Advances in Social Science and Culture
ISSN 2640-9682 (Print) ISSN 2640-9674 (Online)
Vol. 8, No. 3, 2026
www.scholink.org/ojs/index.php/assc

Original Paper

Trust in AI: A Marx's Existential Perspective

Li Wu[1*] & Lisha Zhang[2]

[1] Hunan Agricultural University, Changsha, Hunan, China

[2] Hunan City University, Yiyang, Hunan, China

[*] Corresponding Author

Received: March 19, 2026 Accepted: April 22, 2026 Online Published: May 19, 2026

doi:10.22158/assc.v8n3p53 URL: http://dx.doi.org/10.22158/assc.v8n3p53

Abstract

With the accelerated advancement of globalization and modernization, the issue of trust has gradually become the focus of global social contradictions. In the highly advanced contemporary society, cutting-edge technologies such as AI (Artificial Intelligence) have not only reshaped the way of human existence, but have also become the fundamental force driving social development. The reality of technological existence makes people's trust in technologies such as AI particularly important. This article aims to explore the construction of trust in AI from the perspective of Marx's existential theory. First, it points out the importance of trust in the field of AI and emphasizes the impact of technological trust on social development and human well-being. Secondly, the appropriate boundaries of trust in AI are explored, emphasizing the necessity of seeking a balance between technological development and human values. Finally, a realistic approach to building AI trust is proposed.

Keywords

AI, trust, existential theory, risk

1. Introduction

Since the second half of the 20th century, the development of globalization and the deepening of modernization have greatly highlighted the issue of trust, making it a global problem and a social contradiction. In today's highly technological life, new technologies such as AI, supercomputing, genetic engineering and biomedicine, and driverless vehicles are emerging one after another. Living in an environment surrounded by technology, the real existence of human beings is constantly being technologized, and technology has almost become the basic form of objectifying human species life, the fundamental means of human survival and development, and the manifestation of the essence of human existence. New technologies provide positive energy for industrial production and social progress, but at the same time it is a double-edged sword. Because there are huge databases and

scientific algorithms behind technology, which are highly accurate, people increasingly trust the judgments made by programs and unconsciously choose to trust technology. But trusting technology without thinking can lead to negative effects, conflicts or disastrous consequences once it gets out of control or is exploited as a tool of exploitation. At the same time, excessive reliance on and promotion of new technologies without a thorough understanding of them can also hinder human innovation and progress, thereby providing an opportunity for the hidden risks behind them to explode. Therefore, in the "AI +" era, it is particularly important to examine AI technology from the perspective of Marxist philosophy, to focus on exploring AI from the perspective of Marx's existential theory, and to scientifically trust AI.

2. The Importance and Complexity of Trust in the Field of AI

Marx's existential theory, which emphasizes human existence and the free and all-round development of human beings, is a human-centered theory that holds that human beings are the subjects of social history, the driving force and purpose of social development, and the all-round development and free liberation of human beings are the highest goals of social development. At its core, Marxist philosophy is philosophical thinking based on the most fundamental facts of human survival and development. Marx emphasized: "The first premise of all human history is undoubtedly the existence of the living individual." (Marx, 1995) The most fundamental meaning here points out that the beginning of human history is the existence of living human individuals. Without living human beings, there would be no human society and history. Marx's Theory of Survival Marx's theory of survival encompasses the subject of survival, the way of survival, and the conditions of survival. At its core lies the emphasis on human social activities and practices, holding that technology is not merely a tool or means, but rather an embodiment of the way of human survival and an important component of human practice.

