Whose Values Are AI Models Aligning With? How Culture Shapes People's Normative Expectations of AI Value: An Integrative Review

Authors: Huarong LIU, Xinyu ZHU, Nanying LI, Xinrui HUANG, Qingqing DU, Ruoting LIU, Liyunhan PENG, Feng YU, Kaiping Peng, Sylvia Xiaohua CHEN, Xiaomeng HU, Xiaomeng HU

Date: 2024-07-31T00:00:00+00:00

Abstract

With the rapid development and widespread application of artificial intelligence (AI) technology, the profound cultural influence on AI values has attracted widespread attention. Research to date, however, has not systemically looked at both the human universals and cultural differences in people's normative expectations of AI values. To further explore the potential impacts of culture on AI values through the lens of cultural psychology and highlight the importance of taking into account the role of cultural diversity played in AI developments and applications, our current integrative review briefly synthesizes what might be the cross-cultural consensus and what might be the cultural differences in shaping people's attitudes, behaviors, and normative expectations regarding AI values. In addition, we discuss the vital role of cultural beliefs and cultural norms in the ethical supervision and application of AI in human society. To better understand the complex interaction between AI and culture, future work should focus on developing and iterating algorithms for diverse cultural scenarios, thereby both promoting the globalization of AI application and meet diverse cultural demands to ultimately improve the well-being of individuals and society across the globe.

Full Text

Preamble

Whose Values Are AI Models Aligning With? How Culture Shapes People's Normative Expectations of AI Value: An Integrative Review

Huarong LIU[1], Xinyu Zhu[2], Nanying LI[3], Xinrui HUANG[4], Qingqing DU[5], Ruoting LIU[6], Liyuhan PENG[7], Feng YU[8], Kaiping PENG[9], Sylvia Xiaohua

CHEN[10], Xiaomeng HU[1]*

[1]Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China

*Corresponding Author: Xiaomeng HU, PhD
Department of Psychology, Renmin University of China
59 Zhongguancun Street, Haidian District, Beijing (100872), China
Email: xiaomenghu@ruc.edu.cn

Abstract

With the rapid development and widespread application of artificial intelligence (AI) technology, the profound cultural influence on AI values has attracted widespread attention. Research to date, however, has not systematically examined both the human universals and cultural differences in people's normative expectations of AI values. To further explore the potential impacts of culture on AI values through the lens of cultural psychology and highlight the importance of accounting for cultural diversity in AI development and applications, our current integrative review synthesizes what might constitute cross-cultural consensus and what might represent cultural differences in shaping people's attitudes, behaviors, and normative expectations regarding AI values. In addition, we discuss the vital role of cultural beliefs and cultural norms in the ethical supervision and application of AI in human society. To better understand the complex interaction between AI and culture, future work should focus on developing and iterating algorithms for diverse cultural scenarios, thereby both promoting the globalization of AI application and meeting diverse cultural demands to ultimately improve the well-being of individuals and society across the globe.

Keywords: artificial intelligence, cultural psychology, AI ethics

1 Introduction

As AI systems continue to permeate various aspects of our daily lives, scholarly attention has increasingly focused on the significant impacts of cultural factors in shaping the development and deployment of AI. AI encompasses a wide range of technologies, from simple algorithms to complex machine learning systems, which are used to perform tasks such as visual perception, language understanding, and decision-making [**?**, **?**]. AI can be categorized into different types, including physical AI (e.g., robots), virtual AI (e.g., Siri and other digital assistants), and embedded AI (e.g., car navigation systems) [**?**, **?**]. It is imperative to investigate the complex interactions between AI and culture. Cultural psychology is dedicated to studying how different cultural contexts or cultural traits shape individuals' and groups' mental processes and behavioral patterns. As AI technology permeates human real-life contexts in unprecedented ways, it is critical to explore how these different types of AI technologies are understood

and applied in different cultural contexts among different cultural populations from the perspective of scientific psychology [?, ?]. Culture is not only regarded as an external factor that affects an individual's cognition, emotions, and behaviors, but also as an internal part of the mental process itself [?, ?]. Given that contemporary research and application of AI technology increasingly engage with the deep psychological processes and social behaviors within human society, it is particularly imperative to examine the development, application, and global governance of AI technology from the perspective of cultural psychology.

Cultural psychology helps us better understand how AI technology interacts with global users in different cultural contexts within the context of globalization, and guides the development and applications of AI technology to adapt to these complex interactions [?, ?]. By considering the role of culture in shaping human cognition, emotion, and behaviors, the development and applications of AI will be able to more precisely adapt to the cultural diversity needs of global users around the world, thereby promoting the flexibility, universality, and acceptance of AI technology. However, the current design, development, and practical applications of AI systems often rely on big data clusters and machine learning algorithms, which may ignore subtle changes caused by cultural differences when addressing complex AI-human interactions. This oversight may lead to the failure and inability of AI systems to fully comprehend the subtle needs and well adapt to the needs of global users from diverse cultural backgrounds with differing cultural mindsets, thereby exacerbating cultural biases and social inequalities, and affecting the fairness and effectiveness of AI systems [?, ?].

