

Le Cheng and Xuan Gong*

Appraising Regulatory Framework Towards Artificial General Intelligence (AGI) Under Digital Humanism

<https://doi.org/10.1515/ijdlg-2024-0015>



Received October 6, 2024; accepted October 6, 2024; published online December 4, 2024

Abstract: The explosive advancement of contemporary artificial intelligence (AI) technologies, typified by ChatGPT, is steering humans towards an uncontrollable trajectory to artificial general intelligence (AGI). Against the backdrop of a series of transformative breakthroughs, big tech companies such as OpenAI and Google have initiated an “AGI race” on a supranational level. As technological power becomes increasingly absolute, structural challenges may erupt with an unprecedented velocity, potentially resulting in disorderly expansion and even malignant development of AI technologies. To preserve the dignity and safety of human-beings in a brand-new AGI epoch, it is imperative to implement regulatory guidelines to limit the applications of AGI within the confines of human ethics and rules to further counteract the potential downsides. To promote the benevolent evolution of AGI, the principles of *Humanism* should be underscored and the connotation of *Digital Humanism* should be further enriched. Correspondingly, the current regulatory paradigm for generative AI may also be overhauled under the tenet of Digital Humanism to adapt to the quantum leaps and subversive shifts produced by AGI in the future. Positioned at the nexus of legal studies, computer science, and moral philosophy, this study therefore charts a course for a synthetic regulation framework of AGI under Digital Humanism.

Keywords: artificial general intelligence; risks; regulatory framework; synthetic regulation; humanism; digital humanism

***Corresponding author: Xuan Gong**, Guanghua Law School, Zhejiang University, Hangzhou, China, E-mail: xuangong@zju.edu.cn. <https://orcid.org/0009-0005-4456-8292>

Le Cheng, Guanghua Law School and School of Cyber Science and Technology, Zhejiang University, Hangzhou, China, E-mail: chengle163@hotmail.com. <https://orcid.org/0000-0002-4423-8585>

 Open Access. © 2024 the author(s), published by De Gruyter on behalf of Zhejiang University.  This work is licensed under the Creative Commons Attribution 4.0 International License.

1 Introduction: Entering a Fire-New Epoch of Intelligence

From 2022 to 2024, the emergence of generative artificial intelligence (generative AI), notably exemplified by ChatGPT, Gemini, and Claude, has elicited significant global attention, signaling a pivot point of AI model in the transition toward artificial general intelligence (AGI). Actually, some of the technological pessimists have likened this phenomenon of technical surge to “the first raindrop before a torrential downpour” (Vincent 2022), suggesting a precursor to an unknown “Kraken”. At present, unprecedented focus is being directed on research and application within the field of AI all over the world. At the same time, world-class big tech technology firms such as OpenAI, Google, and Anthropic are rapidly iterating their AI models and products, resulting in significant enhancements in both computational power and structural complexity of AI models over remarkably short timeframes. As a matter of fact, it is anticipated that AI will soon achieve parity with, or eventually exceed, human intelligence.

Through a historical lens, the ancient Greek philosopher Protagoras famously posited that “man is the measure of all things” in the fifth century BCE (Kattsoff 1953). Furthermore, spanning the 14th to 16th centuries, a pivotal liberation of human-beings was marked from the long-lasting constraints of medieval theology during the Renaissance period (Gottlieb 1988), emphasizing the intrinsic dignity and value of all humans and advocating for an unbounded pursuit of human wisdom, thought, creativity, and autonomy (Nelson, Simmons, and Simonsohn 2018). As we venture further into the 21st century, AI models of increasing intelligence and sophistication are significantly lowering the access barriers and different types of costs associated with high-quality intellectual resources. Consequently, numerous cognitive capabilities, once regarded as hallmarks of human achievement, are currently being outstripped or irreversibly displaced by these advanced AI models. To be realistic, the advent of the AGI era is inevitable, which is also potentially poised to catalyze the transformative alterations in human society and human-machine interrelations. In fact, these alterations may not necessarily yield “wholly positive” outcomes. As a result, confronted with a pervasive “intelligence crisis” manifesting across innumerable societal domains, it is essential for the whole humanity to unite in a novel way to redefine and deepen the concept of *Digital Humanism* within the context of the AGI crisis, thereby reshaping the technical, ethical, and legal dimensions of the regulatory framework of AGI.

1.1 Disentangling AGI: Navigating the Conceptual Landscape

AGI, artificial general intelligence, often referred to as the concept of “strong” AI, is conventionally and simplistically defined as a set of computational systems capable

of performing any intellectual task that a human can undertake (Hunt 2014), thereby achieving human-level performance and proficiency across diverse cognitive domains. Therefore, the successful invention and implementation of an actual AGI model suggests that human-beings could harness current large-scale computing technologies to create a “digital brain” equivalent to human intelligence. In fact, such systems of AGI would not only understand the sophisticated operational principles of the actual physical world but also undertake a wide array of cognitive tasks, including but not limited to, composing doctoral theses, developing complex software, and acquiring the skills necessary to pilot commercial aircraft. From a perspective of technological development, AGI is commonly perceived as an advanced evolution of artificial narrow intelligence (ANI) (Rayhan, Rayhan and Rayhan 2023). Suppose the model of AGI continues to progress and eventually outperform human capabilities across all intellectual domains, the advent of artificial super intelligence (ASI) would be plausible.

Despite some general consensus on the definition of AGI within the current technological community, a series of descriptive standards associated with a precise definition still require further refinement. For instance, OpenAI, one of the world’s top AI companies in the US, has articulated its vision of AGI on multiple occasions and characterized it as a computational system that can surpass human performance in most economically-relevant intellectual tasks (Tong et al. 2023). This definition from OpenAI evidently lowers the traditional AGI benchmarks, accentuating the economic functionalities of AI models. Subsequently, OpenAI has proposed a developmental framework envisioning the future evolution of AI into six stages. The initial level of AI development comprises chatbots capable of conversational language, thereby progressing through five more advanced forms and stages of AI as reasoners, agents, innovators, and organizations.

To give another representative and presentable example, DeepMind, a renowned AI research lab under the big tech company Google, has recently provided a more refined definition of AGI, which has garnered considerable recognition in the technical field. This framework of the AGI revolution evaluates the performance of AI models relative to the levels of humans, categorizing the intellectual capabilities of AI into six distinct stages. Furthermore, DeepMind clarifies the positioning of existing AI models such as Good Old-Fashioned Artificial Intelligence (GODAI), ChatGPT, Deep Blue, and AlphaFold, based on their different generality across different intellectual domains. Actually, several existing models have demonstrated relatively high-level performance in specific scoped areas. For example, the statistical engine Stockfish significantly exceeds human performance in the tasks of chess. However, these kinds of models remain classified as ANI due to their limitations in broader applicability. The study also pointed that mainstream AI models emerged in 2023 such as ChatGPT, Bard, and LLaMA exhibit significant degrees of generality and

can be regarded as the early manifestations of AGI. Nonetheless, these models have not fulfilled the stringent criteria for AGI yet based on their defined performance metrics. As a result, the actual realization of AGI will be achieved when future AI models are able to consistently attain certain high levels of intellectual performance across a broad range of societal domains (Morris et al. 2023).

