GOVERNANCE, ACCREDITATION, ETHICS, AND TRUST IN FUTURE AI SYSTEMS

Asst. Prof. Dr. Abdulkadir Taşdelen
Ankara Yıldırım Beyazıt University

Attribution: Taşdelen, A. (2026). Governance, accreditation, ethics, and trust in future AI systems. In H. Arslan, A. Taşdelen, E. Kuzucu Hıdır, & O. A. Çetin (Eds.), *Artificial intelligence and national technology reflections* (pp. 527–557). Turkish Academy of Sciences. https://doi.org/10.53478/TUBA.978-625-6110-86-1.ch22

10.53478/TUBA.978-625-6110-86-1.ch22

GOVERNANCE, ACCREDITATION, ETHICS, AND TRUST IN FUTURE AI SYSTEMS

Asst. Prof. Dr. Abdulkadir Taşdelen
Ankara Yıldırım Beyazıt University

Abstract

The rapid spread of Artificial Intelligence (AI) across domains such as public services, healthcare, finance, and security shows that these technologies are no longer just technical tools. AI systems have become complex socio-technical systems that shape decisions, institutions, and everyday life. As a result, they require effective governance, ethical oversight, accreditation mechanisms, and, critically, public trust. This chapter presents a holistic framework for AI governance that spans the entire lifecycle of AI systems. It highlights the close interdependence between technical robustness, ethical principles, institutional accountability, and societal expectations. Drawing on the OECD AI Principles, the NIST AI Risk Management Framework, and international accreditation practices, the chapter argues that trustworthy AI depends on continuous monitoring, transparent decision-making, risk- and opportunity-based regulation, and sustained cooperation among multiple stakeholders. The analysis demonstrates that contemporary AI governance rests on four core pillars: technological design and lifecycle management; stakeholder roles and institutional responsibility; regulatory and compliance mechanisms; and ethical, value-based principles. Within this framework, the chapter identifies key challenges, such as algorithmic uncertainty, poor data quality, model drift, cybersecurity risks, discrimination, and bias, alongside emerging opportunities. It shows how these issues can be addressed through well-designed governance structures, systematic auditing, and ethics-by-design approaches. Looking ahead, the chapter emphasizes that AI governance must adapt to a rapidly evolving landscape. Topics such as digital sovereignty, flexible and adaptive regulation, security testing and auditing of large-scale models, public–private collaboration, democratic oversight, and stronger human-centered ethical frameworks are expected to grow in importance. Together, these priorities offer a strategic roadmap for managing AI in a way that is safe, inclusive, and sustainable. Ultimately, the chapter argues that technical excellence alone is not sufficient to achieve trustworthy AI. Meaningful trust requires transparent and proactive governance, ethical responsibility, the protection of human rights, and inclusive stakeholder participation. This holistic approach provides a solid foundation for ensuring that AI systems develop as fair, accountable, and sustainable technologies that contribute positively to societal well-being.

Keywords

Artificial Intelligence Governance, Trustworthy AI, AI Ethics, Digital Sovereignty, Accreditation

Introduction

Over the past decade, AI has evolved from a purely technical innovation into a transformative force affecting economic structures, public services, national security, and social life. In particular, the widespread use of automated decision-making systems in contexts that directly affect human lives has turned AI into a critical policy issue, not only in terms of efficiency and speed, but also with respect to ethical, legal, and societal dimensions (Batool et al., 2025).

In recent years, discriminatory outcomes produced by AI systems, non-transparent and non-verifiable decision-making processes, data privacy violations, misclassifications, and security vulnerabilities have increasingly attracted the attention of both the public and policymakers. Platforms such as the OECD AI Incidents and Hazards Monitor (*OECD AI Incidents Monitor*, 2025), the AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC) Repository (*AIAAIC - AIAAIC Repository*, 2025), the MIT AI Incident Tracker (*MIT AI Incident Tracker*, 2025), and the Artificial Intelligence Incident Database (*Artificial Intelligence Incident Database*, 2025) provide growing empirical evidence that many AI-related failures stem from systemic governance deficiencies. These developments suggest that AI technologies should not be treated as tools that are addressed only after problems emerge; rather, they should be understood as complex socio-technical systems that must be governed proactively throughout their entire lifecycle.

The rapid pace of technological development further demonstrates that governance needs cannot be limited to technical standards alone. Effective AI governance requires a multi-layered and integrated structure that includes legal regulations, institutional responsibility mechanisms, and international cooperation. Globally influential cases such as the Cambridge Analytica scandal have made clear the need for strong ethical principles, transparency standards, and accountability mechanisms within AI and data ecosystems (Hadzovic et al., 2024). In this context, digital sovereignty has emerged as a key axis of governance, referring to states' efforts to protect their data infrastructures, national interests, and strategic autonomy. As a result, digital sovereignty has become a key factor shaping national AI policies.

This chapter aims to examine AI governance, ethical principles, and accreditation processes at institutional, national, and international levels. Particular attention is given to national governance models and debates on digital sovereignty due to their increasing strategic importance. The chapter first presents the conceptual foundations of AI governance, then discusses ethical principles and trust mechanisms, international accreditation and certification practices, and finally future-oriented governance strategies.

This holistic approach seeks to evaluate the norms, processes, and governance tools required for building trustworthy, fair, and accountable AI ecosystems in an integrated manner. At the same time, it offers a guiding framework for the sustainable development of AI systems that support social welfare.

1. Conceptual Framework of AI Governance

AI governance can be broadly defined as the set of policies, processes, and standards that ensure the ethical, secure, transparent, accountable, and lawful development, deployment, and monitoring of AI systems (Batool et al., 2025). This framework goes beyond technical requirements and encompasses legal regulations, societal values, economic impacts, and political consequences. Therefore, AI governance should be understood as an integrated field located at the intersection of technical systems and socio-political structures.

At the institutional level, AI governance is closely linked to risk management, ethical responsibility, model validation, data integrity, and compliance processes. Prior to 2016, the absence of comprehensive national strategies or policy documents governing AI indicated a significant governance gap despite the rapid diffusion of the technology. The publication of an AI-related framework report by the United States in 2016 marked an early turning point, followed by a sharp increase in national AI strategies, ethical guidelines, and governance frameworks in subsequent years. Globally influential incidents, most notably the Cambridge Analytica scandal, played a critical role in accelerating this shift. The scandal exposed the risks of algorithmic manipulation for democracy, privacy, and public trust, making the need for transparent and accountable AI governance visible at a global scale (Hadzovic et al., 2024). As a result, AI governance emerged as a central policy concern rather than a peripheral technical issue.

A comprehensive meta-analysis conducted by Corrêa et al. (2023) demonstrates that nearly 200 AI-related policy documents, ethical guidelines, and governance frameworks published worldwide converge around a common set of 17 principles. Among these, privacy, non-discrimination, transparency, data protection, security, and accountability stand out as dominant themes.

This convergence indicates a high level of normative alignment across countries and institutions, despite differences in legal systems and political priorities. It also provides a strong foundation for future regulatory efforts by clarifying which values are most widely accepted as core elements of trustworthy AI. From a governance perspective, this alignment reduces fragmentation and supports the development of interoperable regulatory and certification mechanisms.

The healthcare sector represents one of the most illustrative examples of sector-specific AI governance challenges. Between 1995 and 2025, the United States Food and Drug Administration (FDA) approved a total of 1,247 AI-driven and machine learning-based medical devices. Approximately 77% of these devices are concentrated in the field of radiology, highlighting the governance sensitivity of clinical decision support systems and medical imaging applications (FDA, 2025).

