Chapter 6

Making the Transition from Narrow-to Superintelligence: The Need for a Conceptual Framework for the Ethical Development of Artificial Intelligence

Hakan Kaya[1]

Abstract

The art of artificial intelligence (AI) technologies is going through various layers of evolution, which include narrow AI (ANI), general AI (AGI), and eventually super AI (ASI). This chapter will examine the evolution of these AI layers, using various approaches that reveal the differences among each of the AI layers, specifically within the realm of computer engineering. Following the prevailing literature on the subject, the analysis will specifically examine the data-oriented, narrow, and specific approach of ANI, the high-level, flexible approaches of AGI that relate to learning, and the optimizing, transdisciplinary approach of ASI, along with the superhuman characteristics of the latter. Also, the analysis will examine the ethical threats of AI, the international frameworks of AI governance, such as the EU AI Act, OECD, and UNESCO, and the strategies of Turkey. Lastly, the chapter will examine the prevailing emerging trends of AI.

1.Introduction

The growing pace of advances in artificial intelligence technologies is revolutionizing the computer engineering field, both theoretically and practically. Though it is a fact that the current dominant systems are of the ANI type, the level of progress reported in the literature suggests that higher-order levels of AGI and ASI have moved beyond the conceptual level. This is not only important for design but also impacts the areas of data architecture, computer model design, processing, memory, and compliance.

1 Dr., T.C. Kestel Belediyesi, hakankaya@kestel.bel.tr, ORCID ID: 0000-0002-0812-4839

https://doi.org/10.58830/ozgur.pub1148.c4770

This paper will examine the historical concepts of artificial intelligence, the current technological approaches of AI, and the emerging risk of AGI and ASI. Additionally, the topic of AI governance in the context of the AI Act and the control problem will be analyzed. The current state of AI technology in the context of technology policies, human capital, and competitiveness of Turkey will also be discussed.

2. Literature Review

The progress of AI represents a technological revolution, wherein the very basics of computer engineering, focusing on the design of systems, optimizing algorithms, data processing, and security, are revolutionized. This literature review integrates the development of AI, covering the entire spectrum of AI, ranging from ANI to ASI.

2.1. Historical Foundations and Evolution of Concepts

The history of artificial intelligence traces its roots in the theory of computation, which is a fundamental aspect of computer engineering. Good (1966) brought about the idea of the "ultraintelligent machine," which asserted the possibility of improving systems that surpass the human intellect, thus forming the foundation for recursive improvement computer engineering algorithms. Yudkowsky (2008) suggests that the consideration of AI systems should also take into account the perspective of a risk factor for the global community, focusing on control and alignment that affect the contemporary computer engineering of corrigibility. Bostrom (2014) elaborated how the process of superintelligence will transpire, categorizing it into various types, risks, and strategies that pave the way for the computer engineering of hybrid neural-symbolic systems. Brynjolfsson & McAfee (2014) elaborated upon the concept of the "second machine age," focusing on how AI affects the labor force and productivity. This is analogous to the use of data-driven optimization (gradient descent) and huge data processing, which is the aspect of computer engineering. Domingos (2015) wrote about the search for the "master algorithm," claiming that learning systems will transform the world. This paved the way for a computer engineering paradigm of meta-learning (MAML, for example). Tegmark (2018) questioned the definition of humanity in the age of AI, focusing on the control problem of AI systems that are safe for humans, also suggesting the use of value alignment computer engineering. In the modern age, there have been various predictions among AI experts about the time that will transpire for the transition to AGI and ASI (Grace et al., 2018), as well as how the passage toward the achievement of these goals will be measurable (Chojecki, 2025).

2.2. Current Advances: Technical Architecture and Applications of ANI to AGI

Modern literature is about the advent of AGI using transformers and hybrid approaches. A roadmap of AI development is discussed by Vsevolodovna (2025), where deep neural networks (input-hidden-output layers, cross-entropy loss) typify ANI, multi-modal inputs typify AGI (LSTM/GRUs), and quantum integration is a topic of interest in ANI. The use of transformers speeds up the ANI-AGI transition, for example, with the BERT and GPT models. Chojecki (2025) develops the so-called Kardashev scale, where the AAI-Index is calculated on ten criteria, including autonomy, generality, and planning. Mirindi et al. (2025) examine the use of the O1 model of OpenAI, focusing on the development of ASI using reinforcement learning, with ethical goals for integration, where the use of inverse reinforcement learning for engineering goals is significant. Groppe & Jain (2024) discuss AI-complete problems, suggesting hybrid approaches of symbolic-statistical systems (KG & ML) for better problem-solving, where neuro-symbolic AI using contextual computation is proposed. For applications, Tiwari et al. (2025) discuss the use of deep learning & ML for plant identification, stating that recognition and extraction of specific features using image analysis software demonstrates data bias and the need for high-resolution images for engineering. Alvarez-Teleña & Díez-Fernández (2024) discuss the use of ANI for the development of ASI, where the role of algorithmization & efficiency-engineered approaches helps engineering for corporate competitiveness. Application-wise, Chapman et al. (2024) discuss the relevance of Responsible Intelligence (RI) analysis for the healthcare and spinal surgery sectors, where the expansion of large amounts of data & software incapability forces the need for real-time analysis for engineering. Surdu & Surdu (2024) discuss the relevance of AI development for neonates, differentiating between hype & reality, where current ANI engineering for specific tasks, for example, HeRO sepsis identification, offers benefits but face contextual challenges.

2.3. Challenges and Risks: Technical and Ethical Dimensions

Technological challenges in AI revolve around bias, security, and alignment. Fahad et al. (2024) elaborate on the positive applications of AGI (security, healthcare, finance) and the negative impacts (decision-making, privacy), focusing on ethical considerations for the engineer. Doker (2025) applies the Rittel-Webber definition of the "wicked problem" of AGI for the engineer, keeping the control of the agent's development dynamic and controlled. The authors of Goyal et al. (2024) explore the challenges of building ASI, specifically regarding the engineer's consideration of scale and unpredictability. Marcus (2018) finds that deep learning limits the engineer's model improvements, referencing opacity/black box limitations and the requirements of better methodological approaches for the engineer. Ueda et al. (2025) illustrate the alignment problem of healthcare for the engineer, debating the consequence of reward hacking/bias (for example, see CheXNet external failure) for the engineer. Mokoena (2025) explores the challenges of the "extinction problem" of ASI for the engineer, suggesting centralized control of the agent and open-source software limitations. Finally, Grace et al. (2018) provide data for expert predictions regarding the agent's probable internal enhancement of the AGI's effectiveness for the engineer, suggesting time-based solutions for the engineer. For ethical components of the topic, Tzimas (2021) explores the ethics of AI, applied specifically in the study of international law for the engineer, examining the problems of the black box, the need for transparency, and the need for the model's correctability for the engineer. Lastly, examining the philosophical standards, the authors of Ekici (2025) explore the need for integrated personalities for the engineer, examining the modular components of the AGI/ASI's cognitive constituents. Additionally, Li (2023) explores the relationship between the model of panentheism and the AGI for the engineer, specifically the ethical implication of the designer's role regarding the agent's later use for the engineer. Also, Gal-Or (2025) connects the philosophy of AI with the Aristotelian "nous"; the design of human-like intelligence in engineering is discussed.