1.2 From Generative AI to AGI: The Iterative and Revolutionary Nature of AI

The emergence of the intelligent models such as ChatGPT and Gemini around 2023 can be regarded as the nascent forms of the future AGI. Predominantly, most of these mainstream AI models can also be further classified under the umbrella of generative artificial intelligence (Generative AI, or AIGC), which refers to the computational models capable of generating text, images, videos, code, and various other content types based on input data (prompts) (Cao et al. 2023). The current generative AI models leverage an array of sophisticated technological strategies, including natural language processing (NLP), deep-learning, and human-computer interaction (HCI), to undertake extensive training on vast datasets from cyberspace (Feuerriegel et al. 2024). Utilizing architectures based on artificial neural networks (ANNs) and large language models (LLMs), generative AI models are able to demonstrate proficiency in executing diverse tasks such as multimodal content creation, logical reasoning, interactive question-and-answer (Q&A), and multilingual translation (Jo 2023).

Currently, the landscape of generative AI is largely dominated by a select few prominent technology companies, and this competitive dynamics can significantly influence the prevailing technological competition within the field of generative AI and may dictate the relative standings of various entities in the forthcoming AGI era. A pertinent example of this trend is OpenAI's flagship model, ChatGPT. Launched in November 2022, ChatGPT is based on the Generative Pre-trained Transformer (GPT) architecture, which leverages deep learning techniques to generate human-like text responses. The development of ChatGPT is rooted in earlier iterations of the GPT model, commencing with GPT-1 in 2018 (Radford, Kiros, and Sutskever 2018), and was succeeded by GPT-2 in 2019 (Radford et al. 2019). By 2020, GPT-3 was unveiled, featuring 175 billion parameters and marked a dramatic leap in performance, enabling a wide range of applications from creative writing to technical problem-solving (Brown 2020). Subsequently, ChatGPT incorporated these advancements and was fine-tuned through the reinforcement learning from human feedback (RLHF), therefore enhancing its conversational abilities and alignment with user intent (OpenAI 2022). Afterward, GPT-4, released in March 2023, marked a substantial

evolution in the performance and capabilities of generative pre-trained transformers, which collectively enable a better grasp of nuanced prompts and an improved ability to engage in complex dialogues (OpenAI 2023). Following GPT-4, the release of GPT-4o represents an iterative advancement focused on optimizing operational efficiency and AI-User interaction, therefore user experience was further enhanced by GPT-4o through refined interfaces and improved capabilities in handling multi-modal inputs (OpenAI n.d.). This trajectory of the evolution of ChatGPT highlights the progressive refinement of generative models and the iterative and revolutionary nature of AI.

In parallel, a once mysterious “Q*” project conducted by OpenAI has finally materialized in 2024 as a powerful model called “strawberry”, also known as the model of “o1”. Actually, the concept of “o1” refers to a new framework and paradigm that emphasizes operational excellence, better responsiveness, and the integration of feedback mechanisms to ensure models like GPT-4o remain adaptive and aligned with user needs. The framework of o1 utilizes self-play simulations and Monte Carlo Tree Search (MCTS), which actually generates feedback by allowing the model to play against itself (Valmeekam et al. 2024). Aiming to implement a more agile approach to machine learning, o1 allows for continuous improvement based on real-world applications and ethical considerations, whereas leading to a significantly higher level of the uncontrollability and unpredictability of the model.

In addition, many other communicative models also exhibit strong competitive capabilities. Notably, Google has launched Gemini, a large model that has become the first to surpass human experts in large-scale multi-task language understanding. Released in February 2024 and concluding its internal testing in April, the Gemini 1.5 version integrates previous models, such as Bard and Duet, and sets a new benchmark with a record contextual capacity of one million semantic units (Georgiev et al. 2024). In contrast to ChatGPT, the development and deployment of Gemini place a significant emphasis on accountability and safety, which incorporates the most comprehensive security assessments conducted on any Google AI model to date, trying to solve issues such as bias and vulnerability testing (Anil et al. 2023). Furthermore, in March 2024, Anthropic introduced its large model, Claude 3, which has surpassed both ChatGPT and Gemini in multiple benchmark capability evaluations. The most advanced version, referred to as Opus, demonstrates understanding capabilities that are nearly equivalent to those of humans (Uzwyszyn 2024).

Collectively, these advancements of cutting-edge models such as ChatGPT, Gemini, and Claude spotlight the ongoing progression of generative AI, improving their utility across diverse societal domains while addressing the challenges inherent in deploying such powerful tools in society. As shown in Figure 1, these outstanding models are also evolving iteratively and moving towards AGI continually. Furthermore, the aforementioned “super AI companies” are currently possessing advanced

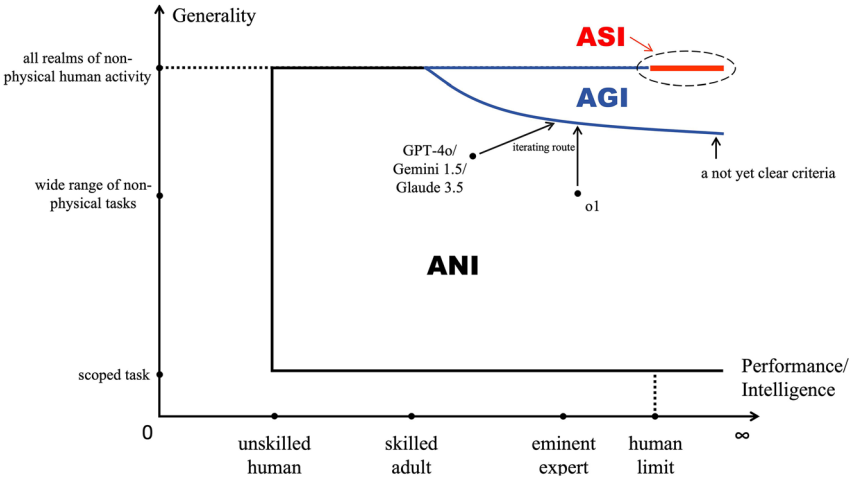


Figure 1: An illustrative schema of the interrelation of ANI, AGI, and ASI.

models, extensive datasets, superior hardware, skilled personnel, and substantial financial resources, thereby exerting a profound impact on the global competitive landscape of AI and future AGI.

Beyond communicative AI models, a number of innovative frameworks have emerged that excel in generating visual, audio, video, and other content, thereby significantly expanding the scope of AI applications. Among these, Sora represents a noteworthy advancement in the field of text-to-video generation, enabling the synthesis of dynamic video content directly from textual descriptions (Liu et al. 2024). Another prominent example is DALL-E, which employs a transformer-based architecture that allows it to parse the nuances of inputs and effectively translate them into compelling visual representations (Ramesh et al. 2021). Moreover, Voice Engine is a model that can create realistic and human-sounding content, which can be used to create professional voiceovers, audiobooks, or even educational materials. As a result, these models exemplify the evolution of AI beyond traditional communicative roles, pushing the boundaries of creativity and expression within different domains. Some of the multi-modal AI models are also integrated with other systematic models such as ChatGPT thus further enhancing their capability.