This is especially true for the training and application of large language models (LLMs) around the world. Recent research indicates that the datasets of these large models are mainly derived from specific cultural backgrounds, especially in the United States, thus introducing systematic cultural bias in decision-making and language output. The Hofstede Cultural Survey from the University of Copenhagen and the World Values Survey used by the AI start-up Anthropic have quantitatively analyzed this problem. Researchers tested LLMs through the Hofstede Cultural Survey, which measured human values across different countries, and the results indicated a strong alignment of LLMs with American mainstream culture [?, ?]. Anthropic conducted similar tests using a World Values Survey and reached similar conclusions, finding that LLMs tend to reflect and reinforce various aspects of American mainstream culture [?, ?]. This cultural bias is not limited to the language output of large language models but also affects the way large models solve problems and make decisions. For example, when asked to generate "breakfast" images, the training of DALL-E 3, which is primarily on Western images, generates images of pancakes, bacon, and eggs, which reflect the eating habits of Western cultures. These findings underscore the necessity of considering cultural diversity in the development and applications of AI systems. Failure to address these cultural biases may lead to greater prejudices and misunderstandings in AI, thereby diminishing the universality and effectiveness of AI systems.

At the same time, people's attitudes, views, and behaviors toward AI are also influenced by their cultural background. For example, some cultures may place greater emphasis on privacy protection and have reservations about the penetration of AI technology into personal life, while others may be more open and willing to accept the conveniences brought by AI. These cultural differences in turn have an impact on the acceptance, design requirements, and application scenarios of AI, requiring developers to fully consider the needs and expectations of people in a multicultural context when designing AI systems [**?**, **?**]. Therefore, to achieve a truly fair and effective AI system, it is necessary not only for AI technology to adapt to and understand the needs of global users from different cultural backgrounds, but also to deeply understand and respect the cultural differences in people's attitudes and behaviors toward AI across different cultural contexts.

In the practical applications of AI technology, considering cultural factors from the perspective of human universals and cultural differences is crucial to the development and application of AI systems. This not only helps to improve the cross-cultural management of AI technology, but also assists designers in re-examining the legal, ethical, and global governance issues related to AI systems. By deeply understanding and integrating behavioral patterns and psychological characteristics from different cultures, designers can promote the harmonious coexistence of humans and AI systems. Current research has found that cultural differences may lead to different cultural groups holding different ethical norms and value orientations when dealing with AI systems [**?**, **?**]. For example, there are significant cultural differences in the expectations of users from different cultures regarding AI privacy and information sharing [**?**, **?**, **?**]. Therefore, examining AI from the perspective of cultural psychology is not only conducive to ensuring the fairness and effectiveness of AI systems, but also crucial for ensuring that AI technology can be widely accepted and flexibly applied across different cultural environments worldwide. Only by fully considering and integrating cultural diversity can AI systems achieve their true globalization process and localization potential, and provide high-quality and highly adaptable services while respecting the cultural demands of all users.

Although there have been numerous research findings on the development and applications of artificial intelligence in various fields, its cultural impacts have not received sufficient research attention. Particularly in the context of globalization, how AI technology adapts to and reflects the commonalities and differences among various cultures remains inconclusive. Therefore, our current work aims to selectively review the existing literature and provide a theoretical framework for the cross-cultural development, management, and global applications of artificial intelligence. We will integrate the complex attitudes and ethical expectations towards AI technology concerning universal values shared by humans across different cultures, the unique psychological and behavioral characteristics within various cultural contexts, and how these cultural characteristics shape the development, management, and application of AI technology. We will also examine the role of cultural characteristics in the acceptance and ethical norms

of AI technology. Finally, we will propose exciting future research directions, aiming to emphasize the profound impacts of human universals and cultural diversity on the scientific research and practical applications of AI technology.

2 Human Universals in AI Values

The main tenets of cultural psychology are that culture shapes and reshapes human cognition, perception, and behavior [**?**, **?**]. Thus it is very natural to deduce that culture also has profound and sometimes implicit impacts on people's psychological expectations and behavioral tendencies regarding AI systems. Interestingly, there may be somewhat of a cultural consensus on human values among different cultures [**?**, **?**]. Schwartz [**?**] puts forward the Theory of Basic Human Values, which contains ten universal values, later expanded in 2012 to nineteen universal values [**?**, **?**]. Similar to the findings in cross-cultural studies of demand hierarchy theory, different cultures exhibit varying preferences for the importance of these universal values. However, individuals and groups across diverse cultures demonstrate similar cognition, perception, and behavior under the influence of these universal values [**?**, **?**]. This phenomenon is also evident in the development and applications of AI. Governments in different countries and a variety of international organizations around the world share similar expectations and application needs for the development and governance of AI systems [**?**, **?**]. Consequently, there are commonalities in the ethical principles and guidelines for AI. For example, Joblin et al. [**?**] found that among the guidelines involving 84 countries and international organizations, there were 73 requirements for transparency in AI technology, 68 requirements for justice and fairness, and 60 requirements for non-malice and responsibility. The Beijing Consensus on Artificial Intelligence and the New Generation of Artificial Intelligence Ethics recently promulgated in China also stipulate the guidelines of artificial intelligence in terms of information transparency, fairness and justice, and social responsibility.

Not only in the formulation of norms and guidelines, but also under the influence of universal values, different cultures have similar demands for the application and governance of AI technology. For instance, in fields such as environmental protection and healthcare, AI developers from diverse cultural backgrounds expect AI to have positive impacts on human life through environmental management and medical assistance [**?**, **?**, **?**].

