AIS Transactions on Human-Computer Interaction

Volume 15 | Issue 1 Article 5

3-31-2023

From Artificial Intelligence (AI) to Intelligence Augmentation (IA): Design Principles, Potential Risks, and Emerging Issues

Lina Zhou
University of North Carolina at Charlotte, lzhou8@uncc.edu

Cynthia Rudin
Duke University, cynthia@cs.duke.edu

Matthew Gombolay
Georgia Institute of Technology, matthew.gombolay@cc.gatech.edu

Jim Spohrer
ISSIP.org, spohrer@gmail.com

Michelle Zhou
Juji, mzhou@juji-inc.com

See next page for additional authors

Recommended Citation

Zhou, L., Rudin, C., Gombolay, M., Spohrer, J., Zhou, M., & Paul, S. (2023). From Artificial Intelligence (AI) to Intelligence Augmentation (IA): Design Principles, Potential Risks, and Emerging Issues. *AIS Transactions on Human-Computer Interaction, 15*(1), 111-135. https://doi.org/10.17705/1thci.00185

DOI: 10.17705/1thci.00185

From Artificial Intelligence (AI) to Intelligence Augmentation (IA): Design Principles, Potential Risks, and Emerging Issues

Authors

Lina Zhou, Cynthia Rudin, Matthew Gombolay, Jim Spohrer, Michelle Zhou, and Souren Paul

Transactions on Human-Computer Interaction

Volume 15 Issue 1

3-2023

From Artificial Intelligence (AI) to Intelligence Augmentation (IA): Design Principles, Potential Risks, and Emerging Issues

Lina Zhou

The University of North Carolina at Charlotte, lzhou8@uncc.edu

Cynthia Rudin

Duke University, cynthia@cs.duke.edu

Matthew Gombolay

Georgia Institute of Technology, matthew.gombolay@cc.gatech.edu

Jim Spohrer

International Society of Service Innovation Professionals, spohrer@gmail.com

Michelle Zhou

Juji, mzhou@juji-inc.com

Souren Paul

Northern Kentucky University, souren.paul@gmail.com

Recommended Citation

Zhou, L., Rudin, C., Gombolay, M., Spohrer, J., Zhou, M., & Paul, S. (2023). From artificial intelligence (AI) to intelligence augmentation (IA): Design principles, potential risks, and emerging issues. *AIS Transactions on Human-Computer Interaction*, *15*(1), pp. 111-135.

DOI: 10.17705/1thci.00185

Available at http://aisel.aisnet.org/thci/vol15/iss1/5

Transactions on Human-Computer Interaction

Research Commentary DOI: 10.17705/1thci.00185 ISSN: 1944-3900

From Artificial Intelligence (AI) to Intelligence Augmentation (IA): Design Principles, Potential Risks, and Emerging Issues

Lina Zhou
Department of Business Information Systems and Operations Management, The University of North Carolina at Charlotte
Lzhou8@uncc.edu

Cynthia Rudin
Department of Computer Science and Department of Electrical and Computer Engineering, Duke University

Matthew Gombolay
Interactive Computing, Georgia Institute of Technology

Jim Spohrer
International Society of Service Innovation Professionals

Michelle Zhou
Juji

Souren Paul
School of Computing and Analytics, Northern Kentucky University

Abstract:

We typically think of artificial intelligence (AI) as focusing on empowering machines with human capabilities so that they can function on their own, but, in truth, much of AI focuses on intelligence augmentation (IA), which is to augment human capabilities. We propose a framework for designing intelligent augmentation (IA) systems and it addresses six central questions about IA: why, what, who/whom, how, when, and where. To address the how aspect, we introduce four guiding principles: simplification, interpretability, human-centeredness, and ethics. The what aspect includes an IA architecture that goes beyond the direct interactions between humans and machines by introducing their indirect relationships through data and domain. The architecture also points to the directions for operationalizing the IA design simplification principle. We further identify some potential risks and emerging issues in IA design and development to suggest new questions for future IA research and to foster its positive impact on humanity.

Keywords: Intelligence Augmentation, Artificial Intelligence, Design Principle, Simplification, Interpretability, Risks, Human-AI Interaction.

Fiona Nah was the accepting senior editor for this paper.

1 Introduction

Artificial intelligence (AI) applications have grown tremendously in number in recent years, particularly in areas such as medicine, finance, customer service, and online marketing. As the relationships between machines and humans evolve, we need to re-examine how human-machine teaming may impact human work in the future. Scholars have increasingly acknowledged that, in addition to a focus on empowering machines with human capabilities so that they can function on their own, AI also focuses on intelligence augmentation (IA); that is, enhancing and elevating human intelligence, capacity, performance, protection, and quality of life with support from information technology (Zhou et al., 2021). We consider fraud detection in financial statement audits as a scenario to illustrate the different focuses between IA and AI.

The Association of Certified Fraud Examiners (2022) has estimated that organizations lose five percent of their revenue to fraud each year and for losses to fraud on a global scale to exceed more than US$4.7 trillion. Thus, fraud detection has significant economic impacts on organizations, their investors, and other stakeholders. Although financial reporting misconduct (9%) occurs less commonly than other types of fraud, it costs significantly more (median value US$593,000) (Association of Certified Fraud Examiners, 2022). Independent human auditors play an important role in detecting fraud risks, such as material misstatements in financial reporting, which helps support effective internal control, promote good financial reporting practices, and protect investors.

In this scenario, AI primarily automates the detection process by optimizing how well fraud-detection models perform algorithmically. However, state-of-the-art models still face many issues due to the complexity and/or challenges of detecting fraud risks. As a result, those models may fail to identify material misstatements in financial statements, which can cause financial losses to organizations and investors as the above statistics evidence.

On the other hand, IA focuses on how to empower human auditors in detecting fraud risks with AI-enabled detection models. Even though the above models have imperfections, an IA system can:

1) Make inferences in a way that resembles the way auditors would explain to people how to detect fraud
2) Identify precursors or evidence for fraudulent activity in a form that auditors can easily check and leave it to human auditors to decide whether and how to use the precursors or evidence in assessing fraud risks
3) Assist human auditors in complying and keeping up with Public Company Accounting Oversight Board auditing standards, following good auditing practice, and recognizing possible mistakes, which can help human auditors sharpen their auditing skills; and/or
4) Enable human auditors to identify new ways to detect fraud.

We can characterize IA along two main dimensions: 1) a technical dimension that encompasses computer systems that enable IA and 2) a social dimension that describes the stakeholders and environmental factors that IA interacts with via taking inputs and/or exerting impact. Additionally, the two dimensions relate to each other. The technologies that enable IA have advanced rapidly. For example, many real-world applications now use deep learning, which can automatically learn complex patterns from vast amounts of data (particularly unstructured data such as images, sound waves, and text). OpenAI's ChatGPT, which gained one million users in the first five days after it released to the general public, seems to be at or near a tipping point of being generally useful to people across many different domains by enabling "human-machine hybrid work" (Mollick, 2022). In this research commentary, we focus more on IA's technical dimension even though we recognize the social aspect's importance.

This commentary starts with addressing "WH" questions central to IA. We draw on discussion at a panel at the 55th Hawaii International Conference on System Sciences. In answering these questions, we introduce an IA framework that comprises six key aspects important to designing and developing IA systems. In particular, the how aspect highlights four IA design principles (i.e., simplification, interpretability, human centeredness, and ethics), while the what aspect goes beyond the direct interactions between humans and machines by introducing their indirect relationships through data and domain, which leads to the proposed four-component IA architecture. These components also provide directions for operationalizing the IA design simplification principle. In addition, we discuss the potential risks related to developing IA technology, such as privacy, misuse, deskilling, and emotional attachment/detachment. Furthermore, we identify emerging issues in IA research, such as design patterns, IA maintenance, conflict management, IA intervention, data cycle in IA development, re/upskilling, embodiment for IA, and IA use cases. The proposed 6WH framework, the IA component architecture, and the identified potential risks and emerging issues can serve as a guide for researchers, designers, developers, managers, employees, policymakers, and other stakeholders involved in the IA ecosystem to foster IA's positive impacts on humanity.

2 A 6WH Framework of IA

The IA literature has grown exponentially in the past decade (Zhou et al., 2021). Going beyond a conceptual definition, we propose a framework for IA by asking the standard 6WH questions to systematically explain IA. The framework characterizes IA based on six dimensions: why, what, how, who/whom, where, and when. The *why* aspect states the motives behind IA, *what* describes what IA focuses on compared with AI in general, *how* introduces the methods and guidelines for building IA, *who/whom* depicts the roles involved in the IA ecosystem, *where* illustrates where one can apply IA, and *when* addresses when one should use IA. We introduce each of the above dimensions in detail next.