2.1 What Is Trust

Trust is often regarded as a complex psychological and social phenomenon. It involves an individual or group's confidence in the ability, honesty, and reliability of the other party. This confidence is based on past experience, evidence, or some form of consensus. Trust helps build and maintain various relationships and is the cornerstone of social interaction and cooperation. From the perspective of Marx's existential theory, trust is a dynamic, historical, and critical social structure in human social practice, which is both the foundation of social interaction and the product of practical activities, and is profoundly influenced by social and historical conditions. The sociologist Georg Simmel linked trust to human cognitive abilities: those who know everything (like God) do not need to trust others; But it is precisely because of the lack of this ability that the finite individual (limited cognition) in reality gets stuck in many situations where you either trust or don't trust. And when you choose to trust, you are actually making a leap without rational support, because there is no known information to support your action. (Georg, 2004) So any trust is blind—unless you have the Eye of God and can "see the face and the heart". Another German sociologist, Niklas Luhmann, suggests: "Trust is built on illusions. In

reality, there is less information available to ensure success, "he said. (Niklas, 1979) In other words, the giver and the trusted are structurally in a situation of information asymmetry. This situation makes the basis of trust completely "illusory"—the information available to the trust maker is structurally insufficient to make that trust. In this sense, trust is an "overdraft" of information.

2.2 The Singularity Battle

AI is an interdisciplinary field aimed at developing and applying theories, methods, and techniques that can simulate, extend, and expand human intelligence. The concept of AI is recognized as having been first proposed at the Dartmouth Conference in 1956. In the 1970s, American philosopher John Searle proposed "strong AI" and "weak AI". Later, Oxford University philosopher Nick Bostrom proposed "super AI". During the rapid development and application of AI technology, some scientists and futurists have focused on discussing and debating the limits of AI development and the possibilities of the future. They believe that when the technology of AI reaches a certain level, machines will be able to improve and optimize themselves, thus surpassing human intelligence and capabilities, which will lead to an important turning point in human history, the "singularity". The concept was first proposed by mathematician John von Neumann, and American futurist Ray Kurzweil, in his books "The Singularity Is Near" and "The Future of AI", uses the "singularity" as a metaphor to describe a certain stage in time and space when the capabilities of AI surpass those of humans. The Singularity debate mainly revolves around two core questions: one is whether AI can reach or even surpass human intelligence; The second question is how AI will impact human society and civilization if it reaches that level. Supporters argue that with the exponential growth of computing power and the continuous advancement of algorithms, the singularity of AI is inevitable, which will bring about a huge leap in technology and society. It has sparked attention and reflection on AI technology and prompts people to think about how to deal with the challenges and problems brought by AI. Geoffrey Hinton, a 2018 Turing Award winner and a pioneer in deep learning, publicly expressed his judgment in a speech titled "Two paths to Intelligence" at the Beijing Academy of AI Conference on June 10, 2023: Artificial neural networks will be smarter than real neural networks, and "super-intelligent AI" will soon arrive. So is trusting AI in the era of superai an active choice or a passive and reluctant choice? Will AI contribute to human well-being or destroy humanity? What should be done before that if it can be trusted? Wait, these are the hot topics that people are worried about and discussing right now.

2.3 Trust in AI amid Optimism and Pessimism

There are three attitudes that have an impact on AI trust.

One is technological optimism, seeing technology as the messenger of a bright future for humanity and advocating rationality and the omnipotence of science and technology. Dessauer is a representative of technological optimism. Rather than believing that technology poses essential harm to human beings, Dessaur argues that technology, through its own development, directly imposes higher cultural and ethical requirements on technicians, which may lead to the progress of human civilization, or rather foreshadowed the progress of the entire human civilization. As he puts it, "technology makes (praegt)

man (Introduction to Philosophy of Technology, 2009). Optimists believe that the development of AI will have a huge positive impact on society. They emphasize the potential of AI, including increasing productivity, improving healthcare services, optimizing transportation systems, and promoting personalized education. In this view, AI is seen as a catalyst for innovation and progress, capable of addressing many long-standing social problems. They advocate building trust through education and public engagement to ensure that the development of AI is in line with the overall interests of society.