The rapid increase in approvals after 2015 underscores the need for stronger governance structures addressing accuracy, safety, transparency, and explainability. In high-risk domains such as healthcare, AI governance cannot rely solely on post-market interventions; instead, it must incorporate continuous validation, monitoring, and accountability throughout the system lifecycle.

At the national level, AI governance plays a critical role in protecting the public interest, establishing societal trust, strengthening data protection mechanisms, and maintaining digital independence. Governments increasingly perceive AI governance not merely as a compliance instrument, but as a strategic capacity that directly affects economic competitiveness, national security, and social stability. National governance frameworks define how AI systems may be developed, deployed, and monitored within a country's legal, ethical, and cultural context.

National AI governance typically includes the formulation of national AI strategies, data protection laws, sector-specific regulations, and public-sector oversight mechanisms. These instruments aim to ensure that AI systems align with societal values, respect fundamental rights, and operate transparently and accountably. In addition, national governance structures support the development of domestic AI ecosystems, including local data infrastructures, research capacities, and skilled human capital, which are increasingly linked to the concept of digital sovereignty.

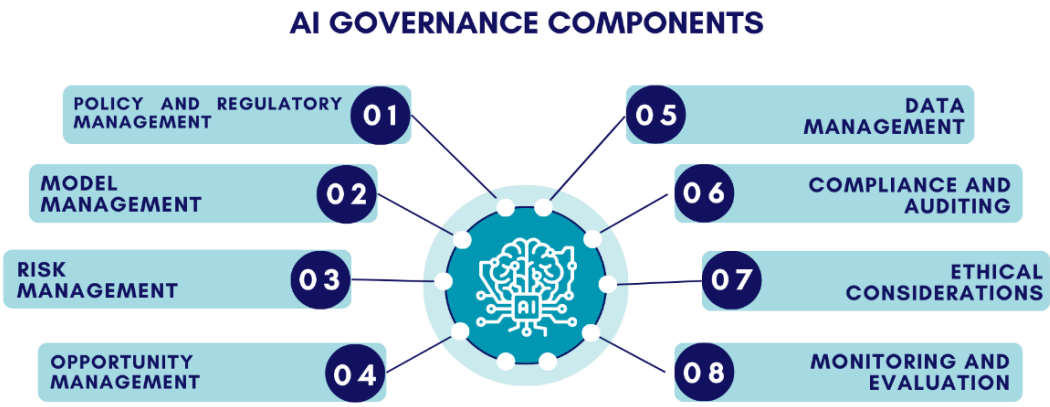
At the international level, the cross-border nature of AI systems makes regulatory alignment and cooperation essential. AI models, data flows, and digital services often transcend national boundaries, creating shared risks related to security, privacy, discrimination, and systemic failures. As a result, international AI governance focuses on harmonizing standards, coordinating risk management approaches, sharing ethical frameworks, and developing joint certification and accreditation mechanisms.

International organizations such as the Organisation for Economic Co-operation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the International Organization for Standardization (ISO) play a central role in shaping global AI governance. Frameworks such as the OECD AI Principles and UNESCO's Recommendation on the Ethics of AI (*OECD AI Incidents Monitor*, 2025) promote common values, including human-centered design, transparency, accountability, and respect for human rights. Technical standards developed by ISO, IEC, and IEEE further support interoperability and comparability across jurisdictions.

Ultimately, trustworthy AI can only be achieved through multi-level governance architectures that integrate national regulatory authority with international coordination. National frameworks provide contextual legitimacy and enforcement capacity, while international cooperation reduces fragmentation, lowers compliance costs, and strengthens collective resilience against global AI risks. As emphasized by governance frameworks developed by the OECD, NIST, and ISO/IEC, effective AI governance must be multi-stakeholder, adaptive, and embedded across both national and international dimensions (Wodi, 2024).

Figure 1

AI Governance Components



In this chapter, the components of AI governance are classified under eight headings (Figure 1). These components represent the core functional domains required to ensure the effective, ethical, and accountable governance of AI systems throughout their lifecycle. Together, they provide a structured and holistic framework that integrates technical, organizational, and societal dimensions of AI governance.

1.1. Policy and Regulatory Management

Policy and Regulatory Management refers to the systematic development, alignment, implementation, and maintenance of the legal, regulatory, and institutional foundations governing AI systems. This component establishes the normative framework within which AI technologies are designed, deployed, and operated, encompassing binding regulations, internal policies, standards, and guidelines. In the context of advanced AI governance, policy and regulatory management ensures that AI applications are developed in compliance with applicable laws, sectoral regulations, and internationally recognized standards, thereby providing legal certainty and institutional accountability.

This component plays a critical role in translating abstract governance objectives into enforceable rules and operational requirements. It includes the formulation of internal AI policies, alignment with external regulatory instruments (such as the EU AI Act and data protection legislation), definition of roles and responsibilities, and continuous updating of policies in response to technological and regulatory change. Effective policy and regulatory management also serve as the foundation upon which compliance, auditing, and risk management mechanisms are built, ensuring coherence across the governance architecture.

At the global level, policy and regulatory management supports the harmonization of international AI principles, regulatory frameworks, and soft-law instruments, facilitating cross-border interoperability and convergence. At the national level, it contributes to the development and implementation of national AI strategies, legal frameworks, and sector-specific regulations. At the institutional/corporate level, it encompasses the establishment of internal AI policies, regulatory mapping, policy ownership structures, and mechanisms to ensure that operational practices remain aligned with evolving legal and regulatory requirements.

1.2. Model Management

Model management refers to the holistic management of the processes through which AI systems are developed, validated, deployed, updated, and retired throughout their lifecycle. This component requires that technical processes, such as model architecture selection, hyperparameter optimization, version control, documentation, model explainability, and performance validation, are conducted in accordance with institutional standards. In addition, regular retraining, calibration with up-to-date data, and consistency testing across different use cases form the foundation of reliable AI applications.

At the global level, model management is supported by international technical standards and best practices that promote interoperability and shared expectations. At the national level, it underpins regulatory requirements for model documentation, auditing, and post-market monitoring. At the institutional/corporate level, it provides the operational backbone for trustworthy AI by linking technical practices with governance obligations (Batool et al., 2025).

1.3. Risk Management

Risk management encompasses the processes of identifying and mitigating the potential technological, ethical, and societal risks associated with AI. Biases in training data, model generalization errors, security vulnerabilities, and the potential to cause harm to users necessitate the responsible governance of AI systems (Batool et al., 2025). Tools such as impact assessments, red-teaming exercises, adversarial robustness analyses, and model drift detection support advanced risk management

practices. Through these processes, organizations enhance their capacity to ensure legal compliance while identifying and mitigating security and ethical failures at an early stage.

This component promotes international risk frameworks and ethical and social risk awareness at the global level. At the national level, it enables the creation of regulations and oversight mechanisms. At the institutional/corporate level, it includes impact assessments, red-teaming tests, adversarial testing, and model drift detection.

1.4. Opportunity Management

The findings of Olsson (2007) on project risk and opportunity management demonstrate that traditional risk management processes are often effective only for "tame" technical problems, while opportunities tend to emerge from complex uncertainties characterized as "messes" and "wicked problems." Because such uncertainties involve behavioral and structural complexity, they cannot be captured by classical risk management models that operate analytically and linearly. According to Olsson, the systematic identification and realization of opportunities require a holistic understanding of the project and stakeholder ecosystem, strong internal communication, and an organizational structure that actively encourages opportunity-seeking behavior. Otherwise, opportunity management remains confined to a narrow space defined merely as what remains after risks are addressed.