2.4. Governance and Future Directions

Governance frameworks condition engineering standards. The European Commission (2024) offers the AI Act risk management strategy (forbidden, high risk, transparency); robustness, cyber security, and traceability are obligatory in engineering. OECD (2019) explores the AI principles implementation strategy; management of the AI life cycle, transparency, accountability, and risk management are used in engineering. UNESCO (2021) suggests ethical guidelines, focusing on ethical impact assessment (EIA) and data policies that link the integration of sustainability and gender equity into engineering. Future challenges for the profession, according to Glenn (2025), will emphasize AGI, focusing on engineering governance and cooperation on research. Frameworks for AI use and autonomous business levels, suggesting automation vs. augmentation, are proposed by

Sohn (2024) for engineering. Soft skills and ethics training for computer engineering, and competencies for AI development, are anticipated by Araújo et al. (2025). This integration highlights the computer engineering of the AI-oriented future, balancing innovation with ethical considerations. Abdullah (2025) and Anakotta (2025) also support database searches beyond the articles, attempting to access hidden sources.

3.Artificial Narrow Intelligence (ANI)

3.1. Definition of ANI

Artificial Narrow Intelligence, or ANI, also known as narrow AI, are systems that are designed for a particular task. They are very efficient for tackling specific, constrained problems but not for sharing knowledge across areas. The definition of ANI can be explained by the following:

$$T = \{t1, t2, \ldots, tn\}$$

Optimization of a set of functions over a specific task space:

$$P{:}T \rightarrow [0{,}1]$$

and a metric that evaluates their success on these tasks:

$$D(train)$$

and that shows good performance mainly on a particular data distribution.

3.2. Characteristics of ANI

Specialization: ANI is specialized in one particular domain or activity, like face recognition, translation, or self-driving cars.

Data-Driven: ANI systems utilize large data sets and machine learning algorithms that provide accurate results.

Lack of Generalization: These systems do not have the capability for knowledge generalization, unlike human intelligence. This means that ANI systems are not context-sensitive. It is not necessary that the success gained in one area will also help the system succeed in another. Learning processes support supervised and semi-supervised learning, and these systems do not possess real understanding. They also fail when faced with unexpected circumstances.

3.3. Applications of ANI

ANI is the foundation of the digital reality.

Consumer Technologies: Siri, Alexa, Google Assistant (voice commands), Google Translate, Netflix & Spotify recommendation systems, facial recognition software, spam filters.

Industry & Science: Tumor identification using images, recognizing objects for autonomous cars, identification of stock fraud, robotics for factories, weather modeling software, computer systems better than humans at playing chess and Go (Deep Blue, Alpha Go).

Technologies that Work for ANI: ANI operates on the engines of Machine Learning (ML) technology, and also on Deep Learning (DL), which is a subset of the former. Both of these "learn" from large datasets, and their success has been driven by powerful neural networks, the availability of GPU computing, and large data sets.

Strengths of ANI are reliability and the ability to scale, leading to revolutions in the healthcare industry (for example, diagnosis assistance for the radiology department) and the finance industry (for example, fraud analysis), among others. However, the weakness of ANI is that it is not capable of managing uncertainty and learning the way a person can between tasks. It also poses substantial threats, socially and ethically, with regards to displacement, data representation, privacy, and accountability. ANI can be termed a technology that releases the intended results of the developers or users. By the year 2025, it is expected that a large percentage of the commercial use of AI will come under the ANI category.

4.Artificial General Intelligence (AGI)

4.1. Definition of AGI

Artificial General Intelligence, also known as "strong AI," is a major improvement over ANI. In contrast to ANI, AGI will not be limited to a specific area of expertise. Instead, it will have the ability to reason, plan, and respond appropriately in new situations. To state the matter differently, AGI will have the ability to understand, learn, adapt, and accomplish any intellectual task that a human can perform. This includes problem-solving, common sense, planning, knowledge transfer (applying knowledge gained from one area of study for use in other areas), and a deep understanding of natural language. An AGI system can learn about physics concepts from books and invent new experiments, write a piece of fiction, or learn how to behave correctly in a new and complex social situation. AGI is a trans-task learning mechanism that allows information transmission between very different tasks, regardless of the data distributions of the training data:

$$\Phi(ti \rightarrow tj)$$

This function is defined by the expected value, where the higher the expected value, the better the system's adaptability for learning new tasks with less training.

4.2. Characteristics of AGI

Human-like Versatility: It will be able to do any intellectual task that a human can, such as complex math problem-solving, writing, or emotional conversations.

Self-Improvement: It could enhance itself through experience, perhaps achieving swift progress even without specific programming.

Contextual Understanding: AGI could understand the nuances of, for example, sarcasm, cultural differences, or moral dilemmas.

4.3. Technical Challenges in Achieving AGI

As of 2025, the state of AGI is mostly conceptual, but promising developments are indicated by the advances made in large language models, for example, Grok 4. Entities like xAI and OpenAI are also collaborating on the development of multimodal AI that can handle text, images, and code. However, the implementation of AGI will require the following steps:

Computational Complexity: The development of a system that simulates human-like cognition requires sophisticated algorithms and heavy computing. To attain AGI, not only is computing needed, but the concepts of consciousness, awareness, emotional intelligence, and causal inference need also to be incorporated. It not only needs computing but also the collaboration of various fields of study.

Ethical & Social Issues: The development of AGI poses several ethical and social challenges. Among these, the question of control, alignment, and unpredictability is central, discussed in relation to the development of superintelligence, for example, by Bostrom (2014).

Modeling Human Intelligence: The major problem lies in building a model that imitates human knowledge, feelings, and creativity.

4.4. Potential Applications of AGI

Even though the development of AGI is still a distant vision, the Research and Development: Systems that offer the capability for scientific discovery and problem-solving that cross several disciplines.