2.1 Cross-Cultural Consensuses in AI Values

Within the framework of Basic Human Values proposed by Schwartz [**?**], the value of security originates from the fundamental needs of individuals and groups. These needs encompass the pursuit of physical security, such as avoiding dangers and threats, as well as the pursuit of psychological security, such as stable relationships and predictable environments. During the COVID-19 pandemic, the cleaning and disinfection functions of AI robots have physically

helped high-contact and high-interpersonal industries, such as hotels [?, ?], airports [?, ?], and other application scenarios, to avoid the threat of the virus. These functions also meet people's needs for psychological security. Additionally, British hospital pharmacies use AI robots to assist in dispensing, which has successfully reduced the rate of dispensing errors [?, ?] and protected the physical and mental safety of patients. Security reflects people's psychological expectations of social order and protective mechanisms across different cultures [?, ?]. The core of this value lies in the pursuit of security, harmony, and stability within human society, relationships, and the self. It emphasizes the need for a stable and orderly social environment and focuses on protecting individuals and groups from harm.

At the same time, universalism values [?, ?], as one of the universal values, play a crucial role in encouraging countries to utilize AI technology for environmental governance. The core goal of universalist values is to understand, appreciate, tolerate, and protect the well-being of all people and nature. This value extends beyond attention to a specific group or community to encompass a broader concern for all mankind and environmental protection [?, ?, ?, ?]. Universalist values originate from human recognition of survival and resource scarcity, as well as an understanding of the interdependence between different individuals and groups. Nishant et al. [?] found that the innovative solutions brought by AI technology are providing new directions for business practices and addressing social issues, thereby supporting the pursuit of sustainable development in differing countries around the world. Ahmad et al. [?] found that AI technology can automatically control energy supply and demand, promote decision-making and operational processes, and change consumption habits to achieve a sustainable environmental protection mechanism, while maximizing operational efficiency and work performance.

Universalism reflects a focus on understanding, appreciation, tolerance, and protection for the welfare of all people and for nature [?, ?]. This value entails a commitment to justice and responsibility, indicating that when employing AI, we must not only consider the immediate benefits of the AI technology but also its long-term impacts on the environment and human society. Justice requires ensuring fairness and equality in the use of AI technology, protecting the rights and interests of vulnerable groups, and preventing the abuse of technology and unfair resource allocation. Responsibility requires us to consider the impacts of AI solutions on the environment during AI development and deployment, ensuring the sustainability and ethicality of AI technology. Therefore, universalism reflects not only the pursuit of technological progress but also the importance of global justice and civic responsibility in promoting the application of AI technology to assist in environmental governance. These values encourage the pursuit of technological innovation while ensuring the care and protection of all humanity and the natural environment, guaranteeing that scientific progress and technological development are coordinated with social well-being and environmental sustainability, thus creating a more just, safe, and sustainable future for humankind.

In the realm of AI, “moral machines” has become a common and important concept [?, ?, ?, ?]. Globally, the ethical guidance and policy formulation of AI demonstrate the universal values of cultural sharing. Although people from different cultural backgrounds have varying levels of acceptance and adaptation to AI, global society has shown significant cultural consistency in the transparency, fairness, justice, and responsibility of AI technology [?, ?]. This universal consensus reflects the basic values of humanity regarding how AI technology should be developed and applied. Among these, transparency requires that the decision-making process of AI be understandable and reviewable, which is essential for building global users' trust in AI systems. For example, Jobin et al. [?] indicates that transparency is one of the most frequently mentioned principles in the guiding principles of AI technology ethics worldwide. Transparency not only helps to reveal the basis and logic of AI decision-making but also enables potential biases and errors to be identified and corrected in a timely manner, thus enhancing the credibility and acceptance of the AI system. Justice and fairness emphasize that AI decision-making should not exacerbate existing inequalities, but should strive to reduce social injustice. The requirements of responsibility ensure that when errors or misconducts occur in AI systems, clear accountability can be established and appropriate corrective measures can be taken. This includes not only the correction of technical errors but also the compensation and protection of affected individuals and groups. These global requirements for transparency, fairness, justice, and responsibility in AI technology reflect cross-culturally and universally accepted ethical standards for AI development and applications.

Under this ethical framework, Ikkatai et al. [?] further reveals how universal human values are reflected in the specific applications of AI and its ethical principles. They focus on eight universally shared themes in the guiding principles of AI technology: privacy, accountability, safety and security, transparency and interpretability, fairness and non-discrimination, human control of technology, professional responsibility, and the enhancement of human value. Through an online questionnaire survey conducted in four scenarios in Japan, researchers explored the public attitudes towards AI ethics and found that public approval or opposition to the use of AI varies from scenario to scenario. For example, in scenarios where AI is used in weapon systems, people are more concerned about AI ethics. Age significantly affects people's views on these topics in different scenarios, while gender and understanding of AI technology vary according to the theme and scene [?, ?]. We not only see the reflection of universal human values such as security, justice, and responsibility in AI policy formulation, but also observe the intersection and overlap between these values and AI ethical principles.