This insight is highly relevant for AI governance. The opportunities created by AI technologies often arise not from simple, technically definable problems, but from the interaction of economic systems, societal expectations, regulatory frameworks, and organizational behavior, characteristics typical of "messes" and "wicked" problems (Olsson, 2007). Consequently, opportunity management in the AI context must go beyond a traditional risk-reduction perspective and instead enable the strategic identification and steering of AI's positive impacts.

Within this framework, opportunity management constitutes the key component that distinguishes advanced governance from traditional risk-focused models. It involves the systematic analysis of AI's potential positive contributions to economic growth, productivity gains, social welfare, and sustainable development. As emphasized in the OECD's 2023 update (*AI Principles | OECD*, 2025), modern AI governance should not focus solely on risk mitigation, but also on consciously directing the technology's capacity to generate development, innovation, and societal value. In this regard, it is critical for organizations to measure the returns of AI applications, identify innovation opportunities, and adopt structured approaches to opportunity development in order to maximize impact.

Opportunity management assumes distinct yet complementary roles across governance layers. At the global level, it supports the use of AI to promote sustainable development, social welfare, and the reduction of global inequalities. At the national level, it contributes to the development of national strategies aimed at maximizing economic and social impact. At the institutional/corporate level, it encompasses the identification of innovation opportunities, the assessment of the value-creation potential of technology investments, and the integration of realized benefits into governance structures.

In conclusion, Olsson (2007) findings offer a strong warning for AI governance: opportunities are not accidental by-products, but a managerial capability that requires a holistic perspective, strong governance, organizational learning, and proactive policy-making. The framework presented in this section positions opportunity management as an integral component of modern governance, making the positive impacts of AI visible and manageable.

1.5. Data Management

As data constitutes the foundation of AI, advanced governance necessitates a comprehensive approach to data management. Data quality, accuracy, integrity, access control, and privacy standards are critical to the reliability of AI systems (Kazim & Koshiyama, 2021). Particularly in high-sensitivity sectors such as healthcare, finance, and public services, processes such as data anonymization, ethical data collection, data source traceability, and legal compliance (e.g., GDPR) must be supported by robust mechanisms. Data management also includes the development of inclusive data policies aimed at reducing potential biases in AI systems.

Data management supports international data standards and cross-border data-sharing principles at the global level. At the national level, it encompasses national data policies and regulations on privacy and access. At the institutional/corporate level, it ensures data quality, integrity, and confidentiality while aiming to reduce bias.

1.6. Compliance and Auditing

The compliance and auditing component ensures that AI systems are operated in accordance with legal regulations, ethical principles, and technical standards. Internal audit mechanisms, external independent audits, and certification processes play a critical role in this context. The emergence of ISO/IEC 42001 represents a significant development in the institutionalization of AI management systems (Benraouane, 2024). The EU AI Act has further established binding global standards by introducing conformity assessments, risk analyses, and technical documentation requirements for high-risk AI systems (Roberts et al., 2024). Through these mechanisms, organizations can demonstrate both legal compliance and ethical responsibility in a transparent and externally verifiable manner.

Compliance and auditing support international standards (ISO/IEC 42001, EU AI Act) at the global level. At the national level, they enforce AI regulations and certification programs. At the institutional/corporate level, they ensure adherence to policies through internal and external audits.

1.7. Ethical Considerations

The ethical component is not merely one element of advanced governance, but a central dimension that intersects with all other components. The ethical framework proposed by Floridi & Cowls (2019) requires a holistic consideration of beneficence, non-maleficence, autonomy, justice, and explainability in AI applications. Ethical considerations must be operationalized through concrete policies and processes rather than remaining at a purely theoretical level. Practices such as ethics review committees, developer and user training programs, mandatory explainability tools, and stakeholder participation in decision-making processes (Society-in-the-Loop, SITL) are increasingly adopted (Wirtz et al., 2022).

At the global level, ethical considerations encompass global ethical frameworks such as beneficence, non-maleficence, autonomy, justice, and explainability. At the national level, they are supported through national ethics committees and training programs. At the institutional/corporate level, they are implemented through ethics committees, developer and user training, and explainability tools.

The key ethical principles identified will be addressed under a separate heading in the subsequent sections.

1.8. Monitoring and Evaluation

The dynamic nature of AI systems necessitates continuous post-deployment monitoring and evaluation. Model performance, security vulnerabilities, drift phenomena, changes in data quality, and the impact of user feedback are core indicators of trustworthy AI governance (Winter et al., 2021). Accordingly, advanced governance approaches promote the implementation of mechanisms such as periodic performance assessments, real-time monitoring dashboards, algorithmic transparency reports, and ethical compliance analyses. These mechanisms enable the sustainable assurance of both technical reliability and ethical alignment.

Monitoring and evaluation encompass global best practices and assessment methods for AI performance and compliance at the global level. At the national level, they support the establishment of national monitoring frameworks. At the institutional/corporate level, they are carried out through continuous post-deployment monitoring, dashboards, and performance and ethical reviews.

Figure 2

The Components of the AI Governance Framework at the Global, National, and Institutional/Corporate Levels

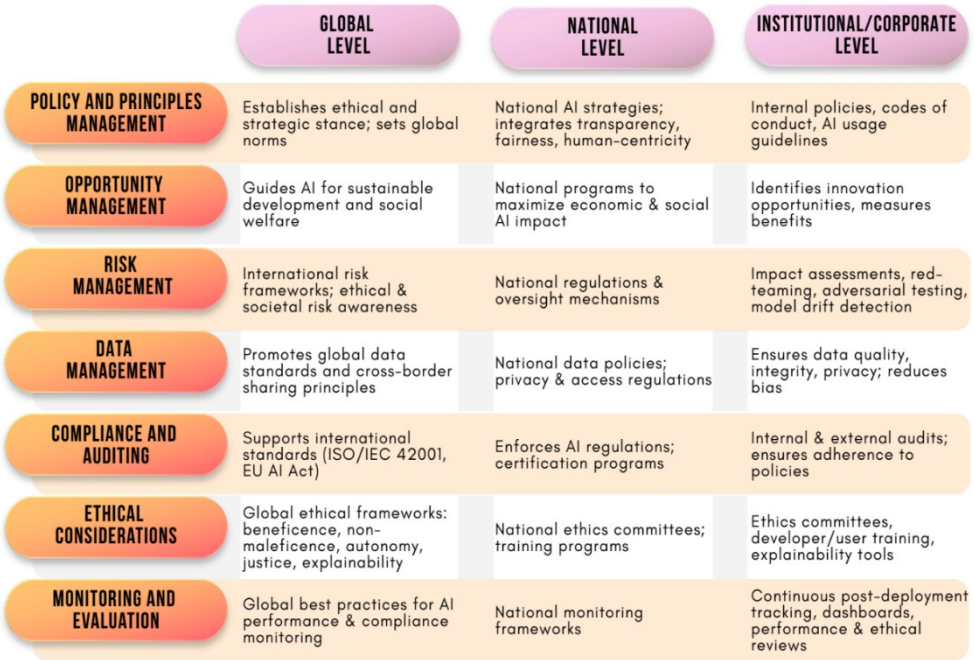


Figure 2 illustrates that AI governance does not operate within a single regulatory or institutional layer, but rather emerges from the interaction of global, national, and institutional governance mechanisms. Each governance component, ranging from policy and regulatory management to monitoring and evaluation, assumes distinct functions and instruments at each level, while remaining interdependent across the overall system.