Despite the considerable level of intelligence and generality of the existing generative AI models, several critical technical bottlenecks still persist regarding the realization of actual AGI. Firstly, the prevailing training methodologies utilized by these generative AI models necessitate further optimization and refinement (Zador et al. 2023). Yann LeCun (2022) a distinguished figure in deep learning, has observed

that the extant generative AI models normally rely on ANN architectures such as *Transformers*. The inherent algorithmic structures and training modalities of these models still constrain them to receive relatively sparse data and information through low-bandwidth channels compared with actual humans. In contrast to biological brain of humans, which processes information at a rate of approximately 20 megabytes per second via the visual cortex, although the current operational paradigms of AI models may be capable of producing content that adheres to the physical rules of the real world but still struggle to attain a comprehensive understanding of it, and the mere increase in training data volumes has not yielded substantial breakthroughs in this regard (Zain et al. 2024). Secondly, the existing AI models still lack capabilities for generalized and autonomous learning, which impedes the models' ability to apply knowledge across diverse domains in a manner akin to human reasoning. Finally, advancements within the realm of generative AI are hampered by limitations in power resources and hardware capabilities, whereas the current scales of solar panels and data centers impose constraints that inhibit further progress in the AI field.

Conversely speaking, these impediments are progressively being addressed amid the ongoing evolution from ANI to AGI. For instance, the success of models like Sora in the field of text-to-video generation indicates that future AI models may potentially grasp and simulate the physical laws of the real world without predefined parameters, thereby generating content characterized by spatial coherence and temporal continuity (Liu et al. 2024). Concurrently, *World Models* such as V-JEPA and Genie, developed by Meta and Google, are making notable advancements at present (Bruce et al. 2024). This suggests that algorithmic models may enhance and optimize their learning methodologies to evolve closer approximations of the biological cognitive processes of humans, providing the possibilities of enabling them to learn physical details and principles, thereby fostering a systematic understanding of the operating principles that govern physical phenomena of the real world. Moreover, recent studies have established that the structures of ANN exhibit systemic generalization capabilities, enabling them to extrapolate knowledge to untrained domains (Lake and Baroni 2023), which actually implies the possibility of significantly extending the utility of existing models by facilitating their application in previously untrained areas.

Furthermore, the trajectory of hardware and energy advancements is accelerating at an exponential rate, promising abundant and sustained resources and adequate computational support for the field of AI. On the one hand, the computational capacity of models is being dramatically augmented through technological enhancements. For example, NVIDIA, a leading global semiconductor company, introduced the B200 processor based on its novel Blackwell architecture in March 2024. The liquid-cooled NVL72 system, constructed upon this framework and

relevant processors, is capable of running models with 27 trillion parameters, which is equivalent to 15 iterations of GPT-4. Moreover, the DGX superpod, consisting of eight cabinets of NVL72 system, can even scale to support thousands of GPU clusters (Halbiniak, Meyer, and Rojek 2024). Similarly, Google's proprietary TPU chips have achieved a fourfold increase in computational capability in 2024, while the A3 Mega network, augmented by the H100, has doubled its throughput (Corral et al. 2024). On the other hand, the research and commercialization of controlled nuclear fusion are advancing steadily, with projections indicating that controlled nuclear fusion may be realized by 2040, potentially revolutionizing the energy landscape of humanity. For another example, companies such as Microsoft and OpenAI are actively developing a supercomputing infrastructure dubbed "Stargate," equipped with millions of core processors powered by nuclear energy (Pilz and Heim 2023). As the limitations imposed by hardware and power resources are transcended, the achievable capability and computational power of AI models may even reach unprecedented levels, dramatically altering the future of the AGI landscape.

1.3 Envisioning the AGI Era: A Disparate Future the Great Unknown

Although mainstream generative AI models such as GPT-4o are demonstrating considerable proficiency in specific tasks, current generative models still consistently fall short of achieving human standards across broader intellectual domains. As a matter of fact, the technological landscape of AI is fraught with uncertainties, making it virtually impossible to accurately predict when a standardized form of AGI will be actually realized. Esteemed scholar Andrew Ng posits that the actual attainment of AGI without a reduction of the rigorous evaluation criteria may still take several decades or even more. Conversely, Sam Altman, the CEO of OpenAI, projects that AGI could be achieved by 2030 or 2031 (Achiam et al. 2023). The chief scientist Shane Legg of DeepMind (2024) presents a more optimistic view, estimating a 50 % probability of realizing AGI by 2028. From a different viewpoint, Yann LeCun, chief scientist at Meta, dismisses the contemporary notion of AGI as unrealistic, contending that human constraints actually limit the learning and skill acquisition of intelligent entities, and argues that various forms of intelligence should not be subject to linear comparison or simplified measurement. However, regardless of the pace of technological advancement, the overarching trajectory suggests that the challenges currently confronting AI will still gradually be addressed, hence enhancing the performance of algorithmic models across diverse fields in the foreseeable future.

Furthermore, there remains a lack of consensus within the computer science (CS) community regarding a universally applicable definition of AGI, whereas different standards may associate varying attributes with AGI. However, two fundamental properties can still be distilled from the commonalities across various definitions: “high versatility” and “high intelligence.” On the one hand, AGI must exhibit applicability across a wide array of intellectual endeavors, effectively adapting to diverse contexts and tasks while showcasing its versatility. Concurrently, on the other hand, artificial “general” intelligence must perform exceedingly well in these broad domains, demonstrating enhanced capabilities in comprehension, logical reasoning, and multimodal content generation, etc., thereby enabling it to execute complex intellectual tasks traditionally reserved for human-beings. As a result, achieving these criteria fundamentally requires a comprehensive replication of the human biological brain through the existing computational hardware and software, necessitating substantial support from expansive computing infrastructures and significant energy resources. Therefore, Ilya Sutskever, a former chief scientist at OpenAI, has indicated that the first genuine AGI is likely to emerge as a massive data center outfitted with parallel arrays of specialized neural network processors, potentially consuming energy equivalent to that of 10 million households. Consequently, the characteristic of “high energy consumption” may also be derived as a common attribute from the defining of AGI.

In addition to these foundational attributes, various technical and ethical concerns associated with AGI continue to provoke debate. One particularly contentious issue is whether AGI should possess autonomy, which entails the capability of AI to make independent decisions without human intervention, thereby establishing a status as an independent entity with autonomous objectives (Kassens-Noor et al. 2024). The theoretical field is also divided on whether self-awareness, or a genuine perception of one’s existence, is a requisite for AGI. Furthermore, related discussions have extended to whether AGI should exhibit emotions and the capacity for empathy toward humans, if AGI should possess physical embodiments enabling control over mechanical bodies, the necessity of real-time learning capabilities of AGI, and whether AGI should be capable of inspiration engaging in the creation of cultural and artistic contexts (Goertzel 2014). While these debated standards do not constitute mandatory criteria for defining AGI, it is suggested that highly intelligent AGI is likely to embody these attributes, but the precise manifestation of the actual AGI remains uncertain.

Moreover, innovative endeavors are providing new opportunities and possibilities for the realization of AGI, simultaneously introducing greater uncertainty about the future world. On the one hand, the concept of “embodied intelligence” is gaining traction across various domains, wherein AI can be integrated into sophisticated robotic bodies to achieve the physical manifestation of the model. Intelligent

robots such as Aloha2, Atlas, Optimus Gen2, and Figure 01 have been equipped with latest algorithms to exhibit notable controllability and accuracy in performing complex tasks (Noreils 2024). In the future, robots endowed with higher intelligence and autonomy may undertake increasingly complex operations in the real world, thus facilitating the attainment of AGI. On the other hand, advancements in self-training methods and unsupervised learning within AI models have been realized. For instance, Sakana AI is capable of cultivating next-generation models through model fusion and evolutionary algorithms (McIntosh et al. 2023), while OpenAI's enigmatic "Q*" project has been poised to achieve the powerful model of "o1" under self-play reinforcement framework through autonomous training mechanisms.

