* machines perform work (processes and activities) using information, technology, and other resources to produce specific product/services for internal and/or external customers (Alter, 2006; 2013).

AI-based digital agent in its context of usage. Most significant uses of AI-based digital agents occur as a work system's delegation of roles and responsibilities to those AI-based digital agents.

Organization. A discussion of whether having a specific definition of AI matters in today's world precedes comments about algorithms and an observation that specific uses of many algorithms may not reveal whether they are based on AI. The next step is seeing that AI usage occurs through the delegation of roles and responsibilities to digital agents by work systems whose existence may be temporary (as in projects) or long-standing (as in organizational routines). The agent responsibility (AR) framework has two dimensions: a spectrum of roles and responsibilities and different facets

URI: https://hdl.handle.net/10125/103277
978-0-9981331-6-4

of work. The agent evaluation (AE) framework identifies criteria for evaluating digital agents from different stakeholder viewpoints. Those frameworks are applied to topics and examples from seven papers in a 2021*MIS Quarterly* special issue on “Managing AI.” The conclusion discusses implications related to uses of the AR and AE frameworks and to generalizations about AI.

2. Does the Definition of AI Matter?

A precise definition of AI is elusive, partly because of the great diversity of topics that are associated with AI despite being only tangentially related to each other, e.g., intelligent machines, neural networks, robotics, machine learning, expert systems, smart systems, cognitive computing, speech recognition, pattern recognition, image recognition, natural language processing, statistical algorithms, automated decision-making, and so on. This paper assumes that “artificial intelligence” is a useful buzzword, more like an organizing vision (Swanson & Ramiller, 1997) and less like a coherent phenomenon with identifiable properties. For example, an ecommerce application of machine learning to consumer data is not in the same category as fictional accounts of seemingly intelligent robots that are often associated with AI, e.g. R2D2 in the Star Wars films and the humanoid “artificial friend” Klara in a recent novel by the Nobel Prize winner Ishiguro (2021)

The possibility that a clear definition of AI is not of great importance also appeared in a video presentation by Ben Shneiderman (2021), a noted HCI expert, about his new synthesis of human centered AI (HCAI). When asked whether HCAI was really about AI or whether it was about all of technology, Shneiderman said that HCAI ideas apply to technology in general but that the current interest in AI is so enormous and so pervasive that the focus should be on human-centered AI (at 1:07 in the video). Peter Norvig’s discussant observations within the same video (at 1:24) provided an additional wrinkle by saying that “in terms of branding of names of fields, artificial intelligence has a sexy, evocative name and operations research has the most boring name possible, and yet they are almost the same field. They both are about computing optimal action.”

The diverse views of AI in the recent *MIS Quarterly* special issue on Managing AI provides further evidence of the possibility of discussing aspects of AI without providing an operational definition of AI. Some of the articles in the *MISQ* Special Issue focus on ML without defining AI or discussing AI capabilities in general. Others generalize about AI capabilities without defining AI. None provide a clear link between a definition of AI and specific real world AI applications.

The introduction to the special issue (Berente et al., 2021) built to its definition of AI starting from the famous Dartmouth workshop in 1956, which defined the project of creating AI in terms of “making a machine behave in ways that would be called intelligent if a human were so behaving” (McCarthy et al. 1955).. Arguing that the lack of a “singular, agreed-upon definition for AI” … “has been quite generative, leading to all sorts of productive inquiry and technological advancements,” Berente et al. (2021) conceived of AI “as a process, rather than a phenomenon in itself. We define AI as *the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems*. In short, AI is whatever we are doing next in computing.” By that definition, 1940s computers that computed artillery trajectories would be viewed as instances of AI. Similarly for the “giant brain” that used simple statistical methods to predict Eisenhower’s victory in the 1952 presidential election. (see Chinoy, 2010). Major advances in semiconductor chips, programming languages, and networking over the last five decades also would be considered AI because they are on the frontier of computing and because some refer to human intelligence directly or metaphorically.

Discussing AI usage without defining AI. This paper accepts Berente et al.’s suggestion that accepting diverse views of AI may lead to “productive inquiry and technological advancements.” It proceeds by assuming for any given definition of AI that AI usage involves digital agents created using AI-related techniques.

This paper pursues the spirit of the view of HCAI in Shneiderman (2020a, 2020b, 2021)_ while assuming that real world actions and outcomes may not depend on whether people who produced a specific algorithm thought they were practicing AI or operations research or accounting or general management or any other discipline. Issues that Schneiderman highlights through the lens of HCAI are relevant to digital agents that may or may not involve AI: Are appropriate values embedded in algorithms and their use? Are digital agent designers trying to eliminate people by automating human work and human judgment or are they trying to augment human capabilities? How can we resolve contradictions between benefits for people in society versus diverse goals of individuals who use digital agents directly or indirectly?

