

6-30-2020

ICIS 2019 SIGHCI Workshop Panel Report: Human– Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence

Lionel P. Robert Jr.

lprobert@umich.edu, lprobert@umich.edu

Gaurav Bansal

University of Wisconsin - Green Bay, bansalg@uwgb.edu

Christoph Lütge

Technical University of Munich, luetge@tum.de

Follow this and additional works at: <https://aisel.aisnet.org/thci>

Recommended Citation

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. <https://doi.org/10.17705/1thci.00130>
DOI: 10.17705/1thci.00130

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in AIS Transactions on Human-Computer Interaction by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



6-2020

ICIS 2019 SIGHCI Workshop Panel Report: Human-Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence

Lionel P. Robert Jr.

School of Information, Robotic Institute, University of Michigan, lprobert@umich.edu

Gaurav Bansal

Austin E. Cofrin School of Business, University of Wisconsin – Green Bay, bansalg@uwgb.edu

Christoph Lütge

TUM Institute for Ethics in Artificial Intelligence, Technical University of Munich, luetge@tum.de

Follow this and additional works at: <http://aisel.aisnet.org/thci/>

Recommended Citation

Robert, L. P., Gaurav, B., & 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), pp. 96-108.

DOI: 10.17705/1thci.00130

Available at <http://aisel.aisnet.org/thci/vol12/iss2/3>



ICIS 2019 SIGHCI Workshop Panel Report: Human-Computer Interaction Challenges and Opportunities for Fair, Trustworthy and Ethical Artificial Intelligence

Lionel P. Robert Jr.

University of Michigan
lprobert@umich.edu

Gaurav Bansal

University of Wisconsin – Green Bay

Christoph Lütge

Technical University of Munich

Abstract:

Artificial Intelligence (AI) is rapidly changing every aspect of our society—including amplifying our biases. Fairness, trust and ethics are at the core of many of the issues underlying the implications of AI. Despite this, research on AI with relation to fairness, trust and ethics in the information systems (IS) field is still scarce. This panel brought together academia, business and government perspectives to discuss the challenges and identify potential solutions to address such challenges. This panel report presents eight themes based around the discussion of two questions: (1) What are the biggest challenges to designing, implementing and deploying fair, ethical and trustworthy AI?; and (2) What are the biggest challenges to policy and governance for fair, ethical and trustworthy AI? The eight themes are: (1) identifying AI biases; (2) drawing attention to AI biases; (3) addressing AI biases; (4) designing transparent and explainable AI; (5) AI fairness, trust, ethics: old wine in a new bottle?; (6) AI accountability; (7) AI laws, policies, regulations and standards; and (8) frameworks for fair, ethical and trustworthy AI. Based on the results of the panel discussion, we present research questions for each theme to guide future research in the area of human–computer interaction.

Keywords: Artificial Intelligence, AI Bias, AI Fairness, AI Trust, AI Ethics, Algorithmic Fairness, Algorithmic Bias.

Fiona Nah was the accepting senior editor for this paper.

1 Introduction

This panel discussion brought together academia, business and government perspectives on the crucial issues of artificial intelligence (AI) fairness, trust and ethics. Despite the potential transformative effects, research on AI with relation to fairness, trust and ethics in the information systems (IS) field is still scarce, and as a result, our knowledge on the impacts of AI is still far from conclusive. Yet, it is very important from both the business and technical perspectives that we research and examine issues of fairness, trust and ethics in AI. Academia, business and government leaders must seek to develop safeguards that periodically ensure that these so-called black-box algorithms adhere to core values: fairness, trust and ethics. This panel discussion helped to highlight challenges and identify potential solutions to address such challenges.

AI is rapidly changing every aspect of our society and has amplified our productivity as well as biases. John Giannandrea, who leads AI at Google, recently lamented in the MIT Technology Review that the dangers posed by the ability of AI systems to learn human prejudices are far greater than those posed by killer-robots (Knight, 2017). Fairness, trust and ethics are at the core of many of the issues underlying the implications of AI. Fairness is undermined when managers rely blindly on “objective” AI outputs to “augment” or replace their decision-making. Managers often ignore the limitations of their assumptions and the relevance of the data that were used to train and test AI models, resulting in biased decisions that are hard to detect or appeal (Robert, Pierce, Marquis, Kim, & Alahmad, 2020). Challenges are presented when decisions and actions by AI lead to further inequalities in our society. Examples include displaced workers and affordable-housing shortages caused by rental apartments and housing units being diverted to higher-paying Airbnb short-term vacationers.

