

Systematizing AI Governance through the Lens of Ken Wilber's Integral Theory

Ammar Younas^{1,2,3,4,*} and Yi Zeng^{1,2,3,4,*}

¹ Institute of Philosophy, Chinese Academy of Sciences, Beijing, China

² School of Humanities, University of Chinese Academy of Sciences, Beijing, China

³ Institute of Automation, Chinese Academy of Sciences, Beijing, China

⁴ Center for Long-term Artificial Intelligence, Beijing, China

Emails: doctorammaryounas@mails.ucas.ac.cn, yi.zeng@ia.ac.cn

Abstract

We apply Ken Wilber's Integral Theory to AI governance, demonstrating its ability to systematize diverse approaches in the current multifaceted AI governance landscape. By analyzing ethical considerations, technological standards, cultural narratives, and regulatory frameworks through Integral Theory's four quadrants, we offer a comprehensive perspective on governance needs. This approach aligns AI governance with human values, psychological well-being, cultural norms, and robust regulatory standards. Integral Theory's emphasis on interconnected individual and collective experiences addresses the deeper aspects of AI-related issues. Additionally, we propose using Integral Theory as a methodology for literature reviews to overcome the fragmented understanding often seen in traditional reviews of AI governance.

Keywords: Integral Theory, AI Governance, Ethical Considerations, Technological Standards, Cultural Narratives, Regulatory Frameworks, Systemic Approaches

Introduction

Ken Wilber proposed his Integral Theory¹ as a framework for understanding the relationship between the individual and the collective. It is based on the idea that all aspects of existence, including the physical, psychological, social, and spiritual, are interrelated and interconnected. Integral Theory is a comprehensive framework that seeks to synthesize various disciplines, theories, and perspectives into a cohesive understanding of human experience and reality. At its core, Integral Theory integrates insights from psychology, philosophy, science, religion, and other fields to provide a holistic approach to understanding complex phenomena.

We think that Integral Theory has significance because it provides an approach that accommodates the complexity and diversity of human experiences and perspectives. It promotes a holistic view, recognizing that no single perspective can fully capture the richness of reality and helps to create a more complete picture. This framework bridges the gap between different disciplines, encouraging dialogue and integration across fields such as science, art, religion, and psychology, fostering innovation and deeper insights.

The application of Integral Theory extends to addressing complex, multifaceted issues such as environmental sustainability, social justice, and mental health². It encourages solutions that

consider multiple dimensions and perspectives, leading to more effective and sustainable outcomes. The theory's emphasis on inclusivity and integration values both subjective and objective realities, ensuring that all voices and aspects of human experience are acknowledged and respected. By highlighting the importance of ongoing growth and evolution, both personally and collectively, Integral Theory encourages a dynamic and adaptive approach to life.

We aim to investigate AI Governance through the prism of Ken Wilber's Integral Theory. Ken Wilber's Integral Theory offers a powerful framework for understanding the complexities of human experience and the world. It provides a holistic, inclusive, and integrative approach that is invaluable for those seeking to grasp the full spectrum of reality and address the multifaceted challenges of contemporary life. There are several compelling reasons to use this integral theory framework. Firstly, the current AI governance landscape necessitates a holistic approach and understanding. AI technologies are not only advancing at a rapid pace but also permeating almost every aspect of human life, from economic systems and healthcare to social interactions and ethical considerations. A holistic framework like Integral Theory allows us to address this complexity by integrating multiple perspectives and dimensions. By considering the Integral Theory as a framework, we can create a comprehensive understanding of how AI impacts individuals, cultures, societies, and systemic structures. This approach ensures that we do not overlook any critical aspect, providing a more balanced and inclusive governance framework.

Secondly, highlighting the importance of a systemic approach to current issues, especially those with global impact, is crucial. Many AI-related challenges are not merely regulatory but have deep roots in human consciousness and culture. Issues such as ethical decision-making, bias in AI algorithms, and the societal implications of autonomous systems require a nuanced understanding of human values, beliefs, and social dynamics. Integral Theory's emphasis on the interconnectedness of individual and collective experiences allows us to address these deeper, often overlooked aspects. By considering both the subjective and objective dimensions of AI governance, we can develop policies and frameworks that are not only technically sound but also culturally sensitive and ethically robust.