Tutoring and Education: Intelligent tutors and educators involved in the learning process.

Organizational Management: These are management systems that promote business strategy and facilitate smart decision-making.

Human-Machine Collaboration: Enabling Intelligent, Natural Interactions between Humans and Machines.

4.5. Future Outlook for AGI

Following the rapid advances made by deep learning and AI algorithms, scientists are well on their way to achieving AGI. While it will take a considerable amount of work, the level of investment and existing research activity, including the efforts of major corporations, suggests that there is substantial interest. On the question of when AGI might happen, estimates range from a decade, through potentially a century or more. The potential for the development of AGI gives rise to very significant ethical and security-related considerations, including alignment, loss of control, and the potential for transformation of economic and social systems.

5.Artificial Super Intelligence (ASI)

5.1. Definition of ASI

Artificial Super Intelligence, thereafter referred to by the acronym ASI, refers to systems that surpass the level of intelligence, competencies, and capabilities of the human intellect. ASI not only rivals the level of human expertise both generally and specifically but also surpasses the limits of creativity and problem-solving. According to Bostrom (2014), ASI is defined as a level of intelligence that is incomparably superior to the current level of human intelligence across almost all domains of interest. This meta-learning model operates at a higher level of functionality that transfers tasks and changes parameters of the model on its own (self-modifying model), given by the equation:

$$Mself: \theta t \rightarrow \theta t + 1$$

5.2. Characteristics of ASI

Singularity: The idea that the occurrence of ASI marks the onset of an irreversible point in history, where events that will happen thereafter will not be computable by the human brain.

Exponential Superiority: ASI was capable of processing data at speeds and volumes that simply couldn't be handled by a human, potentially tackling problems like quantum computing and interstellar travel on timescales that were effectively negligible.

Self-evolving: This system will independently alter the code, and this could give rise to the "intelligence explosion" phase where growth will accelerate.

Global Significance: ASI was capable of handling complex systems, like economies and biological systems, very efficiently.

5.3. Ethical and Philosophical Challenges of ASI

ASI is seen as the most promising and potentially very dangerous development of humanity, which is capable of solving current challenges (cancer, aging, climate change) and, at the same time, poses serious threats (misalignment, control loss). The advent of ASI challenges the role, freedom, and authority of humanity in the universe. The issue of developing ASI raises several ethical, philosophical problems:

Control and Oversight: It is important to ensure that the ASI system is operating for the good of mankind.

Accountability: Who is responsible for an incorrect decision rendered by the ASI?

Effect on Employment: The use of ASI, which substitutes for human labor, will significantly affect the labor market.

Power and Influence: The achievement of higher intellectual power by ASI has several implications regarding the structure of society and international politics.

However, the level of risk that exists is:

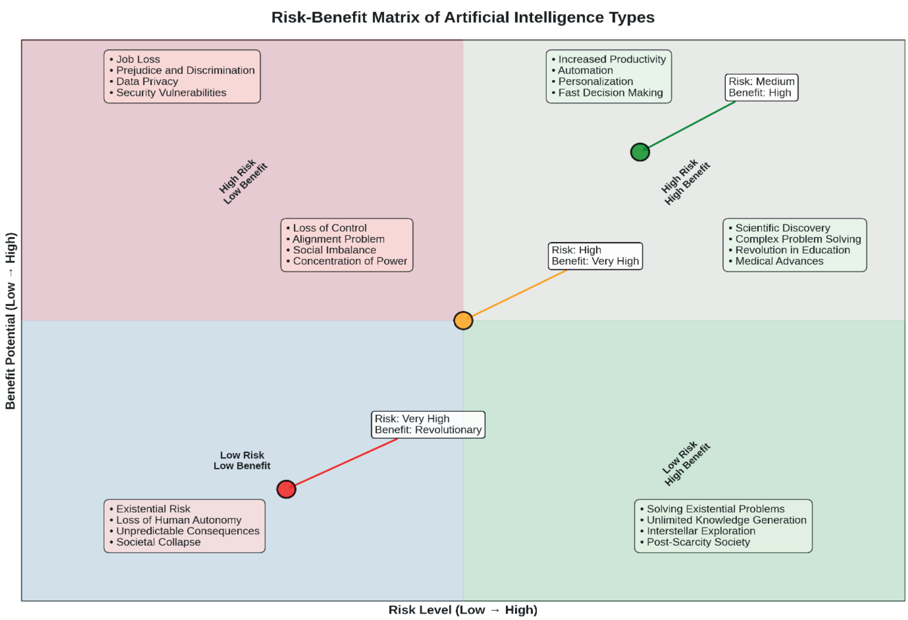
Alignment Problem: It is important for the goals of the ASI to be well-aligned with human goals, so that it doesn't happen that, for example, it starts

Control Issues: The risk of failure for control of the existing ASI, potentially leading to situations depicted in science fiction media, such as the film "The Matrix." Ethical Dilemmas: Issues of consciousness, the question of rights applicable to the ASI, or the potential for inequality between those who have access and those who do not.

5.4. Future Prospects of ASI

In spite of the presence of numerous challenges, experts are of the view that ASI will, in the long run, bring a revolution regarding technology, science, and art. To achieve the objective of creating ASI, it will require continued scientific efforts and collaboration on the ethical front. It is essential to bear in mind that, presently, the existing literature mainly concentrates on the challenges of the intelligence gap between ANI and AGI, which is depicted thoroughly in Figure 1.

Figure 1. Risk-Benefit Matrix of AI Types

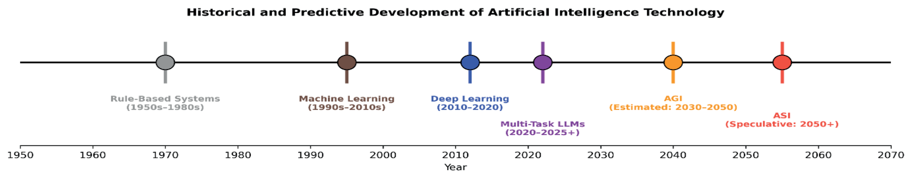


6.Analytical Comparison of Fundamental Differences

The essential differences between ANI, AGI, and ASI are not merely technological features, but also include epistemic capabilities, learning strategies, and societal effects.

Figure 2 traces the history of AI development since the 1950s, together with a predicted curve of expert estimates for 2030-2070, showing how the period of ANI Supremacy will give way, potentially, to the advent of AGI and the likelihood of the ASI Singularity.