1.1 Panel Overview

This panel discussion sought to unpack potential challenges and impacts of AI related to issues of fairness, trust and ethics. Overall this panel sought to improve our understanding about the impacts of AI in organizations and our broader society. More specifically, the panel focused on four major issues:

1. Designing, implementing and deploying fair, ethical and trustworthy AI.
2. Policy and governance for fair, ethical and trustworthy AI.
3. Appropriate and inappropriate applications of AI.
4. Implications of unfair, unethical and untrustworthy AI.

1.2 Panel Structure

Lionel Robert acted as the moderator of the panel. The first half of the panel session was structured. Panelists provided a brief self-introduction. Panelists then were presented with the following questions:

1. What are the biggest challenges to designing, implementing and deploying fair, ethical and trustworthy AI?
2. What are the biggest challenges to policy and governance for fair, ethical and trustworthy AI?

The second half of the panel was unstructured. Panelists took questions from the audience.

1.3 Panelists

Lionel P. Robert Jr. is an associate professor at the University of Michigan School of Information and a core faculty member at the Michigan Robotics Institute. He is the director of the Michigan Autonomous Vehicle Research Intergroup Collaboration (MAVRIC) and an affiliate of the Michigan Interactive and Social Computing Research Group, University of Michigan Robotics Institute, and the National Center for Institutional Diversity, all at the University of Michigan, and the Center for Computer-Mediated Communication at Indiana University. He is currently serving on the editorial boards of DATA BASE for Advances in Information Systems, Journal of the Association for Information Systems, AIS Transactions on Human-Computer Interaction, ACM Transactions on Social Computing, Information and Management, Journal of Computer Information Systems, and Management Information Systems Quarterly. Dr. Robert has published in journals such as Information Systems Research, Journal of the Association of Information Systems and the Journal of the Association of Information Science and Technology as well as top HCI

conferences such as CHI, CSCW, Group, HRI, WSDM and ICWSM. His research has been sponsored by the U.S. Army, Toyota Research Institute, MCity, Lieberthal-Rogel Center for Chinese Studies and the National Science Foundation. He has appeared in print, radio and television for such news outlets as ABC, CNN, CNBC, Michigan Radio Inc., the Washington Examiner, the Detroit News and the Associated Press.

Gaurav Bansal is Frederick E. Baer Professor of Business at the Austin E. Cofrin School of Business at the University of Wisconsin-Green Bay (UWGB), and a distinguished member of the Association for Information Systems (AIS). He is currently serving as incoming editor-in-chief for the Journal of IT Case and Application Research and is on the editorial boards of AIS Transactions on Human-Computer Interaction, Journal of Computer Information Systems, Journal of Global Information Technology Management, and Journal of Midwest AIS. He is also serving on the executive board of the Midwest Association for Information Systems (MWAIS). He is the founding chair and academic director of the Master of Science in Data Science program at UWGB. Dr. Bansal has published in premier MIS journals such as the Journal of Management Information Systems, European Journal of Information Systems, Decision Support Systems, Information & Management, Information Technology & People, Journal of Organizational Computing and Electronic Commerce, and Journal of Computer Information Systems, among others. Before starting his academic career, he worked as a quality assurance engineer for General Motors India and Daewoo Motors India.

Christoph Lütge holds the Chair of Business Ethics at Technical University of Munich (TUM). He has a background in business informatics and philosophy and has held visiting professorships in Taipei, Kyoto and Venice. He was awarded a Heisenberg Fellowship. In 2019, Lütge was appointed director of the new TUM Institute for Ethics in Artificial Intelligence. He has been a member of the Ethics Commission on Automated and Connected Driving of the German Federal Ministry of Transport and Digital Infrastructure, as well as the European AI Ethics initiative AI4People. He has also done consulting work for the Singapore Economic Development Board and the Canadian Transport Commission.

1.4 Panelists' Statements

Lionel P. Robert Jr.

A survey of 1,770 managers from 14 countries found that 86% of managers planned to use AI systems for managing their workers and that 78% of them trust decisions made by AI systems (Kolbjørnsrud, Amico & Thomas, 2016). However, we have learned that AI has been found to produce unfair, biased and unethical decisions (Robert, Pierce, et al., 2020). For example, recent cases involving AI systems screening applicants have been shown to be biased against hiring women and other minorities. One specific case is Amazon's AI-powered recruitment engine that exhibited biases against female applicants (Gonzalez, 2018). For organizations as well as our society, such issues bring AI accountability, transparency and explainability to the forefront of a new set of challenges. Despite this, we know very little about how to design systems to address these issues, which implies that the development and deployment of AI are far outpacing our understanding of the implications associated with AI's use. Going forward, IS scholars need to think big and ahead to begin to grapple with such issues. We hope this panel is just the start.