At the global level, AI governance is primarily normative and coordinative in nature. International principles, ethical recommendations, and technical standards (e.g., OECD AI Principles, ISO/IEC standards) establish shared expectations, reduce regulatory fragmentation, and promote interoperability across jurisdictions. However, enforcement capacity at this level remains limited, relying largely on voluntary adoption and peer pressure.

At the national level, these global norms are translated into binding legal frameworks, sector-specific regulations, and public oversight mechanisms. National authorities define risk classifications, enforcement procedures, certification requirements, and sanctions. In this sense, the national layer functions as the primary bridge between abstract global principles and concrete institutional obligations.

At the institutional (corporate) level, AI governance becomes operational. Organizations are responsible for implementing policies, managing risks and opportunities, ensuring data quality, conducting audits, and continuously monitoring system performance. This level is where governance principles are transformed into measurable practices through internal controls, management systems (such as ISO/IEC 42001), and accountability structures.

Crucially, Figure 2 also implies a bidirectional feedback loop among these levels. Operational experiences and failures at the institutional level inform national regulatory updates, while national regulatory innovations contribute to the evolution of global standards. This dynamic interaction underscores that effective AI governance is not static, but adaptive and learning-oriented across governance layers.

2. AI Ethics and Fundamental Principles

AI ethics defines the principles that determine which moral and societal values AI should be designed and used in alignment with. Although the literature includes ethical guidelines proposed by different institutions, a number of common themes clearly stand out (Corrêa et al., 2023; Huang et al., 2023). The main fundamental principles highlighted in the literature are as follows (Huang et al., 2023; Ryan & Stahl, 2021):

2.1. Transparency

The principle of transparency includes both the transparency of AI technologies themselves and the transparency of the processes through which AI systems are developed and adopted. Transparency requires understanding why an AI system behaves or makes decisions in a particular way within a given context, which is commonly referred to as interpretability or explainability. This requirement is often described using the metaphor of "opening the black box of AI." At the same time, transparency also implies that the design, implementation processes, and outcomes of AI systems should be justifiable and visible.

One of the most striking examples of the systematic injustices that may arise from a lack of transparency is the A-Level grading algorithm used in the United Kingdom during the COVID-19 period. This model "standardized" student performance based on teacher assessments and schools' historical achievement distributions. Due to its non-transparent operation, the system was rapidly perceived by the public as a "black box." Notably, the details of the algorithm were not published until after the results were announced, which created serious concerns that the process was open to political influence (Kelly, 2021).

When the results were released, it became evident that students attending public schools in large and disadvantaged areas were disproportionately downgraded. This occurred because the algorithm applied limited adjustments for small and selective schools, while reproducing historical trends in large public schools, thereby deepening existing socioeconomic inequalities (Kelly, 2021). As a result, the university admissions prospects of thousands of students were directly affected, leading to widespread public protests. Ultimately, the government withdrew the algorithm and was forced to rely on teacher-assessed grades instead.

The A-Level crisis powerfully demonstrates how algorithmic decision-making processes that lack transparency can rapidly erode public trust, and why explainability constitutes an ethical necessity rather than an optional feature.

2.2. Fairness & Justice

The principle of fairness and justice requires that the design, development, deployment, and use of AI systems be fair and equitable, so that AI systems do not produce discrimination or bias against individuals, communities, or groups. In recent years, fairness and justice have been extensively discussed in both the media and academic literature.

One illustrative example demonstrating the serious consequences of neglecting this principle is the Robert Williams case in Detroit. In 2020, Robert Williams was wrongly identified as a criminal by a facial recognition system used by the Detroit Police Department and was subsequently arrested. The incorrect match resulted from biases in the system's training data and its low accuracy rates. This incident led to the violation of the fundamental rights of an innocent citizen and constituted a significant injustice (*Facial Recognition Leads To False Arrest Of Black Man In Detroit : NPR*, 2025).

This case underscores the importance of designing AI systems in a manner that is fair and equitable. AI developers and users have a responsibility to minimize the risk of bias, conduct continuous testing, and implement measures that reduce the impact of false matches and erroneous decisions.

2.3. Responsibility and Accountability

The principle of responsibility and accountability requires that AI systems be auditable. Designers, developers, owners, and users must be responsible and accountable for the behavior of AI systems and for any harm they may cause. Establishing appropriate mechanisms to ensure responsibility and accountability before, during, and after the development, deployment, and use of AI systems is of critical importance.

One of the most frequently discussed domains in relation to responsibility and accountability is autonomous vehicles. Some autonomous vehicle systems have caused fatalities and serious injuries. As emphasized by Uzair (2021), traditional tort liability laws are often insufficient in cases involving autonomous vehicle accidents. When an accident occurs due to a sensor or software failure while a vehicle is operating in fully autonomous mode, assigning responsibility solely to the person inside the vehicle is not fair.

In such cases, multiple actors may be included in the chain of responsibility, including the hardware manufacturer, software provider, system integrator, and insurance companies. Moreover, Uzair (2021) argues that responsibility and accountability should not function only as mechanisms activated after a violation occurs. Instead, they should be embedded within a continuous governance approach that includes risk analysis, testing processes, regular reporting, and algorithmic audits.

This example concretely demonstrates the importance of responsibility and accountability principles in autonomous systems and highlights why ethical and legal frameworks must be clearly defined within complex technological ecosystems.

2.4. Non-maleficence

The principle of non-maleficence fundamentally refers to avoiding harm or refraining from imposing risks of harm on others. AI systems should be safe and secure in order to protect human dignity as well as individuals' mental and physical integrity, and they should be resilient against malicious use. Given the fatal accidents caused by autonomous vehicles and robots, avoiding harm to humans constitutes a major concern in AI ethics.

The failure of Google's AI-based Earthquake Early Warning System to produce an effective warning during the 2023 Türkiye Earthquake represents a striking case that exemplifies a violation of the non-maleficence principle (*OECD AI Incidents Monitor*, 2025). Since disaster early warning technologies inherently operate in high-risk environments, system failure is not merely a technical error; it is an ethical issue due to its direct potential impact on human lives.

Factors such as insufficient regional calibration of the model, inconsistencies in data received from sensor networks, or the inability of algorithmic decision-making processes to adapt to disaster conditions have been identified as elements that increased the risk of harm. In such critical systems, "doing no harm" does not only mean avoiding false alarms, but also minimizing the risk of failing to deliver life-saving warnings in a timely manner.

Accordingly, the system's inability to function effectively during the earthquake prevented individuals from being directed toward safe areas, increased societal risk perception, and undermined public trust in technological infrastructure. This case demonstrates how critical security testing, stress scenarios, regional validation, and human-centered risk assessments are in high-impact AI applications. The principle of non-maleficence evaluates not only direct harm, but also the ethical significance of delayed or unrealized preventive benefits, which themselves may constitute a form of harm.

2.5. Privacy

The principle of privacy aims to ensure respect for privacy and the protection of personal data in the use of AI systems. This includes providing effective data management and governance for all data that are used and generated by AI systems throughout their entire lifecycle. Data collection, use, and storage must comply with applicable privacy and data protection laws and regulations.

One of the most striking examples related to privacy is the previously mentioned Cambridge Analytica scandal (Hadzovic et al., 2024). In this scandal, the personal data of millions of individuals were made available to third-party applications without their knowledge. It was later revealed that these data were ultimately exploited for political purposes.

The debates that emerged following the disclosure of this scandal clearly demonstrated the critical importance of precautionary measures and provided one of the most compelling examples of how AI-enabled systems can violate privacy.

