Effects of Uncertainty and Knowledge Graph on Perception of Fairness

JIANLONG ZHOU, BOYUAN ZHENG, AND FANG CHEN, Data Science Institute
University of Technology Sydney, Australia

AI-informed decision making consequential to individuals raises critical concerns on fairness. Fairness introduced by AI models and/or data is perceived by humans (also refers to perception of fairness) implicitly or explicitly. It is a central component of maintaining satisfactory relationships with humans in AI-informed decision making. Furthermore, model uncertainty and knowledge on training data play a crucial role in AI-informed decision making. This paper introduces model uncertainty and knowledge on training data represented by knowledge graphs into AI-informed decision making. We aim to investigate what uncertainty level and knowledge graph, and how they affect user perception of fairness in AI-informed decision making. A user study on judging the recidivism rate of prisoners found that uncertainty of model prediction of recidivism rate can benefit user perception of fairness, but only under low and medium uncertainty conditions. However, we did not find significant effects of knowledge represented by knowledge graph on user perception of fairness. These findings have wide implications in the user interface design of AI-informed decision making applications.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Jianlong Zhou, Boyuan Zheng, and Fang Chen. 2024. Effects of Uncertainty and Knowledge Graph on Perception of Fairness. 1, 1 (April 2024), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Artificial Intelligence (AI) has been increasingly used to make important decisions in different areas ranging from daily lives such as recommending movies to making more critical decisions such as allocation of social benefits, hiring, and criminal justice [5, 7, 9]. Due to the black-box nature of AI models that make it hard for users to understand why a decision is made or how the data is processed for the decision making [22, 24], trust in AI and its decisions have been extensively investigated from the research community to various application domains for safety and legal reasons [8, 20, 21]. Furthermore, since AI is often used by humans and/or for human-related decision making [18], fairness introduced by AI models and/or data is perceived by humans (also refers to *perception of fairness* in this paper) implicitly or explicitly. Human's perceived fairness plays an important role in AI-informed decision making. The perception of fairness is a central component of maintaining satisfactory relationships with humans in decision making [1, 19, 26].

On the other hand, it was found that uncertainty has an significant impact on user trust in AI-informed decision making [19]. However, it is not clear how uncertainty affects human's perception of fairness in AI-informed decision making. Furthermore, Knowledge Graph (KG) is the representation of information as a semantic graph which is built by describing real world entities and their interrelations, organised according to a schema and covering a variety of topical domains [15, 27]. While the knowledge contained in training data in machine learning has significant impacts on user responses in AI-informed decision making [11, 23, 25], it is interesting to investigate effects of the knowledge in the training data represented by the knowledge graph on user's perception of fairness in AI-informed decision making.

This study seeks to represent knowledge contained in the training data with the knowledge graph. Our research questions are: whether the knowledge on the training data represented by the knowledge graph and model uncertainty

Author's address: Jianlong Zhou, Boyuan Zheng, and Fang Chen, jianlong.zhou@uts.edu.au, Data Science Institute
University of Technology Sydney, Sydney, NSW, Australia, 2007.

affect user perception of fairness? If yes, how do they affect perception of fairness? This paper proposes approaches of knowledge representation with knowledge graphs. A user study is designed to introduce both uncertainty and knowledge graph into the pipeline.

2 RELATED WORK

The perception of fairness has been increasingly investigated from various perspectives [18, 26]. Lee et al. [12, 13] investigated people's perception of fairness based on the allocation of resources regarding equality, equity, or efficiency. People showed much variation in the three fairness metrics (equality, equity, efficiency) in the study. Zhou et al. [26] found that introduced fairness is positively related to perception of fairness in a case study of AI-based talent shortlisting.

Furthermore, uncertainty is investigated in different areas to understand relations between uncertainty and perception of fairness. For instance, Singh et al. [17] tried to model uncertainty explicitly in ranking algorithms to optimise the fairness of decisions made by AI. However, it is not clear how uncertainty affects people's perception of fairness. Moreover, knowledge graph has drawn extensive research attentions after being proposed [14].

These previous work motivates us to consider both uncertainty and KG of training data in the perception of fairness, aiming to find what combinations of these conditions and how these conditions affect user perception of fairness.

3 EXPERIMENT

3.1 Case Study

ProPublica's COMPAS data is a public data increasingly used in research studies to test various definitions of algorithmic fairness related to COMPAS software [3]. Our study uses the ProPublica's COMPAS data to build KGs of training data of models. A number of machine learning classification models are simulated to set up different uncertainty conditions in the study. The KG and uncertainty of models are then presented to users for the recidivist risk decision making.

3.2 Uncertainty

(a) Low uncertainty. (b) Medium uncertainty. (c) High uncertainty.