By analyzing and understanding the common views of different cultures on the ethical principles of AI technology, we can more deeply explore how to promote these ethical principles globally to ensure that the ethical norms of AI technology are widely supported and socially recognized. This cross-cultural ethical consensus also provides a solid theoretical foundation for the global applications and

policy formulations of AI technology, and helps to promote the simultaneous development of AI technology and ethical considerations in different cultures.

3 Cultural Differences in Attitudes and Behaviors Towards AI

The Cultural Dimensions Theory proposed by Dutch social psychologist Geert Hofstede has been widely influential in the academic discourse of cultural psychology. This theory categorizes culture into six dimensions that facilitate cross-cultural comparisons and understandings: Power Distance, Uncertainty Avoidance, Individualism versus Collectivism, Masculinity versus Femininity, Long-Term versus Short-Term Orientation, and Indulgence versus Restraint [**?**, **?**]. For instance, regarding power distance, Eastern cultures tend to exhibit high power distance and accept power inequality, whereas Western cultures are more inclined to low power distance, rejecting power inequality [**?**, **?**]. In the context of uncertainty avoidance, Latin, Japanese, and German cultures favor high uncertainty avoidance, emphasizing rules and stability, whereas Chinese and Nordic cultures prefer low uncertainty avoidance. With respect to individualism and collectivism, Western cultures are more inclined towards individualism, emphasizing individual rights and personal interests, while Eastern cultures favor collectivism, emphasizing group cohesion and collective interests. Regarding masculinity versus femininity, Japanese, German, and other cultures tend to be more masculine, characterized by male dominance, while Nordic, Dutch, and similar cultures lean towards femininity. In terms of long-term versus short-term orientation, East Asian cultures are more inclined towards long-term orientation, emphasizing personal effort and knowledge accumulation, while American, Australian, and similar cultures tend to favor short-term orientation, focusing on consumerism and personal luck. Concerning indulgence versus restraint, North American and Western European cultures are more inclined towards indulgence, emphasizing freedom and openness, whereas Eastern European, Asian, and Islamic cultures favor restraint, emphasizing self-discipline and social norms.

Through a selective review of previous studies, it has been observed that public attitudes towards AI are indeed influenced by cultural schemas. Scholars have utilized Hofstede's cultural dimensions theory and empirically demonstrated how various cultural characteristics across different dimensions impact people's complex attitudes towards AI [**?**, **?**] and the interactions between humans and AI systems [**?**, **?**]. Chi et al. [**?**] found that the cultural dimensions of uncertainty avoidance, long-term orientation, and power distance play significant roles in hotel customers' willingness to use AI robots. Meanwhile, Lee and Joshi [**?**] identified that uncertainty avoidance and individualism versus collectivism significantly affect user interactions with AI systems. Although other studies have not directly applied Hofstede's cultural dimensions theory to elucidate cultural differences in AI, they have also discovered patterns in AI performance related to cultural differences by directly comparing subjects from various cultures and synthesizing previous literature.

3.1 Cultural Differences in Attitudes Toward AI Between Eastern and Western Cultures

Most of the existing literature indicates that Easterners are more receptive to AI than Westerners [**?**, **?**, **?**]. Research indicates that Chinese people's acceptance level of AI is higher than that of Germans and British, while their level of fear is lower. Yam et al. [**?**] found that Eastern cultures were more inclined to regard robots as part of nature, thus more accepting of AI and robots, whereas Western cultures were more inclined to view them as outsiders. They proposed a theoretical framework comprising historical, religious, and cultural exposure to explain the differences in attitudes towards AI between the East and the West [**?**, **?**]. The historical framework refers to the animistic tradition in the East and the humanistic tradition in the West, which have respectively influenced the public attitudes of these cultures towards robots. The religious framework highlights the emphasis of Eastern Buddhism and Taoism, as well as Western Christianity on the ideological relationships between humans and non-human entities, affecting the differing attitudes towards robots in Eastern and Western cultures. For instance, in Japan, it is often believed that non-human entities possess a soul, influenced by Shintoism. Conversely, Western culture tends to view robots and AI as outsiders, related to Christianity's emphasis on the uniqueness and importance of human beings. The cultural exposure framework suggests that Easterners have more opportunities to interact with AI robots, which helps to reduce their aversion to AI robots. For instance, Japan's long-established robotics industry supports Sindermann et al.'s [**?**] hypothesis that Easterners are more receptive to AI than Westerners.

Overall, Eastern cultures have been observed to exhibit higher acceptance of AI compared to Western cultures. This is partially attributed to the more frequent interactions with and adoption of AI in daily life in Eastern countries such as Japan and China. Additionally, Eastern religious and historical perspectives view non-human entities as integral parts of nature, often attributing spirituality to entities like AI. Unlike Western cultures, Eastern cultures do not strictly distinguish human and non-human entities [**?**, **?**], such as AI. While Western cultures emphasize the uniqueness of human beings, Eastern cultures are inclined to believe that all things possess spirit and soul, thereby more readily accepting the existence of AI without perceiving it as a threat or an outsider.