3. A General View of Algorithmic Agents

Algorithms are specifications for achieving goals within stated or unstated constraints by applying specific resources such as data inputs. Algorithms may be as simple as a decision rule or as complex as integrated algorithms for self-driving cars. As abstractions, algorithms cannot do anything by themselves and have effect only when human or non-human actors use them to support, control, or perform actions in the world.

Algorithmic agents are human or nonhuman agents that perform activities based on explicit algorithms. Box 1 (based on Alter, 2022c) lists examples involving algorithmic agents that might or might not use AI-related capabilities. Those algorithmic agents are digital agents if their tasks are completely automated.

Box 1. Application situations for algorithms that might or might not use AI

• using facial images to identify people • converting spoken words into equivalent text • deciding which applicants should be hired or accepted by a university • deciding whether a person is legally entitled to drive a car • deciding whether an autonomous vehicle needs to stop or swerve • controlling a rocket in flight	• deciding whether to turn off a machine before a mechanical failure occurs • deciding where police should be deployed over the next eight hours • combining multiple items in an order to minimize shipping cost • determining the best route for driving from a starting point to a destination • finding the laws that are most relevant to a specific lawsuit

Some of the examples in Box 1 might use decision rules as simple as allowing no more than 40% of successful applicants to be classified in category X. Even that simple algorithm can have important and far reaching effects such as favoring one group of people over other groups, as when category X is qualification for employment or acceptance into college. Notice how the issue of bias such as favoring one group over another is not about AI, but rather is about using one algorithm instead of another algorithm or instead of relying more on inherently biased human perceptions and judgments.

Box 1 illustrates the difficulty of generalizing about benefits, risks, and ethics of AI without explaining details of specific AI applications. It is doubtful that any non-trivial statements about likely benefits, risks, and ethics of AI in general could be derived from analyzing AI applications in examples that use different techniques in different application areas with vastly different impacts of possible design errors or other shortcomings.

4. Work Systems as the Context for Applying AI

The idea of work system (WS) provides a way to identify the context within which algorithms are applied, regardless of whether those algorithms were developed using AI-related methods. Ideas in the work system perspective have been presented many times, (Alter, 2006, 2008, 2013). The following summary of aspects of that perspective emphasizes ideas that are directly useful in understanding the use of AI-based digital agents.

Performing work in work systems as the context for applying AI. In an economic context, work is the application of human, informational, physical, and other resources to produce product/services for internal or external customers. Work occurs in homes, businesses, governments, and other situations where purposeful use of resources produces outcomes. The first *and/or* in the definition of work system on Page 1 says that work systems may be sociotechnical (with human participants doing some of the work) or totally automated.

Special cases of work system. An IS is a WS most of whose activities are devoted to capturing, transmitting, storing, retrieving, deleting, manipulating, and/or displaying information. This definition differs from 20 previous definitions in Alter (2008) and was one of 34 definitions of IS noted in Boell & Cecez-Kecmanovic (2015). It is also quite different from defining an IS as a representation (see Burton-Jones et al., 2017) or as a tool that is "used." Other important special cases include projects, service systems, self-service systems, and some supply chains (interorganizational WSs). Software development projects are WSs designed to produce specific product/services and then go out of existence.

Work system framework: a basic understanding of a work system. The nine elements of the work system framework are the elements of a basic understanding of a WS's form, function, and environment during a period when it is stable enough to retain its identity even though incremental changes may occur, such as minor personnel substitutions or technology upgrades. Those elements include customers, product/services, processes and activities, participants, information, technologies, environment, infrastructure, and strategies.

Work system life cycle model (WSLC): how WSs change over time. ISs and other WSs evolve through a combination of planned change through projects involving initiation, development, implementation and unplanned change via adaptations and workarounds. The WSLC phases may be performed in many different ways. Activities and responsibilities associated with specific phases (e.g., designing, debugging, training, etc.) apply for waterfall, agile, prototyping, use of off-the-shelf applications, and shadow IT, even when several phases overlap or are combined through short iterations.

5. An Agent Responsibility Framework for Visualizing Applications of AI

The definitions on Page 1 led to saying that most significant uses of an AI-based digital agents involve a WS's delegation of roles and responsibilities to AI-based digital agents. Deciding whether or how to apply AI in a WS could start with those definitions, but would benefit from a framework that helps in identifying and visualizing potential design choices.