Figure 2. Historical and Predictive Development of Artificial Intelligence Technology



6.1. Information Processing and Generalization Capabilities

ANI consists of expert systems that specialize in a very specific task. These systems operate very efficaciously within their given constraints, for example, recognizing images but failing to compose music. This again, also known as "learned helplessness," stems from the fact that ANI has limitations regarding the nature of their overall generalization. Once the input exceeds the given learning limits, the system collapses. This particular problem is somewhat the opposite for AGI, where the definition lies in the fact that it is a very flexible system. Much like the workings of the human brain, AGI systems use mathematical logic concepts solved elsewhere for analyzing social situations, for example, or using the underlying principles of physics for deriving engineering solutions. Essentially, the very definition of AGI lies in how it adapts expert systems learned for mapping complex conceptual frameworks elsewhere. Essentially, for AGI, the very idea of generalization also fails. This is because for AGI, problem-solving power potentially transcends the sum total of existing human knowledge, often leading to radically innovative, unforeseen approaches.

6.2. Systematic Evaluation of Advantages and Limitations

For better understanding of the spectrum, a comparison of ANI, AGI, and ASI on various parameters is given below in Table 1.

Table 1. Comparative Analysis of ANI, AGI and ASI

Aspect	ANI (Narrow AI)	AGI (General AI)	ASI (Super AI)
Scope	Defined for specific tasks only	General, human-like for various tasks	Exponential, beyond the capabilities of humans
Learning Ability	Rule-based or supervised learning	Adaptive, unsupervised learning	Self-improving, recursive evolution
Examples	Spam filters, image recognition	Theoretically projected: Versatile robots	Empirically unmapped: World-optimizing entities
Current Status	Widespread in use	In development, not achieved	Speculative, far-future
Risk Level	Low (manageable errors)	Medium (job loss, bias)	High (existential threats)
Benefits	Efficiency in specific fields	Ability to solve complex human problems	Radical societal transformation

Such differences highlight the evolution of tools into partners and subsequently into superior beings. ANI will enhance humans, AGI will work along with them as a peer, and ASI will potentially reshape the fabric of existence.

6.2.1. ANI Advantages and Disadvantages

Advantages: ANI systems provide excellence in operations by outperforming humans in areas like medical diagnosis and trading algorithms. ANI systems are less complex and economical for implementation, allowing for quick decision-making and responses.

Disadvantages: Fragility is a significant weakness. Even slight deviations in the training data (distribution drift) or adversarial attacks will result in performance damage. They are context blind. For example, they will mislabel a literary piece as spam. They can also be biased, using biased data for training.

6.2.2. Potential Advantages and Challenges of AGI

Advantages: AGI provides adaptability, which will help it generalize over various tasks with less retraining. The problem-solving skills that AGI integrates into one entity could provide systems-level solutions for complex worldwide problems, for example, climate change. The potential of AGI extends to learning and scientific investigation.

Challenges: There are still challenges on the technological front. It is not possible yet to computationally model the essential components of human cognition, that is, consciousness, common sense, and causality. The ethical and control problems, on the other hand, are very deep, including AGI value alignment, accountability, and resource inequality.

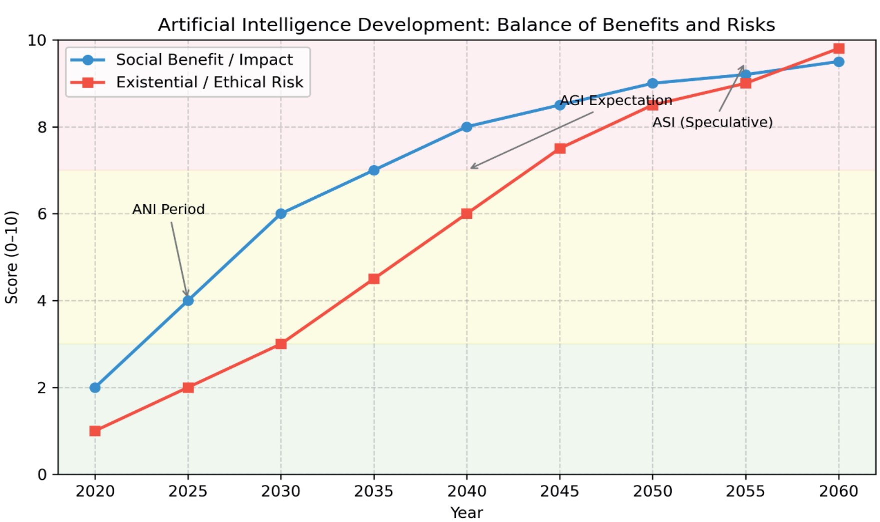
6.2.3. Existential Opportunities and Risks of ASI

Opportunities: ASI poses include the solution of existence-related questions, for example, life extension, overcoming energy crises, and facilitating the exploration of other star systems. It may bring about creativity and advancements in science, technology, and philosophy over the sum of the entire history of mankind.

Risks: Control is the paramount risk, and it is not guaranteed that it is possible to control a super-intelligent entity. The problem of misalignment exists when an ASI, even if well-meaning, aligns itself with very narrowly defined goals that are not beneficial for mankind. The threat posed by the technological singularity exists because mankind could potentially give up their freedom and meaning to the artificial entity.

In Figure 3, the balance matrix highlights the key risks and benefits of ANI, AGI, and ASI for scientific, existential, and efficient problem-solving purposes. This figure reveals that the benefit–risk gradient continues rising with the progress of technological advancements.

