



ACM Transactions on Knowledge Discovery from Data (TKDD)

Special Issue on Trustworthy Artificial Intelligence

Guest Editors:

- **Wenqi Fan**, The Hong Kong Polytechnic University, wenqi.fan@polyu.edu.hk
- **Shu Zhao**, Anhui University, zhaoshuzs2002@hotmail.com
- **Jiliang Tang**, Michigan State University, tangjili@msu.edu

In the past few decades, Artificial Intelligence (AI) techniques have undergone rapid development and revolutionized modern society, including the economy, healthcare, education, and transportation, bringing significant convenience to our daily lives. However, many recent studies indicate that AI can pose serious threats and harm to users and society, for example, making unreliable&opaque decisions in safety-critical scenarios, amplifying bias&unfairness toward under-represented groups or individuals, and disclosing user privacy and business secrets. As a consequence, there has been growing attention on AI techniques' trustworthiness for mitigating the potential adverse effects caused by AI.

The aim of this TKDD special issue is to promote the trustworthiness of AI, so as to enhance the public's trust towards AI techniques. Meanwhile, the European Union (EU) has provided an ethics guideline for AI, stating that Trustworthy AI should be grounded in four ethical principles: respect for human autonomy, prevention of harm, fairness, and explicability. These principles guide the design and operation of AI systems from various perspectives. Technically, AI should be accurate, robust, and explainable. From a user's perspective, it should be available, usable, safe, respect privacy, and uphold user autonomy. Socially, AI should be law-abiding, ethical, fair, accountable, and environmentally friendly. Adherence to these principles ensures that AI systems are beneficial, safe, and respectful of users and society. Establishing Trustworthy AI is a crucial step towards realizing its full potential and allowing humans to fully reap its benefits and conveniences.

Topics

We welcome submissions on the following topics, including but not limited to:

- **Robustness and Safety.** AI techniques need to be robust to against adversarial attacks and ensure secure decision-making.
- **Non-discrimination and Fairness.** AI systems are expected to avoid discrimination and unfairness toward certain groups or individuals in making decision process.
- **Transparency and Explainability.** It suggests that the AI's decision mechanism system should be able to be explained and understandable to various stakeholders, and the inference process should be transparent.
- **Privacy.** Protecting users' private information carried by the data is a vital requirement for AI systems.
- **Environmental Well-being.** Due to tremendous energy consumption on the environment, sustainable and eco-friendly practices are important for AI systems to promote environmental algorithms and benefit our future generations.
- **Auditability and Accountability.** Ensuring that an AI system is assessed by a third party and, when necessary, assign responsibility for an AI failure, especially in critical applications.
- **Interactions among Different Dimensions.** An ideal trustworthy AI should simultaneously meet different dimensions by considering their accordant and conflicting interactions.
- **Trustworthy Large Language Models (LLMs).** The rapid development of LLMs in AI has been revolutionizing

many fields. Understanding the trustworthiness of LLMs has emerged as a critical concern.

- Trustworthiness of AI-powered Applications. AI techniques have been widely used in various real-world applications, such as recommender systems, natural language processing, computer vision, and so on. Achieving different applications' trustworthiness has received increasing public awareness.
- Surveys, evaluation metrics, new dimensions, or benchmarking on state-of-the-art research in achieving trustworthy AI.

Important Dates

- Submissions deadline: September 15, 2023
- First-round review decisions: January 30, 2024
- Deadline for revision submissions: March 10, 2024
- Notification of final decisions: April 30, 2024
- Tentative publication: Summer 2024

Submission Information

Authors can submit their manuscripts via Manuscript Central (<https://mc.manuscriptcentral.com/tkdd>). Submissions to this special issue will follow the regular TKDD submission guidelines (<https://dl.acm.org/journal/tkdd/author-guidelines>). Submissions must be accompanied by a cover letter containing all of the following: (1) Confirm that the paper is not currently under submission at another journal or conference. (2) Confirm that the paper is substantially different from any previously published work. (3) Confirm that none of the co-authors is a Guest Editor for this special session. (4) Disclose possible conflicts of interest with Guest Editors. The review process will be single-blind. Strict policies will be followed for plagiarism, submission confidentiality, reviewer anonymity, prior and concurrent paper submission based on the guidelines.

Papers with a "Major Revision" or "Minor Revision" decision should be resubmitted within 1 months. Revised submissions must be accompanied with a detailed response to reviewers explaining what revisions were implemented.

For questions and further information, please contact Dr. Wenqi Fan (wenqi.fan@polyu.edu.hk).