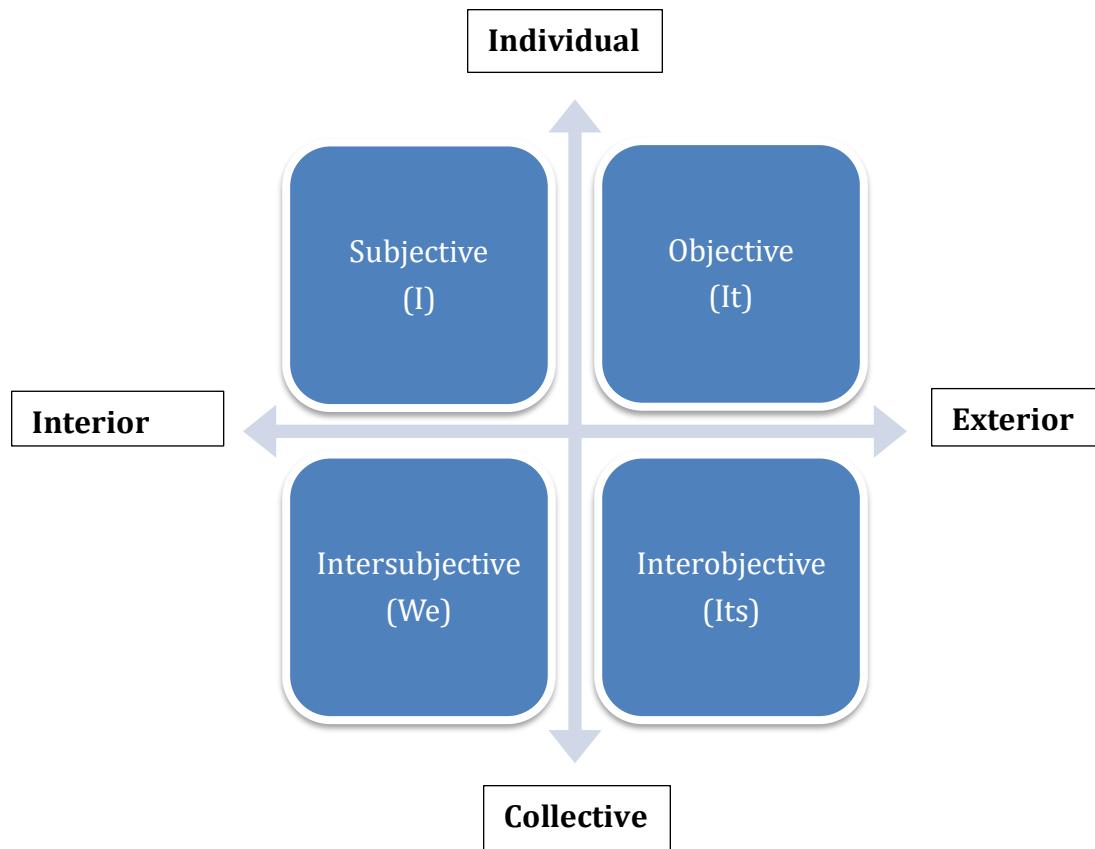
Finally, we propose to extend the knowledge of literature review by introducing Ken Wilber's Integral Theory as a methodology. Traditional literature reviews often focus on specific disciplines or viewpoints, which can lead to a fragmented understanding of AI governance. By employing Integral Theory, we can systematically examine the literature across different quadrants and developmental lines. This methodology enables us to integrate insights from diverse fields such as ethics, technology, law, psychology, and sociology. By mapping out the developmental stages and states of consciousness related to AI, we can identify gaps in the existing literature and highlight areas where interdisciplinary research is needed. This comprehensive review process not only enriches our understanding but also paves the way for more innovative and effective governance strategies.

Ken Wilber's Integral Theory

Ken Wilber's Integral Theory is a comprehensive framework that seeks to synthesize various disciplines, theories, and perspectives into a cohesive understanding of human experience and reality. At its core, Integral Theory integrates insights from psychology, philosophy, science,

religion, and other fields to provide a holistic approach to understanding complex phenomena. The theory is built around the idea that reality can be understood through multiple lenses, each offering a unique perspective that contributes to a fuller picture.

A central component of Integral Theory is the concept of the Four Quadrants, which divides reality into four distinct but interconnected dimensions^{1,3}. The Interior-Individual quadrant represents subjective experiences, including thoughts, emotions, and personal consciousness, focusing on the inner world of individuals. The Exterior-Individual quadrant covers objective, observable aspects of individuals, such as behaviors, physical health, and measurable actions. The Interior-Collective quadrant encompasses shared cultural values, worldviews, and social norms, examining how groups of people collectively make sense of the world. The Exterior-Collective quadrant involves the systems, structures, and environments that influence and are influenced by groups of people, including societal institutions, economic systems, and ecological systems. These quadrants provide a comprehensive map for exploring the complexity of human life and interactions.



Integral Theory also incorporates developmental levels, or stages, which identify various phases of human development both individually and collectively. These stages range from basic survival and egocentric stages to more advanced, integrative stages that encompass broader perspectives and deeper understandings. By mapping out these stages, Integral Theory provides a roadmap for personal and collective evolution, helping individuals and societies understand where they are and how they can progress to higher levels of consciousness and complexity. Additionally, the theory identifies developmental lines, which are specific areas of growth such as cognitive, emotional,

moral, and spiritual lines, each progressing through different stages reflecting growth and complexity in that area.

Another critical element of Integral Theory is the consideration of different states of consciousness, such as waking, dreaming, and meditative states. These states offer various ways of experiencing reality and are integral to understanding the full spectrum of human experience. By acknowledging and incorporating these states, Integral Theory captures the diverse ways in which people perceive and interact with the world. Furthermore, the theory recognizes the importance of types, referring to different categories that people can belong to, such as personality types and gender types. These types help in understanding the diversity and individuality within broader developmental stages and lines.