2.6. Beneficence

The principle of beneficence states that AI should provide benefits to people and contribute positively to humanity. AI technologies should be used to generate beneficial outcomes for individuals, society, and the environment. Accordingly, the objectives of AI systems should be clearly defined and properly justified.

A notable example related to this principle is the use of AI and blockchain integration in halal certification processes with the aim of increasing social benefit. As demonstrated by Sunmola et al. (2025), the combined use of AI-supported analytics and distributed ledger technology can enhance transparency, traceability, and security in food supply chains. Through AI-based content verification, anomaly detection, and automated classification models, the halal compliance status of products can be assessed more rapidly, reliably, and objectively. At the same time, blockchain infrastructure protects these data through an immutable record-keeping system, thereby reducing fraud.

As a result, consumer trust is strengthened, producer responsibility is increased, and the overall integrity of halal certification processes is supported. This approach constitutes a concrete example of AI's capacity to generate beneficence, as it both facilitates access to food that complies with religious sensitivities and enhances overall food safety. Such applications demonstrate that the principle of beneficence is not merely an abstract ideal, but a practical governance objective that, when properly designed, can deliver direct benefits to broad segments of society.

2.7. Freedom and Autonomy

Freedom and autonomy refer to an individual's ability to make decisions in accordance with their own goals and preferences. It is essential that the use of AI does not undermine or restrict human autonomy. This principle requires maintaining a balance between the decision-making authority individuals retain for themselves and the authority delegated to AI systems.

The SyRI (System Risk Indication) case in the Netherlands constitutes a striking example of the effects of AI use on autonomy within the digital welfare state. The District Court of The Hague ruled that the use of SyRI, an algorithm designed to detect social welfare fraud, and the related legislation violated the right to respect for private life under Article 8 of the European Convention on Human Rights (ECHR). The Court held that the SyRI legislation and its implementation failed to meet the requirements of necessity and proportionality (Rachovitsa & Johann, 2022).

2.8. Solidarity

The principle of solidarity requires that the development and deployment of an AI system be compatible with the preservation of solidarity boundaries among individuals and across generations. In other words, AI should promote social security and social cohesion and should not endanger social bonds and relationships.

In this context, in December 2023, Human Rights Watch (HRW) reported that peaceful content supporting Palestine or documenting human rights violations in Palestine was systematically removed or its visibility suppressed on Meta's Instagram and Facebook platforms (*Meta: Systemic Censorship of Palestine Content | Human Rights Watch*, 2025). Such practices directly contradict the principle of solidarity, as they undermine social cohesion and the visibility of vulnerable groups. In these cases, AI systems suppress voices systematically instead of preserving social bonds and ensuring fair representation.

Within this principle, AI systems should be designed and governed in a way that guarantees the protection of social solidarity and equitable representation, particularly in contexts involving vulnerable communities.

2.9. Sustainability

The principle of sustainability represents the requirement that the production, governance, and deployment of AI systems be sustainable and avoid causing harm to the environment. AI technologies should meet the requirements of sustaining long-term human well-being and possess the capacity to preserve a healthy environment for future generations.

In this context, the rapidly increasing electricity consumption of Large Language Models (LLMs) and the associated carbon emissions clearly demonstrate how critical the sustainability principle has become in practice. The training and widespread use of large-scale models place growing pressure on energy infrastructure, while the insufficient use of renewable energy sources further intensifies environmental impacts. Consequently, increasing energy efficiency across all processes in which AI systems are developed, hosted, and operated, prioritizing low-carbon hardware and algorithms, and transparently reporting electricity consumption have become essential sustainability requirements.

The future development of LLMs in more precise, efficient, and environmentally responsible ways has emerged as a fundamental necessity, both to protect the benefits enjoyed by current users and to leave a livable environment for future generations (Ji & Jiang, 2026). Similarly, the application of certain training methods that shorten training duration can reduce energy consumption and thereby produce solutions that support sustainability (Tasdelen, 2025). Numerous similar examples can be identified. Through such practices, sustainability should be actively supported, and, where possible, these considerations should be taken into account at every stage of the AI system lifecycle.

2.10. Dignity

Human dignity refers to the belief that all individuals possess intrinsic value solely by virtue of being human. This intrinsic value should not be diminished, endangered, or suppressed by other individuals or by technologies such as AI. It is therefore essential that AI systems do not violate or harm the dignity of end users or other members of society. AI systems should respect, support, and protect individuals' physical and mental integrity, as well as their personal and cultural identities.

Examples such as the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm, which has been used by courts in various parts of the United States, illustrate how AI can challenge human dignity. This algorithm, designed to assess the risk of criminal recidivism, was shown to incorrectly predict that Black Americans were nearly twice as likely to reoffend compared to White Americans. Due to these high false-positive rates, Black Americans were disproportionately harmed, revealing that the system's discriminatory outcomes

constituted an assault on the self-worth of individuals and groups. This case demonstrates that human dignity is directly linked to the interpretation of protecting vulnerable populations.

Moreover, the presentation of such proprietary algorithms to judges in the form of risk scores creates a lack of transparency in decision-making processes. By influencing individuals in ways that are difficult to detect, these systems undermine both individual autonomy and accountability, which are core requirements of human dignity (Teo, 2023).

2.11. Trust

Trust is one of the most critical factors determining the societal acceptance, widespread adoption, and long-term sustainability of AI systems. Trustworthy AI is defined not only by technical performance, such as accuracy, speed, and reliability, but also by the system's alignment with ethical principles, transparency, accountability, and its operation within an appropriate socio-technical context (Thiebes et al., 2021). Trust in AI systems provides the psychological and social foundation necessary for users, institutions, and society to adopt and interact with these technologies.

The construction of trust is a multidimensional process that encompasses technical, institutional, and societal layers. At the technical level, correct, secure, and stable system performance, explainable decision-making, and the protection of data privacy form the foundation of trust (Afroogh et al., 2024). Especially in high-risk contexts-such as healthcare, justice, and transportation-the potential impact of system failures on human life or fundamental rights demonstrates that trust is not merely a "preference," but a necessity. For example, fatal accidents involving autonomous vehicles (Uzair, 2021) or wrongful arrests caused by facial recognition systems reveal how a lack of trust can lead to tangible and difficult-to-reverse consequences.

At the institutional level, trust is closely linked to the transparency, ethical commitments, and accountability mechanisms of organizations that develop, deploy, and operate AI systems. Users and stakeholders expect to understand how decisions are made, how data are used, and whom to contact when problems arise. Management system standards such as ISO/IEC 42001 and regulatory frameworks like the EU AI Act provide organizations with roadmaps for institutionalizing trust (Benraouane, 2024; Roberts et al., 2024). Independent auditing and certification processes further enhance credibility by enabling third-party verification of trustworthiness.

At the societal level, trust requires a collective belief that AI technologies are developed and used in alignment with societal values, norms, and interests. This can be supported through mechanisms such as SITL approaches, civil society representation in ethics committees, and public oversight of algorithmic decision-

making processes (Wirtz et al., 2022). Trust-eroding incidents, such as Microsoft's Tay chatbot producing racist content (Afroogh et al., 2024) or the Cambridge Analytica scandal (Hadzovic et al., 2024), have demonstrated how fragile societal trust can be and how long it may take to rebuild once it is lost.

Maintaining trust is not limited to initial design and testing phases; it requires continuous post-deployment monitoring, evaluation, and improvement. Model drift, changing data patterns, emerging security threats, and shifts in social context can challenge system trustworthiness over time (Winter et al., 2021). For this reason, real-time monitoring dashboards, regular ethical audits, and transparent performance reporting are vital for managing trust as a dynamic and evolving process.