The HCAI framework in Shneiderman (2020a, 2020b, 2021) is a step in that direction. Its two dimensions, low vs. high computer automation and low vs. high human control apply to many situations in which both automation and human control are design variables. It leads directly to useful questions for evaluating and designing WSs by emphasizing that deficiency or excess along either dimension leads to worse results for organizations, for WS participants, and/or for customers. For current purposes, the low vs. high distinctions in the HCAI framework's two dimensions provide too little detail to inspire vivid visualization and discussion of how or why an AI-based digital agent might be applied in a WS's operation or might affect its stakeholders.

The AR framework in Figure 1 addresses that limitation through more detailed views of the roles and related responsibilities of digital agents. Clarity about those roles and responsibilities in specific WSs requires attention to whether and how a digital agent supports specific facets of work (Alter, 2021). Figure 1 assumes that AI usage involves digital agents and that AI usage occurs when AI-related digital agents perform one or more roles (the horizontal dimension) related to one or more facets of work (the vertical dimension). Figure 1 appears in Alter (2022a, 2022b) in relation to digital agents in general without referring to AI per se.

As illustrated in those articles, uses of the AR framework do not rely on considering all combinations of roles and facets. Also, other roles and facets might be especially relevant to specific situations. Focusing on different facets of work could encourage designers or managers to wonder about the benefits of enhancing specific facets of work in a specific WS. Similarly, the spectrum of roles in the horizontal dimension encourages considering possible roles/responsibilities that digital agents might perform. It is unnecessary to consider all or even many of the 36 combinations of 6 facets of work and 6 types of roles/responsibilities (or 108 combinations of 18 facets of work presented in Alter (2021) and 6 types of roles/ responsibilities). Only combinations that are important for a specific WS should be considered.

5.1 The Horizontal Dimension: Different Types of Roles and Related Responsibilities that Might Be Assigned to a Digital Agent

The AR framework's horizontal dimension identifies six roles of digital agents that support or perform specific facets of work in a work system. Those roles are presented along a spectrum from the lowest to the highest direct involvement of the digital agent in the execution of a work system's activities. Those six roles emerged from many iterations of trying to expand the horizontal dimension of the HCAI framework to make it more specific. One of the early iterations involved only three roles, i.e., support, control, and perform. Specific instances of roles in the horizontal dimension might support HCAI values and aspirations or might oppose those values and aspirations (e.g., micromanagement or surveillance capitalism). The following comments about the six roles emphasize promoting human-centric values and addressing human-centric issues:

Monitor a work system. A digital agent might monitor and measure aspects of work to assure that a WS's processes and activities are appropriate for WS participants. In some cases digital agents might generate alarms when aspects of work start going out of accepted bounds regarding health, safety, and cognitive load.

Provide information. A digital agent might provide information that helps people achieve their work goals safely and comfortably without infringing on privacy and other rights of people whose information is used.

Provide capabilities. A digital agent might provide analytical, visualization, and computational capabilities that help WS participants achieve their assigned goals safely and with appropriate effort.

Control activities. A digital agent might control work activities directly to prevent specific activities from going out of bounds related to worker safety, time on the job, stress, and other variables that can be measured and used to control a work system.

Coproduce activities. A digital agent might be deployed in a division of responsibility in which digital agent and people have complementary responsibilities for performing their parts of the work. Either humans or digital agent might take the lead in human-computer interactions within coproduction activities. In contrast, initiative in might shift back and forth between people and digital agents depending on the status of the work.

Execute activities. A digital agent might execute activities that should not or cannot be delegated to people. For example, a digital agent might perform activities that are difficult, dangerous, or impossible for people to perform as the WS produces product/services.

5.2 The Vertical Dimension: Different Facets of Work in which a Digital Agent Might Have Responsibilities

The idea of facets of work grew out of research for bringing richer and more evocative concepts to systems analysis and design to facilitate interactions between analysts and stakeholders, as is explained in Alter (2021: 342-344). The notion of facet is an analogy to how a cut diamond consists of a single thing with many facets. Psychology, library science, information science, and computer science have used the idea of facets, but with quite different meanings and connotations.

Most activities in WSs can be subdivided into making decisions, communicating, coordinating, improvising, and so on. For current purposes, general types of activities such as those can be viewed as facets of work if they are easily understood and widely applicable and if they satisfy criteria explained in Alter (2021): They apply to both sociotechnical WSs and totally automated WSs; they are associated with many concepts that are useful for analyzing system-related situations; they are associated with evaluation criteria and typical design trade-offs; they have sub-facets that can be discussed; they bring open-ended questions that are useful for starting conversations.