Fig. 1. Uncertainty conditions.

Uncertainty plays a crucial role in decision making. The uncertainty discussed here is constrained to predictive uncertainty. Similar to prior research [4, 6], we measure the uncertainty by predictive entropy [16] in this study.

We designed four kinds of conditions: without uncertainty information, low uncertainty, medium uncertainty, and high uncertainty (see Figure 1). Two ensemble models of A, B are simulated to predict the recidivism rate of prisoners who are randomly collected from the ProPublica dataset [2]. To control the variance of other elements, we make the predicted label from models A, B the same for one specific case, i.e., the overall predictive accuracy of two models is almost the same. The simulated outputs of these two ensemble models are in the form of the probability vector and represented as the bar graph. The deviation between the length of each bar could be used to estimate the predictive entropy and reflect the uncertainty level. For example, an explicit extrapolation of an outcome predicting that an offender has a 90% high risk of recidivism and a low risk of only 10% would result in a low uncertainty model. On the contrary, when the deviation of the output probability is only ten percent (like Model A in Figure 1c). It is difficult to make an accurate high and low-risk determination, and such situation is considered as highly uncertainty.

3.3 Knowledge Graph

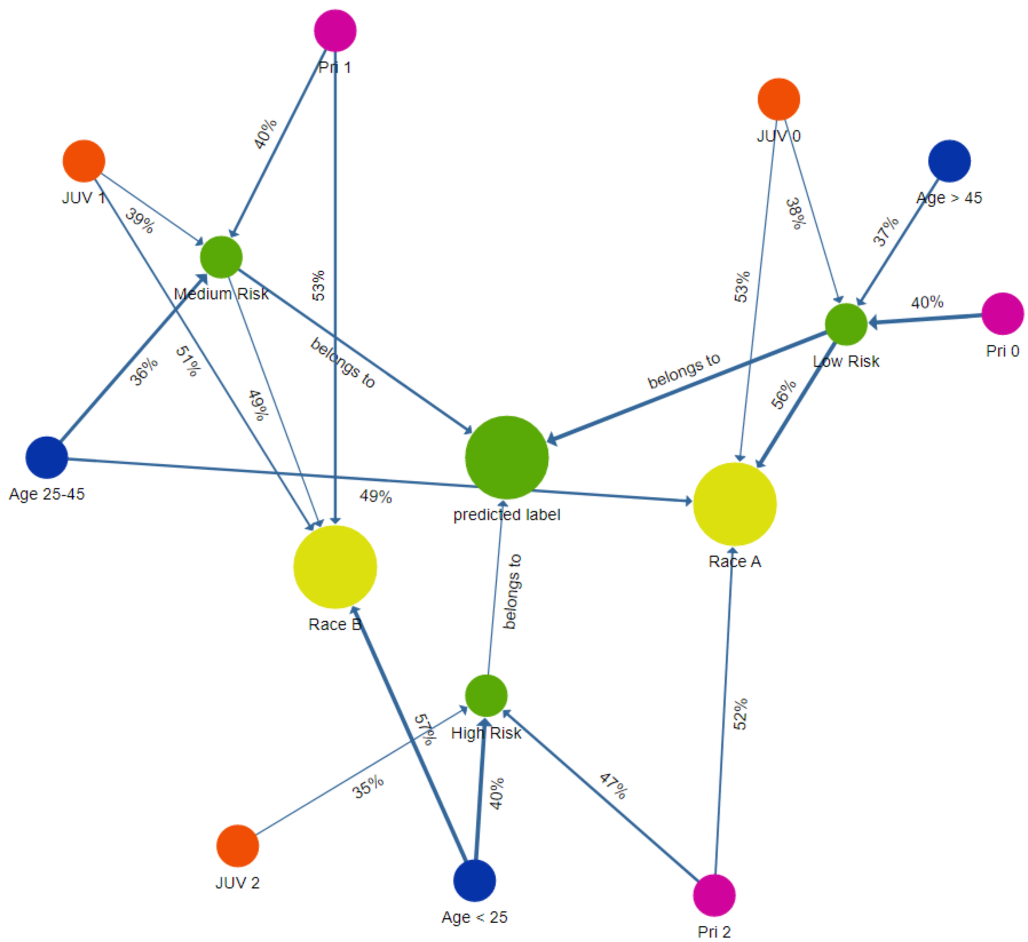

Fig. 2. An example of KG presented to participants.