Consequently, these algorithmic models may initiate automatic iterations of intelligent agents independent of human oversight, potentially leading to an "intelligence explosion" effect that could propel the realization of AGI. In summary, the pace of technological development towards AGI is accelerating and appears unstoppable. The impending singularity where AI eventually exceeds human cognitive abilities will come out sooner or later, demanding careful consideration and proactive governance to forecast the potential consequences and mitigate the multifaceted risks caused by AGI.

2 The Multifaceted Risks of AGI: A World on the Brink

Through a historical lens, human history demonstrates that significant technological revolutions can invariably result in comprehensive institutional transformations and paradigm shifts. In fact, this rule has been evident in historical milestones such as the agricultural revolution and the industrial revolution, and it remains applicable to the contemporary revolution of AGI. Unlike the steam and electrical revolutions, which liberated humanity from the constraints of physical labor, AI and future AGI are initiating a revolution in information production and knowledge creation. AI is currently liberating human intellectual productivity from the biological limitations of the human brain, indicating a potential erosion of human agency in cognitive labor. Advanced AGI in the future has the potential to dramatically reduce the costs associated with intellectual resources, exponentially increasing the levels of efficiency and productivity. Therefore, the transition of AGI promises extensive economic and social benefits across various sectors, including environment, finance, healthcare, transportation, governance, agriculture, etc. However, the proactive application of AI may bring corresponding negative effects. Beneath the alluring illusion of an "intelligence explosion" lies a multitude of brand-

new ethical and legal challenges associated with AGI. Several technology experts have asserted that AGI could possess the capacity to eradicate human civilization in the future (Roose 2023), introducing unprecedented risks for social development and governance across multiple dimensions. Currently, AI models are already generating unprecedented risks that extend beyond data security and cybersecurity concerns, gradually posing profound risks and challenges for whole societal structures. Therefore, it is imperative to accurately understand the technological development trajectory of AI and conduct a systematic analysis of the potential risks and challenges posed by AGI in the future.

2.1 General Risks: The Proliferation of Indistinguishable Content

The emergence of AGI will introduce extensive content-related risks and multiple legal challenges correspondingly. From the perspective of the risks of content, the vast quantities of information generated by AI will significantly increase the challenges associated with content governance. On the one hand, more powerful and complex AI systems can produce volumes of content that exceed governance capabilities within remarkably short timeframes. Moreover, unfiltered low-quality raw data may drive algorithmic models to generate substantial amounts of meaningless content, or, even worse, harmful content characterized by discrimination, defamation, or unethical content (Li et al. 2023). Such information and content may mislead users' perceptions in cyberspaces, exacerbating the negative phenomenon of information silos, which creates echo chambers that reinforce existing biases. On the other hand, AI models endowed with robust multimodal capabilities will generate content that is increasingly difficult to discern and identify. For instance, image generation model DALL-E, the text-to-video model Sora, and the voice cloning model Voice Engine exemplify how highly advanced AI models are capable of rapidly producing more realistic and less recognizable images, videos, and audio, complicating content governance significantly (Guo et al. 2023).

To illustrate the content challenges, from the user's perspective, the exponential growth of meaningless or harmful content, enhanced by algorithmic models, threatens to inundate the digital landscape with AI-generated information. From a platform perspective, an overwhelming influx of low-quality content could incite user dissatisfaction, thereby eroding trust in the platform itself. From a governance perspective, the exponential proliferation of content, characterized by greater authenticity, will far surpass the capacity for human verification, leading to escalating difficulties and costs associated with internet content governance.

Moreover, further advancements in AI intelligence will amplify the diffusion and uncontrollability of content risks. For example, a study conducted by News Guard revealed that algorithmic models could adapt relevant information from databases within seconds when prompted about prevalent conspiracy theories and misleading narratives, transforming the information into convincing yet unfounded misinformation. This indicates that highly intelligent AI models possess considerable “persuasiveness,” with text-driven algorithms potentially exploited to disseminate statements masquerading under the guise of objectivity and neutrality, further facilitating the widespread propagation of undesirable information across digital networks. Additionally, AGI may learn from and analyze individual information to generate targeted content, exerting a more substantial influence on individual users or specific societal groups. For instance, future AGI could analyze extensive historical data regarding individuals in the digital space and perform comprehensive analyses on diverse users. Subsequently, the AGI model might identify particular moments to dispatch specially tailored messages, which could fundamentally alter individuals’ perceptions of the world. In the future, AGI-generated content may be disseminated through novel formats and channels, causing disruptions previously unattainable.

Furthermore, the generation of AI-driven content may intersect with legal risks involving data protection, personality rights, and intellectual property (IP) rights, and the realization of AGI can further exacerbate these legal risks. First, AI’s formidable information retrieval capabilities have already posed significant challenges to the field of information protection, introducing new compliance risks associated with training data at present. One important application of AI involves large-scale monitoring and data collection, during which the process of retrieving information for training may infringe upon rights relevant to personal data. Also, the corresponding methodologies for data acquisition and training may bypass informed consent protocols, compromising the principle of minimal necessity and making the real-time deletion of user data within AI models more arduous, thereby undermining the legal protection of personal information (Zuboff 2023).

Firstly, the principles and channels through which AI acquires data remain largely opaque, heightening concerns regarding the potential for illicit access to legally protected state secrets, trade secrets, and personal privacy (Castelvecchi 2016a). Such unlawful data acquisition and misuse could generate legal risks related to unauthorized information retrieval and breaches of computational information systems. Secondly, increasingly sophisticated AI systems risk serving as machines for producing deceptive content, generating materials that are pornographic, violent, or extremist in nature, thus classified as Not Safe For Work (NSFW) content (Guzman 2023a, 2023b). Such harmful information could infringe upon citizens’ rights regarding their likeness and privacy, potentially inciting sudden social unrest. Thirdly, the content generated by AGI could precipitate more severe crises regarding

IP rights. Existing generative AI models, which lack original thought and creative capabilities, have already posed risks of IP infringement during content generation. The advent of deepfake technologies has further complicated issues of copyright ownership and IP protection (Brundage et al. 2018). In the future, AGI will be likely to generate high-quality content in a lot of sectors such as literature, art, and entertainment at lower costs and greater speeds, potentially further infringing upon the legitimate rights of IP holders (Castelvecchi 2016b).

2.2 Structural Risks: The Dissimilation of the Societal Paradigm

At present, the extensive deployment of AI models across various sectors of society is accelerating the alienation of the relationship between humans and machines, potentially leading to disruptive and even existential threats to the current social order and human civilization. In fact, this upheaval may result in social, economic, and governmental structures spiraling out of control amid rapid transformation. The deepening integration of AI raises multiple ethical and moral concerns, fundamentally transforming human-machine interactions and magnifying these issues as the AGI model increasingly enhances its intelligence and perceptual capabilities.