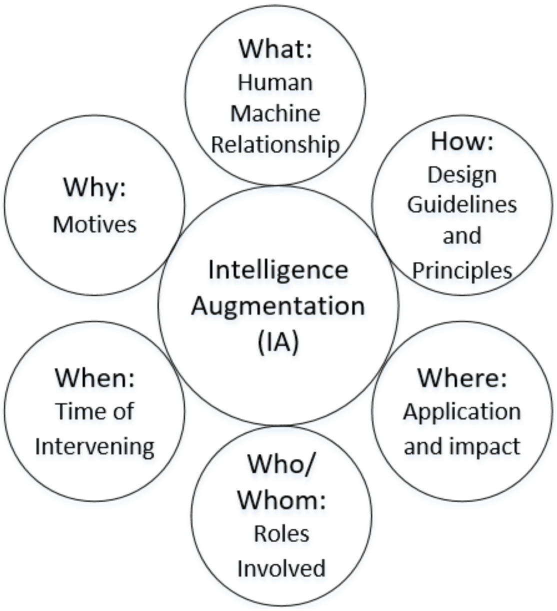

Figure 1. A 6WH Framework for IA

2.1 Why: Motives of IA

Many people have come to expect increasingly more from AI due to evolving computing hardware and software technologies and ever-growing big data. However, a reality check (Cross, 2020) suggests that many people feel AI has yet to deliver its grand promises and still struggles with reasoning and generalization across different tasks. Some scholars even think that AI remains in its very early stages and that it fails to reach its potential in many areas, particularly in some real-world applications. For instance, output quality (how well a system performs job-related tasks, such as improving the clinical outcomes for patients and enhancing clinical efficacy) constituted a contributing factor for why surgeons adopted technology (Reynolds, 2020). Today, most surgical robotics have a limited ability "to perform procedures and make decisions automatically without major human intervention" (Stumpo et al., 2021, p. 2680). Poor output quality can result from the difficulty in addressing model issues such as complexity, robustness, and adaptation (to the environment and tasks) and data issues such as availability, quality, and representativeness. These challenges may impose fundamental constraints on what AI can do. IA has the potential to give machine learning models a more useful alternative by drawing on humans' knowledge and experience and keeping humans in charge.

In many ways, AI already should be IA because AI focuses on "serv[ing] human needs by way of comprehensible, predictable, and controllable tools, appliances, and user experiences" (Shneiderman, 2020, p.113). In other words, humans should control or command AI since AI should not perform its tasks and solve problems separately. In AI's early days, when the first computer program demonstrated the ability to converse with humans in natural language (a pioneer in chatbot), many people perceived the program to have the potential to improve the quality of people's lives, such as patients with mental health issues (Colby et al., 1966). Such potential further inspired the U.S. Government's interest in a machine that could transcribe and translate spoken language (Anyoha, 2017) to keep it competitive with other superpowers. One could argue that IA systems have had more practical successes than fully automated decision-making AI tools—not only in speech transcription and translation but also in many other areas such as Internet search and recommender systems.

IA faces several critical practical challenges. The lack of model interpretability for human understanding and evaluation along with the imperfection and under-delivery of AI for real-world application results in humans' lack of trust in most AI technologies developed in research labs, which further hampers their widespread adoption. We can perceive this poor trust as warranted in many cases since AI model engineers do not often test their models well enough before releasing them, and problems with their training and datasets could cause harm in high-stakes settings. For instance, although AI in medicine has paved the way for smart operating rooms, where robots play a major role in carrying out surgical steps while minimizing human intervention (Stumpo et al., 2021), neurosurgeons often face problems in trying to adopt the technology in their clinical practice (Reynolds, 2020). As another practical issue, machine learning models for computer vision and language continue to increase in complexity and, thus, have led to an increase in their engineering cost and extra cost associated with their optimization. As a result, many organizations cannot afford to build powerful AI models, test them, and deploy them in real applications. One possible solution to create opportunities for businesses involves putting humans in the center and building simpler, less costly models that can deliver value for users. Building these simpler models that interface better with humans can drive future research and development in IA and lead to more trustworthy and practical AI system deployments.

2.2 What: The Relationship between Humans and Machines

One needs to understand the relationship between humans and machines to grasp the IA concept. On one hand, machines have a wide range of capabilities that can complement or extend certain capabilities that humans have. Machines currently outperform humans in some intelligence dimensions, such as computational efficiency, storage, and throughput efficiency. For example, machines can communicate with tens of thousands of people in an interactive way simultaneously, which far surpasses what any human can achieve. Machines can also appear to be empathetic and shield emotions without having actual emotional burdens that typical humans do. On the other hand, humans can broadly understand the world at a system level in a way that current AI systems cannot replicate. In addition to humans' excellence in perceptual, soft, and some cognitive skills, identity can fundamentally distinguish human intelligence and machine intelligence. Klein and Nichols (2012) show that people can derive their personal identity from the memories about past events and mineness ("the mode of existence of experiences and does not presuppose a subject, but rather constitutes it" (Fasching, 2009, p. 133)). Despite the exponential growth in computer memory and retrieval efficiency, AI systems typically cannot (or their creators did not design them to) exhibit mineness or human-like memory that can tell a story about its own experience or having surprises and performing commonsense reasoning and generalization. Therefore, it can be a win-win situation for humans and machines to team up.

IA emphasizes the symbiotic relationship between humans and machines (Licklider, 1960). In the relationship, humans define the goals for machine intelligence, while machines incorporate human knowledge to refine their models and translate their learned information to humans to form new knowledge. On one hand, human intelligence defines machine intelligence and ways of achieving that so that machines can best help rather than compete with (or replace) humans. Furthermore, humans define the goal, build machine intelligence, and drive the socio-economic implications of machine intelligence to help human intelligence.

The symbiotic relationship fundamentally addresses the issue of how humans can trust AI, especially when the latter lacks full capability. By using machines to augment human intelligence, humans can become more intelligent in some ways than they have ever been. AI can also help scale human operations and make scarce human resources much more accessible. Alternatively, humans may offload mundane tasks to machines and trust machines to do them in exchange for the time and effort to do something more interesting or creative or tasks they can perform better. In both cases, machines can help humans gather information and lay out different alternatives but leave it to the human users to make the final decisions.

In addition to the direct relationship between humans and machines, they could also have an indirect relationship through data and/or domain. To this end, we extend MLBiD (a framework of machine learning on big data) (Zhou et al., 2017) to IA by introducing a four-component architecture that comprises human, model (machine), data, and domain (see Figure 2).

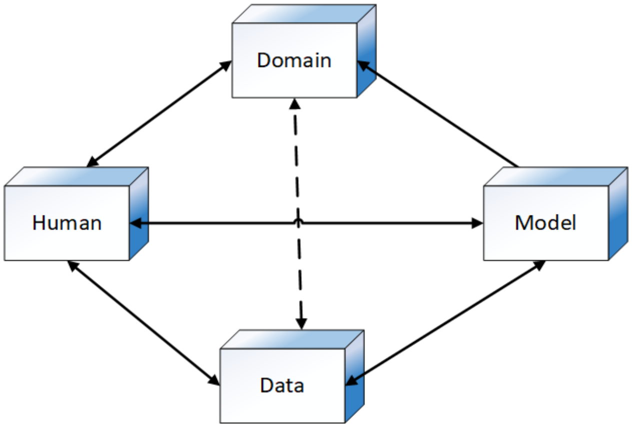

Figure 2. A Four-component IA Architecture

- The *human* component covers developers, end users, and other roles involved in the IA ecosystem (see Section 2.3); their knowledge, experience, and preferences in terms of the problem domain and models; and the application context such as organizational culture.
- The *model* component covers building and maintaining machine learning and statistical techniques that can learn patterns from data with guidance from human and domain knowledge.
- The *data* component covers general or domain-specific data that comes from various modalities, such as numbers, text, image, and video, and can appear in various forms, such as raw data, rules, and patterns.
- The *domain* component serves both as a knowledge source for IA and as a context in which one deploys IA. In addition, different tasks may rely on different types of intelligence (Zhou et al., 2021) and have different automation potentials (Vimalkumar et al., 2021).

These components intertwine and mutually enhance each other. For instance, data strengthens the relationship between humans and machines by both serving as model inputs/outputs and capturing/improving human insights. Likewise, the domain also enhances the relationship between humans and machines because humans can provide both domain knowledge and feedback on domain outcomes that models generate; meanwhile, machines can consider domain requirements that they solicit from humans and generate outcomes to solve humans' problems in the domain. The domain and data share an obvious connection since the data are generated from the domain.

2.3 Who/For Whom: The Roles Involved in IA

The roles involved in IA can range from end users, developers, super users, analysts, analytical modelers, and managers to individuals subject to the decisions. We must consider who should build IA and who should maintain it (e.g., debugging, model tuning, and enhancement) when something goes wrong. Efforts to build IA involve two related yet conflicting trends.

The first trend concerns efforts to democratize AI by making it available and accessible to a broader user population even though people may have limited AI experience or domain knowledge. IA and AI differ in how much AI (or machine learning) expertise they assume the user to have. Unlike AI, which often requires AI knowledge or expertise (or excludes user input entirely), IA gives more and special emphasis on AI democratization because not everyone has AI expertise, understands machine learning, or has programming skills. As a result, IA ensures everyone can interact with AI in the real world (e.g., ChatGPT (Mollick, 2022)). As more people develop and apply machine learning to more applications, it impacts more people with low AI literacy or numeracy skills. In a hypothetical world, users whose intelligence IA will augment and whom IA's decisions will affect constitute the right people to answer the questions about building and maintaining IA. It is a desirable situation because these people understand and care about those applications. Further, the IA ecosystem should also involve policymakers and companies who develop IA systems.