The Four Quadrants of Integral Theory

Wilber's Integral Theory is built around four quadrants, which are:

Upper-Left (UL) Quadrant: Individual-Interior (Subjective)

Wilber's Integral Theory encompasses four quadrants, with the Upper-Left (UL) quadrant focusing on the inner experiences, consciousness, and psychological aspects of individuals. This quadrant includes thoughts, emotions, beliefs, and values, representing the individual's subjective component of consciousness, often described using first-person "I"-language⁴. The UL quadrant emphasizes personal experiences and the inner voice, reflecting an individual's thoughts, sentiments, beliefs, and consciousness. It evaluates an individual's deeper understanding and thought processes, particularly regarding ethics, beliefs, and decision-making.

Ethical awareness and personal values play a crucial role in this quadrant, highlighting the decision-making process through individual subjectivity⁵. Other models also emphasize that emotional reactions and conscious thought contribute to moral judgment⁶. Researchers have explored the relationship between ethical consciousness and individual actions in society, noting the significant impact of emotions on human behavior, including feelings of depression, isolation, and reduced empathy^{7,8}. The uncertainty and stress individuals face suggest that reducing psychological pressure can aid in decision-making^{9,10}.

Personal experiences and values influence decision-making, reinforcing Wilber's view that subjectivity involves collaborative judgment by consciousness and action¹¹. It is important to mitigate bias and promote inclusive policies through multiple perspectives¹². The UL quadrant provides a dynamic framework for understanding individual subjectivity, emphasizing the interplay between personal experiences, emotions, and ethical considerations in shaping human behavior and decision-making.

Upper-Right (UR) Quadrant: Individual-Exterior (Objective)

This quadrant represents the external, observable behaviors and physical aspects of individuals, encompassing brain functions, technological interactions, and measurable actions. The "it" language used in the third person refers to material and objective explanations of phenomena¹³. This aspect of Wilber's theory provides perspectives on the interior-exterior dimension, focusing

on observable behavior, an individual's actions, and the psychological features of behavior and performance.

Evaluating performance to ensure the reliability and effectiveness of brain responses involves various assessments of recall and precision¹⁴. Techniques for optimizing performance and efficiency through reinforcing language highlight the innovation of the objective aspect¹⁵.

Distinctions that separate therapists and researchers, such as those between interior and exterior, qualitative and quantitative, individual and collective, may also act as bridges between them. While phenomenology might not disclose the neurological basis of experience, strict empirical approaches will never reveal the characteristics of lived experience or what makes life worthwhile. A metatheoretical scaffolding those respects various perspectives and epistemologies, acknowledges their limitations, and offers a methodical way to arrange them can help different approaches complement each other¹⁶.

These characteristics are unavoidable and irreducible through reconstructive science; together, they constitute the "Big Three," a kind of foundation-less framework¹. Habermas explains that the objective world correlates with all true assertoric sentences, while normative rightness and subjective truthfulness relate to legitimately regulated interpersonal relationships and subjective experiences¹.

The third-person pronoun "it" refers to the objective, observable exterior form of any occurrence, associated with the UR quadrant¹³. The UR quadrant pertains to job explanations, salaries, and individual duties in organizational contexts¹⁷. When it comes to people, behaviorism, defined by bodily acts and physiological changes, characterizes the UR quadrant. For instance, thoughts might appear externally as smiles, tears, or blushes. Additionally, changes in hormone levels (such as testosterone and dopamine) and brain activity (such as beta-brain activity) can be evaluated objectively.

Lower-Left (LL) Quadrant: Collective-Interior (Inter-subjective)

This quadrant deals with the shared values, culture, and collective consciousness of groups, including social norms, ethics, worldviews, and cultural narratives. The language expressing common significance and collective interaction is found in this quadrant, often using the word "we" (second person)⁴.

The way science is taught in high school reflects an antiquated positivistic understanding of the subject¹⁸. Much research in scientific education has historically been conducted in Wilber's model's right quadrants. In the scientific education sector, which often promotes the idea of value-free science, the tenets of Descartes' dualistic view of the universe and Newton's mechanical worldview remain prevalent¹⁹.

The LL quadrant represents the inter-subjective knowledge of collective culture, encompassing shared worldviews, values, emotions, and language¹³. These ideas can only be understood and interpreted within their specific contexts. The usage of the second-person pronoun "we" indicates context-sensitive communication of shared significance and communal interpretation²⁰. As noted,

"A kiss can mean 'I love you' or 'hello,' but the full meaning of the action can only be understood by interpreting the words, movements, looks, and sounds"¹⁷. This highlights the tension between perception and context specificity.

In a study, students were asked to explain their "understanding of unspoken and written practices, organizational goals and shared values, ethics and morals within the company and how it relates to their personal development" (Jowdy et al., 2004, p. 228). This reflects the intersection of the UL and LL quadrants, representing the left-hand path of the AQAL framework. Wilber argues that "the interior thought itself... only makes sense in terms of my cultural background"¹.

Students participated in an in-class learning exercise covering topics such as people orientation (the degree to which management decisions consider employee impact), team orientation (the extent to which work is organized around teams versus individuals), aggressiveness, and outcome orientation (the extent to which management prioritizes product outcomes over processes employed). They used components for identifying and changing an organization's culture as a guide²¹.