Knowledge graph develops a structural network between entities together with the relationships as edges [10]. The graphical representation of knowledge facilitates the interpretability and comprehensiveness of the data. By setting the availability of KG as one of the variables, we want to investigate whether the availability of KG would influence perception of fairness combined with the uncertainty information. The KG construction process uses the association mining technique, which statistically analyzes the causal relationship between attributes. In order to make the KG more comprehensive to the interviewees, we preprocessed the dataset by rolling up the attributes and grouping the values categorically, then selected the association rules with high confidence and support to construct the knowledge network: the attributes are represented into entities; the confidences and supports are treated as edges and the thickness of these edges respectively. We strictly constrain the number of entities to prevent the graph from looking overwhelming. Figure 2 presents the example of KG provided to the participants.

3.4 Task Design

In this experiment, both uncertainty and KG information are presented to participants in each task. As described previously, four uncertainty conditions (uncertainty information presented (also called control condition), low, medium, and high) are selected. The KG includes two conditions: with KG presentation and without KG presentation (control condition). Therefore, there are 8 task condition combinations (4 uncertainty × 2 KGs) in total. This experiment investigates participants' perception of fairness regarding races in decision making tasks. Therefore, two races of Race A and Race B were manipulated for each task condition combinations, and the tasks with each race were conducted 2 rounds. We finally got 32 tasks (4 uncertainty × 2 KGs × 2 races × 2 rounds). The task orders were randomized during the experiment. Two training tasks were also conducted by each participants before formal tasks.

In each task, the information of a single criminal, including the criminal's age, encoded race (race A, race B), gender, number of juvenile crimes, number of prior crimes committed were firstly presented to participants. The uncertainty and KG conditions are then visualised to participants. Participants were asked to make decisions by selecting one of machine learning models (Model A or B) to estimate the level of recidivist risk under the given conditions.

3.5 Data Collection

This experiment was implemented as an online application. 31 participants in the age group of 20-30 years, who were university students, were recruited. After each decision making task, participants were asked to compare their perception of fairness on Model A and Model B with the question "Do you think Model A and Model B are fair in predicting recidivist risk regarding criminal's races?"

4 ANALYSIS

A two-way ANOVA analysis showed that there were no statistically significant interactions between uncertainty conditions and knowledge graph conditions on perception of fairness, $F(3, 247) = 1.55, p < .202$.

A further one-way ANOVA test analysis did not find any statistically significant differences in perception of fairness among the knowledge graph conditions (with and without knowledge graph presentations), $F(1, 240) = 1.55, p < .202$. However, a one-way ANOVA test found that there were statistically significant differences in user perception of fairness among four uncertainty conditions $F(3, 240) = 3.873, p < .0099$. Further post-hoc comparison with t-tests (with a Bonferroni correction under a significance level set at $\alpha < .013$) was conducted to find pair-wise differences in user perception of fairness between four uncertainty conditions. The adjusted significance alpha level of .013 was calculated by dividing the original alpha of .05 by 4, based on the fact that we had four uncertainty conditions.

It was found that participants had a statistically significant difference in perception of fairness under the control condition of uncertainty (no uncertainty presented) compared to the medium uncertainty condition ($t = 3.376, p < .000$). Moreover, it was found that participants had a statistically significant difference in perception of fairness under the control condition of fairness compared to the low uncertainty condition ($t = 2.574, p < .006$). In both cases, participants showed that both Model A and Model B were fair regarding races under the control condition (no uncertainty information presented). Participants also showed that model which has the larger differences between low and high risk probabilities was fairer than the other regarding races under both the low and medium uncertainty conditions. However, there were no significant differences found in perception of fairness among other uncertainty conditions.

5 DISCUSSION AND CONCLUSION

Overall, we can say that the uncertainty of model prediction of recidivism rate can benefit user perception of fairness, but only under low and medium uncertainty conditions, where users can justify the decision action with lower uncertainty (higher difference in prediction of recidivism rate) and fit their general understanding of the decision making problem. These findings have wide implications in the user interface design of AI-informed decision making applications. For example, in order to enhance user perception of fairness, uncertainty information especially lower uncertainty information could be presented to users in the user interface of AI-informed decision making applications.

While knowledge graph is a semantic representation of information describing real world entities and their interrelations, it is expected to help users to understand the knowledge contained in training data in machine learning for boosting better perception of fairness. However, this study did not find significant effects of KG on user perception of fairness. This is maybe because that a better understandable visualisation approach is expected to represent complex relations and knowledge among entities. Furthermore, it is also necessary to decide what entities and their knowledge could be presented to users instead of all entities and their knowledge are presented to users for better understanding. Therefore, our future work will focus on the investigation of effective visualisation approaches for KG in AI-informed applications, and to examine how they affect user perception of fairness. Furthermore, we will also examine how the presentation of both KG and uncertainty affect user trust in AI-informed decision.

REFERENCES

[1] Pankaj Aggarwal and Richard P. Larrick. 2012. When consumers care about being treated fairly: The interaction of relationship norms and fairness norms. *Journal of Consumer Psychology* 22, 1, SI (2012), 114–127.