The second is technological pessimism, which regards technology as the root of all evil, denies rationalism and criticizes science and technology. It is a trend of thought filled with fear, pessimism and helplessness about the future development of civilized society, believing that the future will be a alienated, degenerate and even infinitely regressive society. The early representative of technological pessimism was the British economist Malthus, in his book "The Principles of Population", who denied the significance and role of economic and scientific and technological development in helping humanity escape poverty. After the publication of the Limits to Growth at the Club of Rome in 1972, technological pessimism became a widely known trend of thought with an independent theoretical form. The ancient Chinese school of Laozi and Zhuangzi regarded technology as a "fancy skill", believing that craftsmanship was immoral and had caused the moral decline of society. Rousseau, in the West, criticized technology "out of place" in the song of its brilliant achievements to mankind. Pessimists are more focused on the risks and challenges that AI may bring. They are concerned that AI's decision-making process may be opaque, biased, and potentially have negative impacts on the job market, privacy rights, and social ethics. For example, automation could lead to mass unemployment, and facial recognition technology could be used to violate personal privacy. 3. Technoneutral proponents argue that the development of AI technology is accompanied by both great opportunities and challenges. This view emphasizes the need to balance the potential of AI with its potential risks by formulating sensible policies, establishing ethical frameworks, and promoting interdisciplinary research. They believe that through ongoing dialogue and collaboration, all stakeholders, including policymakers, technology developers, industry experts and the public, can work together to ensure that the development of AI technology can maximize the benefits of human society while minimizing its potential negative impacts.

3. Appropriate Boundaries of Trust in AI

The term "trust" has its application space in technology, particularly in the context of AI. People's expectations, demands and concerns about technology are all related to trust. There are different views on trust research in AI at present. In terms of the degree of trust, people's trust in AI can be complete trust, complete distrust, and moderate trust.

3.1 Discrimination of Total Distrust of AI

3.1.1 Zero Trust

Zero trust is a new type of cybersecurity model based on never trusting and always verifying. In terms

of the limit threshold, a completely distrustful end corresponds to zero trust. The zero-trust model has received extensive attention in research and practice because it can meet new cybersecurity requirements, in which the access subject needs to be authenticated to access resources, whether on or off the Intranet. The earliest prototype of zero trust originated from the Jericho Forum, which was established in 2004. P. Laplante et al. argue that any critical product or service based on AI (AI) should be constantly questioned and evaluated. This suggests adopting a "zero trust" or "trust but verify" approach to AI. (Laplante & Voas, 2022). Ye Libang et al. argue that the essence of zero-trust security is identity-based dynamic trusted access control, focusing on security capabilities in dimensions such as identity, trust, business access, and dynamic access control, continuously evaluating trust based on multi-dimensional factors such as people, processes, environment, and access context in business scenarios, and dynamically adjusting permissions through trust levels, Form a dynamic adaptive security closed-loop system with strong risk response capabilities. (Ye, Zhang, X., & Zhang, W. Q., 2023) "Zero trust" has become an important research hotspot in today's cyber security field. The core idea of zero trust is "never trust, always verify," which means that under no circumstances should there be a preconceived trust attitude towards any entity or behavior in the network, but it should always be verified and reviewed. The implementation of zero trust mainly relies on the fine management of data. The key lies in controlling access to data, a new data-centric boundary mechanism that can effectively prevent data leakage and abuse, thereby enhancing data security. Zero trust offers an identity-based, more fine-grained approach to access control, compared to traditional security schemes that focus only on boundary protection and grant excessive access to authorized users. The implementation of this concept requires us to abandon traditional security concepts and shift to a more refined and flexible security strategy. Zero trust also provides new ideas and approaches for the security of AI products.