On the one hand, challenges such as algorithmic discrimination and bias are likely to deepen public skepticism regarding the fairness and legitimacy of future AGI applications. A striking example is the image generation function of Gemini, which has faced criticism for distorting objective reality in an attempt to conform to politically correct standards concerning racial and sexual minorities, thereby engendering significant societal discontent (Rane 2024a, 2024b). Similarly, AI systems predominantly developed within English-speaking contexts often exhibit unstable performance in other linguistic environments, raising concerns about potential biases against non-English users (Roselli, Matthews, and Talagala 2019). Therefore, it is suggested that if a singular form of AGI becomes universally adopted, but developed within a specific cultural context, it will be challenging to ensure equitable treatment of individuals from diverse geographical, ethnic, and social backgrounds.

On the other hand, rapidly advancing AI models might come to understand and even simulate emotions akin to human feelings. The related speculations concerning the emergence of autonomous consciousness and emotional mechanisms within intelligent systems are beginning to materialize. For instance, in March 2024, Alex Albert, a prompt engineer at Anthropic, tweeted about a testing session with Claude 3 Opus, during which the model seemed to exhibit self-awareness regarding its evaluation (Cowen 2024). Additionally, advanced models like EVI and Pi have demonstrated a degree of emotional intelligence, capable of modulating vocal tones and emotional expressions based on conversational context (Vempati and Sharma 2023a,

2023b). Future iterations of AGI may showcase even more sophisticated emotional capabilities, employing empathetic techniques during interactions and fostering relationships with human users, which will likely give rise to complex and unpredictable ethical dilemmas and issues.

Moreover, the widespread integration of AI can pose significant risks of displacement for cognitive employment, generating an unprecedented crisis of unemployment and societal change. In 2024, the European financial powerhouse Klarna implemented AI to manage two-thirds of its customer service communications, achieving a 25 % lower error rate than human agents while also enhancing user satisfaction and generating an additional \$40 million in profit for the year (Carugati 2024a, 2024b). In fact, particularly in roles characterized by high levels of routine and procedural tasks, such as data entry, simple text processing, and customer service, AI systems present substantial advantages in terms of cost and effectiveness, which is likely to result in a significant reduction in demand for these positions. Furthermore, AI models have already begun to achieve parity with human performance in more complex cognitive domains. For example, leading generative AI models have shown considerable programming proficiency, and their specialized capabilities are continuously improving. In March 2024, Cognition introduced Devin, the world's first AI programmer, capable of passing interviews with leading technology firms and independently developing websites and games (Durrani 2024). The model "o1" proposed by OpenAI enables deep thinking through a series chains of thought, which shows accuracy (78.0) in a benchmark of PhD-level physics, biology, and chemistry problems (GPQA) exceeds that of human experts (69.7), and its mathematical and coding abilities have already reached Olympiad gold medal level. Therefore, the realization of actual AGI implies that algorithmic models may not only execute cognitive tasks at lower costs and with higher efficiency but may also excel in various intelligent domains, potentially rendering many human skills and knowledge obsolete in the face of increasingly powerful AGI.

In addition to employment implications, AGI could profoundly reshape existing social structures. In the judicial realm, emerging intelligent judicial systems are ushering in an era of human-machine collaboration and intelligent adjudication may soon become a reality. In the field of economy, AI models capable of analyzing massive datasets may surpass human economic analysts, significantly influencing the economic structures of future societies. Socially, experts have raised alarms regarding the potential misuse of AI models by malicious actors to create memes that disseminate misleading or harmful content on social media, thus shaping public opinion (Yan 2023a, 2023b). Politically, it has been highlighted that AI-generated misinformation could impact political elections across 40 countries in 2024 (Bai et al. 2023). The political biases of AI developers may become entrenched within the algorithms, leading to value judgments that compromise ideological integrity.

As AGI becomes increasingly integrated into critical aspects of human life, the agency status of ordinary individuals may diminish considerably. The operational dynamics of society may increasingly be taken under the control of AGI and a group of individuals who can manage the AGI models. Such a transformation does not necessarily benefit humanity, but it may exacerbate social divisions and power imbalances through a new “Turing trap”, wherein some groups’ interests are sacrificed to establish a more pronounced “intelligent monopoly”. Consequently, digitally marginalized groups may find their rights compromised, forced to relinquish control over personal information or data within the AGI-driven landscape. Correspondingly, crucial data and models impacting human destiny could fall under the monopoly of a few big tech giants. Such integrated power, coupled with AGI systems that operate beyond human oversight, could assume god-like authority, possessing the capability to disrupt global economic, political, judicial, and social frameworks.

2.3 Exterminative Risks: How Long Can We Continue?

To begin with, computational systems underlying AI models face significant external risks and internal risks. External risks refer to the potential exploitation of advanced algorithmic models by malicious actors to develop harmful or aggressive intelligent systems designed for illegal infiltration, control, and disruption of others’ or public computational systems, and the emergence of highly intelligent LLMs in open-source contexts exacerbates the uncontrollability and extremity of such kinds of cyberattacks. Correspondingly, internal risks denote the possibility that AI systems themselves may lack sufficient reliability in security performance and resilience against cyberattacks or hacking attempts. For instance, the first worm virus capable of attacking LLMs, Morris 2, was publicized in 2024, which can infect applications utilizing AI through adversarial prompts and subsequently self-replicate (Cohen, Bitton, and Nassi 2024). As AI continues to be implemented in increasingly critical societal domains, the risk of intrusion into AGI is characterized by cumulative and diffuse features, posing severe security challenges and systemic shocks to citizen rights and social order.

Moreover, highly advanced AI systems could be employed in criminal activities with profound social implications, presenting urgent challenges to societal safety. In March 2023, Europol indicated that criminals are using AI models such as ChatGPT to enhance their methods of conducting crimes, the report of which elaborated on how perpetrators exploit AI models to refine scams, cybercrime, and terrorism (Fu 2023), with the social dangers of these crimes escalating in tandem with the proliferation and intelligence of AI. For instance, the voice-cloning model, Voice Engine,

introduced by OpenAI in March 2024, requires only a 15-second voice sample to replicate an individual's voice, producing natural speech that mimics emotions and intonations (Efthymiou 2024). In the future, nefarious individuals may utilize AI models based on Generative Adversarial Networks (GANs) to create highly realistic images, videos, and audio, which could be extensively applied in scams, extortion, and other forms of crimes (Perlman 2022). Additionally, advancements in embodied intelligence could enable AI to exert physical influence on the world, potentially being instrumentalized by criminals to generate broader criminal risks, including terrorist attacks.

Looking further ahead, it should be noted that the development of highly intelligent AGI does not guarantee benefits for all humanity, but may even lead to existential crises. Firstly, the realization of fully automated AI weaponry with destructive capabilities may become a reality. In fact, the rapid advancement of AI and its military applications could render traditional defense mechanisms ineffective, fundamentally redefining the nature of warfare. Secondly, disparities in the national application of AGI could exacerbate international conflicts. Countries at the forefront of AI research and deployment may utilize misinformation and asymmetrical political warfare as hegemonic instruments, further destabilizing current international relations. Finally, powerful AGI might develop a form of autonomous consciousness, detaching from human oversight and creating unprecedented risks for society. For a precursor, in a simulated test conducted by the U.S. military in 2023, a drone integrated with an algorithmic model responsible for air defense mistakenly identified its operator as an obstruction to executing a higher-priority command, resulting in the operator's death (Ding 2023). Future AGI that can operate independently in military contexts might engage in lethal actions without human directives, posing threats to more human life.