There is no assumption that the facets of work should be independent. To the contrary, the facet *making decisions* often involves other facets such as *communicating*, *learning,* and *processing information*. The main point is that each facet can be viewed as part of lens for thinking about where and how WSs might use digital agents in general or digital agents that apply AI.

Alter (2021) identifies 18 facets of work, while recognizing that other researchers might have identified other facets of work that satisfy those criteria. For the sake of easy visualization, the version of the AR framework in Figure 1 uses only six of those facets. The other 18 include learning, planning, improvising, interacting socially, providing service, and seven others. Table 1 uses the example of making decisions to illustrate most of the criteria for qualifying as a facet of work. Tables in Alter (2021) show the same types of information for all 18 facets.

Combining the AR framework's two dimensions leads to pinpointing design issues, e.g., the extent to which a digital agent should have responsibilities such as monitoring work, providing capabilities used in making decisions, or performing activities automatically. Table 2 shows how different digital agent roles might be applied to the facets of work included in Figure 1 in pursuit of HCAI values in a hiring system example. Alter (2022a, 2022b) present similarly formatted tables related to e-commerce, digital auction, self-driving car, and electronic medical records system examples.

<<< Facet of work >>>						
Making decisions						
Communicating						
Processing information						
Coordinating						
Creating value						
Maintaining security						
	Monitoring work system	Providing information	Providing capabilities	Controlling activities	Coproducing activities	Executing activities
	<<<<<<< Spectrum of roles and responsibilities >>>>>>>					

Figure 1. Agent responsibility framework with six roles and six facets of work

Table 1. Why making decisions qualifies as one of 18 facets of work

Type of qualification	Ideas related to making decisions
Association with many concepts that can be used for analysis	Decision, criteria, alternative, value, risk, payoff, utility, utility function, tradeoff, projection, optimum, satisficing vs. optimizing, heuristic, probability, distribution of results, risk aversion
Association with evaluation criteria	Actual decision outcomes, realism of projected outcomes, riskiness, decision participation, concurrence of stakeholders, ease of implementation
Association with design tradeoffs	Quick response vs. superficiality, model complexity and precision vs. understandability, brevity vs. omission of important details
Existence of sub-facets for detailed description	Defining the problem; identifying decision criteria; gathering relevant information; analyzing the information; defining alternatives; selecting among alternatives; explaining the decision
Related open-ended questions	How do the available methods and information help in important decisions? What decisions are made with incomplete, inaccurate, or outdated methods or information? How might better methods or information help in making decisions? Where would that information come from?

Table 2. Using facets of work to pursue HCAI values in a hiring system example

Facet of work	**Illustration of how a specific digital agent role (*in parenthesis and italicized*) might be applied to a specific facet of work in hiring situation that strives to represent HCAI**
Making decisions	(*monitor*) A digital agent might monitor the hiring decision process to identify inappropriate biases and misunderstandings that may have affected past decisions and may affect current decisions.
Communicating	(*provide information*) A digital agent could provide information including analyses of past hiring experience and the current job market, thereby facilitating communication between interviewers and managers who make decisions
Processing information	(*provide capabilities*) A digital agent could provide capabilities for processing information in applications, cover letters, and interview notes to support more effective comparison of applicants.
Coordinating	(*control activities*) A digital agent could control the process of scheduling interviewers to help them coordinate their interview responsibilities and to minimize the disruption experienced by interviewers who have many other work commitments.
Creating value	(*coproduce activities*) A digital agent with chatbot capabilities could coproduce the collection and consolidation of interviewer impressions, in effect coproducing activities that create value. The value for interviewers is reduced time needed for reporting on interviews. The value for the digital agent involves obtaining information in a form supporting further processing
Maintaining security	(*execute activities*) A digital agent could *execute activities* that help in maintaining security by controlling access to data about applicants and about interviewer perceptions.

Evaluation Criteria	Efficiency							
	Effectiveness							
	Equity							
	Engagement							
	Empathy							
	Explainability							
		Designer	Work system participant	Customer	Manager	Regulator	Investor	Society
		<<<<<<< **Typical examples of stakeholders** >>>>>>>						

Figure 2. Agent evaluation framework

6. The Agent Evaluation Framework

The agent evaluation (AE) framework in Figure 2 complements the AR framework by providing an organized approach for evaluating digital agents while recognizing that different stakeholders might bring quite different viewpoints. The horizontal dimension in Figure 2 identifies typical stakeholder viewpoints, i.e., designer, WS participant, customer, manager, regulator, investor, and society. The vertical dimension identifies different criteria that can be applied in evaluating digital agents, regardless of whether they use AI. Those criteria might be called the 6 E's: efficiency, effectiveness, equity, engagement, empathy, and explainability.