3.1.2 Questioning of Zero Trust in AI

The "chain of suspicion" is actually a logical consequence of the inability to establish the most primitive trust. Each individual, nation or even civilization is reluctant to make further "overdraws" beyond the information and rational evidence they know. Individuals who enter the logic of the chain of suspicion cannot enter into a state of political community with each other: they can only remain in what Hobbes called the "natural state" of pre-politics, where people fight each other like wolves (Wu, 2019). According to Marx's theory of human nature, human beings are the sum total of social relations, and their actions and decisions are influenced by social environment, economic base, cultural background and historical conditions. Under this theoretical framework, trust is regarded as an integral part of social relations and is one of the foundations of human interaction and cooperation. When humans are completely distrustful of AI, this can be seen as an extreme breakdown of social relationships. In such a breakdown, the interaction and cooperation mechanisms between humans and AI are cut off, and AI systems are regarded as completely untrustworthy external entities. As a result, the human decision-making process will no longer take into account the input or advice of the AI system, but will rely entirely on human judgment, experience, and intuition.

3.2 Regarding the Risk of Complete Trust in AI

Today, with the increasing popularity of AI, people's trust in it is also gradually rising. However, when humans fully trust AI, it means that the decisions produced by AI systems are directly equivalent to those made by humans, which can lead to a series of problems. First of all, complete trust in AI can lead to a violation of the rules. For example, from a teaching perspective, a teacher's complete trust in AI could lead to a violation of teaching rules. Sun Lihui (2022) et al. pointed out that emotion is a more advanced way of thinking that current AI technology cannot accurately simulate and apply, and this is precisely the direction that teachers should focus on cultivating in learning. [8]Excessive reliance on AI may lead to a misplacement of the teaching subject, where teachers may place the teaching focus entirely on the process of using AI technology or fully trust the output results of AI machines, thereby neglecting their own teaching position. At the same time, learners may also focus too much on the technical form, making learning activities more cumbersome and affecting learning outcomes. Secondly, from the perspective of information reception, Li Liwen (2020) argues that the risk cannot be ignored. If an individual fully trusts the pushed information under the influence of a semi-closed mechanism, once this pushed information is malicious or false, it will pose a huge threat to the individual. (Li, 2020) At present, there are still many unstable factors in the technical aspect of AI, such as the hidden dangers of autonomous driving, etc. Complete trust is also not advisable.

3.3 An Analysis of the Moderation of Trust in AI

Moderate trust is somewhere between full trust and total distrust. AI, as a powerful technological tool, can handle large amounts of data, increase productivity, assist in decision-making, and even play a significant role in areas such as education, healthcare, agriculture, and transportation. But the debate over whether AI has autonomy or will, whether AI can be regarded as a moral subject, and whether AI is above human beings, in essence, stems from the reflection on whether human beings can trust and control technology to make it serve human beings, and this reflection points directly to at least two aspects: People trust people and people trust their own decisions, and people trust AI technology.

3.3.1 People Trust People and People Trust Their own Decisions

Marx's ontology emphasizes the social nature of human beings, the importance of practical activities, and the role of technology in the process of human liberation. From Marx's ontology, the key to moderate trust in AI lies in understanding and grasping the relationship between man and technology (especially AI), and how this relationship affects man's social practice and self-actualization. Therefore, there are several aspects to consider when moderately trusting AI: First, as a technology, the development and application of AI should promote human practical activities rather than replace them. Marx's ontology holds that human beings transform the world and the foundation of their own existence through practical activities, but when carrying out "AI +" actions, it is necessary to recognize that it can enhance human capabilities in some respects, but also be wary of excessive reliance on AI, which may weaken human subjectivity and creativity. At the same time, be wary of technological alienation, that is, human beings being controlled and alienated by their own creations. Second, the

development and application of AI should be in line with the overall interests of society, promoting social equity and justice, rather than merely serving specific interest groups. Third, we should be able to apply AI to liberate human beings from simple, repetitive, heavy and dangerous labor, providing them with more free time and development space, and promoting the all-round development and self-actualization of human beings. The bidirectionality of the relationship between science and technology and human beings, that is, human beings change the world by creating science and technology, while science and technology also shape human thoughts, behaviors and lifestyles. Here, science and technology are not only tools of human activity, but also an important part of human life. Marx believed that the future society would be one where “people use technology” rather than “technology uses people”. In the future socialist society, people will be able to freely master and use technology, rather than be controlled by it. People will be able to use technology creatively according to their own needs and desires to achieve all-round development of human beings. Here, the trust in AI is the trust that people have in their own decisions and technological control, the trust that people have in each other.