In conclusion, trust in AI systems is not a one-time certification or a purely technical feature. Rather, it is a phenomenon that must be continuously nurtured through technical excellence, institutional responsibility, ethical compliance, and ongoing societal dialogue. Future trustworthy AI ecosystems will be shaped around transparent, fair, and accountable systems that embrace this multi-layered approach.

3. International Accreditation and Certification Practices

In order to ensure the global adoption of AI systems in a trustworthy, safe, and ethically compliant manner, standardized accreditation and certification processes are becoming increasingly critical. These processes aim to build trust among users, regulators, and society by independently verifying that AI systems meet defined standards related to quality, safety, transparency, and ethical compliance.

Particularly in high-risk domains such as healthcare, finance, justice, and transportation, certification is no longer a "luxury," but rather an operational and legal necessity.

3.1. International Standards

International standards constitute the technical and administrative foundation of AI governance. By providing a common language and methodology for developers, manufacturers, and auditors, these standards facilitate compliance, interoperability, and comparability across different jurisdictions and sectors.

ISO/IEC 42001: Artificial Intelligence Management System (AIMS)

ISO/IEC 42001, the AIMS standard, aims to institutionalize the policies, procedures, and processes required for organizations to manage AI systems in a responsible and trustworthy manner. Adopting a risk-based approach, this standard emphasizes transparency, traceability, and human oversight throughout the entire lifecycle of AI projects. The possibility of certifying an organization's compliance with this standard through independent audits constitutes a significant trust signal in the marketplace *(Benraouane, 2024)*.

ISO/IEC 23894: Artificial Intelligence – Risk Management Guidance

ISO/IEC 23894 provides organizations with guidance on identifying, analyzing, evaluating, and treating AI-related risks, including ethical, operational, legal, and reputational risks. The standard places particular emphasis on risks such as algorithmic bias, data privacy violations, and security vulnerabilities (*ISO/IEC 23894:2023 - AI - Guidance on Risk Management*, 2026).

IEEE 7000 Series

The IEEE 7000 series, developed by the Institute of Electrical and Electronics Engineers (IEEE), primarily focuses on the design of systems that are aligned with ethical values. For example, IEEE 7000 defines process standards for addressing ethical concerns during the design and development of ethically aligned systems and software. IEEE 7010, on the other hand, provides guidance for measuring the impact of AI systems on human well-being, proposing metrics to assess the societal benefits of such systems (*IEEE SA - IEEE 7000-2021*, 2026).

3.2. Certification Mechanisms

The mere existence of standards on paper is not sufficient; how compliance with these standards is tested, verified, and certified in practice is of critical importance.

Third-Party (Independent) Certification

In this model, an accredited independent body (certification body) audits either the AI system itself or the organization's AI management system in accordance with the relevant standard (e.g., ISO/IEC 42001). If compliance is confirmed, a certificate is issued. This approach represents the most robust model, as it minimizes bias and maximizes credibility.

The EU AI Act mandates third-party conformity assessments for high-risk AI systems. Within this process, elements such as technical documentation, risk analysis, data quality reports, and fundamental rights impact assessments are examined.

Self-Declaration

For low-risk AI systems, a provider's self-declaration of compliance with relevant standards may be considered sufficient. However, due to the absence of independent auditing, this model offers lower reliability and is typically supplemented by market surveillance mechanisms.

Sector-Specific Certification Programs

Certain sectors may develop certification schemes tailored to their specific needs in addition to general standards. For example, in the field of medical devices, FDA approvals-such as 510(k) clearance or Premarket Approval (PMA)-or CE marking conformity in Europe function as de facto certification mechanisms for AI-based diagnostic tools.

3.3. The Role of Accreditation

Accreditation constitutes the cornerstone of the overall credibility of certification processes. An accreditation body, such as TÜRKAK in Türkiye or the ANSI-ASQ National Accreditation Board (ANAB) in the United States, assesses and approves the technical competence, impartiality, and operational quality of certification bodies. In other words, for an organization to obtain an ISO/IEC 42001 certificate, the certification body issuing this certificate is also expected to be authorized by a recognized accreditation body.

This mechanism of "auditing the auditors" enables the mutual recognition of certificates at the global level and thereby facilitates cross-border compliance processes. Accreditation ensures that independent verification activities are conducted on a reliable foundation in terms of both technical capacity and procedural rigor.

3.4. The Challenge of Global Harmonization

AI regulation and standardization remain emerging and evolving fields. Although the EU AI Act, China's AI regulations, the sectoral and state-based approach of the United States, and the national strategies of other countries share similarities in their underlying principles, they differ significantly in terms of implementation details, risk classification schemes, and compliance pathways. This problem of regulatory fragmentation increases compliance costs for globally operating companies, slows innovation, and complicates international cooperation.

To overcome this challenge, international cooperation is of vital importance. Multilateral initiatives such as the OECD AI Principles, UNESCO's Recommendation on the Ethics of AI and the G20 AI Dialogue aim to promote convergence around shared values and principles. At the same time, international standardization organizations such as ISO, IEC, and IEEE provide critical platforms for the harmonization of technical standards.

In the future, progressing toward a "certified once, accepted everywhere" approach will represent one of the most important steps for strengthening both global safety and the digital economy.

However, given the increasing polarization of the global landscape, how such global harmonization can be achieved remains uncertain. In this context, Türkiye should not wait for full international convergence, but instead define its own national roadmap and complete these processes in a timely manner. This issue constitutes a critical component of AI governance in particular, and of digital sovereignty more broadly.

4. Future-Oriented AI Governance Agendas and Strategies

The future agenda of AI governance is becoming increasingly multidimensional and dynamic, driven by the acceleration of technological advancement, the diversification of national strategic priorities, and the emergence of AI as a decisive factor in global competition. In this process, not only technical safety but also ethical, legal, political, and socio-economic dimensions are moving to the center of governance strategies (Batool et al., 2025).

From this perspective, Türkiye's National AI Strategy (2021–2025), the Digital Türkiye Strategy, the 11th and 12th Development Plans, Personal Data Protection Law (KVKK) regulations, and the activities of the TÜBİTAK AI Institute constitute the core documents shaping the national framework for AI governance. As emphasized in these documents, Türkiye regards AI as a critical component of economic growth, efficiency in public administration, societal benefit, and national security.

The following issues encompass the key areas that are expected to gain prominence in AI governance in the coming years.

4.1. AI Sovereignty and Digital Independence

In the coming period, states are expected to intensify their efforts to control their own data infrastructures and to reduce external dependencies in critical digital components.

AI sovereignty has become an increasingly prominent concept in Türkiye's national strategy documents in recent years. The National AI Strategy explicitly identifies digital sovereignty as a strategic objective, emphasizing data security, the strengthening of the domestic software ecosystem, and the reduction of strategic dependencies. Digital sovereignty is particularly critical for sectors such as defense, healthcare, finance, and public administration, and it is aligned with initiatives under the National Technology Initiative (Milli Teknoloji Hamlesi) that aim to enhance Türkiye's independence in these domains.

To strengthen Türkiye's digital sovereignty, the National Data Space project should be elevated to a strategic priority. This approach, similar to the European Union's GAIA-X model, can enable the establishment of secure data-sharing infrastructures among the public sector, private sector, and academia. In addition, the objective of developing domestic LLMs, as outlined in the National AI Strategy, should be accelerated, and a legal framework should be established to ensure that public data are used ethically and securely in the training of domestic models.