The second trend concerns the increasing complexity that one faces in building machine learning applications from both development and operational perspectives. Machine learning continues to become more complex in terms of both its algorithms and the number of parameters. Even though machine learning frameworks with simple APIs (e.g., Keras (Keras Team, 2015)) help reduce developers' cognitive load in implementing algorithms and make it easy to develop models, it would be inadequate to put humans in the center of an IA ecosystem without providing them with systematic guidance on how to develop IA from developers or engineers and interfaces to help them understand the implications that their programming choices may have. Successful IA deployments require a close partnership between the business/users and AI experts.

2.4 How: Design Guidelines and Principles

Drawing on decades of research and experience in machine learning, human-robot interaction, software engineering, and design science, we introduce four IA design principles: simplification, interpretability, human centeredness, and ethics.

2.4.1 Simplification

This principle focuses on simplifying humans' decision-making process and the need for human knowledge. The IA architecture components point to the directions for the simplification principle (see Figure 2):

1) First, take complex **models** and turn them into simple tools that people can use for some applications. Despite a complex tradeoff between simplicity and accuracy, the context of the discussion is to improve simplicity without compromising accuracy (Rudin, 2019). For instance, studies (Paleja et al., 2020; Silva et al., 2020; Sun et al., 2020) have proposed techniques to simplify the training process and resulting deep neural network models without incurring accuracy loss.

2) Second, reduce **data requirements**. We can approach data simplification from two perspectives: data quality and data preprocessing. Generally, higher-quality data is more difficult to obtain and takes more time in preprocessing for model building. We can capture data quality in four dimensions (Wang & Strong, 1996): intrinsic (in conformance with the true data values, such as accuracy and reputation), contextual (pertinent to the users' task or decision-making process at hand, such as value added, relevancy, timeliness, completeness, and an appropriate amount of data), representational (concerning the format and meaning of data, such as interpretability, consistent and concise representation, and ease of understanding), and accessibility (available or obtainable to data consumers). Obtaining quality data incurs various costs, such as data-acquisition costs and labeling costs (Turney, 2002). Poor data quality can result in poor model performance (Sanders, 2017), delayed project deliverables, lower customer satisfaction, and increased costs (Redman, 1998). On a related note, researchers have claimed data preprocessing or preparation (e.g., data cleansing, normalization, and labeling) to take up to 70 percent of the total time in machine learning or data-mining projects (see (Lean et al., 2006)). Thus, simplifying data by lowering the data-quality requirements and data-preparation efforts can have significant practical and managerial implications. Nevertheless, it is difficult to operationalize the data-simplification principle in isolation; rather, it requires one to interact with other IA architecture components. For instance, supervised machine learning models require labeled data (high representational data quality and data preparation effort) for training and testing, and the latter often remains practically challenging and costly to obtain. In dataless classification (Chang et al., 2008) and zero-data learning (Larochelle et al., 2008), a machine learning technique uses world knowledge, such as the meaning of class labels or a description of classes, to induce classifiers. Inspired by the above types of methods, zero-shot learning (Xian et al., 2019) allows a model to recognize data such as image objects from classes that it may not have observed during training. Moreover, combining reinforcement learning and natural language processing can lead to improved success rate in zero-shot learning and improved efficiency in transferring knowledge from solved tasks to new tasks (Silva et al., 2022). Few-shot learning (Li et al., 2006) can generalize to new tasks with only a few samples by leveraging prior knowledge. In addition, one can alleviate the need for contextual and representational data quality by leveraging foundational models, such as BERT (Devlin et al., 2019), DALL.E2 (see https://openai.com/dall-e-2/), and GPT-3 (Floridi & Chiriatti, 2020), which use massive amounts of unlabeled data for training yet can serve as the foundation for training models for and

transferring knowledge to a wide range of downstream tasks. Foundation models can "facilitate NLP research and model development at all scales" (Wiggins & Tejani, 2022, p. 1). One can adapt these general models to specific tasks with much smaller task-specific data via fine-tuning. Another way to achieve simplified data is by focusing on parts of data (e.g., parts of an image that one wants to compare with prototypical parts of images from a given class (Chen et al., 2019)). Furthermore, combining simplified data and simplified models would be even more powerful.

3) Third, echoing the idea behind AI democratization (see Section 2.3), lower the AI literacy that **human users** require to interact with AI.

4) Fourth, simplify the **domain** by drawing on a multi-dimensional comparison between human and machine intelligence (Zhou et al., 2021). A domain or task would often be considered simpler if it is structured (vs. unstructured), static/certain (vs. dynamic/uncertain), repetitive (vs. non-routine), specialized (vs. general), and so on.

2.4.2 Interpretability

Users need to be able to interpret AI algorithms or machine learning models and such algorithms and models require transparency for users to adopt them in high-stakes decisions. For instance, people have made significant attempts at developing machine learning models for medical imaging. It would be very helpful to explain to a radiologist why a machine learning model based on the imaging data suggests that they should biopsy a lesion (Barnett et al., 2021). As one possible explanation, a part of the lesion under evaluation could look like part of another patient's already diagnosed lesion. Such an explanation would help a radiologist make an accurate diagnosis. Developing interpretable IA methods also helps close the control loop on human-machine interaction. If AI users or clients do not understand the reasoning process that a model followed to arrive at a particular decision, they would not be able to adopt the model nor help machine learning engineers troubleshoot or improve it. Interpretable machine learning, which comes in different forms as we illustrate below, addresses the above issues. Note that interpretable machine learning (Rudin et al., 2022) differs from explainable machine learning (Rudin, 2019); while explainable machine learning explains the important variables in black box models, interpretable machine learning designs inherently interpretable models that reveal their reasoning processes. Some example interpretable machine learning models include:

- **Scoring systems or risk scores**: in a scoring system that awards points for each feature, one can use these points in sum to predict risks. For instance, physicians in intensive care units in hospitals use AI-generated risk scores to predict patient risks for seizure (Ustun & Rudin, 2019). Historically, people have not designed risk scores using AI. Traditional methods for producing them rely on manual feature elimination in a post-processing step, which does not lead to optimal solutions. To learn optimized scores, researchers have formulated the risk scoring problem as a mixed integer non-linear program and proposed optimization-based methods for producing scoring systems from data. The methods "allow practitioners to address application-specific constraints without parameter tuning or post-processing" (Ustun & Rudin, 2019; Xin et al., 2022).

- **Sparse decision trees**: decision trees offer interpretability by producing decision rules that humans can easily comprehend. Nonetheless, complex decision trees with many leaves and much depth go against the interpretability principles. Sparse decision tree models maximize the accuracy while minimizing the number of leaves (Lin et al., 2020). Researchers have extended these models to work for reinforcement learning domains in which these models learn a decision-making policy through trial and error (Paleja et al., 2020; Silva et al., 2020).

- **Dimension reduction and data visualization**: these models take high-dimensional data, such as biological data, and project it down to a two-dimensional space so that people can try to understand the cluster or manifold structure in the data. One related challenging question concerns how to preserve the original data's local and global structure (Wang et al., 2022a).

- **Estimated causal effects**: to perform observational causal inference, which goes beyond prediction, one can mimic a randomized controlled trial that randomly assigns participants to one of two groups: an experimental group that receives the evaluated target treatment and the control group that receives a conventional (or no) treatment. In addition, the participants in both types of groups match as closely as possible, and the target treatment constitutes the only major difference between the groups. If one finds any difference in the outcomes between the two groups, under appropriate assumptions, one can infer that a cause-effect relation between the treatment and the outcome exists (e.g., see Wang et al., 2021).

Since interpretability makes it easier to understand the reasoning behind predictions and decisions, it can also help one operationalize the simplification principle.

2.4.3 Human-centeredness

By putting humans at the center of systems-design thinking, human-centered AI focuses on designing systems that “support human self-efficacy, promote creativity, clarify responsibility, and facilitate social participation” (Shneiderman, 2020). The emphasis on humans encourages technology design to consider goals, such as privacy, security, social justice, and gender equality. The 10 levels of automation represents the canonical taxonomy for understanding a human’s (or machine’s) role in a human-AI system (Sheridan & Verplank, 1978). In particular, the taxonomy ranges from manual control/no AI assistance (level 1) to full automation (level 10). However, this taxonomy only defines and does not prescribe: which level of automation should a system have? Further, the taxonomy helps address how humans can or should trust AI, especially when an automated system has different competencies. Research has shown that humans can have trouble self-regulating their trust in automated systems, which can result in inappropriate compliance or reliance on their advice or actions (Gombolay et al., 2018; Natarajan & Gombolay, 2020). To this end, Shneiderman (2020) proposed a framework for creating new technology designs that comprise two dimensions (i.e., human control and computer automation) that each range from low to high levels. Combining these two dimensions can lead to four ways to think about new designs. For instance, high human control and high automation could co-exist (e.g., washing machines and cruise control in cars). In addition, one can improve low human control and automation design to high human control design either with or without higher computer control. To prevent against excessive computer control or human control and their associated risks, every design process for an improved system should include a step to design the coupling between the two dimensions (Shneiderman, 2020). Another key idea behind human centeredness concerns a trade-off between researchers’ and designers’ desires to make computers humanlike and human users’ desire to be in control.

