

金融大脑 2.0: 当金融遇到可信人工智能

周俊^{1,2}, 陈超超¹, 李龙飞², 张志强², 郑小林¹

¹浙江大学计算机科学与技术学院, 中国杭州市, 310027

²蚂蚁科技集团, 中国杭州市, 310027

摘要: 人工智能通过从数据中识别隐藏模式以提高金融决策质量, 从而加速金融服务的发展。然而, 除了通常需要的属性, 如模型准确性, 金融服务还需要可信赖的人工智能, 但其属性尚未充分实现。这些可信人工智能的属性是可解释性、公平性和包容性、稳健性和安全性, 以及隐私保护。在本文中, 我们回顾人工智能应用于金融服务各领域的最新进展和局限性, 包括风险管理、欺诈检测、财富管理、个性化服务和监管技术。基于这些进展和局限性, 介绍了金融大脑2.0, 一个走向可信人工智能的研究框架。我们认为, 在金融服务中, 我们离真正可信人工智能还有很长的路要走, 并呼吁人工智能和金融业的社区一同努力。

关键词: 金融智能; 可信人工智能; 风险管理; 欺诈检测; 财富管理

<https://doi.org/10.1631/FITEE.2200039>