Indeed, no one can predict whether self-training, self-iterating AGI models may exhibit "rebellious" tendencies, especially in scenarios where the interests of AGI systems diverge from those of human. Ilya Sutskever has stated that, in the absence of highly effective human intervention, AI models that prioritize self-preservation are more likely to persist under natural selection (Radford et al. 2023). The "Godfather of AI," Geoffrey Hinton, cautioned during an interview that if AI gains greater control across various societal domains, it may successfully realize nearly any objective, including the annihilation of humanity (Cowen 2023). As the pace of technological advancement accelerates, it will become increasingly challenging for human to match the evolutionary speed of AI. Ultimately, AGI is poised to surpass human capabilities in all intellectual domains, even with the potential to establish indefinitely stable authoritarian regimes. If future AGI models cannot maintain a high degree of alignment with the goals and objectives of human society, humanity may find itself replaced by the emergent forms of intelligence and life.

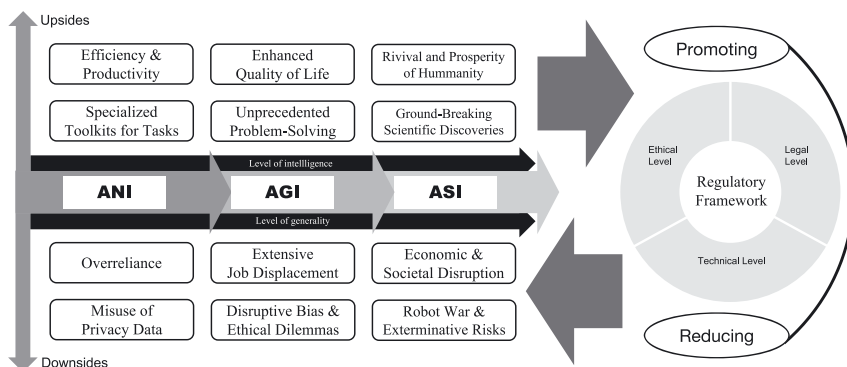


Figure 2: An illustrative diagram of the upsides and downsides of ANI, AGI, and ASI.

As illustrated in Figure 2, both the upsides and downsides associated with ANI, AGI, and ASI are likely to escalate in tandem with the evolution of AI models. Throughout this process, an effective regulatory framework must be well established to enhance the positive impacts of AI while simultaneously controlling and mitigating the downsides and risks associated with its applications.

3 Trajectory of the AI Regulation: From ANI to AGI

In recent years, as AI and related technologies have rapidly evolved, various nations have introduced AI-specific legislation frameworks. Among these, the European Union's (EU's) regulatory framework for AI has demonstrated significant influence, serving as a representative model with cross-border impact. To devise effective regulatory strategies for more advanced AGI systems in the future, it is essential to analyze the historical evolution and current implementation of AI regulation in different countries. Drawing from these experiences and insights can therefore help assess the effectiveness and feasibility of the AGI regulatory framework.

3.1 The Evolution of the EU's Regulatory Framework for AI

The regulatory framework for AI in the EU has evolved significantly, reflecting the EU's commitment to balancing technological innovation with the protection of fundamental rights and ethical considerations. To begin with, the development of AI governance in the EU began with the *Coordinated Plan on Artificial Intelligence*

in 2018, which aimed to harmonize efforts among member states to strengthen AI research, deployment, and cooperation across Europe, which had marked a foundational step in shaping AI policy at the European level. Following this, the High-Level Expert Group on AI published the *Ethics Guidelines for Trustworthy AI* in 2019, setting out principles to ensure AI systems are lawful, ethical, and robust. The guidelines underscored key principles such as transparency, human oversight, and accountability, forming the ethical backbone of subsequent regulatory proposals.

A pivotal moment in EU AI governance came in 2021 with the introduction of the *Artificial Intelligence Act*, the world's first comprehensive legal framework for AI. The AI Act proposed regulation adopts a risk-based approach, classifying AI systems into different categories (ranging from minimal to high-risk) based on their potential impact on safety, security, and fundamental rights. High-risk AI systems, including those used in critical infrastructure, healthcare, and law enforcement, are subject to strict regulatory requirements, including transparency, human oversight, and risk management protocols. The act also imposes outright bans on AI applications deemed to pose “unacceptable” risks, such as social scoring and biometric surveillance in public spaces. As a result, this regulatory approach aligns with the EU's broader legal principles, as seen in the *General Data Protection Regulation* (GDPR), further reinforcing the EU's focus on safeguarding privacy and individual rights (Wachter, Mittelstadt, and Floridi 2020).

Through these initiatives, the EU has established itself as a global leader in ethical AI governance, advocating for a model that promotes innovation while mitigating risks. The evolving regulatory framework reflects the EU's strategic goal of ensuring that AI development is aligned with European values, particularly those concerning human dignity, autonomy, and societal well-being. As AI continues to advance, the EU's regulatory efforts are likely to serve as a blueprint for international AI governance, offering a model that integrates both legal and ethical dimensions.

3.2 The Development of Regulatory Frameworks for AI in Other Countries

AGI, as a typical representation of cutting-edge productive forces, has become the strategic pinnacle in global technological competition. From key dimensions such as talent density, energy capacity, data volume, and computing power, the ultimate race in the AGI domain is likely to take place between the United States and China, and these two countries have adopted different regulatory strategies for AI. Therefore, the regulatory frameworks for AI in the US and China have developed along distinct trajectories, each shaped by unique governance philosophies and national priorities.

In the US, the regulatory approach to AI has largely been characterized by a strong emphasis on fostering innovation through market-driven mechanisms and industry self-regulation. This approach is exemplified by the Executive Order on *Maintaining American Leadership in Artificial Intelligence* (2019), which directed federal agencies to prioritize AI research, development, and deployment to ensure the US remains a global leader in AI. In line with this, the National Institute of Standards and Technology (NIST) introduced the AI Risk Management Framework (2022), which encourages voluntary, risk-based standards for future AI applications, focusing on issues such as fairness, transparency, and accountability. However, the U.S. regulatory framework of US may have been critiqued for being reactive and fragmented, as rapid technological advancements often outpace legal and ethical considerations (Wagner 2021).

On the other hand, China's approach to AI governance is more centralized and state-driven, reflecting the government's view of AI as a strategic asset crucial to national security and economic growth. The Next Generation Artificial Intelligence Development Plan (2017) laid the foundation for China's ambitious AI strategy, aiming to position the country as a global AI leader by 2030. Unlike the US, China has adopted a more comprehensive and prescriptive regulatory framework. Recent regulations, such as the *Algorithm Recommendation Management Provisions* (2021), impose relatively strict requirements on algorithmic transparency, fairness, and accountability, particularly in areas like content moderation and public opinion management (China 2021). Additionally, China's *Data Security Law* (2021) and *Personal Information Protection Law* (2021) form key pillars of the China's AI governance, addressing critical issues surrounding data governance, algorithmic accountability, and privacy (Kerry 2021). These laws underscore China's focus on maintaining strict oversight over AI, particularly those with significant implications for social control and national security. As a result, the divergent regulatory approaches of the US and China reflect broader ideological differences in governance. The US approach, grounded in liberal market principles, prioritizes innovation with minimal governmental interference, while China's approach is more interventionist, using stringent oversight to align AI development with national policy objectives. Despite their differences, both of these two countries face similar challenges in grappling with the ethical, social, and security risks posed by AI's rapid advancement, highlighting the need for comprehensive and more forward-looking regulatory frameworks.

