Guest Editorial: Special Issue on Responsible AI in Social Computing

ARTIFICIAL intelligence (AI) continues demonstrating its positive impact on society and successful adoptions in data-rich domains including social computing systems. There are serious ethical and legal concerns about AI's ability to make decisions in a responsible way. Many principles and guidelines for responsible AI (RAI) have been issued by governments, research organizations, and enterprises. For instance, the Institute for Ethical Machine Learning provides various RAI resources [1], including higher level guidelines and frameworks, tools, standards, regulations, course, and so on. However, high-level principles are far from ensuring the trustworthiness of AI systems.

Social computing investigates the intricate interrelationships among individuals, organizations, and social systems [2], [3]. This concept encompasses computing theories and techniques along with a wide array of social behaviors, organizational dynamics, and media interactions. In recent years, a substantial amount of projects have been conducted to develop novel social applications and provide services that facilitate data collection and generation based on this theory.

However, it is observed that there is a significant gap between high-level principles and low-level actionable practice for developers. Therefore, this special issue is dedicated to provide cutting-edge technologies and novel studies and promising developments, which can improve trustworthiness of AI in social computing systems. Important research directions include the following:

1) fairness in AI for social computing;
2) privacy and security protection for AI-driven social computing;
3) reliability and safety in AI for social computing;
4) transparency and explainability in AI for social computing;
5) contestability in AI for social computing;
6) accountability in AI for social computing;
7) methods, models, and tools for RAI in social computing.

What Do We Cover in This Special Issue?

In this special issue, ten articles have been accepted, which represent the most recent research of RAI in social computing.

In [A1], Lu et al. develop an incentive mechanism to address the challenges of information self-disclosure and social utility maximization in vertical federated learning. Specifically, the proposed mechanism consists of payment and sample size selection rules, while the authors also analyze the relationship between individual rationality and budget balance when self-disclosure and efficiency are satisfied. The proposed solution is evaluated via comparative analysis regarding social utility, area under curve, and mean-square error using both synthetic and real-world datasets.

In [A2], Zhang et al. present a data-driven responsible framework to formalize the high-level RAI principles and guidelines into operational rules in cyber–physical–social systems. In particular, the framework perceives users' daily activities and then analyzes and predicts users' future intentions to responsibly enhance the systems. The authors design three tensor-based Baum–Welch algorithms to improve the performance of activity analysis, and evaluate the proposed solution in terms of precision, recall, and $F1$ measure.

In [A3], Zeng et al. propose a feature-contrastive graph solution for federated learning, to improve the robustness and alleviate the weight divergence. Specifically, they design a feature-contrastive graph federated learning system architecture, which consists of three main components: communicator, data loader, and core. In addition, the authors propose a weight similarity constraint approach, which can combine the losses of model representation, weight similarity, and supervised learning for model training. The performance (i.e., accuracy and time consumption) of this novel approach is compared with two baseline methods on multiple datasets.

Djenouri et al. address the issue of insight retrieval and time consumption for analyzing fake news, in [A4], through deriving patterns from fake news and integrating decomposition and MapReduce mining techniques, respectively. The authors first categorize collected data into fake and real news datasets via a multiobjective k-means method. Afterward, the decomposition and MapReduce mining strategy conduct pattern mining, while the patterns of fake news are matched with real news data to identify actionable insights. Experiments are carried out to examine the quality of decomposition and time consumption of the proposed solution on four databases.

In [A5], Wang et al. first prove the discrete tendency of predicted item ratings and then present a framework to minimize accuracy loss when realizing popularity fairness in recommendation systems. The proposed framework consists of three main modules: 1) community exploration, which utilizes the random walk strategy to organize multiple overlapping

Digital Object Identifier 10.1109/TCSS.2023.3331781

communities; 2) local recommendation, which generates local recommendation lists for each constructed community; and 3) fair decision fusion, which combines the local recommendation lists for users by calculating discrete ratings and generating edge reranking. The authors validate their design on three public datasets and compare the results with eight related studies.

In [A6], Khan et al. conduct a survey to understand how engineers and lawyers in the AI industry comprehend related AI ethics principles. The authors collect, synthesize, and analyze survey data from 99 participants, and the results indicate that the following hold: 1) AI practitioners and lawmakers regard transparency, accountability, and privacy as the most three significant AI ethics principles, while lack of ethical knowledge, no legal frameworks, and monitoring bodies are three common challenges; 2) the participants consider that conflicts between different principles are the most serious factor for realizing AI ethics principles; and 3) domain-specific principles and guidelines should be constructed for different stakeholders.