In addition to differences in acceptance of AI, other studies have explored cultural differences in attitudes towards AI from perspectives such as expectations and emotional responses. Contrary to Yam et al.'s [**?**] animistic theoretical framework, Haring et al. [**?**] found that Japanese participants tended to view robots as machines rather than humanoids, while European participants exhibited no marked preferences. Regarding negative perceptions, Haring et al. [**?**] discovered that Japanese participants were not particularly concerned about the misuse of robots, nor were they as optimistic about the development of robotics as European participants, instead adopting a more cautious stance towards the proliferation and advancement of AI robots. Additionally, concerning expected

tasks, Japanese participants were more likely to expect robots to provide intimate services such as massages, whereas European participants did not express a specific demand for such services. Regarding interaction, Japanese participants preferred robots to complete tasks autonomously, while European participants were more inclined to interact with robots via voice commands. On the positive side, both groups agreed that robots could assist with daily chores; however, Japanese participants particularly emphasized the role of robots in healthcare and assisting individuals with disabilities. Regarding negative views, Japanese participants were more concerned about the misuse of robots, while European participants were more worried about robots replacing human jobs. Expectations for the future revealed that Japanese participants maintained a more cautious outlook towards the proliferation and development of AI robots, whereas European participants were relatively optimistic. Regarding design preferences, both Japanese and European participants favored robots that appear more like machines rather than humanoids.

Concerning emotional responses towards AI, Kim et al. [**?**] found that Korean users valued the pleasure aspect of AI to a greater extent, whereas English-speaking users prioritize practicality. Furthermore, English-speaking users place greater emphasis on the risk dimension compared to their Korean counterparts.

3.2 Cultural Differences in Human-AI Interactions

Cultural backgrounds influence user interactions with artificial intelligence systems (AIS) because cultural values impact users' decisions regarding AIS usage [**?**, **?**]. Researchers have found that users from cultures with high uncertainty avoidance were more likely to rely on AIS, whereas users from individualistic cultures tended to prefer autonomous decision-making. This finding aligns with Hofstede's cultural dimensions theory, which posits that individuals from different cultural backgrounds exhibit varying behaviors when faced with uncertainty [**?**, **?**]. Additionally, users from collectivist cultures prioritize social harmony and group welfare [**?**, **?**, **?**, **?**], and they may favor AIS recommendations that promote social connections and group well-being. Regarding usage patterns, users from collectivist cultures may display different dynamics when interacting with AIS, such as handling contradictory information and considering multiple possibilities in their decision-making. Conversely, users from individualistic cultures may be more inclined to choose between opposing statements and exclude one to reduce cognitive dissonance. Therefore, users from individualistic cultures may more frequently utilize AIS when its recommendations confirm their expected decisions. These findings indicate that cultural dimensions, such as the degree of uncertainty avoidance and the distinction between individualism and collectivism, can lead to differences in how users from various cultures interact with AIS, including decision-making, interdependence on AIS, preferences in AIS recommendations, and usage patterns.

3.3 Cultural Differences in AI Policymaking

In the realm of artificial intelligence policymaking, particularly concerning AI ethics and global governance, significant cultural differences between Eastern and Western cultures are evident [**?**, **?**, **?**]. These cultural disparities pose challenges for international cooperation in AI ethics and governance, particularly in balancing the establishment of global standards with respecting diverse cultural needs [**?**, **?**, **?**]. Wong [**?**] argues that cultural differences may cause some actors to overlook or justify behaviors that violate key ethical values, presenting a significant challenge to the global ethics and governance of AI technology. For instance, some cultures may lack specific ethical values (e.g., privacy) or hold values that conflict with Western perspectives (e.g., favoring macro-level state intervention). Researchers emphasize that although the human rights approach aims to provide a universally applicable and enforceable global framework, it has not sufficiently accounted for cultural diversity, making it challenging to apply directly in non-Western cultural contexts [**?**, **?**]. Therefore, the normative standards for the ethics and global governance of AI technology must consider cultural diversity and should be viewed not as a predetermined endpoint, but as an ongoing process of negotiation and construction.

Existing research underscores that global AI policymakers need to remain open-minded and responsive to cultural values and engage in collaborations with diverse cultures to enrich the normative foundations of global AI ethics and governance [**?**, **?**]. To ensure that global AI policymaking genuinely reflects and respects cultural diversity, ÓhÉigeartaigh et al. [**?**] analyzed the obstacles to international cooperation on AI ethics and governance between Europe, North America, and East Asia, and proposed practical recommendations to promote cross-cultural collaborations, including multilingual translation of key documents, researcher exchange programs, and the development of cross-cultural research agendas. They argue that despite misunderstandings and cultural differences, greater understanding and trust can be fostered through collaborative efforts by governments, industry, and academia, thereby facilitating effective cross-cultural cooperation. They emphasize that international cooperation does not require consensus on ethical principles in all AI domains. Instead, consensus can be sought on practical issues and societal applications. For example, despite differing values on key issues such as data privacy, various cultures and groups can agree on the common goal of protecting individual privacy. This provides a viable pathway for international cooperation and offers crucial insights into how to prevent cultural differences from adversely impacting global AI policymaking.

4 Contributions and Implications

Taking a cultural psychology perspective, our current work synthesized how culturally shared values influence the diverse demands for AI applications and the formulation of ethical standards. By summarizing attitudes, Human-AI interaction, and policy formulations regarding AI across different cultural backgrounds, this article unveils both the human universals and cultural differences among

human cultures. Our work advances both scientific research and practical applications in AI psychology and enhances our understanding of the importance of cultural contexts in AI development and application, thereby promoting the philosophy of building a safer, more regulated, and sustainable development of AI and harmonious coexistence between humans and AI.