2.4.4 Ethics

While algorithmic automation can help an organization achieve economic gains by improving process efficiency, it also raises ethical concerns (Vimalkumar et al., 2021). One can extend the ethical principles for design science research (Myers & Venable, 2014) to guide efforts to design and develop IA to help address potential AI-related risks to individuals and society as a whole. These principles include the public interest (explicitly identify all stakeholders and critically consider what benefit or harm may result for/to such stakeholders), informed consent (obtain informed consent from any person involved in the project), privacy (ensure adequate safeguards to protect the privacy of the people who are directly involved in the current project and who might use the artifact or be affected by its use), honesty and accuracy (acknowledge inspiration from other sources and honestly report research finding), property (agree about the ownership over the IP and collected information), and artifact quality (ensure an artifact’s quality and that one can use it safely). However, addressing these ethical principles presents significant research challenges due to their implementation complexity. To this end, researchers (Benke et al., 2020) point out two pathways that one can extend to IA design: 1) articulate the next-generation ethical principles using prescriptive knowledge structures from AI and 2) extend established AI conceptualizations with an ethical dimension.

2.5 When: The Time of Intervening

Given that IA focuses on augmenting human intelligence, an important question concerns when machine intelligence should intercede to augment human intelligence and when humans should make a decision on their own. If machine intelligence intervenes too much (e.g., the old Microsoft paperclip that the company intended to act as the human user’s assistant), many people will choose to turn it off. In addition, letting humans decide when to use machine intelligence also helps to keep AI under control. A related question concerns when humans should trust machine intelligence.

Some moments when machine intelligence or an AI system might intervene include:

- At key moments when humans feel uncertain or confused about making the right decision or coming up with a solution to a problem. For instance, decades of research and practice in situational awareness suggest that humans perform poorly at vigilance tasks because they require hard mental work and are stressful (Warm et al., 2008). Yet, humans should know if something important will come in the future.

- When machine intelligence can perform a task more easily and cost effectively than a human. In the case the model performs poorly, it may be important to be able to diagnose what went wrong. Examples include self-driving cars, face ID, and digital voice assistants.
- When machine intelligence can better ensure the quality measures that human users value most. For instance, once human users decide on accuracy as the most important quality measure for a specific task, they should leverage machine intelligence to achieve better accuracy since machine intelligence generally outperforms human users in such tasks (humans generally need to check high-stakes decisions before making them). We have seen much growth in the number of documents and regulations regarding AI ethics such as fairness (vs. bias) in recent years (Schiff et al., 2021; Robert et al., 2020). If humans or businesses consider fairness as a key quality measure and perceive that AI models could help eliminate human biases in decision making, then they could leverage machine learning models for assistance. Criminal justice represents one example: one can more easily fix an algorithm than human judges given the latter constitute biased black boxes (Mullainathan, 2019).
- When transferring the power that humans do not want to have or may not mind forgoing to machines. This issue differs from whether AI or humans can perform a task more capable. Examples include smart thermostats and spelling checkers. In a business context, a human customer service representative may not want to answer the same questions hundreds and even thousands of times a day when interacting with customers. Conversational bots, on the other hand, can do that constantly and instantaneously, and a transition to a human representative would only occur whenever necessary. Such a design allows humans to focus on less routine and more interesting tasks (Gagné et al., 2022), which might also help boost their productivity.
- When an operation becomes dangerous (e.g., when natural disasters, war and regional instabilities, and other humanitarian crises occur) (Walsh et al., 2019, xxiii).
- When one lacks resources and humans cannot perform a task due to limited or nonexistent availability. An example includes bedside monitors in hospitals. Humans cannot feasibly look at each bedside monitor manually to flag possible emergencies; thus, machines should perform it automatically.

Researchers have also considered hybrid or "mixed-initiative" schemes that essentially automate the problem of deciding whether a human or machine should do which task and when. For example, Johnson (2010) developed a "meta" AI that dynamically determined which tasks to allocate to a human pilot or an autopilot for landing on the moon to achieve a workload balance between the human and autopilot systems. In contrast to such discrete allocation schemes, some researchers have tried to develop approaches that blend human and AI system inputs, which could be a state or time-dependent blending (Bradshaw et al., 2004).

Transitioning to machine intelligence that enables decision support goes beyond just showing that the underlying algorithms make good decisions. It is important to contextualize and culturize the human-machine relationship concerning the rules and cultural norms of an organization, institution, particular field, and so on that IA exists in. In some contexts, humans have been and will always engage in making decisions regardless of whether humans or machines make better decisions. Prominent examples include high-stake and complex decision-making settings such as healthcare, criminal justice, and the military, which may involve complex planning and scheduling problems.

2.6 Where: IA Application and Impact

IA has received research attention and application across a wide range of disciplines (Zhou et al., 2021), such as information systems, computer science, medicine, business, telecommunication, education, architecture, information science, law, materials science, and so on. The deployment context and the end goal of IA for that specific context have emerged as significant factors in deploying AI (Paul et al., 2022). IA has begun changing the way people work, study, and live and has important implications for work in the future at all levels (i.e., the individual, group, community, organizational, societal, national, and international levels). IA has the potential to increase the effectiveness and productivity of human work significantly and, thus, to drive economic growth and development.

Typical IA applications use machine intelligence to improve an individual's or team's productivity and scale up the number of customers who can receive high-quality service in a fixed amount of time and budget. Examples include automating high-touch student services by giving personalized learning advice to each student based on the student's learning style and personality at scale, collaborative decision making in enterprises, personalized medicine and healthcare services at scale (Gombolay et al., 2018), human-computer collaborative driving, and cloud robotics (Zheng et al., 2017). According to the Gartner 2020 CIO Agenda (Panetta, 2019), 40 percent of infrastructure and operations teams in large enterprises will use AI-augmented automation by 2023, which will lead to increased productivity with greater agility and scalability. In addition, researchers have also highlighted cybersecurity and counterterrorism as IA application areas (Jain et al., 2021).

Emerging AI techniques such as foundation models and stable diffusion (generating high-resolution or realistic images conditioned on text descriptions (Rombach et al., 2022)) have enabled new use cases and resulted in much excitement. For instance, ChatGPT (Mollick, 2022) has the potential to extend human intelligence, creativity, and problem solving (e.g., composing marketing messages or generating python code to perform data analytics). One can potentially use these solutions directly or they can serve as a basis for humans to perfect. These technologies, while promising, may have less impact than one might expect given that one cannot easily control them; in other words, it is difficult to troubleshoot these models and constrain them to provide a domain-specific result. ChatGPT, for instance, can write a convincing scientific-looking paper; however, it might contain blatantly false content, and we lack a clear way to fix this issue as yet.

3 Potential IA Risks

While developing IA, one should recognize its potential risks. These risks pose challenges in building IA and signal its potential negative implications.

3.1 User Privacy

Training models for IA can benefit from using users' private information. For instance, users' demographic information can be invaluable for customer relationship management in business, such as online marketing and personalized recommendations. Many studies over the past decade have inferred users' demographics or attributes from their data, such as name, gender, ethnicity (Wood-Doughty et al., 2018), age, religious and political views (Bi et al., 2013; Kosinski et al., 2013), education levels, whether they have children or not, income, life satisfaction (Volkova & Bachrach, 2016), sexual orientation, marital status, blood type and zodiac sign (Zhong et al., 2013), location and social strategies (Dong et al., 2014), personality (Wang, Guo, Lan, Xu, & Cheng, 2016), and even intelligence, happiness, whether they use addictive substances, and parental separation (Kosinski et al., 2013). The data that researchers have used to make such inferences ranged from search queries (Bi et al., 2013), social media text (Volkova et al., 2015), network structure (Dong et al., 2014), emotion tone and contrast (Volkova & Bachrach, 2016), social images (Wu et al., 2017), purchase history (Wang et al., 2016), online behavior (Kosinski et al., 2013), and so on. Moreover, cognitive AI assistants can further infer a person's potential strengths and weaknesses by analyzing user behavior.

While personal information can enhance IA's effectiveness, access to it also raises serious privacy concerns. Indeed, the ongoing use of privacy-invading technology has become problematic (Noorden, 2020). Few locations in the US currently restrict someone from using facial recognition technology as well, which could lead to pervasive monitoring practices. A notable exception to the lack of is the Biometric Information Privacy Act (BIPA) (2008), an Illinois state bill that the Governor of Illinois signed into law in 2008. This bill restricts private entities from collecting, using, or storing biometric data, such as fingerprints, voiceprint, and iris scans, without written, informed consent. However, the bill has some key exceptions; for example, it places no restrictions on government entities. Further, courts continue to debate the law's implications, such as whether facial recognition constitutes "biometric" (e.g., *Fredy Sosa v. Onfido Inc.; Monroy v. Shutterfly, Inc*. (Kracht, Mueller, Sotto, & Sterns, 2018)). Five other states have introduced bills like BIPA with only New Hampshire successfully passing their bill, New Hampshire House Bill 523 (2018). At the national level, Senator Jeff Merkley introduced the "National Biometric Information Privacy Act of 2020", but the bill never received a vote (GovTrack, 2020). While the US has been slower to adopt privacy measures, the European Union has been more proactive and adopted the General Data Protection Regulation (GDPR) in 2018, which provides many safeguards for consumer data privacy.