Gaurav Bansal

There is a wide-scale adoption of AI, and every industry and company is trying to adopt AI to gain efficiencies, lower costs, and increase revenues. The industry-wide adoption of AI could be understood with the help of three mechanisms—coercive pressures, normative pressures and mimetic pressures (DiMaggio & Powell, 1983). Coercive mechanisms are enforced through rules and government regulations. Normative mechanisms are enforced through professional values and technical norms. Mimetic mechanisms are enforced by the objectification of values, norms and beliefs created by the leading organizations in the sector. These forces have been argued to pressure organizations into isomorphism, making them increasingly alike in how they use data and AI. However, in the case of AI, in particular, isomorphism has been perpetuated very rapidly. AI can learn and predict failures much faster, more accurately and at a lower cost than most of the domain-specific experts (Anderson & Rainie, 2018). Moreover, the skills required to develop and implement AI-based solutions do not require much domain-level knowledge, either, implying that the same (AI) skill set can be carried over from one industry to another.

Christoph Lütge

AI ethics is one of the key topics in digital technologies in our time. It is becoming more and more clear to companies, as well as other organizations and to the NGO sector, that AI systems and solutions will only

gain the necessary acceptance if they are equipped with ethical guidelines, standards and rules. A number of group efforts around the globe have recently developed some guidelines, among them the EU High Level Expert Group, the AI4People and other more specific groups such as the German Ethics Commission for Automated and Connected Driving. However, the task for the future will be to get beyond these highly abstract efforts and down to more specific rules for concrete AI systems. The Institute for Ethics in AI at the Technical University of Munich is one of the institutions where this task is currently being undertaken.

In the following section, we elaborate on the research themes that emerged from the panel discussions. For each research theme, we first articulate related issues and ways to address them. Then, we present relevant research questions.

2 Question 1: What are the biggest challenges to designing, implementing and deploying fair, ethical and trustworthy AI?

Five research themes related to identifying and addressing AI biases will be covered. Research questions associated with each of them will also be highlighted.

2.1 Theme 1: Identifying AI Biases

First, scholars discussed the importance of defining AI biases and the difficulty of defining it in a given context. How do you define biases? This was one of the first questions asked by the panelists. The discussion led to some common agreement that a bias should be viewed as an unfair assessment in favor of or against some person or group even though a more formal definition would be helpful.

Panelists and audience members agreed that assessments of biases would depend on societal norms and they might not be transferable across societies. Questions emerged around fears that U.S. and Western society views would dominate what is and is not considered fair or biased by AI. Yet others were concerned that countries like the U.S. might be too quick to adapt non-U.S. and Western society views of fairness to accommodate societies that were viewed as not respecting human rights.

Panelists and audience members discussed how a given context might change what is or is not considered a bias. Everyone agreed that context is important but struggled to identify specific situational dimensions that could be used to better understand why or when a given context would matter. For example, when would the same assessment be viewed as more or less biased based on a given context?

Given this discussion, we uncovered the following research questions related to this theme for researchers to explore:

- RQ1.1:** How should AI biases be defined?
- RQ1.2:** What criteria should be used to determine whether the decisions by an AI system are fair or unfair?
- RQ1.3:** Should fairness be based on goals of equity or equality?
- RQ1.4:** How do biases change relative to a given context?
- RQ1.5:** What are the specific situational dimensions that determine when AI biases change?
- RQ1.6:** How do societal norms impact definitions of AI biases?

2.2 Theme 2: Drawing Attention to AI Biases

Second, scholars highlighted the need to develop systems to draw attention to AI biases. This discussion centered on how systems could be designed to identify and draw attention to biases. Interfaces could be designed to highlight potential biases by providing an alternative design to the user. The interface might point out potential problems inherent in its decisions. For example, a hiring AI system might alert the human resources (HR) users that its selection criteria produced an interview list with no women. Panelists and audience members wondered whether the organization should set the parameters based on its definition of fairness or allow the user to set the parameters. Other questions emerged around how aggressive the AI system should be in this process. For example, more passive AI interfaces could be designed to only identify and alert users of potential biases. However, more aggressive AI interfaces could be designed to default to

non-action until the user has acknowledged the biases or offered a solution to the problem. Even more aggressive AI interfaces could be designed to intervene or nudge fair actions.

