

前 言

《庄子》有言：“有机械者必有机事，有机事者必有机心。机心存于胸中，则纯白不备，纯白不备，则神生不定；神生不定者，道之所不载也。”本书旨在深入探讨随着科学技术，特别是人工智能的高速发展，其在人类生产生活中扮演的角色愈发重要，同时引发的伦理道德问题。随着人工智能智能、性能和自主性的不断提升，对其的伦理争论也愈发激烈。本书从多个角度分析了人工智能的道德规范治理问题，包括其必要性、治理方法、人类治理的局限以及人工智能道德规范自治的可能性等。特别关注了人工智能道德建模的技术路径，尤其是语义分析技术在其中的应用。全书共分为十八章，从多个维度对人工智能的道德问题进行了全面、深入地剖析，旨在为推动人工智能的健康、可持续发展提供理论支持和实践指导。

在科技日新月异的今天，人工智能(AI)的崛起无疑是人类文明进程中的一大里程碑。它以其强大的智能、卓越的性能和不断提升的自主性，正在逐步改变着我们的生产生活方式，成为推动社会进步的重要力量。然而，正如任何一项新技术的出现都会伴随着一系列挑战和争议，人工智能的迅猛发展也引发了关于其伦理道德的广泛讨论。随着人工智能技术的不断突破，其在医疗、教育、交通、金融等诸多领域的应用日益广泛，其自主决策和行动的能力也日益增强。这种趋势使得人工智能在人类社会中的角色越来越重要，同时也带来了前所未有的伦理道德问题。如何确保人工智能在为人类带来便利的同时，不侵犯人类的基本权益，不破坏社会的和谐稳

定，成了一个亟待解决的问题。

对于人工智能的伦理争论，已经不再是简单的“是与否”的问题，而是涉及了更为复杂和深刻的“如何”问题。随着人工智能自主性的提高，如何确保其决策和行为符合人类的道德标准，如何避免其可能带来的道德危机，成为人们关注的焦点。这种关注不仅体现在学术界对人工智能伦理的深入探讨，也反映在了政策制定者和公众对人工智能发展的审慎态度上。

在这场关于人工智能伦理的争论中，有两种观点尤为引人注目。一种观点认为，随着人工智能自主性的增强，其可能会走向“有智无心”的极端，即拥有高度智能但缺乏道德情感，从而可能导致一系列道德问题。另一种观点则主张，人工智能应该朝着“有智有心”的方向发展，即不仅拥有智能，还要具备道德情感和道德判断能力，以更好地适应人类社会。然而，从目前人工智能的前沿研究来看，人工智能朝着自主性更强的“有智有心”和“强人工智能”甚至强于人类的“超人工智能”的发展方向已经不可阻挡。这一趋势使得我们不得不面对一个现实：遏制人工智能的发展路径以解决其可能带来的道德问题，显然是不可取的。相反，我们应该积极寻求通过技术手段来解决这一问题，其中，人工智能道德建模技术便是一个重要的方向。

在探讨人工智能道德建模技术之前，我们有必要对现有的伦理道德理论进行梳理和反思。传统的伦理道德理论大多基于人类的情感、经验和价值观，而人工智能作为一种非生物智能体，其情感和价值观的缺失使得传统的伦理道德理论难以直接套用。因此，我们需要为人工智能量身打造一套新的道德规范体系，以确保其在自主决策和行动时能够遵循人类的道德

标准。

语义分析技术作为人工智能道德建模的一种重要手段，近年来得到了广泛的关注和研究。通过对自然语言的理解和分析，语义分析技术能够帮助人工智能理解和识别人类的道德规范和价值观，从而为其决策提供道德依据。此外，语义分析技术还可以通过对大量道德相关文本的学习和处理，提取出其中的道德规范和道德原则，为人工智能的道德建模提供丰富的理论支撑。

本书的前几章将从理论和实践两个方面深入探讨人工智能道德规范治理的必要性和方法论。我们将分析人工智能在各个领域的应用中所带来的伦理道德挑战，以及传统的治理手段在面对这些挑战时的局限性。同时，我们将探讨人本主义方法在人工智能道德规范治理中的应用，以及如何通过算法干预和生物法则等手段来实现对人工智能的道德规范治理。

在后续的章节中，我们将重点关注人工智能道德规范自治的可能性及其实现路径。我们将分析人工智能在自主决策和行动中所表现出的道德自主性，以及这种自主性对传统的道德规范治理方式所带来的冲击。同时，我们将探讨如何通过构建道德规范自治模型来赋予人工智能自主道德决策的能力，并分析这种模型在实际应用中的可行性和局限性。

此外，本书还将对人工智能道德建模技术的地位和作用进行深入剖析。我们将分析人工智能道德建模技术在解决人工智能伦理道德问题中的独特优势，以及其在推动人工智能健康发展中的重要作用。同时，我们还将探讨如何进一步完善和发展人工智能道德建模技术，以更好地适应人工智能技术的快速发展和变化。

最后，本书将总结人工智能道德规范治理的最新进展和未来展望。我们将回顾人工智能伦理道德领域的研究成果和实践经验，分析当前面临的挑战和问题，并展望未来的发展方向和趋势。我们相信，通过不断深入的研究和实践，人工智能道德建模技术将在解决人工智能伦理道德问题中发挥越来越重要的作用，为人工智能的健康发展提供有力的支撑和保障。

综上所述，本书旨在全面、深入地探讨人工智能道德规范治理的问题和挑战，以及如何通过技术手段来解决这些问题。我们希望通过本书的研究和探讨，能够为推动人工智能的健康、可持续发展提供有益的参考和借鉴。

特别说明：本书第一至九章、第十四章共 28 万字由第一作者何冠岐完成；第十至十三章八万字由第二作者钟丹丹完成。