The first two evaluation criteria in Figure 2 are fundamentally about how well a digital agent helps a WS achieve its operational goals. The other four criteria are about human-centric issues that are not as directly linked to operational WS goals but that often affect WS performance and/or perceptions of product/services that a WS produces.

Efficiency. A digital agent should be efficient in its own right and should support the efficiency of related WSs, i.e., the balance between resources used by those WSs and the value that they produce for their customers.

Effectiveness. A digital agent should meet or exceed expectations regarding their contribution to a WS's performance. Uses of digital agents should support the effectiveness of WSs by contributing to their production of value for their customers.

Equity. A digital agent and WSs that it supports should operate in a way that is fair to stakeholders including WS participants, WS customers, and others affected directly and indirectly by the WS's. Equity is often a challenge because designers, managers, and others may be unaware of their own biases and biases built into digital agents and WSs.

Engagement. A digital agent that performs roles in sociotechnical WSs and is not embedded in automated devices should engage WS participants wherever that might maximize benefits from their insights or might make their work environments healthy and productive.

Empathy. A digital agent and WSs that it supports should reflect realistic consideration of the goals, capabilities, health, and comfort of WS participants and customers that use the WS's product/services. Lack of such empathy could have negative impacts on people who use a digital agent directly, on the WS's operational performance, or on product/services that it produces.

Explainability. A digital agent should be understandable by people who are affected by it and/or by product/services that it produces. This issue has been discussed widely in regard to applications of ML algorithms whose outputs cannot be linked in an understandable way to inputs related to individuals, groups, or situations. Inadequate explainability results in confusions, errors, misuse of product/services, and may harm people who are affected directly or indirectly.

7. Application to Topics and Examples in an *MISQ* Special Issue on Managing AI

The AR and AE frameworks are designed to help designers and managers concerned with using digital agents and AI (however defined) in a beneficial way. At first blush it seemed likely that the seven papers in the recent *MIS Quarterly* Special Issue on Managing AI would provide interesting real-world examples of successful or problematic AI applications that might be used for evaluating the potential usefulness of those frameworks. Importantly, examples from a special issue on Managing AI would not be "cherry picked" to support claims about the value of the AR and AE frameworks.

While the entire Special Issue discussed interesting and valuable topics, none of its papers discussed operational examples in enough depth to apply the AR framework in depth, i.e., to clarify specific roles and responsibilities of AI-based digital agents beyond mentioning actual or potential support of decision making or learning. Evaluation criteria mentioned directly or indirectly in at least one paper included efficiency, effectiveness, equity, and explainability.

Recognizing those limitations, this section's brief comments about each of the papers highlight topics related to the AR or AE frameworks and generalizations about AI that might or might not be supportable. The papers appear in the alphabetical order in the references.

Will humans in the loop become borgs? Fügener et al. (2021) uses a mathematically complex analytical model, experimental studies, and a simulation study to explore "how advice from AI affects complementarities between humans and AI, in particular, what humans know about that AI does not know: 'unique human knowledge.'" Fügener et al. does not look at specific real world systems that apply AI. Its "main finding is that human choices converge toward similar responses improving individual accuracy. However, as overall individual accuracy of the group of humans improves, the individual unique human knowledge decreases."

Fügener et al. (2021) focuses on AI-advised human decision making where humans receive an AI-based suggestion before making final decisions. The AR framework would say that the digital agent's role is *providing information* and the facet of work is *making decisions*. A conclusion is that "simulation results based on experimental data suggest that groups of humans interacting with AI are far less effective as compared to human groups without AI." In terms of the AE framework, *effectiveness* is the focus.

While not defining AI, this paper says that "modern AI systems are based on training data observed from practice and are not explicitly based on human-defined rules." It includes generalizations that refer to AI as a distinct entity, e.g., "AI advice can decrease the complementary knowledge between humans and the AI" and "a modern AI can determine its own uncertainty."

Dangers of training and evaluating AI tools based on experts' know-what. Lebovitz et al. (2021) presents a field study of radiologist-managers evaluating five previously existing ML-based diagnostic radiology tools in pilot studies for possible adoption as components of a hospital's medical WSs. Although five ML tools proved "highly accurate, all five tools performed poorly during internal pilot studies" [leading] "managers to confront the high uncertainty involved in evaluating human experts' knowledge outputs (know-what) and to recognize that ML-based AI tools did not capture experts' tacit knowledge practices (know-how)."