Furthermore, ongoing efforts in the defense industry led by institutions such as ASELSAN, HAVELSAN, and TUSAŞ should be expanded, and national chip and semiconductor initiatives should be actively supported.

4.2. Adaptive and Dynamic Regulations

Static legislative models are insufficient for rapidly evolving technologies such as AI. In the future, dynamic regulatory models are expected to become increasingly prevalent.

Türkiye's current regulatory framework is primarily shaped by the KVKK, Information and Communication Technologies Authority (BTK) regulations, sector-specific guidelines on personal data processing, and the National AI Strategy. However, the rapid transformation of AI necessitates a more dynamic, risk- and opportunity-based regulatory approach. The national strategy also includes objectives aimed at updating the legal infrastructure governing AI.

In Türkiye, the establishment of a "National AI Regulation and Supervision Authority" or a "Dynamic Regulation Office" would enable the development of a risk-based, sector-specific, and continuously updatable framework. In particular, the creation of regulatory sandboxes in high-risk sectors such as finance, healthcare, transportation, and justice would support the safe innovation of domestic AI startups. Additionally, updating the KVKK to expand its scope to cover algorithmic decision-making, automated processing, and model transparency is necessary. By adopting an approach aligned with the EU AI Act, Türkiye can enhance its global competitiveness.

4.3. Infrastructure Security and Oversight of Large-Scale Models

As advanced AI models increase societal risks, stricter testing requirements and international oversight mechanisms are expected to become more prominent.

Large-scale AI models have the potential to generate significant impacts on national security, public order, and critical infrastructures. For this reason, the National AI Strategy emphasizes the necessity of security and ethical testing. While Türkiye's investments in cybersecurity, such as the structures of USOM, SOME (*Ulusal Siber Olaylara Müdahale Merkezi - USOM*, 2026), and cybersecurity solutions developed by HAVELSAN (*HAVELSAN - Siber Güvenlik*, 2026), provide an important foundation in this domain, specialized testing laboratories are required specifically for large-scale models.

A "National AI Security Testing Center" (Red Teaming & Validation Lab) should be established. This center could operate through the cooperation of institutions such as TÜBİTAK BİLGEM, BTK, USOM, ASELSAN, and HAVELSAN. Its responsibilities should include hallucination and security testing of large language models, robustness assessments against adversarial attacks, algorithmic transparency testing, autonomous system security evaluations, and ethical and fundamental rights impact assessments.

In addition, compliance with ISO/IEC 42001 and the submission of an independent audit report should be made mandatory for AI systems procured for public-sector use.

4.4. Public–Private Sector Collaboration

AI governance has evolved into a field that requires the joint efforts of governments, the private sector, and international organizations, rather than being the sole responsibility of states.

Türkiye's National AI Strategy defines public–private collaboration as a strategic objective and particularly recommends the development of joint platforms in sectors such as healthcare, agriculture, manufacturing, energy, and transportation. One of Türkiye's key strengths in this context is its access to extensive public datasets and a governance structure capable of rapid decision-making.

In Türkiye, the establishment of a "National AI Ecosystem Consortium (TR-AI Alliance)" could be considered. This structure should carry out joint research and development projects among universities (such as Ankara Yıldırım Beyazıt University, Middle East Technical University, and Istanbul Technical University), major technology companies (ASELSAN, HAVELSAN, TUSAŞ), startups, and public institutions (TÜBA, TÜBİTAK, YÖK, BTK, and the Ministry of Health).

Public procurement mechanisms should be updated to incentivize domestic AI solutions, and tax incentives for startups operating in the AI domain should be expanded.

4.5. Societal Participation and Democratic Oversight

In the future, trends toward increased citizen participation, strengthened ethics committees, and enhanced public oversight mechanisms are expected to become more prominent.

In domains where AI systems directly affect citizens' lives, such as social welfare, justice, education, and healthcare, Türkiye needs to reinforce its transparency and accountability mechanisms. The National AI Strategy emphasizes that the trustworthiness and ethical use of AI should be grounded in a society-centered approach.

A "Public Algorithm Transparency Platform" should be developed. Through this platform, public institutions should publish information about the algorithms they use, including impact assessments, transparency reports, data sources, and performance metrics. In addition, a National AI Ethics Council involving civil society organizations, academia, and citizen representatives should be established. This body should work in interaction with European ethics committees to ensure the continuous updating of Türkiye's ethical standards.

Furthermore, regular independent audits should be conducted for algorithms used in social welfare, education, and justice services.

4.6. Strengthening the Human-Centered and Ethical Perspective

Beyond technological benefits, the integration of human rights and ethical values into all AI processes will constitute one of the most important governance strategies of the future.

Ethical governance represents a core component of Türkiye's vision for developing trustworthy AI. While the KVKK provides an important legal foundation for the protection of personal data, AI-specific ethical challenges, such as the risk of discrimination, the justification of automated decisions, and violations of autonomy, require a more comprehensive approach. The National AI Strategy emphasizes that ethical principles should be integrated into all stages, ranging from education to certification.

In Türkiye, an "AI Ethics Impact Assessment" could be made a legal requirement. In high-risk domains such as healthcare, finance, justice, and transportation, AI systems should not be deployed without the completion of such assessments. AI Ethics Certification Programs should be introduced for universities and public institutions, and ethics modules should be made mandatory within engineering curricula.

In addition, a "National AI Carbon Monitoring Program" could be launched to track the environmental impacts of large-scale models. Environmentally friendly methods that support sustainability, such as the energy optimization techniques proposed by (Tasdelen, 2025) and (Ji & Jiang, 2026), should become standard practices in both the public and private sectors throughout the design, development, and deployment of AI models.

Conclusion

The rapid integration of AI systems into nearly every domain of life has made it imperative to address these technologies not merely as technical tools, but through their economic, societal, ethical, and governance dimensions. The framework presented in this chapter demonstrates that governance, accreditation, ethical principles, and trust mechanisms are complementary components that collectively enable the trustworthy design, deployment, and societal interaction of AI systems. AI therefore requires a socio-technical infrastructure grounded not only in technical optimization processes, but also in normative values such as transparency, accountability, explainability, and sensitivity to human rights.

AI governance can no longer be understood as a linear or static process. The rapid evolution of models, changes in data sources, contextual risks, and emerging security threats necessitate governance mechanisms that are dynamic, proactive, multidimensional, and embedded across the entire AI lifecycle. In this context, all stages of model development, testing, deployment, monitoring, and updating should be supported by technical standards, ethical assessments, risk and opportunity

analyses, and stakeholder participation. Recent initiatives by international standardization bodies reflect this trajectory, as comprehensive AI management systems, independent audit models, and compliance frameworks are becoming increasingly influential.

From a Türkiye-specific perspective, the findings of this chapter indicate that the coming period will require decisive institutional, regulatory, and infrastructural actions. Strengthening AI sovereignty and digital independence through national data spaces, domestic large language model development, and reduced reliance on external critical digital components will be central to this effort. The establishment of dynamic and risk-based regulatory mechanisms, including regulatory sandboxes and a dedicated national AI regulatory authority, will be essential to keep pace with technological change. In parallel, the creation of national AI security testing and validation centers for large-scale models, mandatory compliance with standards such as ISO/IEC 42001 in public procurement, and independent audit requirements will play a key role in safeguarding national security and public trust. Furthermore, institutionalized public–private collaboration, expanded support for domestic AI startups, enhanced societal participation and democratic oversight, and legally mandated AI Ethics Impact Assessments will be critical for ensuring that AI development in Türkiye remains human-centered, transparent, and socially legitimate. These measures collectively position AI governance not only as a regulatory necessity, but also as a strategic instrument of digital sovereignty and long-term national competitiveness.