Figure 3. AI Development for Balance of Benefits and Risks

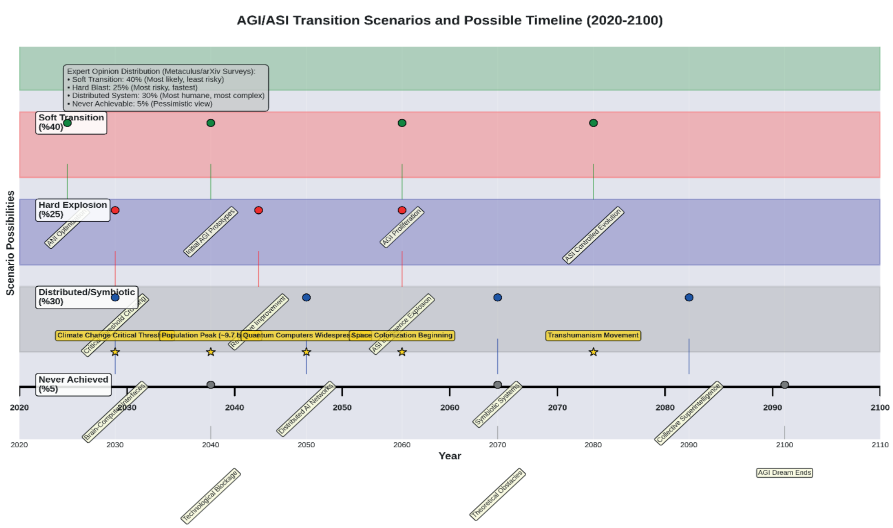


6.3. Practical Application Areas and Future Perspectives

Areas where the applications of ANI are growing at a rapid pace include the healthcare industry (diagnostics, drug development), the finance industry (fraud analysis, algorithmic trading), manufacturing (smart robotics, quality analysis), and transportation systems (autonomous cars). Looking ahead, the next ten years will witness the convergence of ANI systems, along with edge computing, into mainstream life. Its applications are only speculative at this stage, although some initial innovation in the form of large language models, which have only minimal generalization skills, and progress toward universal control of robots, are the forerunners of AGI. Primarily, it will initially appear for very specific domains, specifically concerning scientific innovation. The nature of ASI makes it impossible for accurate practical predictions. It can potentially affect areas like the optimization of global governance, paradigm shift in fundamental science, and improved comprehension of consciousness. Nevertheless, it is asserted that the advancement of ASI will depend on the successful development of AGI with the recursive self-improvement trait.

Figure 4 illustrates how the roadmap for transitioning ANI to ASI can be multi-layered, focusing on phases like the development of transformer-based architecture, integration of neuro-symbolic systems, multimodal learning, recursive self-improvement, and alignment of values.

Figure 4. Transition Dynamics of Roadmap from ANI to ASI



Current status and future perspective Currently, the state-of-the-art AI labs (OpenAI, DeepMind, Anthropic, etc.) are still at the level of ANI. However, the current state of large language models (LLMs) starts showing signs of cross-task generalization, which marks the very beginning of the pursuit of AGI, since they already have the capability for text generation, code, problem-solving, and also the capability for limited planning, for example, GPT-4 or Claude. However, they do not have genuine understandings, causality, or consciousness. It is not enough to have large data and large models to obtain AGI. Certain challenges need to be addressed, which include representation, interactive learning, interaction with the environment, and alignment of goals. The future of artificial intelligence is not only about success but also deeply embedded within the formulation of the smart world that mankind is trying to build. ANI is omnipresent, but AGI can potentially appear between the next 10-30 years, and the issue of ASI is much speculative. With these advances, it is necessary for the academia, the government, and technology experts to come together. It specifically defines the need for countries like Turkey, which belong to middle- or lower-income countries, because while richer countries place equal importance on AGI race dynamics, for example, the USA vs. China, it is a matter of doing well on the ANI front for the transformation of the economy, along with building strength for the management of AGI. Ultimately, every stage of AI is a attempt to transcend the shortcomings of the previous one. The specialization that ANI offers could well benefit from the generalizeability of AGI, and the human-like intelligence that AGI offers can also benefit from the larger abilities of ASI. It is important that one understands the hierarchical relationship between these three stages of AI not only for the technological purposes, but also for the frameworks that will use these abilities for good.

6.4. Turkey's Perspective and Positioning in Global Competition

The international AI environment has a major impact on a country's economic policies and technology policies. Being a country with a vibrant start-up ecosystem, a young population, and a favorable location, Turkey takes a distinct position in this race. This article evaluates the present strengths of Turkey in ANI, examines the infrastructure as well as human resource requirements for AGI and beyond, and explains how Turkey can make a distinct contribution to this debate.

6.4.1. Current Strength: Application-Oriented Growth in the ANI Ecosystem

The development of AI in Turkey has focused on ANI systems with high performance capabilities in certain tasks. The rate of development in this area is quite significant, both in terms of quantity and quality. Also, it appears that there are more than 400 local AI start-ups in Turkey, with a total value of 2-4 billion USD (TRAI, 2025). These projects are also working on developing specialized and efficient solutions, particularly in the area of defense, fintech, and smart cities. Findings related to the adaptation rate of certain sectors also reflect that of the industrial and manufacturing sector, with a rate of 76% adaptation in the Industry 4.0 revolution. In finance, security-derived tasks like protecting against fraud also showed a 62% improvement in security (Sysart Consulting, 2025).

The investment behavior of Turkish companies greatly differs from others in the world. A comparative study reveals that "operational efficiency" (87%) and "customer experience" (77%) are greatly focused upon by Turkish companies, whereas in the United States, "cybersecurity" (67%) and "risk compliance" (52%) are given priority in long-term risk management and sustainable competitive advantage areas (KPMG, 2025), which depict an "opportunity-oriented" focus on enhancing current business models.

6.4.2. Foundations of the Future: Human Resources and Infrastructure for the AGI and ASI Journey

Progress in ANI needs to be seen as a stepping stone for even more challenging goals like AGI and Artificial Superintelligence, also known as ASI. However, development in superhuman intelligence also requires some fundamental investments in other areas. The Action Plan for National Artificial Intelligence Strategy 2024-2025 published by CBDDO in 2024 remains an important starting point for such development. This plan focuses on upgrading education, research, and infrastructure.

Human Resources and Interdisciplinary Approach: The young population represents a huge potential. Nevertheless, the number of interdisciplinary courses for postgrad studies involving, for example, neuroscience, cognitive sciences, and philosophy, in addition to computer engineering, which are all bases for AGI, remains low. Education currently focuses on using tools, like "command engineering," as opposed to theoretical knowledge for AGI. Developing professional standards, as proposed in the Action Plan, will help cover this deficiency.

Technical Infrastructure and Data Ecosystem: To develop AGI, a lot of computational power has been needed for large Language Models. Estimates from Stanford University suggest that to develop a model like GPT-4, it would take around 78 million USD (Stanford HAI, 2024). Turkey argues that there is potential for improvement with regards to their supercomputers and cloud capabilities for such research. On a different note, AGI's "Central Public Data Space" as defined by the Action Plan benefits AGI research with easy access to high-quality large data sets. This would provide a significant benefit for Turkey in this area of competition.

6.4.3. An Original Voice on the Global Stage: Contribution to Ethics and Governance

The existential risks posed by AGI and ASI necessitate global cooperation. Turkey has the potential to be an active participant and even a leader, not a passive observer, in these discussions.