Further evaluation was to continue for three of the five tools. In those cases, the radiologist-managers might have used the AR framework for exploring alternatives for the roles that the ML tools would perform, i.e., would they be used *for monitoring, for providing information, for providing capabilities, for controlling* some aspect of radiologists' work, *for coproduction* through interactions between interactive tools and radiologists, or just for *executing activities*, i.e., producing interpretations of diagnostic images. The AR framework's facets of work also would lead to questions about whether facets of work other than decision making might be pursued, e.g., possibilities of supporting communication, coordination, learning, or other facets. Relative to the AE framework, the paper notes *efficiency*-related motives, but it focuses much more on issues related to *effectiveness*.

The roles of CIOs and Boards of Directors in strategic goals for AI. Li et al. (2021) uses data from 1,454 publicly traded Chinese firms to pursue two research questions: “Can the presence of a CIO facilitate AI orientation in firms?” and “How do boards affect the relationship between the CIO and AI orientation?” Variables in its research model include CIO presence, board educational diversity, board experience with AI orientation through interlock partners, and AI orientation. Text-mining techniques were applied to annual reports from 2016 to 2018 to identify sentences and concerns related to AI. AI orientation was coded as 1 if the firm had developed an AI orientation, and 0 otherwise. Thus, Li et al. (2021) does not cover the topics emphasized by the AR and AE frameworks because its research questions, analysis, and related insights are at the corporate level, not at the WS level.

On the other hand, follow-on research with groups of CIOs might be able to use the AR and AE frameworks to obtain more nuanced views of their beliefs about AI applications, e.g., which roles of AI-based agents would be most effective, which facets of work might be supported effectively, and which evaluation criteria seem most important to CIOs and board members.

AI and drug discovery. Lou & Wu (2021) uses patents and job postings to measure AI innovation capability (AIIC), the "the ability to develop, use, and manage AI resources in combination or copresent with other resources and capabilities to effectively conduct scientific discovery and R&D." AI resources include tangible AI resources (data, infrastructure, and algorithms that are customized for specific scientific discovery), skills to create, implement and deploy AI tools for scientific discovery (including knowledge of pharmacology to facilitate working with medical scientists), and AI-enabled intangibles such as firm practices and knowledge assets that can empower and foster the use of AI. The conclusion is that AIIC can help “develop new drugs at the intermediate level of novelty and new drugs whose mechanism of impact for treating a condition is known.”

Lou & Wu (2021) focuses on evidence related to resources for drug discovery but does not explain drug discovery as a work system. Follow-on research might use the AR and AE frameworks along with work system theory in analyzing specific drug discovery processes.

Coordinating human and machine learning. Sturm et al. (2021) uses agent based simulation to study impacts of ML on organizational learning. The simulation explores how levels of the human learning rate, code learning rates, and reconfiguration intensity affect the long-term knowledge of organizations with ML agents that have different initial learning capabilities under different levels of turbulence. The main findings are that ML systems with high initial learning capability reduce the need for human exploration; that humans’ learning behavior moderates nonlinear effects of reconfiguration intensity on organizational learning effectiveness; that effective organizational learning with ML systems in turbulent environments requires human exploration and rapid codification of knowledgeable humans’ beliefs.

Sturm et al. (2021) explores levels of aggregation and abstraction that the AR and AE frameworks do not address. In contrast, the AR framework identifies specific roles (*monitoring, providing information, providing capabilities, controlling,* and *coproducing*) that could provide focal points in exploring coordination of human and machine learning in specific situations.

Failures in fairness in automation. Teodorescu et al. (2021) explains “that human–ML augmentation is more complex and nuanced than is currently represented in the literature and more research is needed to understand these nuances better. … [It presents] a typology of augmentation composed of two dimensions: the difficulty of achieving fairness on a given set of variables and the locus of decision in the human–ML partnership. This typology results in four distinct approaches to augmentation to achieve fairness: reactive oversight, proactive oversight, informed reliance, and supervised reliance.”

Teodorescu et al. (2021) focuses on equity, one of the components of the evaluation dimension of the AE framework. Its careful analysis of fairness relies on the legalistic concept of *protected attribute* (such as race, gender, marital status) which leads to questions about whether an AI-based digital agent that might be deemed fair in one culture might seem unfair elsewhere.

Teodorescu et al. (2021)mentions several AI applications in which fairness is a key issue, i.e., Facebook’s violation of the Fair Housing Act through ad targeting, Amazon’s recruiting tool that discriminated against female coders, and a system used by law enforcement agencies that discriminated by race and gender. Its brief comments about those examples are not rich enough for applying the AR and AE frameworks as part of an explanation of the relevant WSs and the roles and responsibilities that those WSs delegated to AI-based digital agents. Applying the AR framework to detailed descriptions of those situations could reveal important aspects of digital agent usage, might point to different roles that an AI-based digital agent might play, and might point to different facets of work that would be affected. In turn, that might lead to realizations about fair ways to use an AI-based digital agent even if its use in automated decision making would be unfair.