Lower-Right (LR) Quadrant: Collective-Exterior (Inter-objective)

This quadrant focuses on the systemic and structural aspects of societies, including institutions, laws, economic systems, and technological infrastructures. It represents the collective's external shape in third-person "its" language (Landrum and Gardner, 2005; Wilber, 2006). In the LR quadrant, institutions, technologies, and production forces interact to form a functional system of objective social action.

Organizations have predictable and controllable systems, financial systems, and information systems (Cacioppe, 2000). All four quadrants, representing the inside and outside of both the individual and the collective, must be included for an integral perspective. No quadrant is superior; all are essential for understanding human behavior and solving issues (Wilber, 2006).

Interns should report and record observed behaviors, issues (e.g., power, leadership, communication), and how these fit with the formal systems within an organization. They assess observed behaviors to determine their suitability to the given situation or problem, involving all relevant individuals. They also evaluate the situation's nature, defining the issue, interpreting it, and considering potential solutions²².

The LR quadrant highlights the importance of understanding the systemic and structural dimensions of society. By examining how institutions and technologies interact within this quadrant, we gain insight into the collective mechanisms that shape societal behavior. This understanding is crucial for developing effective organizational systems and addressing complex social issues. Integrating observations from internships and real-world experiences further enriches this perspective, providing practical insights into the alignment between individual behaviors and formal organizational structures. This comprehensive approach ensures a balanced and integral understanding of human and collective dynamics, facilitating more effective solutions to societal challenges.

Integrating Wilber's AQAL Framework Across Various Disciplines

In 1995, Wilber introduced the AQAL (All Quadrants All Levels) framework, expanding his theory to include a fourth quadrant that covers individual-collective and interior-exterior dimensions¹. This framework incorporates theories explaining mental and spiritual development, the evolution of collective awareness, and the structure of the brain and society into layers or holons. Many scholars believe that the integral theory aims to be a metatheory connecting all academic disciplines, knowledge categories, and life experiences.

Ken Wilber's theory extends beyond integral psychology into business and other fields. In an interview, Wilber reflected on the humility he gained from three years of part-time dishwashing work before writing his first book. He described this experience as an "extraordinary education" in grounding and engaging with the world in a concrete, tangible way, which balanced his Zen meditation practice²³. Wilber continues to emphasize a well-rounded approach to personal development, involving physical exercise, intellectual work, social interaction, and meditation²⁴. He advocates for "Integral Transformative Practice," which combines practical, everyday activities with transformative practices²⁵.

Wilber's philosophy promotes a balanced education that values practical, humble experiences alongside academic and intellectual pursuits. He famously advised, "Forget the degrees, forget the books and articles, forget the titles, forget everything really, and wash dishes for two years," emphasizing the importance of grounding oneself in everyday reality²³. This approach aligns with Deweyian pragmatism, advocating for basic education as a fundamental service essential for healthy, comprehensive human development.

Wilber's AQAL framework posits that no single perspective is ontologically superior; rather, all perspectives should be considered simultaneously. This principle of "simultaneity" asserts that perceptions construct reality¹³. By applying this holistic approach, we can integrate insights from multiple disciplines to address complex issues more effectively. For instance, in business, this means combining strategic thinking with empathetic leadership and ethical considerations. In education, it involves blending theoretical knowledge with practical experiences and emotional intelligence. In healthcare, it requires balancing clinical expertise with patient-centered care and mental health support.

The AQAL framework's versatility makes it applicable across various disciplines, fostering a more interconnected and comprehensive understanding of human development and societal progress. By embracing Wilber's integral approach, we can create more holistic and effective solutions to the challenges faced in different fields. For example, in AI governance, the framework can help balance technological advancement with ethical considerations and societal impacts, ensuring that AI development aligns with human values and serves the greater good.

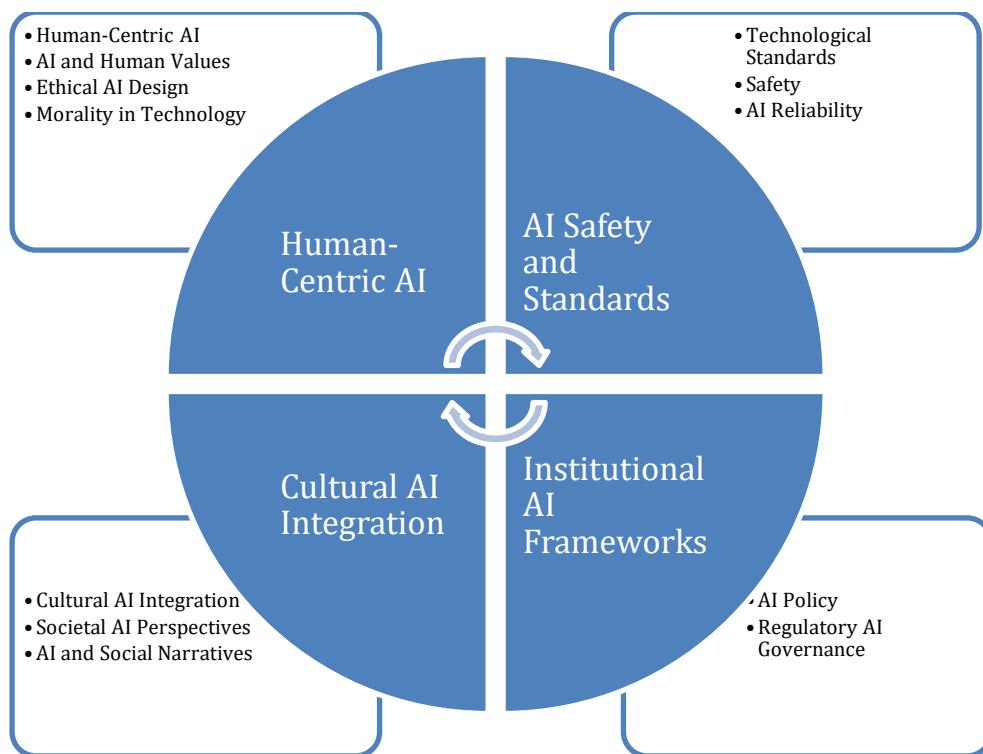
Applying Ken Wilber's Integral Theory to AI Governance

