Chapter 2
The Trust Paradox:
A Study on the Overlapping of Trust and Distrust Constructs

Pedro Brazo
https://orcid.org/0000-0002-8858-6813
University of Seville, Spain

Felix Velicia-Martin
https://orcid.org/0000-0003-2300-6256
University of Seville, Spain

Pedro R. Palos-Sanchez
https://orcid.org/0000-0001-9966-0698
University of Seville, Spain

ABSTRACT

This chapter discusses the complicated interplay between trust and distrust in artificial intelligence. It argues against the prevalent view of trust and distrust as two contrary states, providing theoretical foundations for the existence of overlapping states in decision-making processes. This revolutionary view is corroborated with both theoretical analysis and methodological novelties by applying a probabilistic model based on the Beta distribution. Thus, it becomes clear that trust in artificial intelligence has to go much beyond just the parameters of credibility and benevolence, which is comparable to human relationships, and include ethical, social, and technical components that are interlinked in the collaboration between humans and machines.

DOI: 10.4018/979-8-3693-9894-4.ch002

INTRODUCTION

The rapid advancement of artificial intelligence (AI) has brought about a new era in technology and decision-making, unparalleled in opportunities and complex ethical and moral issues. AI presents several major ethical concerns for society: bias and discrimination, privacy and surveillance, and the risk of system malfunctions and misuse. These issues raise difficult questions about the role of AI in decision-making. While AI offers significant potential for economic and social development, its implementation also exposes users to risks, from reinforcing social prejudices (Köbis et al., 2021) to the dangers of mass surveillance (Hu et al., 2024; Mobilio, 2023; Sedky et al., 2005), and privacy violations in areas like healthcare (Borda et al., 2022). These challenges highlight the need for careful ethical consideration of AI's impact, as it becomes a powerful tool full of both commitment and jeopardy, much like the duality of Dr. Jekyll and Mr. Hyde.

Given AI's worldwide reach and impact, evaluating the concepts of trust and distrust in these technologies is critical for its ethical development and implementation (Bécue et al., 2021; Gasser, 2023). Indeed, despite the widespread acceptance of trust in all types of relationships, the potential for trust betrayal has received little attention. While trust often denotes a collaborative engagement between agents in a relationship that is featured by shared expectations of ethical behavior, however, distrust denotes a break from such collaborative standards, often appearing as opportunistic behavior (Arrow, 1974). Inside the trust construct appear credibility and benevolence, two foundational sources of trust, stressing this contrast. According to the literature, credibility refers to a seller's attributes, especially in the areas of expertise, honesty, and reliability. On the other hand, benevolence points to the vendor having a personal stake in the customer's satisfaction (Garbarino & Lee, 2003; Sirdeshmukh et al., 2002). Under these precedents, distrust might be seen as the exact opposite of trust, as the absence of trust (Cho, 2006).

As AI increases in importance, expanding to encompass increasingly diverse partnerships, buyers' and suppliers' views become obsolete and limited, due to the entrance into the scene of other agents. The use of artificial intelligence has led to the existence of new relations and types of trust. Traditionally, the literature identifies four groups of trust-based relationships: (1) between individuals or between groups, (2) between individuals and organizations, (3) between organizations, and (4) between individuals and technology (Söllner et al., 2016). However, with the advent of artificial intelligence, a new model has evolved, creating a new type of agent that interacts transversally with all the previous agents: the machine learning agent. This new agent interacts autonomously, transforming the previous four groups of trust-based relationships into (1) individuals-individuals-AI, (2) individuals-organizations-AI, (3) organizations-organizations-AI, and (4) technology-AI. This