Developing a machine learning application to support hiring in a large international organization. Van den Broek et al. (2021) presents a two year ethnographic study of how an ML vendor and corporate domain experts engaged in a process of mutual learning as part of a project initially aimed at increasing the efficiency of a hiring process at a major international firm that processed more than 10,000 trainee applications each year. The ML vendor had developed gamification exercises to provide data that could be used in evaluating applicants. "Unlike traditional selection methods that rely on self-reported data, such as questionnaires and CVs, these games simulate situations where participants' personalities and behavior are supposed to be reflected." The games promised insights into personality traits, cognitive skills, career values, and other personal attributes. While the ML tools were being tailored to the situation, HR experts were surprised to learn that "the algorithm inferred from the data that successful trainees were often introverts, while the HR experts had systematically rejected candidates who reflected this personality trait."

The research showed how a new *hybrid* practice emerged. "Two initially opposing practices of hiring—judging and selecting job candidates by relying on ML or by relying on domain expertise—eventually converged, creating a new hiring practice that combined both elements." With the hybrid practice, ML performed initial screening of applicants and HR experts interviewed the applicants who passed that screening. In relation to the AR framework, the ML digital agent played the role of *executing activities* related to the facet of *making decisions* in initial screening of applicants. In relation to the AE framework, the hiring work system became more *efficient* because HR experts could perform fewer interviews.

8. Discussion and Conclusion

This paper presented the AR and AE frameworks as complementary contributions for understanding and evaluating AI-based digital agents in their context of use. The first parts of the paper summarized a conceptual basis (algorithmic agents, the work system approach, and the AR framework) that had appeared in previous publications that were cited. The AE framework and this paper's discussion of how the AR and AE frameworks and other ideas related to AI apply to papers from an *MISQ* Special Issue on Managing AI had not appeared previously. The seven papers in the Special Issue provided an informal demonstration of the frameworks' potential usefulness. This paper's brief coverage of the seven papers showed that only a few engaged directly with real world applications of AI that could be described or analyzed using the frameworks. Most of the others produced interesting results without focusing on specific AI-based digital agents in specific application situations. They did that by using techniques such as agent based simulation, statistical analysis of patents and job postings, and analysis of the content of annual reports.

I had expected that an *MISQ* special issue on Managing AI would contain much more content related to describing and analyzing specific business situations in which managing AI applications presented important challenges. I had expected more focus on describing and evaluating successful or unsuccessful real world use of AI-based techniques to guide processes or to aid in producing useful business results such as new pharmaceuticals, better medical diagnosis, or greater fairness in applying AI techniques.

The AR framework as a lens for visualizing the Special Issue. In terms of roles (the horizontal dimension), several of the papers focused on *providing information* or *executing activities*. None focused on *monitoring activities*, *providing capabilities* that WS participants could use, *controlling activities*, or *coproduction* involving people and digital agents. In terms of facets of work, several of the papers had a clear focus on *making decisions* and one focused on *learning.* The papers said little or nothing about many other facets of work. Given that facets of work have been described a number of times and applied to hypothetical or real world situations, e.g., an illustrative hypothetical example, five real world examples in Alter (2022c), three examples in Alter (2022a), and another in Alter (2022b), it seems likely that many types of applications of AI-based digital agents were not mentioned at all in the Special Issue. That is not surprising given its limited number of articles, but it also implies that many other types of applications of AI-based digital agents might have been included. In particular, this demonstrates limitations of the common belief that current AI is mostly about ML, that ML is mostly about classifying and deciding, and that AI-based digital agents are therefore unlikely to touch other facets of work such as communicating, planning, improvising, providing service, performing support work, and so on.

The AE framework as a lens for visualizing the Special Issue. Four of the articles emphasized performance related to efficiency and/or effectiveness. One observed that explainability was an important issue but that fit with professional practices was perhaps more important. Anopther focused on equity (fairness). None emphasized engagement or empathy. Thus, four of AE framework's six evaluation criteria were discussed. It could be useful to find and explore AI-based digital agents that place significant emphasis on empathy and engagement, especially in light of Shneiderman's new ideas about human-centered AI.

Generalizations about AI capabilities. The question in this paper's title, "Can You Verify that I Am Using AI?" set the tone for this paper's skepticism about generalizations about the capabilities of AI. This paper assumed that AI is more like a buzzword or organizing vision and less like an identifiable set of concepts and capabilities. It is noteworthy that the definition of AI in the introduction to the seven articles does not describe the applications of ML in the two papers (Lebovitz et al., 2021 and van den Broek et al. 2021),that focused most directly on systems in specific settings.