In [A7], Wang et al. provide a comprehensive overview of state-of-the-art research of data hiding techniques based on deep learning. In particular, the authors exclusively explore deep learning-based data hiding architectures, watermarking, and steganography approaches. Furthermore, the authors discuss a series of open questions and research agenda, including data hiding removal attack mitigation, watermarking for machine learning model protection, and steganography-based malware.

Hao et al. mathematically formulate the problem of absolute fair cliques enumeration in attributed social networks, in [A8], and prove it to be NP-hard. To address this problem, the authors represent the topological structure and attributes information of a cohesive subgraph via a new concept of attributed equiconcept, based on which the detection methodologies and algorithms for identifying absolute fair maximal cliques are developed for multiple types of attributed social networks. The proposed solution is evaluated via comparative analysis with four related work.

In [A9], Cai et al. investigate data privacy protection during the machine learning process via designing a hardware-assisted framework. In the proposed framework, practitioners can first develop machine learning models in a conventional scheme. Second, the framework splits the whole given dataset into multiple enclaves. Third, the framework can ensure a secure execution environment, since Gramine is integrated as a Software Guard Extension (SGX) Library Operating System (LibOS), and a global model is trained across different enclaves. The authors perform experiments to evaluate the model accuracy, loss, and training time of the proposed solution.

In [A10], Cai et al. explore the indirect social relationships for modeling entities' interest preferences, while laying emphasis on the explainability, accuracy, novelty, and content quality of social recommendation results. Specifically, a model is proposed to analyze user similarity and explainability of social recommendation results through entity and embedding vectors. Furthermore, the authors employ the partition deletion strategy to design a many-objective recommendation algorithm to ensure model efficiency. Experiments are carried out with four evaluation metrics (i.e., accuracy, explainability, novelty, and quality of the content).

ACKNOWLEDGMENT

The guest editors would like to thank all the authors who submitted their work to this special issue. In addition, they want to thank all the reviewers for their great efforts. Finally, they sincerely appreciate the support from the Editor-in-Chief, Prof. Bin Hu for his support and guidance in administrative issues. The guest editors wish all readers enjoy the studies in this special issue.

QINGHUA LU, *Guest Editor*
Data61, CSIRO
Eveleigh, NSW 2015, Australia
e-mail: qinghua.lu@data61.csiro.au

WEISHAN ZHANG, *Guest Editor*
Department of Intelligence Science
China University of Petroleum
Qingdao 266580, China
e-mail: zhangws@upc.edu.cn

ZHEN WANG, *Guest Editor*
School of Cyberspace and ZhuoYue Honors College
Hangzhou Dianzi University
Hangzhou 31000, China
e-mail: wangzhen@hdu.edu.cn

QUN JIN, *Guest Editor*
Faculty of Human Sciences
Waseda University
Tokyo 169-8050, Japan
e-mail: jin@waseda.jp

VINCENZO PIURI, *Guest Editor*
Depatimento di Information
University of Milan
20133 Milan, Italy
e-mail: vincenzo.piuri@unimi.it

APPENDIX: RELATED ARTICLES

[A1] J. Lu, B. Pan, A. M. Seid, B. Li, G. Hu, and S. Wan, "Truthful incentive mechanism design via internalizing externalities and LP relaxation for vertical federated learning," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2909–2923, Dec. 2023.

[A2] S. Zhang, L. T. Yang, Y. Zhang, Z. Lu, J. Yu, and Z. Cui, "Tensor-based Baum–Welch algorithms in coupled hidden Markov model for responsible activity prediction," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2924–2937, Dec. 2023.

[A3] X. Zeng et al., "Feature-contrastive graph federated learning: Responsible AI in graph information analysis," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2938–2948, Dec. 2023.

[A4] Y. Djenouri, A. Belhadi, G. Srivastava, and J. C.-W. Lin, "Advanced pattern-mining system for fake news analysis," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2949–2958, Dec. 2023.

[A5] B. Wang, S. Song, S. Liu, and X. Deng, "MultiFDF: Multi-community clustering for fairness-aware recommendation," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2959–2970, Dec. 2023.

[A6] A. A. Khan et al., "AI ethics: An empirical study on the views of practitioners and lawmakers," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2971–2984, Dec. 2023.

[A7] Z. Wang et al., "Data hiding with deep learning: A survey unifying digital watermarking and steganography," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 2985–2999, Dec. 2023.