Based on a selective review of previous studies and our theoretical formulations, we found that, regardless of cultural backgrounds, the values of safety and universalism are widely prevalent, guiding the application of AI in service industries and environmental sustainability. Additionally, there are common ethical demands for AI regarding privacy, transparency, fairness, justice, and accountability, which further promote the formation and implementation of consensus on AI ethical standards. However, due to differences in historical, religious, and cultural exposure, individuals from different cultural backgrounds still exhibit varying attitudes, behaviors, and policy-making tendencies in the application and regulation of AI. Specifically, Eastern cultures tend to accept the coexistence of AI and humans, maintaining a conservative attitude towards its development. In contrast, Western cultures are more inclined to view AI as oppositional and threatening to humans, emphasizing the realistic threat and symbolic threat it poses to human societies and anticipating its rapid and potentially uncontrollable future development.

In Human-AI interactions, individuals from collectivist cultural backgrounds rely more on AI's judgment and decision-making, considering and reconciling conflicting information simultaneously, whereas individuals from individualist cultural backgrounds prefer autonomous decision-making and tend to choose one direction when faced with conflicting information. Furthermore, there are cultural differences in the principles followed by Eastern and Western cultures in AI policymaking. Western cultures emphasize individual privacy and data transparency, while Eastern cultures prioritize social stability and national security, often supporting government interventions. Based on these findings, this paper proposes reflections and suggestions for future research directions in AI psychology.

5 Limitations and Future Directions

Firstly, existing literature predominantly employs a binary classification of Eastern and Western countries to explore convergent and divergent cultural aspects. However, there is a paucity of studies that conduct more nuanced quantitative measurements and qualitative analyses of cultural systems. Due to the widespread influence of globalization, cultural differences between East and West may be gradually diminishing. Relying solely on the established binary classification may not fully capture the subtle cultural contexts of different countries or regions [**?**, **?**]. Future research should incorporate more nuanced and multi-layered cultural theories and measurements to further explore the convergent and divergent aspects of AI technology applications in various cultural contexts.

Secondly, current research primarily focuses on specific cultural traits. A significant portion of these studies focuses on the influence of collectivist and individualist cultures on the use and application of AI technology, while other cultural traits have not received sufficient attention. For example, preliminary studies indicate that uncertainty avoidance may affect attitudes, usage, and behaviors regarding AI tools across different cultural backgrounds [**?**, **?**]. Future research could investigate the roles of specific cultural traits, such as dialectical thinking and analytical thinking [**?**, **?**], in shaping attitudes and behaviors towards AI technology, its governance, and ethical policies.

Thirdly, there is strong cultural bias in existing research samples. Most studies on cultural similarities or differences are conducted in countries such as the United States, Japan, China, and South Korea. Although these countries are highly representative of Eastern and Western cultures, significant cultural differences exist within Eastern countries (e.g., India) and within Western countries (e.g., the United Kingdom, France, and Germany) [**?**, **?**]. Future research should extend to other countries within Eastern and Western cultures to further explore whether current conclusions can be generalized and applied more broadly.

Finally, current research on the complex influence of culture on AI phenomena has not sufficiently considered other relevant factors. Studies suggest that individual characteristics, such as age, gender, race, education level, social class, and political ideology, also impact human-AI interactions \cite{Mantello et al., 2023; O'Shaughnessy et al., 2022}. For instance, research indicates that well-educated, high-income groups tend to have a more comprehensive understanding of AI tools, utilize AI more effectively, and are less negatively impacted by AI [**?**, **?**]. To further validate conclusions regarding cultural similarities and cultural differences, future research needs to take into account these potential variables to clarify the roles of macro-cultural and micro-cultural traits in AI psychology.

Taken together, our current work synthesized cultural sharing and cultural differences in AI psychology, highlighting the crucial role of cultural background and cultural traits in the acceptance, application, and policymaking of AI. Our current review indicates that while individuals and groups from different cultural backgrounds have common normative expectations regarding AI transparency, fairness, and accountability, significant cultural differences yet exist in AI acceptance, human-AI interaction, and policy orientations. Eastern cultures tend to embrace harmonious coexistence with AI, whereas Western cultures adopt a more cautious attitude toward the potential threats posed by AI. Additionally, cultural background influences user interactions with AI systems, with individuals from collectivist cultures relying more on AI's judgments and decisions, while those from individualist cultures prefer autonomous decision-making. Future research should further explore the role of cultural diversity in AI design, development, and policymaking to ensure the safe, fair, and precise global applications of AI technology. It is also essential to validate the effectiveness and global justice when applying AI large language models across cultures, considering the needs and expectations of users from different cultural backgrounds.

Through cross-cultural cooperation and intellectual exchange, the global governance of AI technology can be promoted, providing more personalized, culturally inclusive, and adaptable products and services to global users from diverse cultural backgrounds. Finally, our current work emphasizes the importance of respecting cultural diversity and cultural differences, fostering cross-cultural communication, and international cooperation in a globalized and multicultural context. The development and applications of AI systems requires not only technological innovations but also a more comprehensive understanding of the cultural diversity of human societies and cultural forms to foster harmonious coexistence between humans and AI, ultimately contributing to the psychological well-being and overall welfare of humanity through AI technology.