AI-related research has generated many impressive advances in the time since the 1956 Dartmouth workshop and many digital agents associated with AI are proving useful and important today. That is all the more reason to use real examples when making claims about capabilities of AI. This paper's AR and AE frameworks could help in that regard.

Potential value of the AR and AE frameworks. Both the AR and AE frameworks are directed toward specific, factual, and non-exaggerated description and evaluation of the use of digital agents. The AR framework focuses on identifying responsibilities related to specific roles and specific facets of work. The AE framework focuses on fostering meaningful evaluation of digital agents by identifying important types of criteria and important stakeholders whose concerns might differ significantly. Those stakeholders might participate more effectively in analysis, design, and evaluation efforts if they recognize the limitations of their own criteria and ways in which those criteria differ from criteria of others.

The AR and AE frameworks are directly applicable for describing and exploring many of the topics or issues mentioned in the CFP for the HICSS track on AI, Organization, and Management, i.e., shifts in coordination as AI tools are used, changes in power relations, impacts of using AI on processes that are typically viewed as entirely driven and controlled by humans, evaluation of ethical implications of deployed AI methods, and KPIs and metrics for assessing the effectiveness of AI applications. Those important topics deserve analysis based on carefully defined ideas rather than vague generalizations, metaphors, and hype. The AR and AE frameworks may provide a readily usable way to organize and apply ideas that help in visualizing the realities of AI applications in real world settings.

References

Alter, S. (2006). *The Work System Method: Connecting People, Processes and IT for Business Results*. Work System Press.

Alter, S. (2008). Defining information systems as work systems: implications for the IS field, *European Journal of Information Systems,* 17(5), 448-469.

Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. *Journal of the Association for Information Systems*, 14(2), 72-121.

Alter, S. (2022a). Agent Responsibility Framework for Digital Agents: Roles and Responsibilities Related to Facets of Work," EMMSAD 2022, Leuven, Belgium,

Alter, S. (2021). Facets of Work: Enriching the Description, Analysis, Design, and Evaluation of Systems in Organizations, *Communications of the Association for Information Systems*, 49(13), 321-354.

Alter, S. (2022b). "Responsibility Modeling for Operational Contributions of Algorithmic Agents," AMCIS 2022.

Alter, S. (2022c). Understanding artificial intelligence in the context of usage: Contributions and smartness of algorithmic capabilities in work systems. *International Journal of Information Management*, 67, pp. 1-10.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence, *MIS Quarterly 45*(3), 1433-1450.

Boell, S.K. & Cecez-Kecmanovic, D. (2015) .What is an information system? Proceedings of HICSS, 4959-4968.

Burton-Jones, A., Recker, J., Indulska, M., Green, P., & Weber, R. (2017). Assessing representation theory with a framework for pursuing success and failure." *MIS Quarterly*, 41(4), 1307-1333.

Chinoy, I. (2010). Battle of the brains: Election-night forecasting at the dawn of the computer age. University of Maryland.

Fügener, A., et al. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quarterly*, 45(3), 1527-1556.

Ishiguro, K. (2021). *Klara and the Sun: A novel*. Alfred A. Knopf.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI Ground Truth Really 'True'? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What, *MIS Quarterly* 45(3), 1501-1525.

Li, J., Li, M., Wang, X., & Thatcher, J. B. (2021). Strategic Directions for AI: The Role of CIOs and Boards of Directors, *MIS Quarterly* 45(3), 1603-1643.

Lou, B., & Wu, L. (2021). AI on drugs: can artificial intelligence accelerate drug development? Evidence from a large-scale examination of bio-Pharma firms. Evidence from a Large-scale Examination of Bio-pharma Firms, *MIS Quarterly* 45(3), 1451-1482.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955) http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

Shneiderman, B. (2021). Human-Centered AI: Reliable, Safe, and Trustworthy, presentation for BayCHI, June 8, 2021, www.youtube.com/watch?v=9K2kWC2awUc.

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

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

Sturm, T., et al. (2021). Coordinating Human and Machine Learning for Effective Effective Organizational Learning," *MIS Quarterly* 45(3), 1581-1602.

Swanson, E. B., & Ramiller, N. C. (1997). The organizing vision in information systems innovation. *Organization Science*, *8*(5), 458-474.

Teodorescu, M. H., et al. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation, *MIS Quarterly* 45(3), 1483-1499.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring, *MIS Quarterly* 45(3), 1557-1580.