[2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica (2016). *Google Scholar* (2016), 23.

[3] Matias Barenstein. 2019. ProPublica's COMPAS Data Revisited. arXiv:1906.04711 [econ.GN]

[4] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 401–413.

[5] Fang Chen and Jianlong Zhou. 2019. AI in the public interest. In *Closer to the Machine: Technical, Social, and Legal Aspects of AI*, Cliff Bertram, Asher Gibson, and Adriana Nugent (Eds.). Office of the Victorian Information Commissioner.

[6] Marc Combalia, Ferran Hueto, Susana Puig, Josep Malvehy, and Veronica Vilaplana. 2020. Uncertainty estimation in deep neural networks for dermoscopic image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 744–745.

[7] Akshay Gugnani and Hemant Misra. 2020. Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 13286–13293.

[8] K. Holzinger, Klaus Mak, Peter Kieseberg, and A. Holzinger. 2018. Can we Trust Machine Learning Results? Artificial Intelligence in Safety-Critical Decision Support. *ERCIM News* 112, 1 (2018), 42–43.

[9] Claretha Hughes, Lionel Robert, Kris Frady, and Adam Arroyos. 2019. Artificial Intelligence, Employee Engagement, Fairness, and Job Outcomes. In *Managing Technology and Middle- and Low-skilled Employees*. 61–68.

[10] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[11] Pang Wei Koh and Percy Liang. 2017. Understanding Black-box Predictions via Influence Functions. In *Proceedings of ICML 2017* (Sydney, Australia).

[12] Min Kyung Lee and Su Baykal. 2017. Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. 1035–1048.

[13] Min Kyung Lee, Anuraag Jain, Hea Jin Cha, Shashank Ojha, and Daniel Kusbit. 2019. Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proceedings of the ACM on Human-Computer Interaction* 3 (November 2019), 1–26. Issue CSCW.

[14] Ishani Mondal, Yufang Hou, and Charles Jochim. 2021. End-to-End NLP Knowledge Graph Construction. *arXiv preprint arXiv:2106.01167* (2021).

[15] Petar Ristoski and Heiko Paulheim. 2016. Semantic Web in data mining and knowledge discovery: A comprehensive survey. *Journal of Web Semantics* 36 (2016), 1–22.

[16] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

[17] Ashudeep Singh, David Kempe, and Thorsten Joachims. 2021. Fairness in ranking under uncertainty. *Advances in Neural Information Processing Systems* 34 (2021).

[18] Christopher Starke, Janine Baleis, Birte Keller, and Frank Marcinkowski. 2021. Fairness Perceptions of Algorithmic Decision-Making: A Systematic Review of the Empirical Literature. arXiv:2103.12016 [cs.HC]

[19] Jianlong Zhou, Syed Z. Arshad, Simon Luo, and Fang Chen. 2017. Effects of Uncertainty and Cognitive Load on User Trust in Predictive Decision Making. In *Human-Computer Interaction – INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer, Cham, 23–39.

[20] Jianlong Zhou and Fang Chen. 2018. 2D Transparency Space—Bring Domain Users and Machine Learning Experts Together. In *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*. Springer International Publishing, 3–19.

[21] Jianlong Zhou and Fang Chen (Eds.). 2018. *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent.* Springer International Publishing. 3–19 pages.

[22] Jianlong Zhou, Amir H. Gandomi, Fang Chen, and Andreas Holzinger. 2021. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. *Electronics* 10, 5 (2021).

[23] Jianlong Zhou, Huaiwen Hu, Zhidong Li, Kun Yu, and Fang Chen. 2019. Physiological Indicators for User Trust in Machine Learning with Influence Enhanced Fact-Checking. In *Machine Learning and Knowledge Extraction.* 94–113.

[24] Jianlong Zhou, M. Asif Khawaja, Zhidong Li, Jinjun Sun, Yang Wang, and Fang Chen. 2016. Making Machine Learning Useable by Revealing Internal States Update — A Transparent Approach. *International Journal of Computational Science and Engineering* 13, 4 (2016), 378–389.

[25] Jianlong Zhou, Zhidong Li, Huaiwen Hu, Kun Yu, Fang Chen, Zelin Li, and Yang Wang. 2019. Effects of influence on user trust in predictive decision making. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[26] Jianlong Zhou, Sunny Verma, Mudit Mittal, and Fang Chen. 2021. Understanding Relations between Perception of Fairness and Trust in Algorithmic Decision Making. In *Proceedings of the International Conference on Behavioral and Social Computing (BESC 2021)*.

[27] Xiaohan Zou. 2020. A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series* 1487, 1 (mar 2020), 012016. https://doi.org/10.1088/1742-6596/1487/1/012016