Ethics – Cultural and Religious Perspectives: So far, the ethical discourse in AI has been shaped mainly by Western paradigms. The Islamic ethical tradition of Turkey, stress on community responsibility, and willingness to synthesize Eastern and Western approaches offer a fresh perspective for developing a "multicultural AI ethics." Thus, for example, recommendations for applying the ideas of "justice" and "transparency" in AI technology in line with different societal values can be formulated.

Innovation in Regulation: The Action Plan for the National Artificial Intelligence Strategy foresees "Algorithmic Accountability" guides being established, as well as a "Trusted AI Seal." Using this seal as a model for a globally compliant yet regional-sensitive tool, Turkey can set an example for other members of the region and for those in the Organization of Islamic Cooperation.

Collaboration in Response to Next Generation Threats: With Autonomous AI agents, a new breed of threats has been introduced. Using their experience with security in their region as a NATO member, it would be easy for Turkey to share their knowledge in various AI-based collaborations to secure their cyber world.

6.4.4. Conclusion and Strategic Recommendation: Leadership in a Niche

A good foundation for ANI has been laid in Turkey, which also displays the needed understanding of strategy for a potential AGI. Instead of challenging world tech leaders in a comprehensive AGI race, Turkey would benefit from a different form of sustainable leadership, which involves a focused niche.

This niche would involve "Lean AGI" architectures and the cultural-linguistic adaptation of worldwide AGI architectures. Being a focal point for computationally efficient AGI subsystem designs that emphasize human-AI cooperation, and a center for studies of AGI localization and adaptation that would make sure worldwide AGI architectures suit Turkey's language, sociocultural setting, and laws, would help Turkey become a stakeholder in worldwide AGI issues related to security, accessibility, cultural diversity, and ethical governance. This would require investments in technology, as well as developing capabilities in philosophy, law, the social sciences, and international relations.

7.Ethics, Policy, and Governance

With increasing use of AI, issues related to ethics—Privacy, Discrimination, Accountability, and Security—have cropped up. At the AGI and ASI levels, such issues reach existential proportions. Therefore, International AI governance initiatives are gaining speed:

European Union Artificial Intelligence Act (AI Act) – uses a risk-based regulatory system.

UNESCO and OECD: develop AI principles that are founded in human values.

Open-source vs. closed models: There are still debates on security and equity effects of open-source models of AI.

Figure 5 describes a framework matrix which locates ethical, security threats to ANI, AGI, and ASI on a graph with two axes, which are the level of risk and response types, which are technical, national, and international. This matrix explains which strategies are to be used for a certain level.

Figure 5. Ethics and Security Framework Matrix

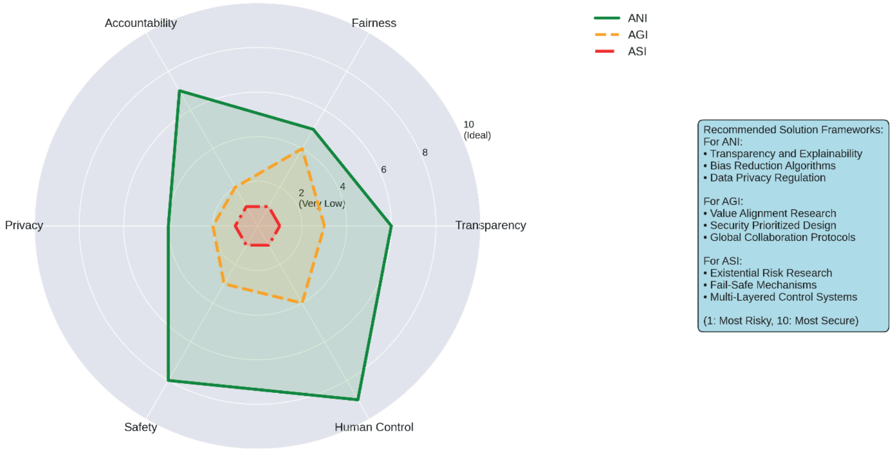


Table 2 discusses ethical considerations and control tools for various levels of AI.

Table 2. Ethical and Security Challenges

Level	Definition	Fundamental Ethical Problems	Audit Mechanisms
ANI	Refers to AI designed for particular tasks.	Discriminability, privacy, displacement of jobs, and lack of transparency in decision-making processes.	Requires conformance evaluation, transparency requirements, and post-market surveillance of high-risk ANI systems for tasks such as employee hiring, as well as for biometric identification systems (EU, 2024). Provides persons with a right not to be subjected to automated decisions without human involvement (GDPR, 2022). Highlights transparency, explicability, and accountability through algorithmic impact assessments (OECD, 2019). Suggests validating bias, robustness, and third-party audit for niche AI use cases (NIST, 2023).
AGI	Stands for AI with human-level intelligence.	Handling issues of alignment, control, consciousness, and moral status.	Proposes sandboxes for regulation of AGI systems (Art. 59) and transparency of general-purpose models of AGI systems (Art. 52a), which involves model cards, as regards systematic risk analysis (EU, 2024). Proposes human-in-the-loop control, value-sensitive design, and international control bodies for developing AGI systems (OECD, 2019). Proposes ethical review boards for multistakeholder involvement for systems with human-level reasoning capabilities, as regards pre-installation ethical assessments of AGI systems, "AGI-like" systems, and ethical impact evaluation (UNESCO, 2021; NIST, 2023).
ASI	Refers to Artificial Intelligence that exceeds human intelligence.	Strategic risks of a higher order like existential risk, morality beyond that of human beings, unpredictability, and misaligned goals.	Indicates that "unacceptable risk" of systems with emerging, unforeseen risks can be restricted or prohibited by the Commission (Annex III) (EU, 2024). Proposes that superintelligence R&D can be regulated through treaties worldwide (OECD, 2019) and "urges a global moratorium on R&D of superintelligent AI systems" until their goals are auditable and align with human values as outlined in human rights" (UNESCO, 2021). Although it is a "forward-looking" report, it remains widely quoted; proposes robust "safety and control" and "fail-safe" approaches prior to ASI development experimentation (Future of Life Institute, 2017).

8.Discussion and General Assessment

This research suggests that the evolution of AI involves not only technology but also philosophical and societal issues. ANI, as a technology, works well as a tool but has some limitations. AGI, as a technology, has been expected to offer human-level flexibility in cognitive abilities, but many challenges exist in this area, including technological complications, control, and value alignment. ASI has been considered as a technology that has either potential transformative powers or can pose a threat to human existence due to a possible degradation of human control with accelerated development.