3.2 IA Misuse

As humanity continues to democratize AI, some people or organizations will unavoidably take advantage of and even weaponize it to cause serious harm to humanity. After all, humans still supervise IA and can use it as they wish. Studies have shown how people may behave less ethically and be more willing to deceive when acting through AI agents across a wide range of social tasks (Gratch & Fast, 2022) partly because, by introducing a new agent into the equation, some new forms of power in interpersonal tasks could bypass existing social and regulatory checks on unethical behavior (Gratch & Fast, 2022). For instance, people have used social text bots to generate text for social media platforms. Advances in AI such as transformer-based machine learning techniques enable generating synthetic text that mimics human-created news stories in style and substance (Kreps et al., 2022). Research has even found this AI-generated text to be hard to distinguish from human-generated text. Due to the power that social media can exhibit in general and the influence that individual social media “users” can wield, bad actors can disseminate misinformation and disinformation rapidly to shape public perception and opinions on controversial political, economic, and social issues (Najee-Ullah et al., 2022). Despite ChatGPT’s great potential in assisting human tasks, its biggest impact may be in creating phishing emails or even malware, which would pose risks to cybersecurity (Lee et al., 2022). Moreover, businesses could use users’ private attributes to manipulate human behavior in a far more effective and less detectable way than traditional manipulative marketing strategies because the information that AI algorithms detect enables businesses to personalize addictive strategies and to exploit human biases, emotion vulnerability (Petropoulos, 2022), or even decision making vulnerabilities (Dezfouli et al., 2020). As a general concern, IA misuse could erode AI trustworthiness.

3.3 Deskilling

IA may enhance skills by fostering people to use “high-performance work practices” more frequently (Holm & Lorenz, 2022). For instance, increasing access to IA allows humans to look up information and even obtain suggestions very quickly. On the other hand, IA may also lead to reduced critical thinking skills in people. Deskilling refers to a loss in one’s skills and/or a decline in one’s performance after AI automates a manual task (Cabitza et al., 2017). For instance, overreliance on automated decision support can lead to automation bias, which can cause clinicians to stop looking for medical evidence after receiving AI output and to make technology-dependent reasoning rather than informed decision making (Ross & Spates, 2020). Consequently, deskilling can result in reduced clinician autonomy, decision-making quality, diagnostic reasoning, and communication with patients (Ross & Spates, 2020). In their survey study on the relationship between AI use in daily activities and job skill requirements, Holm and Lorenz (2022) found that individuals who used AI each day to take orders (receiving orders or directions automatically generated by machines) had some negative effects on jobs across all skill levels. In addition, AI use negatively affected mid-skill workers (who experienced experienced decreased learning and increased monotony) more strongly than high- and low-skill workers (Holm & Lorenz, 2022). Further, technology dependence could contribute to adverse safety events, and health practitioners need to avoid the dependence to prevent medical errors (Ross & Spates, 2020).

3.4 Emotional Attachment and Detachment

As AI becomes more human and emotionally intelligent, it has a greater potential to induce consumers’ attachment to it (Hermann, 2022). Researchers have designed social robots, which possess socially intelligence in a human-like way (Breazeal, 2002), to support personal interactions that involve emotions and feelings. As affective AI (Scheutz, 2012) and the reliance on automation increase, people start to develop some degree of social psychological attachment to machines. On the one hand, such an attachment could provide humans with a sense of security (support for exploration and self-development) and safety (comfort in times of distress) (Rabb et al., 2022). On the other hand, the bond that humans establish through interacting with machines may lead to separation anxiety (Rabb et al., 2022). A recent study (Gillath et al., 2021) found that priming attachment anxiety can lead to reduced trust in AI due to a preoccupation with thoughts about rejection and abandonment. Additionally, if affective artificial agents (that can have affective states of their own) act in an irresponsible manner or do not get the social aspects of the affective behavior right, they can cause harm to human users (Scheutz, 2012). For example, if an artificial agent knows that a person had a high chance to become addicted to playing a game, an irresponsible agent might empower another person or business to use such knowledge to exploit the person’s affective reactions and dependencies and take advantage of that person. Furthermore, AI companions may negatively impact their users' social interactions with other humans (Skjuve et al., 2021). Detachment from human partners can cause psychological harm to humans (Schrum et al., 2021).

4 Emerging Issues

With guidance from the 6WH IA framework, we identify several emerging issues in IA to suggest new questions for future research, address IA's potential risks, and enhance its positive impacts.

4.1 Design Patterns for IA

AI lives on software. Accordingly, in designing and developing IA, one can find enlightenment from software engineering practices. For traditional software development, software engineering methodologies and principles provide guidelines on designing, developing, evaluating, and maintaining computer software. The field has created design patterns that represent reusable solutions and best practices for addressing some recurring design problems in a certain context (Gamma et al., 1995).

We can reimagine the software engineering principles and systematic guidelines for IA, which include how to design and evaluate IA and how to choose and adapt design patterns in building IA. For instance, after comparing some machine learning methods, AI engineers/developers typically present the best model to users. However, it would be interesting to understand how users explore different good models so that they can decide on the best one themselves (e.g., see Xin et al., 2022; Dong & Rudin, 2020; Fisher et al., 2019; Wang et al., 2022b). For instance, doctors and other model users do not just want one single model but a set of models that they can explore so they can select the one they most prefer based on their assessment. Thus, presenting not just one but many good models could be a potential IA design pattern. We can expect these design patterns and guidelines to speed up the development process, detect failure, make it easy to correct system errors, and allow the system to fail gracefully. Identifying and adapting design patterns could represent an important and interesting direction for building IA.

4.2 IA Maintenance

Although it has never been easier for someone to build machine learning models, it remains difficult to understand whether the model they use represents the right or even a good one, how they can fix the model when something goes wrong, and/or how they can prevent it from going wrong in the first place. The fact that not all users or clients have engineering knowledge makes the maintenance issue even more important and challenging.

Based on software development experience over many decades, software engineering goes through lifecycles and constitutes an ongoing process. In particular, software maintenance, where developers constantly fine-tune system performance and debug and fix errors, represents a major issue. It may even reach a point where the developers have to discard the old system and start again from scratch. As IA becomes increasingly available, efforts to develop an IA system will likely follow a similar path to meet changing requirements and support trustworthy IA, which poses challenges for developing IA technology ecosystems.

Good maintenance practice, which in itself may require guidelines, has the potential to address the challenges. As with the how aspect that we discuss in Section 2.4, the maintenance issue concerns not only developers but the end users as well. Developers/organizations should inform IA end users about this issue as they start to adopt IA systems. Developers/organizations need to continuously feed knowledge to their IA systems and keep them current. Otherwise, the IA systems will face survival risks and their clients will eventually abandon them.

A related issue concerns maintenance costs. Owing to AI in medicine, one can use robotic technology not only to carry out surgical steps based on established protocol but also to make adjustments according to the environment and take preventative actions against errors (Stumpo et al., 2021). A survey study on global robotic-technology adoption (Stumpo et al., 2021) found that neurosurgeons who had never used robotics in clinical practice identified the inherent acquisition/maintenance costs as the most important factor that prohibited them from adopting robotics in their clinical practice. Despite the availability of various libraries, tools, and frameworks for developing IA solutions, the cost to implement (US$20,000 to US$1,000,000 (Sanyal, 2021)) not to mention maintain complete IA solutions remains not low or even high. This is partly because an off-the-shelf solution needs customizing and tuning to perform well in a specific organization or business context.

4.3 IA Conflict Management

An IA system does not always work in isolation; sometimes, it interacts with other IA systems. The latter can also benefit from design principles. Inter-IA relationships will likely be dynamic rather than static in nature, which would cause conflicts. For instance, consider if two family members who used personalized IAs (e.g., listening to their favorite music through their house) suddenly come to the same room at the same time. Addressing such a situation requires design principles. The individual IAs could follow different principles to resolve the conflict, such as negotiation (having the two IAs negotiate to arrive at a mutual agreement), avoidance (switching off IA to avoid conflict), and inaction (continuing to use both IA at the same time).

To help different IAs reach agreement regarding a certain quality of interest when negotiating to resolve conflicts, one can draw on the protocols for consensus or synchronization problems in multi-agent systems. Multi-agent systems, such as coordination between agents, security, and task allocation, have faced similar challenges (Dorri et al., 2018). Depending on individual agents' constraints and self-dynamics, the protocols for consensus can come in different types, such as consensus with constraints, event-based consensus, consensus over signed networks (either cooperative or competitive), and consensus among heterogeneous agents (Qin et al., 2017). On the other hand, the process to resolve conflict may lead to understanding and, thus, conflict can also be beneficial. Further, in a world where people had 100 machine agents that worked for them, they would need to act as an (parallel and not just serial) entrepreneur to manage the agents effectively and efficiently.

4.4 IA Intervention

The intervention issue extends to the when aspect of IA, and it could result from a combination of push and pull between AI automation and human involvement. We can expect the relationship between humans and machines to vary depending on the context. A more fundamental issue concerns how to ensure an IA intervention can recognize the context to ensure that it steps in at the right time and intercede in the right ways. One potential direction posits that IA should intercede when humans do something unethical regardless of their intention; for instance, the algorithm could intercede if humans try to spread misinformation (whether or not they recognize it as such) or information that could harm other people, such as hate speech and profanity. Such interventions could help prevent others from weaponizing the information. Platforms on which people spread misinformation could leverage IA interventions to moderate or even correct content to protect users from harmful information. For example, researchers have developed speech censorship chatbot systems with reinforcement learning techniques that comprise an aggressive speech censorship model and a speech purification model (Cai et al.,, 2022) to both detect aggressive speech and respond to its rapid evolution. Researchers have also developed an analytical pipeline with transformers and generative models to both detect and correct online misinformation (Meyer et al., 2022). Moreover, Gonçalves et al. (2021) showed that people perceive AI content moderation as more transparent than human content moderation, especially in situations when one cannot feasibly provide users with explanations or additional information for content removal. On the other hand, some have expressed concerns about using IA in policing, law enforcement, and judicial proceedings, such as predictive policing (using historic crime data to identify individuals or geographic areas with elevated risks for future crimes), which can have implications for discriminatory policing (Asaro, 2019).