The current landscape of AI governance is becoming increasingly complex due to the multitude of factors that need regulation, including ethical considerations, technological standards, cultural narratives, and regulatory frameworks. Ken Wilber's concepts of holarchy and hierarchy provide

a valuable framework for navigating this complexity. Wilber makes two crucial, interrelated points using these concepts. Firstly, he notes that as systems become more complex, they undergo differentiation through increased self-organization. Simultaneously, the preceding subsystem is integrated into the new, more complex system. This natural process of differentiation and integration ensures the coherent evolution of systems¹.

Wilber distinguishes between natural holarchies and pathological hierarchies. Pathological hierarchies, often criticized by postmodernists, arise when this natural process is disrupted. In such cases, a particular part, or holon, attempts to become an independent whole, neglecting its role within the larger context. This can threaten the overall health and functionality of the system. Applying Wilber's framework to AI governance, we can view the various regulatory aspects as interdependent parts of a larger, evolving system. Ethical considerations (Upper-Left quadrant), technological standards (Upper-Right quadrant), cultural narratives (Lower-Left quadrant), and regulatory frameworks (Lower-Right quadrant) must all evolve in a coordinated manner. Disruption in any one area, akin to a holon trying to function independently, can lead to a pathological hierarchy, undermining the effectiveness of AI governance as a whole.

By understanding AI governance through the lens of holarchy, we can ensure that as new regulations and standards are developed, they are integrated into a cohesive system that respects the interdependence of its parts. This holistic approach helps prevent the emergence of pathological hierarchies and supports the creation of a robust, adaptive, and ethical AI governance framework.



Ethical Considerations and Human Values: The Upper-Left Quadrant

According to Ken Wilber's Integral Theory, the Upper-Left (UL) quadrant focuses on the inner experiences, consciousness, and psychological aspects of individuals. This includes thoughts, emotions, beliefs, and values. When applied to AI governance, this quadrant highlights the importance of addressing the ethical implications of AI and ensuring its alignment with human values. It involves understanding the psychological impact of AI on individuals, such as privacy concerns, mental health effects, and the influence on personal autonomy. Therefore, governance frameworks should be developed to ensure that AI respects human dignity and promotes psychological well-being.

The supreme dimension of the subjective in the four quadrants is ethical consideration in AI governance. Autonomy and human agency are critical, and individual experiences must be respected. Ensuring balanced control of decision-making rather than replacing critical choices is essential for maintaining ethical standards²⁶. Obtaining consent in AI applications presents significant difficulties and challenges^{27,28}. Understanding the ramifications of AI systems requires addressing data privacy issues due to the large volumes of personal information involved²⁸. AI transparency is responsible for creating subjectivity within AI systems²⁷. Protecting data privacy mechanisms is crucial for safeguarding human autonomy in the context of AI²⁹.

AI has the potential to reinforce and produce unjust results, even though it cannot make decisions independently³⁰. To produce fair results and reduce prejudice, incorporating a framework is essential^{30,31}. Fairness and comprehensive strategies must be applied in varied social settings³¹. Accountability and responsibility are important for producing valid results that respect human autonomy³¹. Ethical accountability of human choices and AI system behavior ensures responsible ethical implications for using human individuality through technologies³².

Human-centered design in technology should prioritize human accessibility and satisfaction³³. The emotional and psychological aspects of human values regarding AI technology are important for ethical design and the need for empathy. Human values can be governed by AI technologies, which can assist in promoting human-centered developments³⁴.

Technological Standards and Safety: The Upper-Right Quadrant

The Upper-Right (UR) quadrant in Ken Wilber's Integral Theory represents the external, observable behaviors and physical aspects of individuals, encompassing brain functions, technological interactions, and measurable actions. In the context of AI governance, the UR quadrant emphasizes the importance of establishing technical standards and ensuring the safety of AI systems. This involves rigorous testing, validation, and continuous monitoring of AI technologies to prevent malfunctions, biases, and unintended consequences. Regulations should mandate transparency and accountability in AI development and deployment.

