

前　　言

2020年4月,中共中央国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》,提出要加快培育数据要素市场的概念,进一步突出了数据在经济发展、科技进步和生产生活等各方面的重要性。但与此同时,人们对自身产生的数据失去掌控权,以及对大数据决策深度依赖,由此导致用户数据滥用、隐私泄露、数据垄断和决策公平等数据伦理问题层出不穷,严重影响了数据生产者和数据消费者等参与主体的权益。传统的数据治理体系与方法已经不能解决当下数据治理面临的难题,亟待新思想和新方法等突破性进展的出现。目前来看,建立数据透明化的数据治理体系是一条有效途径,即通过增加大数据价值发现过程中的透明性来保障各方参与主体的权益。

鉴于此,《计算机研究与发展》推出“数据治理与数据透明”专题,侧重数据透明、数据伦理、数据决策公平、数据决策可解释、数据垄断、数据隐私保护等方面,探讨当下数据治理的新体系与新方法,展望未来的发展趋势。本专题最终收录了4篇论文,内容涵盖数据伦理判别、数据决策公平、隐私保护新方法和数据透明4个方面,反映了国内学术领域在数据治理方面的主要工作。

在数据透明方面,建立数据透明化的数据治理体系,充分考虑数据生命周期内的多方参与主体的权益,是当前数据治理的有效途径。孟小峰等作者的“基于区块链的数据透明化:问题与挑战”一文提出了数据透明化的概念与框架,深入分析数据透明化与当前大数据生态中的隐私保护、决策可解释和数据垄断的关系。该文指出数据透明化实现需求与区块链特性天然契合,并对目前基于区块链的数据透明化研究现状进行总结分析。最后,指出了基于区块链实现数据透明化可能面临的挑战问题。

在数据伦理问题判别方面,人工智能应用的伦理风险和挑战引起了人们的普遍关注,如何从技术实现角度设计遵守人类价值观和伦理规范的AI系统是亟需解决的重要问题之一。古天龙等作者的“基于社会新闻数据集的伦理行为判别方法”一文基于机器学习对社会新闻中伦理行为的判别方法进行研究。该文基于丰富的新闻语料构建涵盖伦理道德和人类行为的社会新闻数据集,并基于该数据集提出伦理行为判别模型并验证其有效性。

在数据决策公平方面,机器学习在某些应用中可能会强化人类的偏见和歧视,导致决策过程中的不公平现象产生,从而对个人和社会产生潜在的负面影响。陈晋音等作者的“面向深度学习的公平性研究综述”一文围绕深度学习公平性,总结了深度学习应用中的偏见来源、不同类型偏见的去偏方法、评估去偏见效果的公平性评价指标和目前主流的去偏见平台。最后,探讨了现有公平性研究领域存在的开放问题以及未来的发展趋势。

在隐私保护技术方面,除了传统技术的应用发展,也需要新方法的出现来突破现有技术的壁垒。刘峰等作者的“一种基于区块链的泛用型数据隐私保护的安全多方计算协议”一文提出将区块链与安全多方计算相结合,设计泛用性安全多方计算协议。区块链利用安全多方计算提升隐私能力,以便于使用更多的应用场景。安全多方计算可借助区块链技术进行公开透明不被篡改的验证。该协议为加快分布式网络中的数据隐私保护提供了可能。

本专题征文得到研究人员的积极响应,在此感谢广大作者对本专题的鼎力支持!感谢各位审稿专家及时认真的评阅,以及详细中肯的意见与建议。特别感谢《计算机研究与发展》编委会和编辑部,感谢他们对“数据治理与数据透明”这一前沿技术专题的支持,从专题立项到最后的定稿出版,他们都付出了辛勤的汗水。最后,衷心感谢各位作者、审稿专家和编辑部的辛勤工作。

孟小峰(中国人民大学)

冯登国(中国科学院软件研究所)

2021年元月