Based on this discussion, we uncovered the following research questions related to this theme for researchers to explore:

- RQ2.1:** Should AI allow users to determine the parameters of fairness?
- RQ2.2:** How can we design AI interfaces that would alert users that decisions given could be more or less biased?
- RQ2.3:** Should AI default to non-action until the user has acknowledged them and offered a solution to addressing them?
- RQ2.4:** Should AI allow users to determine the parameters of fairness?
- RQ2.5:** Should AI be designed to intervene or nudge fair actions?

2.3 Theme 3: Addressing AI Biases

Next, the conversation shifted to various approaches to addressing AI biases. Panelists and audience members openly wondered whether AI systems should be allowed to address the issues of biases. Most audience members felt that only humans should be allowed to address such issues. The discussion then moved to imagining how an AI would address the issues of biases. One suggestion was to provide financial compensation to individuals for wrongdoing associated with AI biases. It was not clear during the discussion how financial compensation would be determined.

As such, the following research questions for this theme could be explored:

- RQ3.1:** How should AI biases be addressed?
- RQ3.2:** Should AI be allowed to address the issues of biases? If so, when and under what contexts?
- RQ3.3:** How should AI determine how to address the issues of biases?

2.4 Theme 4: Designing Transparent and Explainable AI

The fourth theme discussed by the panel highlighted the need to design transparent and explainable AI. AI systems can collect, store and process large amounts of data in real time. However, the algorithms used to reach decisions are often treated as a black box and lack transparency. Specifically, it is not always clear what datasets nor what criteria the AI system is using to make decisions. It is, in part, because these algorithms are often dynamic, designed to learn and can be highly autonomous. As a result, it is not always clear to a user when or why decision criteria change over time.

Both transparency and explainability have been offered as potential solutions to black-box AI. Transparent AI is defined as the degree to which a user can clearly see how the AI operates. Explainable AI refers to the degree to which the AI can translate and delineate how it operates in human terms to the user. As such, it seems possible that explainable AI can lead to more transparent AI.

Based on this background, we propose the following set of research questions:

- RQ4.1:** What are the most effective ways for the AI to translate its actions into terms the user understands?
- RQ4.2:** What are the most effective ways for the AI to promote transparency?
- RQ4.3:** How can we measure the effectiveness of AI transparency and explainability?
- RQ4.4:** What contextual factors influence AI transparency and explainability?

2.5 Theme 5: AI Fairness, Trust, Ethics: Old Wine in a New Bottle?

The fifth theme focused on whether the issues surrounding AI fairness, trust and ethics are inherently new issues or just old issues in the context of AI. The panelists and the audience discussed whether there is a difference between AI ethics and IS or IT ethics. Questions emerged around whether issues related to AI are any different from the same set of issues related to other information technologies. If so, in what ways are they different and how would we know whether they are meaningfully different? Audience members

wondered whether we should invest in new AI-specific theories or instead focus on contextualizing current theories related to fairness, trust and ethics in information technology.

As such, we propose the following research questions:

- RQ5.1:** Does AI require original theories related to fairness, trust and ethics?
- RQ5.2:** What existing fairness, trust and ethics theories should we leverage?
- RQ5.3:** What characteristics might make AI distinct enough to require original theories related to fairness, trust and ethics?

3 Question 2: What are the biggest challenges to policy and governance for fair, ethical and trustworthy AI?

Three research themes related to fair, ethical and trustworthy AI will be presented along with their associated research questions.

3.1 Theme 6: AI Accountability

The sixth theme focused on issues surrounding AI accountability. The panelists and the audience discussed who should be held accountable for the actions of an AI. Examples of third-party hiring firms using biased AI systems were brought up. Audience members questioned whether the third-party hiring firm or the organization that outsourced to the firm should be held accountable for any hiring biases. Panel members also pointed out that issues of legal and financial accountability vary greatly among and within nations. Current approaches and efforts to solve such problems were acknowledged (World Economic Forum, 2020). In the end, everyone acknowledged that AI accountability is an important but emerging area of study.

Based on this discussion, the following set of research questions are proposed:

- RQ6.1:** How should AI accountability be determined?
- RQ6.2:** Should AI accountability be driven by legal requirements of social norms?
- RQ6.3:** Who should be accountable to whom with respect to global corporations?