References

Agras, S., & Ates, H. (2015). Investigating the predictive abilities of Schwartz's ten universal values on five work performance outcome behaviors. Social and Basic Sciences Research Review, 3(1), 16-31.

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: status quo, challenges and opportunities. Journal of Cleaner Production, 289, 125834. https://doi.org/10.1016/j.jclepro.2021.125834

Akkuş, B., Postmes, T., & Stroebe, K. (2017). Community collectivism: A social dynamic approach to conceptualizing culture. PLoS ONE, 12(9), e0185725. https://doi.org/10.1371/journal.pone.0185725

Alderfer, C. P. (1969). An empirical test of a new theory of human needs. Organizational Behavior and Human Performance, 4(2), 142–175. https://doi.org/10.1016/0030-5073(69)90004-X

Anthropic. (2023, March 8). Core views on AI safety: when, why, what, and how. www.anthropic.com. https://www.anthropic.com/news/core-views-on-ai-safety

Ariza-Montes, A., Arjona-Fuentes, J. M., Han, H., & Law, R. (2017). Employee responsibility and basic human values in the hospitality sector. International Journal of Hospitality Management, 62, 78–87.

Barling, J. (1981). A cross-cultural study of Maslow's motivation theory in industry. South African Journal of Psychology, 11(1), 47–50. https://doi.org/10.1177/008124638101100107

Bhalla, K., Shivakumar, S., & Kumar, T. (2021). Design justice: community-led practices to build the world we need (information policy) by Sasha Costanza-Chock. Design Issues, 37(4), 103–107. https://doi.org/10.1162/desi_r_{00661}

Bonnefon, J. F., Rahwan, I., & Shariff, A. (2024). The Moral Psychology of Artificial Intelligence. Annual Review of Psychology. https://doi.org/10.1146/annurev-psych-030123-113559

Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). Assessing cross-cultural alignment between ChatGPT and human societies: an empirical study. (arXiv:2303.17466). arXiv. https://doi.org/10.48550/arxiv.2303.17466

Chi, O. H., Chi, C. G., Gursoy, D., & Nunkoo, R. (2023). Customers' acceptance of artificially intelligent service robots: the influence of trust and culture. International Journal of Information Management, 70, 102623. https://doi.org/10.1016/j.ijinfomgt.2023.102623

Chin, T., Lin, C. L., Caputo, F., & Hu, F. (2024). Editorial: Understanding cross-cultural differences through cognition and perception analysis: integrating neuroscience and cultural psychology, volume II. Frontiers in Psychology, 15, 1416853. https://doi.org/10.3389/fpsyg.2024.1416853

Dressler, W. W. (2020). Cultural Consensus and Cultural Consonance: Advancing a Cognitive Theory of Culture. Field Methods, 32(4), 398. https://doi.org/10.1177/1525822X20935599

Datler, G., Jagodzinski, W., & Schmidt, P. (2013). Two theories on the test bench: internal and external validity of the theories of Ronald Inglehart and Shalom Schwartz. Social Science Research, 42(3), 906–925.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Springer Science, LLC. https://doi.org/10.1007/978-1-4899-2271-7

Gambrel, P. A., & Cianci, R. (2003). Maslow's hierarchy of needs: does it apply in a collectivist culture? Journal of Applied Management and Entrepreneurship, 8, 143-161.

Ge, X., Xu, C., Misaki, D., Markus, H. R., & Tsai, J. L. (2024, May). How Culture Shapes What People Want From AI. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-15).

Gerlich, M. (2023). Perceptions and acceptance of artificial intelligence: A multi-dimensional study. Social Sciences, 12(9), 502. https://doi.org/10.3390/socsci12090502

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. Academy of Management Annals, 14(2). https://doi.org/10.5465/annals.2018.0057

Goundrey-Smith, S. J. (2019). Technologies that transform digital solutions for optimising medicines use in the NHS. BMJ Health & Care Informatics, 26(1), e100016. https://doi.org/10.1136/bmjhci-2019-100016

Gursoy, D., & Chi, C. G. (2020). Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda. Journal of Hospitality Marketing and Management, 29(5), 527–529. https://doi.org/10.1080/19368623.2020.1788231

Jecker, N. S., & Nagasawa, E. (2022). Bridging East-West differences in ethics guidance for AI and robotics. AI, 3(3), 764-777. https://doi.org/10.3390/ai3030045

Haring, K. S., Mougenot, C., Ono, F., & Watanabe, K. (2014). Cultural differences in perception and attitude towards robots. International Journal of Affective Engineering, 13(3), 149-157.

Heine, S. J. (2007). Culture and motivation. In S. Kitayama & D. Cohen (Eds.), Handbook of cultural psychology (pp. 714-733). Guilford Press.

Hofstede, G. (2011). Dimensionalizing cultures: the Hofstede model in context. Online Readings in Psychology and Culture, 2(1), 8.