Wilber's focus on the technological standards in the UR quadrant underscores the necessity for the reliable, safe, and ethical operation of AI technology. This aspect highlights observable, measurable features of AI technologies, aligning with AI Ethics Principles. Globally, over 100 AI ethics principles have been developed, reflecting the diverse and multifaceted nature of AI ethics.

These principles cater to various specific contexts, including regional, country-specific, sector-specific, and industry-specific guidelines³⁵. Strong technical standards are essential for mitigating risks and preventing unexpected outcomes that could lead to system failures. Ensuring the safety and reliability of AI systems involves evaluating established safety protocols and implementing risk mitigation strategies.

The complexity of AI systems demands proactive governance measures. Key AI safety issues, such as distributional shifts, reward hacking, and safe exploration, have been identified by many scholars³⁶. Addressing these challenges requires rigorous testing and validation under specific conditions to ensure safe and predictable operations. Cybersecurity is another crucial component of AI security. Brundage et al. (2018) explore potential malicious applications of AI that could compromise system stability and user security, such as adversarial assaults. To mitigate security risks and maintain system integrity, comprehensive cybersecurity measures must be developed.

Looking AI governance through the prism of Integral Theory, highlights the critical importance of establishing robust technical standards and ensuring the safety and reliability of AI systems. By focusing on observable and measurable aspects of AI technologies, we can address ethical principles and mitigate potential risks. The development of comprehensive governance frameworks, proactive safety measures, and stringent cybersecurity protocols is essential to safeguard human values and promote the ethical use of AI. Through these efforts, we can create AI systems that not only advance technology but also respect and enhance human dignity and well-being.

Cultural and Social Narratives: The Lower-Left Quadrant

In the Lower-Left (LL) quadrant of Ken Wilber's Integral Theory, cultural and social narratives are crucial in AI governance for shaping public perception and fostering acceptance. This involves inclusive dialogue that goes beyond risk management and safety concerns, addressing misconceptions to promote a cultural understanding of AI technology and its positive societal role. Educational institutions play a vital role in developing a positive mindset and creating a collective and responsible AI generation.

This aspect of Wilber's theory represents the inter-subjective dimension of society's collective interior thoughts regarding AI technology. It emphasizes not only the technological importance of AI but also the cultural and social narratives that influence social perception. These narratives positively affect the integration of social beings with AI systems.

Social settings and cultural values significantly influence how people respond to AI technologies³⁷. For example, Japan's strong cultural narrative surrounding technological development leads to greater acceptance of robotic technologies compared to more skeptical viewpoints elsewhere. Understanding these cultural narratives is essential for ensuring that AI systems align with social norms and values³⁸. By integrating and interpreting these narratives, policymakers can create relevant and necessary policies that reflect the interests and concerns of diverse populations.

Social narratives play a pivotal role in shaping social values, ethical standards, and AI governance. Often, AI-generated discourses instill societal fears about job displacement, privacy violations,

misuse of autonomy, and other concerns. Narratives about surveillance and privacy control, especially in low-income communities, highlight the challenges posed by AI in sensitive social contexts¹⁰. These narratives demand that AI systems be designed and implemented with an understanding of their social impact.

Media portrayals of concepts like 'AI for Social Good' serve as counter-narratives to dystopian views, emphasizing AI's positive potential. Fostering a positive societal impact of AI is challenging, particularly in areas like environmental sustainability, financial systems, healthcare, scientific advancement, and education³⁹. Social narratives help integrate AI governance into society, enhancing the acceptance and reliability of AI technologies as morally and socially responsible tools.

The Lower-Left (LL) quadrant of Ken Wilber's Integral Theory underscores the significance of cultural and social narratives in AI governance. By fostering inclusive dialogue and addressing misconceptions, we can promote a cultural understanding of AI technology and its societal benefits. Educational institutions are key to developing a responsible AI generation. Understanding and integrating cultural narratives into policymaking ensures that AI systems align with social norms and values, thereby enhancing their acceptance and reliability. Addressing societal fears and highlighting AI's positive potential can help create a balanced view of AI, fostering its ethical and responsible use in various sectors. Through these efforts, AI technologies can become integral, accepted, and beneficial components of society.

Institutional and Regulatory Frameworks: The Lower-Right Quadrant

The Lower-Right (LR) quadrant in Ken Wilber's Integral Theory focuses on establishing robust institutional and regulatory frameworks to govern AI effectively. This includes developing comprehensive laws, policies, and standards that guide the ethical use of AI. Achieving authentic acceptance of these regulations requires international cooperation and harmony among global stakeholders. Wilber's framework emphasizes inter-objective coordination for the acceptance and governance of AI technology, underscoring the importance of regulatory frameworks in shaping collective experiences. These inclusive laws are essential for creating transparency and safeguarding human autonomy when using AI technology.