4.5 Data Cycle in IA Development

In the IA component architecture (see Figure 2), humans and machines can interact indirectly through data. In case a machine learning model does not perform well enough, developers and researchers have typically gone through the *model cycle* by continuously tuning the model parameters to improve its quality. Even with robust model training techniques, they still need to cope with imperfect data (Whang et al., 2023). Given the significant influence that data quality and data representativeness have on how well machine learning models perform (e.g., Gennatas et al., 2020), we have seen a recent data cycle trend that involves an iterative process that involves collecting, cleaning, selecting, validating, and integrating data to improve its quality. The data cycle along with the model cycle provides a great way for humans to integrate their domain knowledge. The cycle can also help address the emerging regulatory requirements for AI fairness or ethics in general. Introducing data-centric processes and management for IA can further have significant impacts on organizations' techno-social structure and management priorities. On the other hand, this trend makes data quality essential to IA applications, which would otherwise introduce bias. In contrast to fairness, bias describes "problems related to the gathering or processing of data that might result in prejudiced decisions on the bases of demographic features such as race, sex, and so forth" (Ntoutsi et al., 2020). Bias can manifest in sensitive features and causal inferences, data representativeness, and data modalities (Ntoutsi et al., 2020). It can lead to inequalities in the AI outcomes (Carter et al., 2020). Thus, we need measures to mitigate and account for bias in advancing data-driven AI for augmenting human intelligence.

4.6 Re/Upskilling

AI will not replace service providers, but service providers who use IA to augment their performance will replace service providers who do not. Every person in an organization is a service provider, and, as AI-based digital twin technologies advance, learning many skills will become less costly and also be personalized for the learners (Spohrer et al., 2022). AI-driven education platforms can improve training by helping to educate everyone, which will make a big difference in reskilling (increasing people's ability to switch jobs after they have taken one) and upskilling people (continuous skill-building or education in various fields such as medicine). They have the potential to address the different levels of the digital divide in general (Riggins & Dewan, 2005; Wei et al., 2010) and the AI divide specifically (Carter et al., 2020), which includes inequalities related to access to AI, AI skills or self-efficacy, and outcomes from employing AI. Many expect future IA applications in education to fundamentally transform the domain and allow people to stay relevant and increase their mobility. Since younger generations are more likely to embrace the changes than the older ones, we may see skill shifts in whole industries, which would create new work, play, learning, shopping, and socializing opportunities for everyone.

4.7 Soft-skill Training

In addition to having language skills, one can also empower machines with advanced human soft skills, such as active listening (which enables them to engage with human users empathetically) and reading between the lines (which enables them to automatically infer users' unspoken needs and wants, interests, and personality from a conversation). Such skills can not only enable machines to automate complex tasks but also can augment human intelligence with insights to determine the next best actions. For example, machines can pass on patient personality insights that they infer to human caregivers, who can then best help the patients and, thereby, truly fulfill IA's purpose. Tangentially, researchers have found robots to potentially exhibit the ability to improve social skills in children on the autism spectrum (Scassellati et al., 2018).

4.8 Embodiment for IA

Whether built into a conventional computer, smartphone display interface, or physical robot, AI and IA can have significant impacts on how humans interact with a system. Robinette et al. (2016) have shown that humans in a simulated burning building may inappropriately rely on a robot that should guide the human outside even when that robot displays malfunction behaviors. In a healthcare application, researchers (Gombolay et al., 2018) showed that computer-based IA had harmful effects on humans' reliance on and compliance with the system as compared to a system in a physical robot. Kontogiorgos et al. (2020) showed that physical, robot-embodiment improved a user's willingness to continue work with a failing system. Yet, a potential mitigating factor in whether embodiment matters could be the degree to which the human anthropomorphizes the system (Natarajan & Gombolay, 2020). While it may seem counterintuitive or unproductive to put design and engineering effort into developing a physically embodied IA for only cognitive tasks, research shows that embodiment (or lack thereof) can have significant impacts on interaction fluency.

4.9 Exploring IA Use Cases

IA has abundant potential application. At its full potential, IA will impact our daily life both personally and professionally in many aspects. We envision at least two broad categories for using IA. The first category involves aiding humans in their personal lives. A personal AI advisor could augment every individual to help them learn about themselves (e.g., their strengths and weaknesses) and could gain information to help guide individuals to optimize the decisions they make (e.g., making a career choice or financial planning). The second category involves aiding humans at the workplace in various sectors, such as healthcare, biology, agriculture, climate change, social justice, and scientific discovery. All employees could have their own AI assistant that could augment them in many different ways, such as from automating human engagements to writing project proposals. Moreover, as we continue to democratize AI, every person should be able to customize and manage their own personal AI advisor or professional AI assistant without the need to write code or have AI expertise.

Naturally, using AI to assist human tasks or augment human intelligence in new ways will drive new requirements for IA technologies, and these use cases will become an important part of the IA ecosystem.

5 Conclusion

The 6WH framework and the detailed propositions we present in this paper provide guidelines for future research in IA. The framework, architecture, and risks and emerging issues that we identify can serve as a useful guide for scholars from diverse disciplines to further efforts to research and design IA, for practitioners to understand the IA lifecycle, and even for other stakeholders to evaluate and provide feedback on IA. Developing good IA systems is a challenging task. As IAs become better over time, humans as individuals will gain more capability. A better future will require not only technical improvements to the underlying algorithms but also improved interaction designs between AI and humans that consider the strengths and weaknesses on both sides. In our efforts to find good ways to use IA for humanity, we also need to measure socio-technical interactions, not just the technical part, as we think about IA as a socio-technical system.

Acknowledgments

We gratefully thank Dr. Ce Zhang (ETH Zürich) for serving as one of the panelists in the workshop on Augmented Intelligence and Future of Human Work at the 55th Hawaii International Conference on System Sciences. We greatly thank the Editor-in-Chief, Dr. Fiona Nah, for her constructive comments and suggestions that help improve the paper. We sincerely appreciate Adam LeBrocq’s thoughtful edits and comments.

References

Association of Certified Fraud Examiners. (2022). *Occupational fraud 2022: A report to the nations*. Retrieved from https://legacy.acfe.com/report-to-the-nations/2022/

Anyoha, R. (2017). The history of artificial intelligence. Retrieved from https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Asaro, P. M. (2019). AI ethics in predictive policing: From models of threat to an ethics of care. *IEEE Technology and Society Magazine, 38*(2), 40-53.

Barnett, A. J., Schwartz, F. R., Tao, C., Chen, C., Ren, Y., Lo, J. Y., & Rudin, C. (2021). A case-based interpretable deep learning model for classification of mass lesions in digital mammography. *Nature Machine Intelligence, 3*(12), 1061-1070.

Benke, I., Feine, J., Venable, J. R., & Maedche, A. (2020). On implementing ethical principles in design science research. *AIS Transactions on Human-Computer Interaction, 12*(4), 206-227.

Bi, B., Shokouhi, M., Kosinski, M., & Graepel, T. (2013). Inferring the demographics of search users: Social data meets search queries. In *Proceedings of the 22nd international conference on World Wide Web.*

Biometric Information Privacy Act, 740 ILCS 14/10. (2008). Retrieved from https://www.ilga.gov/legislation/ilcs/ilcs3.asp

Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Taysom, W., & Uszok, A. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Nickles, M. Rovatsos, & G. Weiss, (Eds.), *Agents and computational autonomy* (LNCS vol. 2969). Springer.

Breazeal, C. (2002). Designing sociable machines. In K. Dautenhahn, A. Bond, L. Cañamero, & B. Edmonds (Eds.), *Socially intelligent agents: Creating relationships with computers and robots* (pp. 149-156). Springer.

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA, 318*(6), 517-518.

Cai, S., Han, D., Li, D., Zheng, Z., & Crespi, N. (2022). A reinforcement learning-based speech censorship chatbot system. *The Journal of Supercomputing, 78*(6), 8751-8773.

Carter, L., Liu, D., & Cantrell, C. (2020). Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review. *AIS Transactions on Human-Computer Interaction, 12*(4), 253-275.

Chang, M.-W., Ratinov, L., Roth, D., & Srikumar, V. (2008). Importance of semantic representation: Dataless classification. In *Proceedings of the 23rd National Conference on Artificial Intelligence.*

Chen, C., Li, O., Tao, C., Barnett, A. J., Su, J., & Rudin, C. (2019). This looks like that: Deep learning for interpretable image recognition. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems.*

Keras Team. (2015). *Keras.* Retrieved from https://github.com/fchollet/keras

Colby, K. M., Watt, J. B., & Gilbert, J. P. (1966). A computer method of psychotherapy: preliminary communication. *The Journal of Nervous and Mental Disease, 142*(2), 148-152.

