人工智能工程伦理分析及应对策略

褚灵强 计算机学院 212050175

1. 中文摘要（字数一般不少于150字）。摘要必须反映全文中心内容，内容应包括目的、过程及方法、结论。要求论述简明、逻辑性强、尽量用短句。采用第三人称的写法，并请用过去时态叙述作者工作，用现在时态叙述作者结论。
2. 人工智能；工程伦理；伦理困境

# 引言

随着新一代人工智能算法和产品的快速开发和部署，人工智能在各个领域发挥着越来越重要的作用，并对现代社会结构产生了重大影响。神经网络模型和算法已被广泛应用于各种决策场景，如刑事审判、交通控制、金融贷款和医疗诊断等等。

腾讯研究院院长司晓介绍，全球人工智能创业企业已达2000多家，仅腾讯公司就有四大实验室正在开展人工智能相关研究。事实上，有些人工智能商业场景迟迟不能落地，不是被技术“绊”住了，而是伦理研究“拖了后腿”。司晓举例说，美国2016年的统计表明，自动驾驶的安全性已经明显高于人工驾驶，但自动驾驶中有一个著名的伦理困境，即当它无可避免要撞人时，是撞向人多的一边还是人少的一边。

同时，随着基于人工智能的自动决策系统得到不断的推广和应用，也在安全性和公平性等方面带来潜在的风险，并且已经造成了几例惨剧。比如在自动驾驶系统的安全性方面，美国亚利桑那州一名女性在过马路时被一辆Uber无人驾驶SUV撞倒并最终死亡，这是全球首起自动驾驶汽车撞死人类的事故，引发了广泛关注和担忧。此外，在智能司法场景中，司法人员根据囚犯行为特征，采用智能算法来决定是否授予其假释许可，但这种算法可能会基于种族和文化背景做出有偏见和不公平的决定。

学术界和工业界主机意识到了人工智能应用的安全性和公平性等问题，对人工智能应用的治理也成为了越来越重要的议题。

# 人工智能工程的道德困境

# 应对人工智能道德风险的基本原则

# 中国的人工智能工程伦理原则与治理技术

# 结论