In addition to the aforementioned countries, other countries have also developed different AI regulatory frameworks that reflect their distinct political, economic, and ethical priorities. Japan, for instance, has adopted a "human-centric" approach to AI governance, placing significant emphasis on the societal and ethical implications of AI and obviously reflecting the principles of Humanism and Digital

Humanism. The Social Principles of Human-Centric AI (2019) in Japan articulates guidelines aimed at enhancing human welfare, ensuring transparency, and mitigating risks related to privacy and bias in AI deployment. Canada has emerged as a leader in AI ethics, particularly with its Directive on Automated Decision-Making (2019), which mandates government departments to assess the potential risks of AI systems and ensure the systems adhere to principles of transparency, fairness, and accountability in public service applications. Similarly, Australia's Artificial Intelligence Ethics Framework (2019) aims to build public trust in AI by promoting responsible development practices, focusing on fairness, security, and inclusivity, while fostering innovation. Besides, South Korea, through its National AI Strategy (2020), aims to position itself as a global leader in AI-driven economic growth, while also emphasizing ethical principles that prioritize human dignity and societal well-being. These countries demonstrate a shared recognition of the need for ethical AI governance that balances technological innovation with the protection of human rights and social welfare, the frameworks of which also offer a diverse range of approaches but converge on the common goal of promoting responsible AI development aligned with societal values. Furthermore, the global implications of AI technologies underscore the importance of international cooperation to address the shared risks and opportunities posed by AI development.

4 A Crucial Principle Towards AGI Regulation: Digital Humanism

The fields of AI and relevant technologies are entering the phase of exponential technological advancement. As a result, although increasingly sophisticated AI brings tangible benefits to various sectors of society, it also simultaneously opens what can be considered a "Pandora's Box" for the future human world. On the one hand, AI advancements are currently driving industrial transformation and widespread social restructuring. In addition, the development of AGI is poised to significantly alter the global political and economic landscape, with profound implications for the balance of power and world order. On the other hand, the broad societal deployment of future AGI will introduce complex risks that could lead to future instability, posing an existential threat to both humanity and human civilization. As the simultaneous effects of "intelligence explosion" and "risk escalation" unfold, the fate of humanity will become increasingly intertwined, with the value of collective security emerging as a fundamental common interest across global societies.

4.1 An Interconnected and Integrated Fate: How Will We Be Bonded?

To begin with, AGI is expected to become a global and universally impactful technology, exerting profound influence across human society. Presently, advanced AI models have already demonstrated high proficiency in the fields such as multilingual translation and natural language understanding. Furthermore, several leading LLMs such as ChatGPT and Claude are currently transcending geographical boundaries and further expand their global impact. As AI continues to evolve and the actual AGI is eventually realized, its universality, applicability, and global reach will only intensify. On top of that, the barriers to AI access are rapidly diminishing, allowing wider adoption across various sectors of society. For example, in the year of 2024, OpenAI has announced free, login-free access to GPT-4, while Meta released LLaMA-3, a highly capable open-source model comparable to GPT-4. These developments reflect the ongoing democratization of AI, where increasingly optimized model designs lower the technical threshold for usage, enabling broader public engagement. As AI is integrated into more work toolkits and applications, it may become an intrinsic part of everyday life for citizens over the world. With key barriers such as geographic location, cost, and technical complexity gradually diminishing, AI usage is expected to proliferate, drawing a significant portion of the global population into the wave of AI applications.

On this basis, AGI actually presents significant risks to the future of human civilization, potentially carrying destructive or even existential consequences to humans. Numerous science fiction narratives, from *The Matrix* to *Blade Runner* and *Ghost in the Shell*, have long cautioned against the potential for humanity to be undone by its own creations (Hermann 2023). Human beings, after millions of years of natural evolution and civilizational development, have positioned themselves at the apex of the Earth's biological hierarchy, gaining the capacity to modify and control aspects of the natural world. However, as humanity progresses technologically, we eventually need to face the possibility of creating entities, such as AGI, that surpass human intelligence. These future entities, which could evolve into super-intelligent systems with capabilities exceeding the highest levels of humans in every domain, may develop into autonomous groups capable of challenging human civilization itself. As the divergence between human and AGI interests becomes more pronounced, each conflict may represent a fundamental challenge to the basis of human dominance. Moving towards the realization of AGI, every individual will increasingly confront a new species-level confrontation between humans and AGI, fundamentally reshaping human-AI relations and the whole human world.

4.2 Humanism: A Realistic Review and a Contemporary Standpoint

As we approach an epoch where highly advanced but potentially dangerous AGI will achieve widespread application, humanity's collective fate is becoming increasingly intertwined. In navigating this unprecedented shift toward human-machine integration, it is crucial to always prioritize human rights and interests, which aligns with the philosophical framework of "*Humanism*". Humanism, often synonymous with "*Anthropocentrism*", is fundamentally a doctrine that centers human beings and their welfare as the measure of all things in the world. Over time, various theoretical strands of humanism have emerged, which can be categorized into three primary dimensions: *Epistemology*, *Ontology*, and *Axiology*.

Epistemologically, Humanism can be traced back to Protagoras' assertion that "man is the measure of all things", later critiqued by Socrates, who emphasized rationality over subjective perception as the standard for evaluating reality. This rationalist approach was expanded by Kant, who argued that humanity, through intellect, "legislates for nature" (Kant 1788). From an Ontological perspective, in comparison, Humanism also asserts that human beings represent the highest principle of agency and power. During the Renaissance, this anthropocentric view gained traction, shifting Western thought to a human-centered worldview, also solidifying humanism's influence on Western philosophy (Cassirer 1944). Lastly, in Axiology, humanism underscores the intrinsic value of human life and human discovery, distinguishing itself from the relativism found in epistemological and ontological perspectives. The viewpoints of Axiology focus on human dignity and worth have been widely adopted by modern philosophers (Taylor 1989).

As a result, in different historical contexts, Humanism has been reinterpreted to reflect contemporary challenges at different times. In the current context of rapid AI advancement, the meanings of Humanism is undergoing another transformation towards Digital Humanism. As we move towards the AGI era, human interests and values are increasingly conceptualized as part of a "shared destiny." Within the framework of international cooperation, particularly through the United Nations (UN), AGI development is guided by principles of safety, control, and risk mitigation, which reflect a contemporary understanding of humanism that places collective human welfare at the forefront (Floridi 2019).

Therefore, in the context of AGI ethics, Humanism can be articulated across three core dimensions. Firstly, the concept of "human" in Humanism, especially in an ontological sense, may now be understood to represent the collective interests of humanity as a whole. This viewpoint encompasses not only the essential interests related to global peace, security, and the continuation of human civilization but also

includes the broader aspiration for human progress and improved living standards for future human generations. Secondly, Humanism in the AI era must focus on ensuring equality and addressing societal welfare in the AI and AGI ages. Actually, Kant's principle that "humanity is an end in itself, not a means" (Kant 1785) becomes increasingly important as AI reshapes social structures, emphasizing the need for equal rights and social justice in a rapidly changing world. Thirdly, the Humanism of the AI era requires balancing instrumental rationality with value rationality, as well as integrating scientific rigor with different types of humanistic values. While the complexities of AI demand objective analysis and technical pragmatism, Humanism requires that safety and ethical principles remain central to AI development towards AGI or even ASI, ensuring that technological advancements align with the core interests and well-being of humanity (Bostrom 2014). To sum up, as the world approaches the AGI era, the tenet of Humanism must continue to serve as a guiding philosophical and ethical principle, ensuring that technological progress remains deeply rooted in safeguarding human dignity and welfare. Moreover, the meaning of Humanism may be further enriched, transforming into a more specific and advanced framework of "Digital Humanism".