Cross, T. (2020). An understanding of AI's limitations is starting to sink in. *The Economist*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*

Dezfouli, A., Nock, R., & Dayan, P. (2020). Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences, 117*(46), 29221-29228.

Dong, J., & Rudin, C. (2020). Exploring the cloud of variable importance for the set of all good models. *Nature Machine Intelligence, 2*(12), 810-824.

Dong, Y., Yang, Y., Tang, J., Yang, Y., & Chawla, N. V. (2014). Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining*.

Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access, 6*, 28573-28593.

Fasching, W. (2009). The mineness of experience. *Continental Philosophy Review, 42*(2), 131-148.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research, 20,* 1-88.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681-694.

Gagné, M., Parker, S. K., Griffin, M. A., Dunlop, P. D., Knight, C., Klonek, F. E., & Parent-Rocheleau, X. (2022). Understanding and shaping the future of work with self-determination theory. *Nature Reviews Psychology, 1*(7), 378-392.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns: Elements of reusable object-oriented software*. Addison-Wesley.

Gennatas, E. D., Friedman, J. H., Ungar, L. H., Pirracchio, R., Eaton, E., Reichmann, L. G., Interian, Y., Luna, J. M., Simone, C. B., II., Auerbach, A., Delgado, E., van der Laan, M. J., Solberg, T. D., & Valdes, G. (2020). Expert-augmented machine learning. *Proceedings of the National Academy of Sciences, 117*(9), 4571-4577.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior, 115*.

Gombolay, M., Yang, X. J., Hayes, B., Seo, N., Liu, Z., Wadhwania, S., Yu, T., Shah, N., Golen, T., & Shah, J. (2018). Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research, 37*(10), 1300-1316.

Gonçalves, J., Weber, I., Masullo, G. M., Torres da Silva, M., & Hofhuis, J. (2021). Common sense or censorship: How algorithmic moderators and message type influence perceptions of online content deletion. *New Media & Society*.

GovTrack, S. 4400: 116th Congress: National Biometric Information Privacy Act of 2020. (2020). Retrieved from www.govtrack.us/congress/bills/116/s4400

Gratch, J., & Fast, N. J. (2022). The power to harm: AI assistants pave the way to unethical behavior. *Current Opinion in Psychology, 47*.

Hermann, E. (2022). Anthropomorphized artificial intelligence, attachment, and consumer behavior. *Marketing Letters, 33*(1), 157-162.

Holm, J. R., & Lorenz, E. (2022). The impact of artificial intelligence on skills at work in Denmark. *New Technology, Work and Employment, 37*(1), 79-101.

Jain, H., Padmanabhan, B., Pavlou, P. A., & Raghu, T. S. (2021). Editorial for the special section on humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Information Systems Research, 32*(3), 675-687.

Johnson, A. W. (2010). *An integrated traverse planner and analysis tool for future lunar surface exploration* (thesis). Massachusetts Institute of Technology, Massachusetts. Retrieved from https://dspace.mit.edu/handle/1721.1/59560

Klein, S. B., & Nichols, S. (2012). Memory and the Sense of Personal Identity. *Mind, 121*(483), 677-702.

Kontogiorgos, D., van Waveren, S., Wallberg, O., Pereira, A., Leite, I., & Gustafson, J. (2020). *Embodiment effects in interactions with failing robots.* In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802-5805.

Kracht, T. M., Mueller, M. J., Sotto, L. J., & Sterns, D. (2018). Biometric information protection: The stage is set for expansion of claims. *The Lexis Practic Advisor Journal.* Retrieved from https://www.goodwinlaw.com/-/media/files/publications/the-lexis-practice-advisor-journal-top-10-practice.pdf?la=en

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science, 9*(1), 104-117.

Larochelle, H., Erhan, D., & Bengio, Y. (2008). Zero-data learning of new tasks. In *Proceedings of the 23rd National Conference on Artificial Intelligence*.

Lean, Y., Shouyang, W., & Lai, K. K. (2006). An integrated data preparation scheme for neural network data analysis. *IEEE Transactions on Knowledge and Data Engineering, 18*(2), 217-230.

Lee, R., Orchilles, J., Hoelzer, D., & Skoudis, E. (2022). What you need to know about OpenAI's new ChatGPT bot—and how it affects your security. *SANS*. Retrieved from https://www.sans.org/webcasts/what-you-need-to-know-about-openai-new-chatgpt-bot-and-how-it-affects-your-security-lightning-talks-panel-sessions/

Li, F.-F., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(4), 594-611.

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*(1), 4-11.

Lin, J., Zhong, C., Hu, D., Rudin, C., & Seltzer, M. (2020). *Generalized and scalable optimal sparse decision trees.* In *Proceedings of the 37th International Conference on Machine Learning.*

Meyer, D., Tao, J., & Kris, A. (2022). Tune down the misinformation, please: Generating corrective messages for COVID-19 misinformation. In *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Mollick, E. (2022). ChatGPT is a tipping point for AI. *Harvard Business Review.* Retrieved from https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

Mullainathan, S. (2019). Biased algorithms are easier to fix than biased people. *The New York Times*. Retrieved from https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management, 51*(6), 801-809.

Najee-Ullah, A., Landeros, L., Balytskyi, Y., & Chang, S.-Y. (2022). *Towards detection of AI-generated texts and misinformation.* In *Proceedings of 11th International Workshop on Socio-Technical Aspects in Security.*

Natarajan, M., & Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction.*

New Hampshire House Bill 523, NH HB523. (2018). Retrieeved from https://legiscan.com/NH/bill/HB523/2018

Noorden, R. V. (2020). The ethical questions that haunt facial-recognition research. *Nature*. Retrieved from https://www.nature.com/articles/d41586-020-03187-3

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems—an introductory survey. *WIREs Data Mining and Knowledge Discovery, 10*(3).

Paleja, R. R., Silva, A., Chen, L., & Gombolay, M. C. (2020). Interpretable and personalized apprenticeship scheduling: Learning interpretable scheduling policies from heterogeneous user demonstrations. In *Proceedings of the 34th Conference on Neural Information Processing Systems.*

Panetta, K. (2019). The Gartner 2020 CIO agenda: Resilience during disruption. *Gartner*. Retrieved from https://www.gartner.com/smarterwithgartner/the-gartner-2020-cio-agenda-winning-in-the-turns

Paul, S., Yuan, L., Jain, H. K., Robert, L. P., Spohrer, J., & Lifshitz-Assaf, H. (2022). Intelligence augmentation: Human factors in AI and future of work. *AIS Transactions on Human-Computer Interaction, 14*(3), 426-445.

Petropoulos, G. (2022). The dark side of artificial intelligence: Manipulation of human behaviour. *Bruegel*. Retrieved from https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour

Qin, J., Ma, Q., Shi, Y., & Wang, L. (2017). Recent advances in consensus of multi-agent systems: A brief survey. *IEEE Transactions on Industrial Electronics, 64*(6), 4972-4983.

Rabb, N., Law, T., Chita-Tegmark, M., & Scheutz, M. (2022). An attachment framework for human-robot interaction. *International Journal of Social Robotics, 14*(2), 539-559.

Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM, 41*(2), 79-82.

Reynolds, S. (2020). *Factors influencing surgeon adoption of technology in the medical device industry* (doctoral dissertation). Georgia State University. Retrieved from https://scholarworks.gsu.edu/bus_admin_diss/133/

Robert, L. P., Bansal, G., & Lütge, C. (2020). ICIS 2019 SIGHCI workshop panel report: Human-computer interaction challenges and opportunities for fair, trustworthy and ethical artificial intelligence. *AIS Transactions on Human-Computer Interaction, 12*(2), 96-108.

Riggins, F. J., & Dewan, S. (2005). The digital divide: Current and future research directions. *Journal of the Association for Information Systems, 6*(12), 298-337.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). *Overtrust of robots in emergency evacuation scenarios.* In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction.*

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Ross, P., & Spates, K. (2020). Considering the safety and quality of artificial intelligence in health care. *The Joint Commission Journal on Quality and Patient Safety, 46*(10), 596-599.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys, 16*, 1-85.

Sanders, H. (2017). Garbage in, garbage out: How purportedly great Ml models can be screwed up by bad data. In *Proceedings of Blackhat.*

Sanyal, S. (2021). How much does artificial intelligence cost in 2021? Retrieved from https://www.analyticsinsight.net/how-much-does-artificial-intelligence-cost-in-2021/

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., & Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Sci Robot, 3*(21).

Scheutz, M. (2012). The affect dilemma for artificial agents: Should we develop affective artificial agents? *IEEE Transactions on Affective Computing, 3*(4), 424-433.

Schiff, D., Borenstein, J., Biddle, J., & Laas, K. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society, 2*(1), 31-42.

Schrum, M. L., Neville, G., Johnson, M., Moorman, N., Paleja, R., Feigh, K. M., & Gombolay, M. C. (2021). Effects of social factors and team dynamics on adoption of collaborative robot autonomy. In *Proceedings of theACM/IEEE International Conference on Human-Robot Interaction.*

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Retrieved from https://apps.dtic.mil/sti/citations/ADA057655

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

Silva, A., Gombolay, M., Killian, T., Jimenez, I., & Son, S.-H. (2020). Optimization methods for interpretable differentiable decision trees applied to reinforcement learning. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics.*

Silva, A., Moorman, N., Silva, W., Zaidi, Z., Gopalan, N., & Gombolay, M. (2022). LanCon-learn: Learning with language to enable generalization in multi-task manipulation. *IEEE Robotics and Automation Letters, 7*(2), 1635-1642.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—a study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149*.