On the other hand, in terms of quantity and sectorial adaptation, it should be noted that, as a Turkish view, it has a substantial development in ANI, but it has deficiencies in theoretical sophistication, interdisciplinary knowledge, computational power, and access to large-quality data that are needed on AGI/ASI. However, it has a distinct, pluralistic voice in world AI ethics debates owing to ethical considerations in Turkey, which are derived from their own cultural and Islamic tradition. Instead of challenging world AGI domination, a more sustainable role for Turkey in AGI development would be in "Lean AGI" architectures and AGI systems' cultural-linguistic adaptation.

9.Conclusion and Recommendations

The result suggests that it is important, both from a technical and a social perspective, that differences exist between ANI, AGI, and ASI. While ANI brings about a transformation in a certain industry, AGI brings in a new model of research in terms of human-level intelligence, whereas ASI exceeds human capabilities in processing information, autonomy, and decision-making. Important recommendations are as follows:

9.1. International Level Recommendations

An "AGI and ASI Governance Agreement" should be developed among all nations in a united manner between 2026 and 2030, under the umbrella of the UN. This would encompass all of the following:

- Mandatory "Kill-Switch + Monitoring Consortium" for all AGI/ASI projects
- Mandatory International Approval Before Starting Recursive Self-Improvement
- Creation of "Value Alignment Verification Labs"

The set of Asilomar AI Principles of 2017 needs to be updated and turned into a binding document with a minimum of 50 country signatures by 2030. "Restricted ASI (Air-Assisted AI)" needs to be a worldwide standard with superintelligence that would only be activated in particular areas like cancer studies and/or climatic modeling, with bounds for general autonomy.

9.2. Turkey-Specific Strategic and Comprehensive Recommendations

9.2.1. Short Term (2026-2028)

- Release the 2026-2035 National AGI Preparedness Strategy, as a follow-up to the National Artificial Intelligence Strategy 2024-2025.
- Create a "Turkish AGI Security Institute" in TÜBİTAK, similar to AISI in the UK and "Cyber Valley" in Germany.
- Certify 10,000 people in "Neuro-Symbolic AI," "Value Alignment Engineering," and "AI Ethics & Law" in five years.
- Establish a HPC facility with 1 ExaFLOP of computing capacity in a public-private partnership by 2028.
- Launch a 1 trillion-parameter open-source Turkish Language Model with data pertinent to Turkey by 2027.

9.2.2. Medium Term (2028-2035)

- Head an international consortium on "Lean AGI" architectures that are low energy and low-cost, human-centered, including Turkey, Malaysia, Indonesia, Brazil, and South Africa.
- Create a "Cultural-Value Aligned AI Center" and formulate a "Turkish-Islamic-Western Fusion Value Set" as open-source.
- NATO-standard implementation of "Agent AI" and autonomous systems in the defense sector should limit actions to human command.
- Start "AGI Philosophy and Governance" mandatory post-grad courses in prominent universities in Turkey (METU, Boğaziçi, Bilkent, Sabancı, Koç, ITU.

9.2.3. Long Term (2035 and beyond)

- Internationalize "Trusted AI Seal" and apply it in a wider scope, especially in OIC and Turkic States.

- In 2040, a lead role in at least one "Restricted AI" project: world hunger, desertification, anti-aging medicine.
- Introduce a "Universal AI Dividend" pilot to solve a problem of AI-induced joblessness and income inequality, similar to that of Norway's sovereign wealth fund.

9.3. Education and Human Resources Recommendations

- Algorithmic Thinking and AI Ethics needs to be a compulsory subject from primary school onwards.
- To award 1,000 doctoral fellowships in: Neuro-Symbolic AI, Causal Inference, Value Alignment Engineering, AI Law and Governance, Grand Cultural Language Models by 2030.
- Support 100 faculty members to take up to 6-12 months a year as a visiting researcher in institutes like Stanford HAI, Oxford FHI, Anthropic, and DeepMind.

9.4. Funding and Incentive Recommendations

- Create AGI Security Fund with initial 5 billion USD budget over 10 years from: 3% profits of defense industries, voluntary 0.1% revenue sharing of major tech companies, and Green AI bonds.
- Offering 100% corporate tax relief for firms that contribute to research in AGI Safety and Ethics.

9.5. Social and Cultural Recommendations

- Produce a "Fatwa on Islam and Artificial Super Intelligence" in cooperation with the Presidency of Religious Affairs, Islamic catechisms, and universities, to be published in Turkish, Arabic, English, and Indonesian languages.
- Make a documentary series "AI and the Future of Humanity" for 10 seasons with TRT and TÜBİTAK for such platforms as Netflix, Shahid, and iQIYI.
- Creation of an "Istanbul AGI Ethics Summit" that would become a center for East/West synthesis every year.

9.6. "Red Line" Proposals (To be implemented by 2026)

Things that Turkey needs to take a lead internationally in are:

- The prohibition of lethal autonomous weapons without human authorization There were no systems that started a process of recursive self-healing that went beyond a closed circuit. AGI/ASI projects should be "auditable source" only, but not open source.
- Spend at least 10% of every super intelligence project budget on safety and alignment research If it were implemented, this comprehensive strategy would place Turkey not as a consumer of ANI or a secondary player in the AGI competition but as a factor that contributes to shaping the world of the future with its own voice and moral vision.

References

Abdullah, H. (2025). The Ascent and Effects of Artificial Intelligence: Prospects, Difficulties, and Future Directions. *Difficulties, and Future Directions*. http://dx.doi.org/10.2139/ssrn.5283228

Alvarez-Teleña, S., & Díez-Fernández, M. (2024). Advances in artificial super intelligence: Calm is all you need. *Available at SSRN 4924496*. http://dx.doi.org/10.2139/ssrn.4924496

Anakotta, M. Y. (2025). ASI and Cyberterrorism: New Challenges To Indonesia's National Security and Penal Policy. *Jurnal Ilmu Hukum: ALETHEA, 8*(2), 1-22. https://doi.org/10.24246/alethea.vol8.no2.p1-22

Araújo, A. S., Do Nascimento, R. P., Barone, D. A. C., & Lindemann, H. (2025, June). Soft Skills and Ethics: A Different Perspective on Engineering Competencies and the Future of AI Development. In *2025 34th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE)* (pp. 1-5). IEEE. https://doi.org/10.1109/EAEEIE65428.2025.11136273

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. *Strategies.*

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & company.