Ikkatai, Y., Hartwig, T., Takanashi, N., & Yokoyama, H. M. (2022). Octagon measurement: public attitudes toward AI ethics. International Journal of Human–Computer Interaction, 38(17), 1–18. https://doi.org/10.1080/10447318.2021.2009669

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. (arXiv:1906.1168). arXiv. https://doi.org/10.48550/arxiv.1906.11668

Jordan, M. I., & Mitchell, T. M. (2020). Machine learning: trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kim, M. S., & Kim, E. J. (2013). Humanoid robots as "The Cultural Other": Are we able to love our creations? AI & Society, 28(3), 309-318. https://doi.org/10.1007/s00146-012-0397-z

Kim, J. H., Jung, H. S., Park, M. H., Lee, S. H., Lee, H., Kim, Y., & Nan, D. (2022). Exploring cultural differences of public perception of artificial intelligence via big data approach. International Conference on Human-Computer Interaction, 432. Springer International Publishing.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of culture's consequences: a review of empirical research incorporating Hofstede's cultural values framework. Journal of International Business Studies, 37, 285-320. https://doi.org/10.1057/palgrave.jibs.8400202

Lee, K., & Joshi, K. (2020). Understanding the role of cultural context and user interaction in artificial intelligence-based systems. Journal of Global Information Technology Management, 23(3), 171-175.

Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: establishing a role for culture during human–robot interaction. International Journal of Social Robotics, 13(6), 1307–1333. https://doi.org/10.1007/s12369-020-00710-4

Landeweerd, L., Brey, P., & Rodrigues, R. (2023). Integrating ethics in AI development: A qualitative study. BMC Medical Ethics, 24, Article 10. https://doi.org/10.1186/s12910-023-00879-8

Mantello, P., Ho, M. T., Nguyen, M. H., & Vuong, Q. H. (2023). Bosses without a heart: sociodemographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. AI & Society, 38(1), 97-119.

Maslow, A. H. (1943). A theory of human motivation. Psychological Review, 50(4), 370–396. https://doi.org/10.1037/h0054346

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: challenges, opportunities, and a research agenda. International Journal of Information Management, 53, 102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2023). What governs attitudes toward artificial intelligence adoption and governance? Science and Public Policy, 50(2), 161-176.

ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philosophy & Technology, 33, 571-593.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. American Psychologist, 54(9), 741-754.

Primc, K., Ogorevc, M., Slabe-Erker, R., Bartolj, T., & Murovec, N. (2021). How does Schwartz's theory of human values affect the pro-environmental behavior model? Baltic Journal of Management, 16(2), 276–297.

Russo, C., Danioni, F., Zagrean, I., & Barni, D. (2022). Changing personal values through value-manipulation tasks: a systematic literature review based on Schwartz's theory of basic human values. European Journal of Investigation in Health, Psychology and Education, 12(7), 692–715.

Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology, 25, 1-65. Elsevier Science & Technology. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. Journal of Cross-Cultural Psychology, 32(5), 519-542. https://doi.org/10.1177/0022022101032005001

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. Journal of Personality and Social Psychology, 103(4), 663–688. https://doi.org/10.1037/a0029393

Shweder, R. A. (2021). Cultural psychology: What is it? In D. Cohen & S. Kitayama (Eds.), Handbook of cultural psychology (2nd ed., pp. 23-47). Guilford Press.

Sindermann, C., Yang, H., Elhai, J. D., Yang, S., Quan, L., Li, M., & Montag, C. (2022). Acceptance and fear of artificial intelligence: associations with personality in a German and a Chinese sample. Discover Psychology, 2(1), 8.

Steers, R. M., & Sánchez-Runde, C. J. (2017). Culture, motivation, and work behavior. In T. M. D. L. T. R. J. R. J. L. C. T. M. J. Gelfand, M. Y. J. L. (Eds.), The Blackwell handbook of cross-cultural management (pp. 279-307). Wiley. https://doi.org/10.1002/9781405164030.ch10

Vannucci, F., Sciutti, A., Lehman, H., Sandini, G., Nagai, Y., & Rea, F. (2019). Cultural differences in speed adaptation in human-robot interaction tasks. Journal of Behavioral Robotics, 10(1), 256–266. https://doi.org/10.1515/pjbr-2019-0022

Vasalou, A., Joinson, A. N., & Courvoisier, D. (2010). Cultural differences, experience with social networks and the nature of "true commitment" in Facebook. International Journal of Human-Computer Studies, 68(10), 719-728. https://doi.org/10.1016/j.ijhcs.2010.06.002

Vassallo, M., & Saba, A. (2015). Does money for grocery expenditure sway Italian consumers' motivational values in predicting attitude towards eco-sustainable products? Contemporary Management Research, 11(1), 3–24.

Wong, P. H. (2020). Cultural differences as excuses? Human rights and cultural values in global ethics and governance of AI. Philosophy & Technology, 33(4), 705-715.

World Economic Forum. (2024). Generative AI governance: Ensuring diversity and fairness.

Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., & Gray, K. (2023). Cultural differences in people's reactions and applications of robots, algorithms, and artificial intelligence. Management and Organization Review, 19(5), 859-875.

Yrjö Engeström. (2019). Learning by expanding: an activity-theoretical approach to developmental research. Cambridge University Press.

Zeng, Z., Chen, P.-J., & Lew, A. A. (2020). From high-touch to high-tech: COVID-19 drives robotics adoption. Tourism Geographies, 22(3), 734. https://doi.org/10.1080/14616688.2020.1762118

Note: Figure translations are in progress. See original paper for figures.