A prominent example is the General Data Protection Regulation (GDPR) established by the European Union, which ensures reliability and safety for human autonomy and individual privacy. The GDPR sets strict criteria that enhance reliability and transparency in AI settings⁴⁰. By mandating data protection principles and granting individuals control over their personal data, GDPR has become a cornerstone in AI governance. Another significant regulatory initiative is the proposed Artificial Intelligence Act (AIA) by the European Union. This act aims to ensure the legitimacy of AI systems and safeguard human rights and privacy by classifying AI applications based on their risk levels and imposing stringent requirements for high-risk AI systems⁴¹. These regulations serve as authentic guidelines for transparency in AI systems.

Institutions play a crucial role in developing regulatory frameworks for AI technology. National and international organizations, such as the Organization for Economic Co-operation and Development (OECD), have set AI governance principles to promote inclusivity, human

development, and oversight. The OECD AI Principles, adopted in 2019, emphasize AI's ethical, transparent, and accountable development, ensuring human-centered values in AI deployment⁴². The National Institute of Standards and Technology (NIST) in the USA has also contributed significantly to setting regulations for the accuracy and transparency of AI technology. NIST's AI Risk Management Framework, introduced in 2022, provides guidelines for assessing and mitigating risks associated with AI systems⁴³. This framework helps organizations implement trustworthy AI practices.

Similarly, initiatives like The Global Partnership on AI (GPAI) promote transparency and the development of human autonomy through institutional practices. GPAI facilitates international collaboration and knowledge-sharing to address AI challenges and opportunities, emphasizing ethical AI development⁴⁴. UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in 2021, provides a comprehensive ethical framework for AI that includes principles such as respect for human dignity, human rights, and environmental sustainability⁴⁵. This recommendation aims to guide countries in developing AI policies that align with ethical standards.

These platforms and frameworks foster robust knowledge-sharing and cooperation for effective AI governance, ensuring that AI technologies are developed and deployed responsibly and ethically. Through such international cooperation and harmonized regulatory efforts, we can achieve a balanced and trustworthy AI landscape that respects human values and promotes societal well-being.

Integrating the Quadrants for a Holistic AI Governance Framework

In the current AI governance scenario, where diverse perspectives abound, a systemic approach is essential. A step-by-step framework that makes sense of these complexities is needed. Integral Theory, often described as a "theory of everything," offers a comprehensive lens through which to view AI governance. While its legitimacy can be debated, its pragmatic importance for our purpose cannot be overlooked. AI governance, approached holistically, highlights the significance of regulatory aspects from the four quadrants of Integral Theory. This governance framework not only addresses psychological aspects but also incorporates cultural norms. Such a comprehensive framework is crucial for making institutions reliable and future ready. By embracing the quadrant theory perspective, we can ensure an inclusive and comprehensive approach to AI technology use.

Wilber's theory is noteworthy as it represents the interrelated reality of human psychology and institutional systems. The theory comprises four quadrants: Upper-Left (UL), Upper-Right (UR), Lower-Left (LL), and Lower-Right (LR), which together provide a comprehensive approach to AI governance.

The UL quadrant focuses on the subjective part of individuality, emphasizing human autonomy, beliefs, ethics, and psychology in collaboration with AI technology. This aspect highlights the need for ethical AI designs that safeguard privacy and promote psychological well-being³⁸. The UR quadrant represents the objective, individual-exterior dimension, and emphasizes the necessity

of rigorous safety protocols and a robust, accountable AI system. This ensures that technologies are safe, reliable, and ethically aligned with human values and autonomy³⁶.

The LL quadrant focuses on intersubjective, collective dimensions, emphasizing the importance of social and cultural narratives. This aspect represents public perception and acceptance of AI technology, highlighting the social demands for culturally responsible and acceptable AI systems^{10,30}. The LR quadrant addresses the institutional perspective of regulation, ensuring AI authenticity and safety. Guidelines provided by the OECD and the European Commission emphasize the importance of transparent regulations that align with cultural and social narratives.

By combining these individual and collective dimensions, we generate a holistic AI governance framework that encompasses ethics, culture, social narratives, and institutional regulations and safety. This synthesis makes AI governance more reliable, comprehensive, and authentic, while also adaptable to evolving AI technology with institutional and social impacts.

Investigating AI governance through Ken Wilber's Integral Theory provides a holistic, interdisciplinary, and systemic approach to understanding and addressing the multifaceted challenges posed by AI technologies. This framework enables the integration of diverse perspectives, the conduct of comprehensive literature reviews, and the development of governance strategies that are inclusive and adaptive to the complex realities of our globalized world.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT in order to improve the readability and language of the work. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