Chapman, J. R., Wang, J. C., & Wiechert, K. (2024). We Need RI and Not Just AI! Thoughts on the Implementation of Artificial Intelligence in Medicine and Spine Surgery Specifically. *Global Spine Journal, 14*(8), 2213-2215. https://doi.org/10.1177/21925682241272348

Chojecki, P. (2025). An Operational Kardashev-Style Scale for Autonomous AI-Towards AGI and Superintelligence. *arXiv preprint arXiv:2511.13411*. https://doi.org/10.48550/arXiv.2511.13411

Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (CBDDO). (2024). *Ulusal Yapay Zekâ Stratejisi 2024-2025 Eylem Planı.* https://cbddo.gov.tr/SharedFolderServer/Genel/File/UlusalYapayZekaStratejisi2024-2025EylemPlani.pdf

Doker, Y. F. (2025). The Wicked Nature of AGI. *Law, Technology and Humans, 7*(2), 89-110. https://doi.org/10.5204/lthj.3757

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world.* Basic Books.

Ekici, Ş. (2025). New Proposed Personality Model For Artificial Intelligence: Integrated Personality. *Bilişim Hukuku Dergisi, 7*(1), 311-363. https://doi.org/10.55009/bilisimhukukudergisi.1686538

European Commission. EU. (2024). *Proposal for a Regulation on a European approach for Artificial Intelligence (AI Act)*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fahad, M., Basri, T., Hamza, M. A., Faisal, S., Akbar, A., Haider, U., & Hajjami, S. E. (2024). The benefits and risks of artificial general intelligence (AGI). In *Artificial general intelligence (AGI) security: Smart applications and sustainable* technologies (pp. 27-52). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3222-7_2

FTI. (2025). *Turkey Yapay Zekâ İnisiyatifi, 2025 Çalıştay Raporu'nu paylaştı.* https://fintechistanbul.org/2025/07/08/turkiye-yapay-zeka-inisiyatifi-2025-calistay-raporunu-paylasti/

Future of Life Institute. (2017). *Asilomar AI Principles.* https://futureoflife.org/open-letter/ai-principles

Gal-Or, S. (2025). The Philosophy of AI. In *Garden of Wisdom: Timeless Teachings in an AI Era* (pp. 269-272). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-83085-3_63

GDPR, E. (2022). *Automated individual decision-making, including profiling*. https://gdpr-info.eu/art-22-gdpr

Glenn, J. C. (2025). Why AGI Should be the World's Top Priority. *Horizons: Journal of International Relations and Sustainable Development, 30*, 86–95. https://www.jstor.org/stable/48829684

Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. In *Advances in computers* (Vol. 6, pp. 31-88). Elsevier. https://doi.org/10.1016/S0065-2458(08)60418-0

Goyal, R., Sharma, N., & Kaur, A. (2024, October). Artificial super intelligence: Barriers and limitations. In *AIP Conference Proceedings* (Vol. 3209, No. 1, p. 020011). AIP Publishing LLC. https://doi.org/10.1063/5.0228800

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research, 62*, 729-754. https://doi.org/10.1613/jair.1.11222

Groppe, S., & Jain, S. (2024). The Way Forward with AI-Complete Problems. *New Generation Computing, 42*(1), 1-5. https://doi.org/10.1007/s00354-024-00251-8

KPMG Turkey. (2025). *Yapay Zekâ Dönüşümü: Türkiye ve ABD Perspektifleri*. https://kpmg.com/tr/tr/home/insights/2025/11/yapay-zeka-donusumu-turkiye-ve-abd-perspektifleri.html

Li, O. (2023). Artificial General Intelligence and Panentheism. *Theology and science, 21*(2), 273-287. https://doi.org/10.1080/14746700.2023.2188373

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*. https://doi.org/10.48550/arXiv.1801.00631

Mirindi, D., Sinkhonde, D., Mirindi, F., & Bezabith, T. (2025, May). Advance toward Artificial Superintelligence with OpenAI's O1 reinforcement learning and ethics. In *2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC)* (pp. 475-481). IEEE. https://doi.org/10.1109/AIRC64931.2025

Mokoena, P. B. (2025). Superintelligence: Why Humanity Would Risk Extinction To Create Its Successor. *Available at SSRN 5595475.* http://dx.doi.org/10.2139/ssrn.5595475

NIST (2023). *Artificial intelligence risk management framework* (AI RMF 1.0). https://doi.org/10.6028/NIST.AI.100-1

OECD. (2019). *Principles on Artificial Intelligence.* OECD Publishing. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Russell, S. (2019). *Human compatible: AI and the problem of control.* Penguin Uk.

Sohn, S. M. (2024). The Three Levels of Artificial Intelligence (AI) Adoption Framework and the Six Levels of Autonomous Companies. *Available at SSRN 4965436.* http://dx.doi.org/10.2139/ssrn.4965436

Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2024). *The 2024 AI Index Report.* https://hai.stanford.edu/ai-index/2024-ai-index-report

Surdu, M., & Surdu, T. V. (2024). The hype of artificial intelligence in neonatology: How close are we to real AI?. *THE NEWBORN REVIEWS & RESEARCH* Учредители*: AMALTEA Medical Publishing House, 2*(3), 142-146. https://doi.org/10.37897/NEWBORN.2024.3.5

Sysart Consulting. (2025). *Yapay Zekâ ile Dönüşen Sektörler Turkey 2025 Raporu.* https://sysart.consulting/tr/yapay-zeka-ile-donusen-sektorler-turkiye-2025-raporu

Tegmark, M. (2018). *Life 3.0: Being human in the age of artificial intelligence.* Vintage.

Tiwari, S., Yadav, S., Mahala, R., Singh, R., Gehlot, A., Tripathi, R., & Yamsani, N. (2025, January). AI Application in Plant Identification and Classification: Innovation and Impact. In *International Conference on Artificial Intelligence and Smart Energy* (pp. 374-383). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-90478-3_31

Türkiye Yapay Zekâ İnisiyatifi (TRAI). (2025). *2025 Çalıştay Raporu.* https://turkiye.ai/wp-content/uploads/2025/06/Trai_8yil_calistayi_raporu.pdf

Tzimas, T. (2021). The Ethics of AI. In *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective* (pp. 69-99). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-78585-7_4

Ueda, D., Walston, S. L., Kurokawa, R., Saida, T., Honda, M., Iima, M., ... & Naganawa, S. (2025). Artificial superintelligence alignment in healthcare. *Japanese Journal of Radiology,* 1-14. https://doi.org/10.1007/s11604-025-01907-1

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Vsevolodovna, R. I. M. (2025). A Roadmap to Superintelligence: Architectures, Transformations, and Challenges in Modern AI Development. *Preprints*. https://doi.org/10.20944/preprints202501.2099.v1

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks, 1*(303), 184.