Spohrer, J., Maglio, P. P., Vargo, S. L., & Warg, M. (2022). *Service in the AI era: Science, logic, and architecture perspectives*: Business Expert Press.

Stumpo, V., Staartjes, V. E., Klukowska, A. M., Golahmadi, A. K., Gadjradj, P. S., Schröder, M. L., Veeravagu, A., Stienen, M. N., Serra, C., & Regli, L. (2021). Global adoption of robotic technology into neurosurgical practice and research. *Neurosurgical Review, 44*(5), 2675-2687.

Sun, X., Ren, X., Ma, S., Wei, B., Li, W., Xu, J., Wang, H., & Zhang, Y. (2020). Training simplification and model simplification for deep learning : A minimal effort back propagation method. *IEEE Transactions on Knowledge and Data Engineering, 32*(2), 374-387.

Turney, P. D. (2002). Types of cost in inductive concept learning. In *Proceedings of the ICML Workshop on Cost-sensitive Learning.*

Ustun, B., & Rudin, C. (2019). Learning optimized risk scores. *Journal of Machine Learning Research, 20*(150), 1-75.

Vimalkumar, M., Gupta, A., Sharma, D., & Dwivedi, Y. (2021). Understanding the effect that task complexity has on automation potential and opacity: Implications for algorithmic fairness. *AIS Transactions on Human-Computer Interaction, 13*(1), 104-129.

Volkova, S., & Bachrach, Y. (2016). Inferring Perceived demographics from user emotional tone and user-environment emotional contrast. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

Volkova, S., Bachrach, Y., Armstrong, M., & Sharma, V. (2015). Inferring latent user properties from texts published in social media. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, Austin.

Walsh, S. M., Strano, M. S., & Stanton, S. C. (2019). Approaching robotics and autonomous systems as an integrated materials, energy, and control problem. In S. M. Walsh & M. S. Strano (Eds.), *Robotic systems and autonomous platforms* (pp. xix-xlvi). Woodhead.

Wang, P., Guo, J., Lan, Y., Xu, J., & Cheng, X. (2016). Your cart tells you: Inferring demographic attributes from purchase data. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Wang, T., Morucci, M., Awan, M. U., Liu, Y., Roy, S., Rudin, C., & Volfovsky, A. (2021). FLAME: A fast large-scale almost matching exactly approach to causal inference. *Journal of Machine Learning Research, 22*(31), 1-41.

Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. (2022a). Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *Journal of Machine Learning Research, 22*(1), 1-73.

Wang, Z. J., Zhong, C., Xin, R., Takagi, T., Chen, Z., Chau, D. H., Rudin, C. & Seltzer, M. (2022b). TimberTrek: Exploring and curating sparse decision trees with interactive visualization. In *Proceedings of the IEEE Visualization and Visual Analytics.*

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433-441.

Wei, K.-K., Teo, H.-H., Chan, H. C., & Tan, B. C. Y. (2010). Conceptualizing and testing a social cognitive model of the digital divide. *Information Systems Research, 22*(1), 170-187.

Whang, S. E., Roh, Y., Song, H., & Lee, J.-G. (2023). Data collection and quality challenges in deep learning: A data-centric AI perspective. *The VLDB Journal*.

Wiggins, W. F., & Tejani, A. S. (2022). On the opportunities and risks of foundation models for natural language processing in radiology. *Radiolology: Artificial Intelligence, 4*(4).

Wood-Doughty, Z., Andrews, N., Marvin, R., & Dredze, M. (2018). *Predicting Twitter user demographics from names alone*. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media.*

Wu, B., Jia, J., Yang, Y., Zhao, P., Tang, J., & Tian, Q. (2017). Inferring emotional tags from social images with user demographics. *IEEE Transactions on Multimedia, 19*(7), 1670-1684.

Xian, Y., Lampert, C. H., Schiele, B., & Akata, Z. (2019). Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(09), 2251-2265.

Xin, R., Zhong, C., Chen, Z., Takagi, T., Seltzer, M. I., & Rudin, C. (2022). *Exploring the whole Rashomon set of sparse decision trees*. In *Proceedings of the 36th Conference on Neural Information Processing Systems.*

Zheng, N., Liu, Z.-Y., Ren, P., Ma, Y.-Q., Chen, S.-T., Yu, S.-Y., Xue, J.-R., Chen, B.-D., & Wang, F.-Y. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering, 18*, 153-179.

Zhong, E., Tan, B., Mo, K., & Yang, Q. (2013). User demographics prediction based on mobile data. *Pervasive and Mobile Computing, 9*(6), 823-837.

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing, 237*, 350-361.

Zhou, L., Paul, S., Demirkan, H., Yuan, L., Spohrer, J., Zhou, M., & Basu, J. (2021). Intelligence augmentation: Towards building human-machine symbiotic relationship. *AIS Transactions on Human Computer Interaction, 13*(2), 243-264.

About the Authors

Lina Zhou is a Professor of Management Information Systems at the University of North Carolina at Charlotte. Her research focuses on improving human decision-making and knowledge management through both the design and development of intelligent systems and the understanding of human behavior. She has published in journals such as *MIS Quarterly*, *Journal of Management Information Systems*, ACM and IEEE Transactions, *Information & Management*, *Decision Support Systems*, and *AIS Transactions on Human-Computer Interaction*.

Cynthia Rudin is Earl D. McLean, Jr. Professor of Computer Science, Electrical and Computer Engineering, Statistical Science, Mathematics, Biostatistics & Bioinformatics at Duke University. She directs the Interpretable Machine Learning Lab, whose goal is to design predictive models that people can understand. Her lab applies machine learning in many areas, such as healthcare, criminal justice, and energy reliability.

Matthew Gombolay is an Assistant Professor of Interactive Computing at the Georgia Institute of Technology. He was named the Anne and Alan Taetle Early-career Assistant Professor in 2018. He received a B.S. in Mechanical Engineering from Johns Hopkins University in 2011, an S.M. in Aeronautics and Astronautics from MIT in 2013, and a PhD in Autonomous Systems from MIT in 2017. From 2017 to 2018. Dr. Gombolay served as technical staff at MIT Lincoln Laboratory, transitioning his research to the U.S. Navy and earning an R&D 100 Award. His publication record includes multiple best paper awards and nominations. He was selected as a DARPA Riser in 2018, received the Early Career Award from the National Fire Control Symposium, and was awarded a NASA Early Career Fellowship. He is an associate editor for Autonomous Robots and the ACM Transactions on Robotics.

Jim Spohrer is a retired industry executive, UIDP Senior Fellow, and on the Board of Directors of ISSIP and ServCollab non-profits. At IBM, he was director of Venture Capital Relations, Almaden Service Research, Global University Programs, and Open-Source AI. At Apple, he was a Distinguished Engineer Scientist and Technologist for next-generation learning platforms. After his MIT BS in Physics, he developed speech recognition systems at Verbex (Exxon) before receiving his Yale PhD in Computer Science/AI. With over ninety publications and nine patents, his awards for advancing service science include Christopher Lovelock Career Contributions to Service Discipline, Gummesson Service Research, Vargo and Lusch Service-Dominant Logic, Daniel Berg Service Systems, and PICMET Fellow.

Michelle Zhou is a co-founder and CEO of Juji, an artificial intelligence (AI) company located in Silicon Valley that specializes in building cognitive conversational AI technologies and solutions that enable the creation and adoption of empathic and empathetic AI agents. She is also an Association for Computing Machinery (ACM) Council member and has spoken at conferences including FORTUNE Brainstorm Tech. Her thought leadership has been featured in outlets such as *The New York Times*, *InfoWorld*, *Axios*, *VentureBeat*, and more. Prior to starting Juji, she led the User Systems and Experience Research (USER) group at IBM Research—Almaden and then the IBM Watson Group. Michelle's expertise is in the interdisciplinary area of intelligent user interaction (IUI), including conversational AI systems and personality analytics. She is an inventor of the IBM Watson Personality Insights and has led the research and development of at least a dozen products in her areas of expertise. She has also published 100+ peer-reviewed, refereed scientific articles and 45+ patent applications. She is currently the Editor-in-Chief of *ACM Transactions on Interactive Intelligent Systems* (TiiS), a premier scientific journal on human-centered AI, and an Associate Editor of *ACM Transactions on Intelligent Systems and Technology* (TIST). She received a PhD in Computer Science from Columbia University and is an ACM Distinguished Scientist.

Souren Paul is a Professor of Information Systems at the School of Computing and Analytics of Northern Kentucky University. His research interests are in the areas of virtual teams, collaboration systems, behavioral information security, and augmented intelligence. He has published research articles in *Journal of Management Information Systems*, *Decision Support Systems*, *Information and Management*, and *AIS Transactions on Human-Computer Interaction*. He has served as Conference Co-Chair for the 2015 Americas Conference on Information Systems (AMCIS) and the 2020 International Conference on Information Systems (ICIS).