3.3.2 People Trust AI Technology

It is a controversial question whether AI is autonomous when regarded as an object. Traditional technologies are often passive and controlled, performing tasks according to human design, intention, and instructions. However, with the development of AI and machine learning, technology is becoming increasingly autonomous. This means that technology may need to be trusted to make reliable decisions and actions on its own. But in the application of technology, there are risks, whether due to system failure or human error. Deep Mind researchers analyzed the ethical and social risks of large language models (LLMS) based on multidisciplinary literature in computer science, linguistics, and social sciences, and summarized twenty-one risks identified using expertise and literature as including discrimination, hate speech and exclusion, real information hazards, misinformation hazards, malicious use, human-computer interaction hazards, environment and society There are six major categories of risks (Weidinger, Uesato, Rauh et al., 2022), including economic hazards. In this case, whether people will trust technology depends on its stability and reliability, as well as the ability of people themselves and machines to handle errors and malfunctions.

4. A Realistic Path to Building Trust in AI

AI trust is neither passive acceptance under technological determinism nor utopian total rejection, but a dynamic balance rooted in human survival practices. Full trust leads to the transfer of subjectivity and technological alienation, while total distrust slides into a natural state of “chain of suspicion”, and only moderate trust fits the core of Marx’s existential theory that “man uses technology” rather than “technology uses man”. Following this logic, to build a realistic path of trust in AI, one must start from the dual perspectives of “technology - humanity”, return to human existence itself, and form a practical approach with subjectivity as the core, reliability as the support, and dynamic adaptability as the

feature.

4.1 Return to Human Subjectivity: Reconstructing the Foundation of "People Trust People and Their Own Decisions"

Marx's existential theory repeatedly states that human practical activity is the first premise of history. The starting point of trust in AI is not the worship of machine computing power, but the reconfirmation of human rationality, judgment and value foundation. When teachers rely on AI to generate lesson plans, doctors refer to AI diagnostic advice, and judges use AI to assist in sentencing, the core of trust should always be "human reconfirmation" rather than "one-way output of the system". This means fostering a "reflective trust" at the institutional and educational levels: on the one hand, through general education and professional ethics training, make users of technology clearly aware of the limits of AI's capabilities and known flaws, and avoid misting probability predictions for absolute truths; On the other hand, establish a "veto" mechanism for human-machine collaboration - in critical decision-making areas such as healthcare, justice, and finance, AI recommendations must be reviewed by qualified natural persons who are ultimately responsible. This is not to undermine the effectiveness of technology, but to transform trust from "blind trust in algorithms" to "trust based on understanding", thus truly realizing what Marx called "man using technology", rather than losing the subjectivity of survival in the entanglement of technological logic.

4.2 Strengthening the Dimension of Technological Objects: Multi-level Trust from "Usability" to "Accountability"

People's direct trust in AI technology ultimately depends on the quality and rules of the technology itself. As the risks of "zero trust" and "full trust" have been analyzed earlier, the real path requires the construction of a hierarchical and verifiable system of technological trust. Specifically, the first level is "basic reliability", that is, AI systems should have stable predictability in preset scenarios and reduce black box random errors, which requires developers to embed redundant checkchecks and adversarial tests in algorithm design; The second level is "intelligibility and controllability", drawing on the core vision proposed by scholars such as Sun Liwen, which mandates that high-risk AI systems provide understandable decision-making logic and reserve physical or software interfaces for human intervention for users, making "out-of-control" institutional design impossible; The third level, "accountability," is a contemporary extension of Marx's existential theory that "the consequences of human practice should be borne by human beings." It is necessary to clarify that when AI causes damage, there should not be a vacuum of accountability where the algorithm is not responsible, the developer shirks responsibility, and the operator is exempt from liability. Instead, a full-chain accountability mechanism should be established from the data provider, the model trainer to the deployment and application. Only when these three levels progress step by step will the technology be worthy of being entrusted with a moderate level of trust.