3.2 Theme 7: AI Laws, Policies, Regulations and Standards

The seventh theme focused on related issues to the sixth by examining issues surrounding the role of international, federal, state and local governments in setting policy, regulations and standards. Generally, AI policy was defined as a formal system of principles related to the design and use of AI to ensure a predetermined outcome. AI regulations refer to a set of restrictions pertaining to the design and use of AI systems to prevent specific uses of AI. AI standards were deemed to be the minimum requirements related to the design and use of AI. The discussion seemed to be centered on a view that AI policy is at the highest level, then AI regulations, followed by standards at the lowest level. Current approaches and efforts to solve the problem were acknowledged (Google, 2020).

Panelists and audience members discussed the need for governments at all levels to consider aligning and coordinating their efforts. However, everyone acknowledged that doing so at all levels of governments would be impossible because not only do different national governments have different laws, policies, regulations and standards (LAPRS), but they might also differ within the same nation. Second, the discussion shifted to the potential unintended consequences of LAPRS. Many people were afraid that governments would overreact and pass legislation that could cause more problems than resolve problems. However, others were afraid that governments would do nothing, allowing the markets and private industries to dictate the pace of legislation.

Given this discussion, we outline the following research questions related to this theme for researchers to explore:

- RQ7.1:** Should AI laws, policies, regulations and standards (AI-LAPRS) be driven by a top-down or bottom-up approach?
- RQ7.2:** How can AI laws, policies, regulations and standards (AI-LAPRS) be harmonized to avoid conflict within and across borders?

- RQ7.3:** How can users be made aware of the potential problems associated with AI use to fully understand their implications?
- RQ7.4:** Can awareness among users create enough regulatory and normative pressure on governments and corporations, respectively, to enact ethical AI guidelines?

3.3 Theme 8: Frameworks for Fair, Ethical and Trustworthy AI

Another theme that emerged was the need to develop frameworks for fair, ethical and trustworthy AI for governments and civil society. Panelists and audience members discussed the need to develop a framework to not only help guide policies and practices associated with AI but also help with assessing and evaluating them.

For a framework to evolve, there is a need to consider multiple stakeholders. A multiplicity of stakeholders would favor the adoption of an ethical framework. However, a multiplicity of stakeholders could create multiple conflicting constituent expectations. This phenomenon was argued by Oliver (1991) to be strongly associated with a manipulative response to explore alternatives to the prescribed institutional pressures, i.e. blindly adopting the AI in our case. The term, institutions, refer to regulatory structures, governmental agencies, laws, courts and professions. Ethical frameworks emphasize overall well-being that is in opposition to the model that emphasizes efficiency and profit maximization above all. Hence, in the case of multiple stakeholders, organizations would resist blind adoption of AI and support the adoption of an ethical framework.

Any ethical framework would need to be culturally sensitive to the particular context it is socially embedded within. This would require not only an understanding of the cultural context but also the limitations associated with a particular AI. On one hand, the best case might be to develop AI that recognizes a particular cultural context and adapts accordingly. On the other hand, it simply may not be possible for reasons that range from limitations posed by technology to non-technology (i.e., social, political, etc.) related issues. Therefore, an understanding of the possible trade-offs between being culturally sensitive and being consistent across cultures is warranted.

The following research questions are related to this theme:

- RQ8.1:** What are the key components of an ethical framework?
- RQ8.2:** Who are the key stakeholders of an ethical framework?
- RQ8.3:** What forces would enable development and adoption of an ethical framework?

4 Discussion

4.1 Outcomes

Over and above the research questions, the panel discussion produced four main outcomes. In the following section, we elaborate on the outcomes that emerged from the panel discussions.

4.1.1 Outcome 1: Vital and Understudied Area

Everyone in attendance agreed that AI fairness, trust and ethics are significant areas in need of much work. Going forward, these areas are expected to only increase in importance as AI bleeds over to every aspect of life, including both work and home.

4.1.2 Outcome 2: Need for Global Coordination

There is a need for a global coordinated effort to address issues of AI fairness, trust and ethics. Because AI crosses boundaries, it is difficult to understand how local solutions can be effectively developed and deployed.

4.1.3 Outcome 3: Context Matters

Although scholars all agree that context matters, there is no consensus on what in a particular context matters. Much more thought is needed to fully comprehend the role of context in the development and deployment of AI.

4.1.4 Outcome 4: Need for awareness

Awareness among users is needed to help create a multiplicity of stakeholders and to help governments and corporations realize the importance of ethical AI. This should, in turn, create regulatory pressure as well as normative pressure to act on developing and implementing ethical AI systems.

