IEEE 标准化项目概况

作者：IEEE 自主与智能系统伦理全球倡议项目

如下是一些正在进行的 IEEE 标准化项目的简要信息，这些项目是我们正在制定中的。每个有兴趣的人都可以参加这些项目。

IEEE7000-解决系统设计中伦理问题的建模过程。这个标准给出了在一个系统或者软件程序设计之初，如何识别和分析系统或软件程序中潜在伦理问题。基于价值的系统设计方法通过解决每个开发阶段的伦理考虑，以期在利于创新的同时，帮助避免产生意想不到的负面后果。

这个项目目前正在进行标准草案的审查，预计是在 2017 年底之前收集完成成员国发来的更新草案的反馈意见。

IEEE P7001™ - 自主系统的透明性。这个研究提供了一种用于开发自主技术的标准，依据这个标准可以进行技术的自我评估，并帮助用户了解为什么技术在不同情况下做出某些决定。该项目还提供了一种方法可以使系统实现透明和可问责，以帮助指导和改进系统，例如在自动驾驶的车辆中嵌入事件数据记录器，或从设备的传感器中访问数据。

这个项目的起草工作虽然还处于初级阶段，但工作进展很顺利。目前的工作重点是为不同利益相关方群体提供 AI 技术和自主系统的可度量的透明性定义。

IEEE P7002™ - 数据隐私的处理，这个研究规定了如何管理系统或软件收集的个人隐私数据。它也是通过定义覆盖企业数据收集策略和质量保证的需求来实现。它还为涉及个人信息的应用程序开发的组织提供应用实例和数据模型。该标准提供了识别和衡量系统中隐私信息的隐私影响评估方法，在软件设计过程中为设计人员提供便利。

工作组对标准的初步工作纲要进行了审查和评论。在与现有标准相协调的同时，将工作组成员分为两个小组，分别解决复杂的主题（包括 GDPR 和超出合规性的问题）。

IEEE P7003™ - 算法偏见的处理。这个研究为算法开发人员提供了一种协议，使他们在自主或智能系统的算法开发过程中避免代码中的负面偏见。这些偏见可能包括对数据的主观或不

正确的解释，例如错误的因果关联。该项目为消除算法创建中的偏见问题提供了一些具体的步骤。

该标准还将包括选择验证数据集的基准过程与准则，建立和沟通算法设计的应用程序边界，并防止意外后果。

工作组目前正在起草大纲文件的小节，同时确定小组内的专家在这些小节中的领导作用。

IEEE P7004™ -儿童与学生数据治理标准。这个研究为涉及学生安全性数据的教育机构提供了实现透明和可问责的操作流程和认证。该标准规定了在涉及儿童和学生数据访问、存储和共享的教育机关和机构中，如何访问、收集、共享和删除儿童或学生的数据。

工作组的工作得到法律界和行业专家的支持，所有专家共同研究并提炼出了"标准"要涵盖的问题范围。这个工作组获得大学和法学院的坚定支持，目标是在未来十八个月内完成标准。

IEEE P7005™ - 雇主数据治理标准。这个研究主要是在存储、保护和使用员工数据的伦理和透明性等方面进行了指导和认定。这个标准提供了推荐的工具和服务来确保员工对于个人信息使用情况的知情权。这份标准对于员工个人信息的使用提供了清晰的建议，对于员工和雇主都有益：员工在安全可靠的环境中分享他们的信息，雇主在工作过程中获得工作所需的员工信息。

该工作组正在讨论两个关键主题：数据安全、第 29 条工作组的最佳实践。

IEEE P7006™ - 个人数据的 AI 代理标准。这个标准主要是处理在有人工输入的情形下，机器进行决策所引发的问题。该标准希望告诉政府和行业，当 AI 系统可以自行组织和分享个人信息时，为何要在设计 AI 系统过程中置入伦理考虑机制才能够减少伦理问题。AI 代理作为一种工具，允许任何个人为其数据创建个人的"条款和条件"，因此 AI 代理将为个人提供技术工具，以管理和控制其在数字和虚拟世界中的身份。

该工作组成立了一个多元化、高素质的专业人才队伍。鉴于个人 AI 代理能力的复杂性和细微差别，本研究的重点是制定一系列指导原则和伦理框架，用以指导工作组的研究。这些原则将有助于确定工作分工和全社会所有利益相关者。

IEEE P7007™ - 伦理驱动的机器人和自动化系统的本体标准。这个研究主要是建立一系列具有不同抽象层次的本体，其中包含为机器人和自动化系统设计建立伦理驱动的方法所必需的概念、定义和公理。

IEEE P7008™ -机器人、智能与自主系统中伦理驱动的助推标准。这个标准建立了典型的启示（目前正在使用或可能被创建）的描述，其中包含机器人、智能或自主系统在建立和确保伦理驱动设计方法中所需的概念、功能和优势。机器人、智能或自主系统体现的"启示"被定义为用以影响用户的行为或情绪的显明的或者隐含的建议或操作。

IEEE P7009™ - 自主和半自主系统的失效安全设计标准。这个研究为自主和半自主系统的开发、实施和使用有效的失效安全机制提供具体的方法和工具，建立实用的技术基准。

该标准包括（但不限于）：用于度量、测试和验证系统从弱到强的故障安全能力的程序，以及在性能不令人满意的的情况下可以进行的改进说明。该标准具有鲁棒、透明和逻辑清晰的特点，是开发人员、用户和监管机构建立故障安全机制的强有力基础。

IEEE P7010™ - 合乎伦理的人工智能与自主系统的福祉度量标准。这个研究将构建受智能和自主系统直接影响的人类因素相关的福祉度量，并为系统分析或包含（在编程和运行过程中）的客观和主观数据类型建立基准，从而增加人类福祉。

诚挚感谢：

IEEE 自主与智能系统伦理全球倡议项目中国委员会发起组织了翻译工作，在此特别致谢其翻译工作组的成员对此文做出的特别贡献。对本文翻译工作做出贡献的人员为：

陈鹏： 副教授，北京语言大学信息科学学院。

王亮迪： IEEE 自主与智能系统伦理全球倡议项目中国委员会主席

张文瀚： 项目经理

对翻译 若有反馈和疑问， 或者感兴趣参加中国委员会的工作，包括未来的翻译工作，请联系 IEEE 标准协会中国工作组 IEEE-SA@qq.com 。IEEE 诚挚欢迎广大专家学者阅览《人工智能的伦理准则》（第 2 版）全文。