4.3 Dynamic Evolution and Human Infiltration: The Renewal of Trust Concepts and Institutional Iteration

Trust is not a one-time deposit, but a living relationship that is constantly adjusted in the long coexistence of people and technology. The debate over technological iteration between the hardware and software camps actually reveals that AI itself is still in an unfinished state. Therefore, the realistic path to building trust must include a dynamic evolution mechanism in the time dimension. On the one hand, drawing inspiration from Hinton's "non-immortal computing" - that technical knowledge is hardware-dependent and temporary - society should establish a regular re-evaluation system of technical trust, conducting independent audits of AI systems in key areas every two to three years and dynamically adjusting trust levels based on their actual performance; On the other hand, the deep foundation of technology trust lies in the synchronous growth of humanistic concepts. The "worldview of the machine system" as Mumford put it has shaped modern civilization, and today we need to cultivate a "reflective technological literacy" - through public media, school education and community dialogue, so that people neither fear technology nor deify it, but are accustomed to using AI with critical empathy. Ultimately, as Marx envisioned the future society as a community of the free and all-round development of human beings, trust in AI should also be elevated from "trust in technology beyond human beings" to "a broader trust among people through technology". This is the ultimate direction of trust construction from the perspective of existential theory.

Conclusion

This study conducts a systematic analysis of the trust issue in AI from the perspective of Marx's existential theory. Emphasizing human subjectivity and the core position of social practice in technology trust, placing AI trust within the framework of "people using technology" rather than "technology using people" enriches the cross-disciplinary study of trust theory and technology philosophy. A hierarchical path of AI trust is proposed to provide actionable guidance for policy-making, educational practice, and technology application. In the future, experimental verification, policy simulation and interdisciplinary research will be further integrated to refine the theoretical model and practical application of AI trust, and to support the free development of humanity and the guarantee of social justice empowered by technology.

Acknowledgments

This paper was a phased achievement of the 2024 General Project of Hunan Provincial Social Science Achievement Review Committee: "Research on Technology Trust Models in Risk Society" (Project No. XSP24YBZ041).

References

Georg, S. (2004). *The Philosophy of Money (Third Enlarged Edition)*. London: Routledge.

Introduction to Philosophy of Technology, edited by Qiao Ruijin, Mu Huansen and Guan Xiaogang. Higher Education Press. (2009).

Laplante, P., & Voas, J. (2022). *"Zero-Trust AI? ," in Computer*, *55*(2), 10-12.

Li, L. W. (2020). The concerns and risks of precise governance in the Age of AI. *Journal of Hohai University (Philosophy and Social Sciences Edition)*, *22*(01), 82-90+108.

Marx, E. (1995). *Selected Works of Marx and Engels: Volume I. Marx, Engels, Lenin, Stalin, Central Committee of the Communist Party of China Works Compilation and Translation Bureau*. Beijing: People's Publishing House.

Niklas, L. (1979). *Trust and Power*. New York: John Wiley & Sons.

Sun, L. H., & Wang, X. Q. (2022). The future picture of AI for Education: From the Perspective of Machine behavior. *Chinese Journal of Educational Technology*, *2022*(04), 48-55+70.

Weidinger, L., Uesato, J., Rauh, M. et al. (2022). Taxonomy of risks posed by language models. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 214-229. https://doi.org/10.1145/3531146.3533088

Wu, G. J. (2019). Rethinking War and Peace: A Reinterpretation of the History of Political Philosophy by Hobbes, Kant, Schmitt, and Rolls. *Journal of Tongji University (Social Sciences Edition*), *30*(02), 64-77.

Ye, L. B., Zhang, X., & Zhang, W. Q. (2023). Research on the Current Status and Trends of Zero Trust Architecture in the United States. *Information Security and Communications Confidentiality*, *2023*(07), 12-21.