Chapter 2
Foundations of Artificial Intelligence Ethics

Binastya Anggara Sekti
https://orcid.org/0000-0001-5489-4888
Universitas Esa Unggul, Indonesia

ABSTRACT

Artificial Intelligence (AI) has become an influential force shaping modern society, raising urgent ethical questions about responsibility, fairness, transparency, and human dignity. The discussion explores the conceptual and philosophical foundations of AI ethics as a framework for ensuring that technological progress aligns with moral and social values. It traces the evolution of ethical thought from classical philosophy to contemporary digital governance and defines the core principles guiding responsible AI, including beneficence, non-maleficence, autonomy, justice, and explicability. Emphasizing a human-centered approach, the section examines the challenges of operationalizing ethics into AI design, deployment, and regulation. By grounding innovation in ethical responsibility, it argues that trust in AI depends not on technical superiority alone but on its moral alignment with the public good and social justice.

1. INTRODUCTION

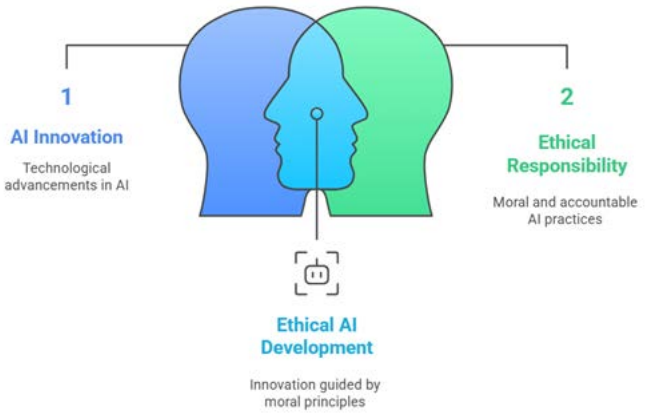
Artificial Intelligence (AI) has emerged as one of the most transformative forces of the 21st century, reshaping economies, societies, and human relationships. From predictive algorithms in healthcare and education to autonomous vehicles and generative language models, AI now permeates every layer of daily life. Yet, as its influence expands, so too does the complexity of the ethical dilemmas it generates. Questions once confined to philosophy—about agency, responsibility,

DOI: 10.4018/979-8-3373-7011-8.ch002

and fairness—have migrated into the realm of technology. The urgency of these questions lies in the fact that AI does not merely mirror human decision-making; it amplifies human choices, biases, and power structures at a global scale. Thus, understanding the ethical foundations of AI has become essential for ensuring that technological progress aligns with moral progress. Competencies for AI ethics require a comprehensive assessment of AI literacy (Knoth et al., 2024).

The development of AI was initially celebrated as a purely technical achievement—a triumph of human ingenuity in automation, computation, and data processing. However, as AI systems began making decisions that directly affected human lives, innovation alone proved insufficient. Autonomous systems can discriminate, misinform, or even cause harm when left unchecked. The conversation has therefore shifted from *what AI can do* to *what AI should do*. This transition marks the birth of AI ethics: an interdisciplinary field that merges moral philosophy, law, and social sciences with computer engineering. It represents a recognition that every algorithm carries moral weight and that the ethical use of AI is as important as its accuracy or efficiency. The justification and rationality of theory, and the basic concepts of AI-based problem-solving can form the core of a good artificial intelligence (AI) society. This is morally acceptable and utilitarian concerns stemming from the beginning and realization of the structural foundations of AI (Paraman & Anamalah, 2023). Figure 1 shows the AI, Ethics, and Social Trust.

Figure 1. AI, Ethics, and Social Trust



At its core, AI ethics is not only about machines—it is about humans and their values. Every dataset reflects human intention, and every algorithm is a projection of human judgment. This makes AI a mirror of our collective moral landscape. Ethical AI must therefore be human-centered, designed to protect dignity, auton-