References

1. Wilber, K. (1995). *Sex, Ecology, Spirituality : The Spirit of Evolution* (Boston : Shambhala ; [New York, N.Y.]).

2. Duffy, J.D. (2020). A Primer on Integral Theory and Its Application to Mental Health Care. *Glob Adv Health Med* 9, 2164956120952733.
<https://doi.org/10.1177/2164956120952733>.
3. Wilber, K., and Hill, R. (2013). AN INTEGRAL THEORY OF CONSCIOUSNESS.
4. Wilber, K. (2001). A Theory of Everything—An Integral Vision for Business, Politics, Science, and Spirituality.
5. Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., Wang, W.W., and Witteborn, S. (2019). Artificial Intelligence, Governance and Ethics: Global Perspectives. Preprint, <https://doi.org/10.2139/ssrn.3414805>
<https://doi.org/10.2139/ssrn.3414805>.
6. Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 108, 814–834.
<https://doi.org/10.1037/0033-295x.108.4.814>.
7. Borenstein, J., and Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *Ai and Ethics* 1, 61. <https://doi.org/10.1007/s43681-020-00002-7>.
8. Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other* (Basic Books, Inc.).
9. Gordon, F. (2019). Virginia Eubanks (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: Picador, St Martin’s Press. Law, Technology and Humans, 162–164.
<https://doi.org/10.5204/lthj.v1i0.1386>.
10. Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. (New York: St. Martin’s Press).
11. Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology* 62, 451–482. <https://doi.org/10.1146/annurev-psych-120709-145346>.
12. Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (PMLR), pp. 149–159.
13. Wilber, K. (2007). Integral Spirituality.

14. Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45, 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>.
15. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks* 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
16. Marquis, A. (2018). Integral Psychotherapy: A Unifying Approach 1st ed.
17. Cacioppe, R. (2000). Creating spirit at work: re-visioning organization development and leadership – Part II. *Leadership & Organization Development Journal* 21, 110–119. <https://doi.org/10.1108/01437730010318200>.
18. Russell, S., and Norvig, P. (2010). Artificial Intelligence: A Modern Approach 4th ed.
19. Settelmaier, E., and Taylor, P. (2001). Ken Wilber's integral philosophy and educational research: Fleshing out the “seventh moment” (and beyond?). In AARE 2001 International Education Research Conference. Crossing Borders: New Frontiers for Educational Research.
20. Landrum, N., and Gardner, C. (2005). Using integral theory to effect strategic change. *Journal of Organizational Change Management* 18, 247–258. <https://doi.org/10.1108/09534810510599407>.
21. Hunsaker, P.L. (2004). Management : a skills approach (Upper Saddle River, N.J. : Pearson Prentice Hall).
22. McDonald, M., Jowdy, E., and Spence, K. (2004). An Integral Approach to Sport Management Internships. *European Sport Management Quarterly* 4.
23. Visser, F. (2003). Ken Wilber: thought as passion (State University of New York Press).
24. Wilber, K. (1997). An integral theory of consciousness - PhilPapers. *Journal of Consciousness Studies*, 71–92.
25. Allwyn, F., and Amalaveenus, D.J. (2022). Understanding the Struggle between Self Consciousness and Cultural Worldview in the Sense of An Ending: Ken Wilber's Perspective. *Journal of Language and Linguistic Studies* 17.
26. Wachter, S., Mittelstadt, B., and Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection

- Regulation. *International Data Privacy Law* 7, 76–99.
<https://doi.org/10.1093/idpl/ixp005>.
27. Doshi-Velez, F., and Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. Preprint at arXiv,
<https://doi.org/10.48550/arXiv.1702.08608>
<https://doi.org/10.48550/arXiv.1702.08608>.
28. Goodman, B.W. (2016). A Step Towards Accountable Algorithms ? : Algorithmic Discrimination and the European Union General Data Protection. In.
29. Solove, D. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154, 477.
30. Noble, S.U. (2018). Algorithms of Oppression How Search Engines Reinforce Racism (NYU Press).
31. Danks, D., and London, A.J. (2017). Algorithmic Bias in Autonomous Systems. 4691–4697.
32. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society* 3, 2053951716679679. <https://doi.org/10.1177/2053951716679679>.
33. Norman, D.A. (2013). The Design of Everyday Things (New York: Basic Books).
34. Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction*.
35. Younas, A., and Zeng, Y. (2024). Proposing Central Asian AI ethics principles: a multilevel approach for responsible AI | AI and Ethics. *AI and Ethics*.
36. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete Problems in AI Safety. Preprint at arXiv,
<https://doi.org/10.48550/arXiv.1606.06565>
<https://doi.org/10.48550/arXiv.1606.06565>.
37. Bryson, J.J. (2018). Patience is not a virtue: the design of intelligent systems and systems of ethics. *Ethics Inf Technol* 20, 15–26.
<https://doi.org/10.1007/s10676-018-9448-6>.

38. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 28, 689–707.
<https://doi.org/10.1007/s11023-018-9482-5>.
39. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., and Floridi, L. (2018). Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach. *Sci Eng Ethics* 24, 505–528. <https://doi.org/10.1007/s11948-017-9901-7>.
40. Voigt, P., and Von Dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR) (Springer International Publishing)
<https://doi.org/10.1007/978-3-319-57959-7>.
41. Commission, E. (2021). Excellence and trust in artificial intelligence - European Commission. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en.
42. OECD (2019). OECD Legal Instruments.
<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
43. NIST (2021). Plan for Federal AI Standards Engagement. NIST.
<https://www.nist.gov/artificial-intelligence/plan-federal-ai-standards-engagement>.
44. OECD (2020). The Global Partnership on Artificial Intelligence.
<https://gpai.ai/about/>.
45. UNESCO (2023). Recommendation on the Ethics of Artificial Intelligence | UNESCO. <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>.