4.2 The Way Forward

In this panel report, we identified eight themes based around the discussion of two questions: (1) What are the biggest challenges to designing, implementing and deploying fair, ethical and trustworthy AI?; and (2) What are the biggest challenges to policy and governance for fair, ethical and trustworthy AI? The eight themes are: (1) identifying AI biases; (2) drawing attention to AI biases; (3) addressing AI biases; (4) designing transparent and explainable AI; (5) AI fairness, trust, ethics: old wine in a new bottle?; (6) AI accountability; (7) AI laws, policies, regulations and standards; and (8) frameworks for fair, ethical and trustworthy AI. Based on the results of the panel discussion, we developed research questions for each theme that can guide future research in the area of human-computer interaction. The themes and research questions are listed in Table 1.

Table 1. Themes and Research Questions

| Topic Questions | Themes | Research Questions |
|---|---|---|
| 1: What are the biggest challenges to designing, implementing and deploying fair, ethical and trustworthy AI? | 1: Identifying AI Biases | RQ 1.1: How should AI biases be defined? RQ 1.2: What criteria should be used to determine whether the decisions by an AI system are fair or unfair? RQ 1.3: Should fairness be based on goals of equity or equality? RQ 1.4: How do biases change relative to a given context? RQ 1.5: What specific situational dimensions determine when AI bias changes? RQ 1.6: How do societal norms impact definitions of AI biases? |
| | 2: Drawing Attention to AI Biases | RQ 2.1: Should these systems allow users to determine the parameters of fairness? RQ 2.2: How can we design interfaces that would alert users that decisions given could be more or less biased? RQ 2.3: Should these systems default to non-action until the user has acknowledged them and offered a solution to addressing them? RQ 2.4: Should these systems allow users to determine the parameters of fairness? RQ 2.5: Should such systems be designed to intervene or nudge fair actions? |
| | 3: Addressing AI Biases | RQ 3.1: How should AI biases be addressed? RQ 3.2: Should AI be allowed to address the issues of biases? If so, when and under what contexts? RQ 3.3: How should AI determine how to address the issues of biases? |
| | 4: Designing Transparent and Explainable AI | RQ 4.1: What are the most effective ways for the AI to translate its actions into terms the user understands? RQ 4.2: What are the most effective ways for the AI to promote transparency? RQ 4.3: How can we measure the effectiveness of AI transparency and explainability? RQ 4.4: What contextual factors influence AI transparency and explainability? |
| | 5: AI Fairness, Trust, Ethics: Old Wine in a New Bottle? | RQ 5.1: Does AI require original theories related to fairness, trust and ethics? RQ 5.2: What existing fairness, trust and ethics theories should we leverage? RQ 5.3: What characteristics make AI distinct enough to require original theories related to fairness, trust and ethics? |
| 2: What are the biggest challenges to policy and governance for fair, ethical and trustworthy AI? | Theme 6: AI Accountability | RQ 6.1: How should AI accountability be determined? RQ 6.2: Should AI accountability be driven by legal requirements of social norms? RQ 6.3: Who should be accountable to whom with respect to global corporations? |
| | 7: AI Laws, Policies, Regulations and Standards | RQ 7.1: Should AI laws, policies, regulations and standards (AI-LAPRS) be driven by a top-down or bottom-up approach? RQ 7.2: How can AI laws, policies, regulations and standards (AI-LAPRS) be harmonized to avoid conflict within and across borders? RQ 7.3: How can users be made aware of the potential problems associated with AI use to fully understand their implications? RQ 7.4: Can awareness among users create enough regulatory and normative pressure on governments and corporations, respectively, to enact ethical AI guidelines? |
| | 8: Developing Frameworks for Fair, Ethical and Trustworthy AI | RQ 8.1: What are the key components of an ethical framework? RQ 8.2: Who are the key stakeholders of an ethical framework? RQ 8.3: What forces would enable development and adoption of an ethical framework? |

5 Conclusion

The panel discussion provided insight into the current and future state of AI fairness, trustworthiness and ethics. The panelists and audience members agreed that the topic of AI is here to stay. The panelists and audience members also indicated that IS as a discipline must employ thoughtful research about how to develop and whether to deploy AI systems. Everyone acknowledged the current efforts being made to address some aspects of the problem (You & Robert, 2018; Robert, Alahmad, Zhang, Kim, Esterwood, & You, 2020). To better address such issues, IS scholars must take care not to engage in isolationism. Therefore, the panel and audience members also encouraged IS scholars entering the area to seek collaboration across disciplines with other scholars. In closing, everyone in attendance agreed that IS scholars have much to contribute to the study of AI fairness, trustworthiness and ethics.

