

Emerging Engineering Ethics for A Era of Artificial Intelligence: Can Humankind Control Superintelligence?

Copyright Statement

© Sakura, 2024. All rights reserved.

This document is protected by copyright law. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law. For permission requests, please contact the author at bili_sakura@zju.edu.cn.

Basic Information

Author: Sakura (陈振源, NO.22338035)

Department: School of Earth Science, Zhejiang University

Latest Update: 2024/5/31

Abstract: ...

Outline

- I. Introduction
 - Introduction to Engineering Ethics
 - Introduction to Large Language Models
- II. Emerging Engineering Ethics for AI
 - Safety of Large Language Models
 - Possible Solutions to AI Safety
- III. Future Directions
- IV. References

I. Introduction

Engineering ethics involves applying moral principles to the practice of engineering. It encompasses the responsibilities engineers have to ensure their work promotes public safety, adheres to ethical standards, and respects both society and the environment ([Harris et al., 2008](#)). Engineering decisions not only impact the physical world but also have long-lasting effects on societal structures and individual lives. Ethical considerations are fundamental to preventing harm and ensuring that engineering projects benefit all stakeholders ([Gail, 2006](#)).

Introduction to Engineering Ethics

Introduction to Large Language Models

II. Emerging Engineering Ethics for AI

Safety of Large Language Models

Possible Solutions to AI Safety

III. Future Directions

IV. References

Harris, C. E., Pritchard, M. S., & Rabins, M. J. (2008). *Engineering Ethics: Concept and Cases* (4th ed.). Wadsworth Publishing. ISBN 978-0495502791.