[A8] F. Hao, Y. Yang, J. Shang, and D.-S. Park, "AFCMiner: Finding absolute fair cliques from attributed social networks for responsible computational social systems," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 3000–3011, Dec. 2023.

[A9] Z. Cai, B. Ren, R. Ma, H. Guan, M. Tian, and Y. Wang, "Guardian: A hardware-assisted distributed framework to enhance deep learning security," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 3012–3020, Dec. 2023.

[A10] X. Cai, W. Guo, M. Zhao, Z. Cui, and J. Chen, "A knowledge graph-based many-objective model for explainable social recommendation," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 6, pp. 3021–3030, Dec. 2023.

REFERENCES

[1] The Institute for Ethical Machine Learning. *Awesome AI Guidelines*. Accessed: Nov. 19, 2023. [Online]. Available: https://github.com/EthicalML/awesome-artificial-intelligence-guidelines

[2] D. C. Dryer, C. Eisbach, and W. S. Ark, "At what cost pervasive? A social computing view of mobile computing systems," *IBM Syst. J.*, vol. 38, no. 4, pp. 652–676, 1999.

[3] J. Riedl, "The promise and peril of social computing," *Computer*, vol. 44, no. 1, pp. 93–95, Jan. 2011.

Qinghua Lu (Senior Member, IEEE) received the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2013.

She is currently a Principal Research Scientist and leads the Responsible Artificial Intelligence (AI) Science Team, Data61, CSIRO, Eveleigh, NSW 2015, Australia, where she is also a part of the OECD.AI's Trustworthy AI Metrics Project Team. She has published more than 150 papers in premier international journals and conferences. Her current research interests include responsible AI, software engineering for AI/generative artificial intelligence (GAI), and software architecture.

Dr. Lu serves as a member of the Australian's National AI Centre Responsible AI Think Tank. She is the winner of the 2023 Asia–Pacific (APAC) Women in AI Trailblazer Award.

Weishan Zhang received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2001.

He was a Post-Doctoral Researcher with the National University of Singapore, Singapore, from 2001 to 2003; a Visiting Scholar with Carleton University, Ottawa, ON, Canada, from 2006 to 2007; and an Associate Professor/a Senior Researcher with Aarhus University, Aarhus, Denmark. He is currently with the Department of Intelligence Science, China University of Petroleum, Qingdao, China, where he is also the Director of the Big Data Intelligent Processing Team. He has authored more than 100 papers in journals and at international conferences. His current research interests include big data processing, artificial intelligence, and middleware for the Internet of Things.

Zhen Wang received the B.Sc., M.Eng., and Ph.D. degrees from Dalian University of Technology, Dalian, China, in 2007, 2009, and 2016, respectively, all in software engineering.

From 2014 to 2016, he was a Research Fellow at Nanyang Technological University, Singapore. He is currently a Professor with the School of Cyberspace and the Vice Dean of the ZhuoYue Honors College, Hangzhou Dianzi University, Hangzhou, China. His current research interests include network security, artificial intelligence security, complex networks, reinforcement learning, and algorithmic game theory.

Qun Jin (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from Nihon University, Tokyo, Japan, in 1992.

He is currently a tenured Full Professor and the Chair of the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Tokyo. He has been engaged extensively in research works in the fields of computer science, information systems, and social and human informatics. He seeks to exploit the rich interdependence between theory and practice in his work with interdisciplinary and integrated approaches. His recent research interests cover human-centric ubiquitous computing, human–computer interaction, behavior and cognitive informatics, big data, personal analytics and individual modeling, massive online open courses (MOOCs) and learning analytics, and computing for well-being.

Prof. Jin is a member of IEEE Computer Society (CS); Association for Computing Machinery (ACM) (USA); and Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ), and Japanese Society for Artificial Intelligence (JSAI) (Japan).

Vincenzo Piuri (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from the Politecnico di Milano, Milan, Italy, in 1984 and 1988, respectively.

From 2007 to 2012, he was the Department Chair with the University of Milan, Milan, where he has been a Full Professor since 2000. He was an Associate Professor with the Politecnico di Milano from 1992 to 2000; a Visiting Professor with The University of Texas at Austin, Austin, TX, USA, from 1996 to 1999; and a Visiting Researcher with George Mason University, Fairfax, VA, USA, from 2012 to 2016. He founded a startup company, Sensure srl, Bergamo, Italy, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010) and was active in industrial research projects with several companies. His main research and industrial application interests are intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, the Internet-of-Things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures.

Dr. Piuri is an Association for Computing Machinery (ACM) Fellow.