Acknowledgments

We would like to acknowledge the audience members as well as the organizations of the 2019 Pre-ICIS workshop sponsored by AIS-SIGHCI.

References

- Anderson, J., & Rainie, L. (2018). Artificial intelligence and the future of humans. PewResearch.org. Retrieved from <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147-160.
- Gonzalez, G. (2018). How Amazon accidentally invented a sexist hiring algorithm. Inc. Retrieved from <https://www.inc.com/guadalupe-gonzalez/amazon-artificial-intelligence-ai-hiring-tool-hr.html>
- Google (2020). Perspectives on issues in AI governance. Google. Retrieved from <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>
- Knight, W. (2017). Forget killer robots—bias is the real AI danger. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/>
- Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review*, 2, 1-6.
- Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, 16(1), 1145-179.
- You, S., & Robert, L. P. (2018). Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots. *Journal of the Association for Information Systems*, 19(5), 377-407.
- Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction*, <https://doi.org/10.1080/07370024.2020.1735391>
- Robert, L. P., Alahmad, R., Zhang, Q., Kim, S., Esterwood, C., & You, S. (2020). A review of personality in human robot interactions. *Foundations & Trends in Information Systems*, 4(2), 107–212.
- World Economic Forum (2020). Model artificial intelligence governance framework and assessment guide. *World Economic Forum*. Retrieved from <https://www.weforum.org/projects/model-ai-governance-framework>

About the Authors

Lionel P. Robert Jr. is an Associate Professor in the School of Information at the University of Michigan and an Association for Information Systems Distinguished Member. He completed his Ph.D. in Information Systems from Indiana University where he was a BAT Fellow and KPMG Scholar. He is the director of the Michigan Autonomous Vehicle Research Intergroup Collaboration (MAVRIC) and also a core faculty member of the University of Michigan Robotics Institute, Center for Hybrid Intelligence Systems and the National Center for Institutional Diversity all at the University of Michigan and the Center for Computer-Mediated Communication at Indiana University. He is currently serving on the editorial boards of *DATA BASE for Advances in Information Systems*, *Journal of the Association for Information Systems*, *AIS Transactions on Human-Computer Interaction*, *ACM Transactions on Social Computing, Information and Management*, *Journal of Computer Information Systems*, and *Management Information Systems Quarterly*. He has appeared in print, radio and/or television for ABC, CNN, CNBC, Michigan Radio, Inc. New York Times and the Associated Press.

Gaurav Bansal is Frederick E. Baer Professor of Business at the Austin E. Cofrin School of Business at the University of Wisconsin-Green Bay (UWGB), and a distinguished member of the Association for Information Systems (AIS). He is currently serving as incoming editor-in-chief for the *Journal of IT Case and Application Research* and is on the editorial boards of *AIS Transactions on Human-Computer Interaction*, *Journal of Computer Information Systems*, *Journal of Global Information Technology Management*, and *Journal of Midwest AIS*. He is also serving on the executive board of the Midwest Association for Information Systems (MWAIS). He is the founding chair and academic director of the Master of Science in Data Science program at UWGB. Dr. Bansal has published in premier MIS journals such as the *Journal of Management Information Systems*, *European Journal of Information Systems*, *Decision Support Systems*, *Information & Management*, *Information Technology & People*, *Journal of Organizational Computing and Electronic Commerce*, and *Journal of Computer Information Systems*, among others. Before starting his academic career, he worked as a quality assurance engineer for General Motors India and Daewoo Motors India.

Christoph Lütge holds the Chair of Business Ethics at Technical University of Munich (TUM). He has a background in business informatics and philosophy and has held visiting professorships in Taipei, Kyoto and Venice. He was awarded a Heisenberg Fellowship. In 2019, Lütge was appointed director of the new TUM Institute for Ethics in Artificial Intelligence. He has been a member of the Ethics Commission on Automated and Connected Driving of the German Federal Ministry of Transport and Digital Infrastructure, as well as the European AI Ethics initiative AI4People. He has also done consulting work for the Singapore Economic Development Board and the Canadian Transport Commission.

Copyright © 2019 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.