The future-oriented perspectives discussed in this chapter indicate that AI governance is evolving into a strategic policy domain. Digital sovereignty, the establishment of national data spaces, oversight of large-scale models, institutionalized public–private collaboration, enhanced societal participation in decision-making processes, and the integration of ethical values across all technical stages will constitute the core governance priorities of AI ecosystems at both national and international levels. These trends aim not only to mitigate risks, but also to ensure that AI develops as an inclusive, sustainable, and socially beneficial technology.

In conclusion, while AI technologies offer significant opportunities for humanity, they also generate substantial responsibilities. Trustworthy AI can be transformed into genuine societal benefit only when technical excellence is reinforced by ethical responsibility, democratic values, the protection of human rights, participatory policies, and continuous oversight mechanisms. The governance framework developed in this chapter articulates the key principles required to build sustainable, trustworthy, and human-centered AI ecosystems at both national and global scales, and provides guiding insights for the design of future AI policies.

Acknowledgements

I would like to express my gratitude to TÜBA, the T3 Foundation, and the KÜME Foundation for providing me with the opportunity to contribute to this chapter. I also appreciate the scholars, colleagues, and institutions whose work, along with the available research, reports, and policy documents on AI governance, ethics, and trustworthy systems, provided a strong foundation for the analysis presented here. Additionally, I am grateful to Türkiye's academic and policy communities for the valuable context they provided. Any remaining errors or interpretations are my own.

References

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications 2024 11:1*, *11*(1), 1568-. https://doi.org/10.1057/s41599-024-04044-8

AI principles | OECD. (2025). Retrieved November 27, 2025, from https://www.oecd.org/en/topics/ai-principles.html

AIAAIC - AIAAIC Repository. (2025). Retrieved November 18, 2025, from https://www.aiaaic.org/aiaaic-repository

Artificial Intelligence Incident Database. (2025). Retrieved November 18, 2025, from https://incidentdatabase.ai/

Artificial Intelligence-Enabled Medical Devices | FDA. (2025). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, *5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Benraouane, S. A. (2024). *AI Management System Certification According to the ISO/IEC 42001 Standard*. Routledge. https://doi.org/10.4324/9781003463979

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10), 100857. https://doi.org/10.1016/J.PATTER.2023.100857

Facial Recognition Leads To False Arrest Of Black Man In Detroit : NPR. (2025). https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Hadzovic, S., Becirspahic, L., & Mrdovic, S. (2024). It's time for artificial intelligence governance. *Internet of Things (Netherlands)*, *27*. https://doi.org/10.1016/j.iot.2024.101292

HAVELSAN - Siber Güvenlik. (2025). Retrieved March 6, 2026, from https://www.havelsan.com/tr/siber-guvenlik-hizmetlerimiz

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, *4*(4), 799–819. https://doi.org/10.1109/TAI.2022.3194503

IEEE SA - IEEE 7000-2021. (2026). Retrieved March 6, 2026, from https://standards.ieee.org/ieee/7000/6781/

ISO/IEC 23894:2023 - AI - Guidance on risk management. (2026). Retrieved March 6, 2026, from https://www.iso.org/standard/77304.html

Ji, Z., & Jiang, M. (2026). A systematic review of electricity demand for large language models: evaluations, challenges, and solutions. *Renewable and Sustainable Energy Reviews*, *225*, 116159. https://doi.org/10.1016/J.RSER.2025.116159

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. In *Patterns* (Vol. 2, Number 9). Cell Press. https://doi.org/10.1016/j.patter.2021.100314

Kelly, A. (2021). A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020. *British Educational Research Journal*, *47*(3), 725–741. https://doi.org/10.1002/BERJ.3705

Meta: Systemic Censorship of Palestine Content | Human Rights Watch. (2025). https://www.hrw.org/news/2023/12/20/meta-systemic-censorship-palestine-content

MIT AI Incident Tracker. (2025). Retrieved November 18, 2025, from https://airisk.mit.edu/ai-incident-tracker

OECD AI Incidents Monitor. (2025). Retrieved November 27, 2025, from https://oecd.ai/en/incidents

Olsson, R. (2007). In search of opportunity management: Is the risk management process enough? *International Journal of Project Management*, *25*(8), 745–752. https://doi.org/10.1016/J.IJPROMAN.2007.03.005

Rachovitsa, A., & Johann, N. (2022). The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case. *Human Rights Law Review*, *22*(2). https://doi.org/10.1093/HRLR/NGAC010

Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: barriers and pathways forward. *International Affairs*, *100*(3), 1275–1286. https://doi.org/10.1093/ia/iiae073

Ryan, M., & Stahl, B. C. (2021). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, *19*(1), 61–86. https://doi.org/10.1108/JICES-12-2019-0138

Sunmola, F., Baryannis, G., Tan, A., Co, K., & Papadakis, E. (2025). Holistic Framework for Blockchain-Based Halal Compliance in Supply Chains Enabled by Artificial Intelligence. *Systems 2025, Vol. 13, Page 21*, *13*(1), 21. https://doi.org/10.3390/SYSTEMS13010021

Tasdelen, A. (2025). Enhancing Green Computing Through Energy-Aware Training: An Early Stopping Perspective. *Current Trends in Computing*, *2*(2), 108–139. https://doi.org/10.71074/CTC.1594291

Teo, S. A. (2023). Human dignity and AI: mapping the contours and utility of human dignity in addressing challenges presented by AI. *Law, Innovation and Technology*, *15*(1), 241–279. https://doi.org/10.1080/17579961.2023.2184132

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

Ulusal Siber Olaylara Müdahale Merkezi - USOM. (2026). Retrieved March 6, 2026, from https://www.usom.gov.tr/

Uzair, M. (2021). Who Is Liable When a Driverless Car Crashes? *World Electric Vehicle Journal 2021, Vol. 12, Page 62*, *12*(2), 62. https://doi.org/10.3390/WEVJ12020062

Winter, P. M., Eder, S., Weissenböck, J., Schwald, C., Doms, T., Vogt, T., Hochreiter, S., & Nessler, B. (2021). *Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications*. https://doi.org/10.48550/arXiv.2103.16910

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, *39*(4). https://doi.org/10.1016/j.giq.2022.101685

Wodi, A. (2024). Artificial Intelligence (AI) Governance: An Overview. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4840769

About the Author

Asst. Prof. Dr. Abdulkadir TAŞDELEN
Ankara Yıldırım Beyazıt University (AYBU) | abdulkadirtasdelen[at]aybu.edu.tr
ORCID: 0000-0003-4402-1463

Dr. Abdulkadir TAŞDELEN received his Ph.D. in Computer Engineering from the Institute of Natural and Applied Sciences at Ankara Yıldırım Beyazıt University. He has authored numerous publications in the fields of computer science and bioinformatics. His research primarily focuses on artificial intelligence and deep learning applications, particularly in areas such as pre-miRNA classification and COVID-19 analysis. In addition, he has presented his work at various international conferences and contributed to several funded research projects. His recent studies also explore blockchain technologies and green computing. Dr. TAŞDELEN has previously worked as a postdoctoral researcher at the Faculty of Engineering, International Islamic University Malaysia (IIUM). Alongside his academic career, he has held various administrative positions, including serving as the Deputy Dean of the Faculty of Engineering and Natural Sciences at Ankara Yıldırım Beyazıt University. He has also worked as an advisor at the Council of Higher Education (YÖK) and continues to serve as an advisor to the Turkish Academy of Sciences (TÜBA).