Editor-in-Chief

<https://aisel.aisnet.org/thci/>

Fiona Nah, Missouri University of Science and Technology, USA

Advisory Board

| | | |
|---|---|--|
| Izak Benbasat University of British Columbia, Canada | John M. Carroll Penn State University, USA | Phillip Ein-Dor Tel-Aviv University, Israel |
| Dennis F. Galletta University of Pittsburgh, USA | Shirley Gregor National Australian University, Australia | Paul Benjamin Lowry Virginia Tech, USA |
| Jenny Preece University of Maryland, USA | Gavriel Salvendy, Purdue U., USA, & Tsinghua U., China | Joe Valacich University of Arizona, USA |
| Jane Webster Queen's University, Canada | K.K. Wei National University of Singapore, Singapore | Ping Zhang Syracuse University, USA |

Senior Editor Board

| | | | |
|---|--|---|---|
| Torkil Clemmensen Copenhagen Business School, Denmark | Fred Davis Texas Tech University, USA | Gert-Jan de Vreede University of South Florida | Soussan Djamasbi Worcester Polytechnic Inst., USA |
| Traci Hess U. of Massachusetts Amherst, USA | Shuk Ying (Susanna) Ho Australian National U., Australia | Matthew Jensen University of Oklahoma, USA | Jinwoo Kim Yonsei University, Korea |
| Eleanor Loiacono Worcester Polytechnic Inst., USA | Anne Massey U. of Massachusetts Amherst, USA | Gregory D. Moody U. of Nevada Las Vegas, USA | Lorne Olfman Claremont Graduate U., USA |
| Kar Yan Tam Hong Kong U. of Science & Technology, China | Dov Te'eni Tel-Aviv University, Israel | Jason Thatcher University of Alabama, USA | Noam Tractinsky Ben-Gurion U. of the Negev, Israel |
| Viswanath Venkatesh University of Arkansas, USA | Mun Yi Korea Advanced Institute of Science & Technology, Korea | Dongsong Zhang U. of North Carolina Charlotte, USA | |

Editorial Board

| | | | |
|---|---|---|---|
| Miguel Aguirre-Urreta Florida International U., USA | Michel Avital Copenhagen Business School, Denmark | Gaurav Bansal U. of Wisconsin-Green Bay, USA | Hock Chuan Chan National University of Singapore, Singapore |
| Langtao Chen Missouri University of Science and Technology, USA | Christy M.K. Cheung Hong Kong Baptist U., China | Cecil Chua Missouri University of Science and Technology, USA | Michael Davern University of Melbourne, Australia |
| Carina de Villiers University of Pretoria, South Africa | Alexandra Durcikova University of Oklahoma, USA | Brenda Eschenbrenner U. of Nebraska at Kearney, USA | Xiaowen Fang DePaul University, USA |
| James Gaskin Brigham Young University, USA | Matt Germonprez U. of Nebraska at Omaha, USA | Jennifer Gerow Virginia Military Institute, USA | Suparna Goswami Technische U.München, Germany |
| Camille Grange HEC Montreal, Canada | Juho Harami, Tampere University, Finland | Khaled Hassanein McMaster University, Canada | Milena Head McMaster University, Canada |
| Netta Iivari Oulu University, Finland | Zhenhui Jack Jiang University of Hong Kong, China | Richard Johnson SUNY at Albany, USA | Weiling Ke Clarkson University, USA |
| Sherrie Komiak Memorial U. of Newfoundland, Canada | Na Li Baker College, USA | Yuan Li University of Tennessee, USA | Ji-Ye Mao Renmin University, China |
| Scott McCoy College of William and Mary, USA | Robert F. Otondo Mississippi State University, USA | Lingyun Qiu Peking University, China | Sheizaf Rafaeli University of Haifa, Israel |
| Rene Riedl Johannes Kepler U. Linz, Austria | Lionel Robert University of Michigan, USA | Khawaja Saeed Wichita State University, USA | Shu Schiller Wright State University, USA |
| Christoph Schneider IESE Business School, Spain | Theresa Shaft University of Oklahoma, USA | Stefan Smolnik University of Hagen, Germany | Jeff Stanton Syracuse University, USA |
| Heshan Sun University of Oklahoma, USA | Chee-Wee Tan Copenhagen Business School, | Horst Treiblmaier Modul University Vienna, Austria | Ozgur Turetken Ryerson University, Canada |
| Dezhi Wu University of South Carolina, USA | Fahri Yetim FOM U. of Appl. Sci., Germany | Cheng Zhang Fudan University, China | Meiyun Zuo Renmin University, China |

Managing Editor

Gregory D. Moody, University of Nevada Las Vegas, USA