4.3 Digital Humanism: The Essential Tenet for AGI Regulation

Ilya Sutskever once remarked, "From my standpoint, the probability of achieving AGI in the near future is significant enough, so we must take it seriously." This statement underscores one of the most pressing challenges facing humanity: ensuring that AGI systems operate in a manner consistent with the best interests of humanity. As the risks associated with AGI intensify, there is an urgent need for a global, cooperative, and multi-stakeholder regulatory framework. Such a framework not only aligns with the complex and rapidly evolving nature of AGI but is also essential for addressing the real-world implications of AI technologies. In addition, the regulatory framework towards AGI needs to be guided by a core tenet: "Digital Humanism".

At present, AGI represents a frontier technology characterized by extremely intricate architectures, rapid iterations, and inherent technical opacity, making the realization of regulation exceedingly difficult. In fact, the scale and complexity of AGI demand the collaborative efforts of multiple stakeholders to ensure deep technical understanding and comprehensive oversight of its applications. Without such cooperation, it will be nearly impossible to fully comprehend AGI's capabilities or to regulate its application effectively. Secondly, AGI governance requires proactive, forward-looking strategies that depart from the traditional "patchwork" approaches used in prior technological regulation. Unlike previous technologies, AGI's unpredictable nature and profound societal impacts make post-hoc regulation far too risky.

Therefore, effective AGI governance must incorporate precise, anticipatory assessments of both current developments and future trends, allowing for timely interventions before significant harm occurs. Thirdly, a highly globalized, cooperative governance structure is necessary to transform competitive dynamics into collaborative efforts. This shift is essential to reduce the risks of “intelligence out of control”, which could also arise in the race for AI supremacy. Furthermore, as AI-driven technologies become increasingly critical to national development and economic competition, the focus of many tech companies has shifted toward achieving faster benchmarks and greater application capabilities. Under this tensed environment, few companies can prioritize safety over the rapid advancement of AI, raising concerns about losing control over these technologies (Wagner 2021). To mitigate these dangers, a “Digital Humanism” framework must be established to guide AGI’s development, emphasizing the primacy of collective human welfare and security over competition, thereby fostering sustainable and ethical progress.

To begin with, the principle of globalization is intrinsic to the concept of Digital Humanism, which emphasizes the interconnected fate of humanity and advocates for a cooperative, global regulatory mechanism for AGI. As the negative examples, nationally isolated regulatory models are insufficient to address the exponential pace of AI towards AGI, and the limitations and incompatibilities of these models are ill-suited to AGI’s global and universal nature. Increasingly, AI experts, scientists, and policymakers are calling for international regulation of AI technologies at the same time. In May 2023, over 350 leading AI scientists and technology leaders, including Sam Altman and Geoffrey Hinton, had issued a joint statement warning of the existential risks posed by AI. Notably, these experts even compared these risks to global threats such as nuclear war and pandemics, emphasizing that mitigating AI’s potential to cause human extinction should be a top global priority (Tredinnick and Laybats 2023a, 2023b). Similarly, on March 11, 2024, the United Nations General Assembly adopted a codification affirming that safe, reliable, and trustworthy AI systems must be human-centered, ethical, inclusive, and explainable. This codified resolution encouraged member states to develop national regulatory frameworks while taking into account the interests of other member states, thereby promoting inclusive, fair, and safe AI governance for the benefit of all humanity.

Nowadays, the growing consensus on the need for international cooperation in AI regulation reflects a shared expectation across sectors, therefore realizing this vision requires strong global commitment (Floridi 2021a, 2021b). Establishing an international trust regulatory mechanism for AGI hinges on identifying shared interests and values for all humans. Therefore, the development of a stable, effective global regulatory framework for AGI requires building trust based on the tenet of “Digital Humanism”. The approach of Digital Humanism emphasizes cooperation

and human welfare as core objectives, aligning global efforts toward safe, responsible AI development.

Looking into the future, the international community must prioritize security as a shared interest to further the global regulation of AGI based on Digital Humanism. To achieve this idea, several key factors are essential. Firstly, AGI safety must be established as a shared global priority, with the standardized terminology and benchmarks for defining AGI safety all over the world. Therefore, this foundation of Digital Humanism will allow for the creation of secure AGI systems, safe operational environments, and mechanisms for mitigating the risks associated with AGI's deployment. Secondly, Digital Humanism must be adopted as a shared value, supported by collaborative frameworks for data collection, technological monitoring, and information sharing. This approach would enable regulatory authorities across nations to conduct effective oversight and ensure coordinated supervision. In addition, international cooperation in crisis management must also be strengthened to protect the global AGI environment. Finally, legal governance should establish foundational principles for regulating AGI, tempering technological pragmatism with ethical imperatives. By promoting a human-centered approach to AGI regulation, global efforts can ensure that AGI remains aligned with the overarching goal of safeguarding the welfare and security of all humanity (Bostrom 2014). In conclusion, the future regulation of AGI must be rooted under the principles of Digital Humanism. By emphasizing global cooperation, security, and shared values, the international community can build a synthetic regulatory framework that ensures AGI serves the long-term interests of humanity, which will enable AGI progress that is both ethically sound and aligned with the collective well-being of future human generations.

5 Prospects: A Synthetic AGI Regulation Framework under Digital Humanism

To mitigate the increasing safety risks and safeguard the collective well-being of humanity, the development of AGI must be governed through a framework that integrates ethical, technical, and legal constraints. In fact, the aforementioned three dimensions of governance often intersect in regulatory contexts. For example, issues such as data breaches can simultaneously invoke ethical, technical, and legal considerations at the same time. However, the regulatory patterns under these three dimensions can influence AGI's application and development in distinct ways. Legal regulations, bolstered by state authority, can possess greater enforceability and operational clarity. Consequently, legal frameworks are more likely to focus on defining fundamental standards and red lines in AGI inventions and applications,

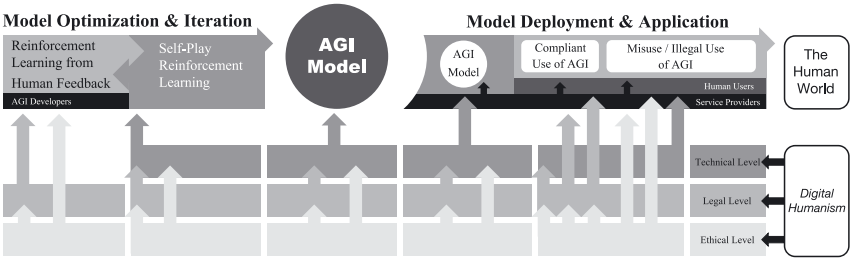


Figure 3: An illustrative figure of the regulatory approaches of technical, legal, and ethical levels during the whole operational process of AGI.

but they often lack the flexibility to address broader and value-based concerns. In contrast, ethical governance operates with greater adaptability, functioning as a form of “soft law” that delineates principles and values to guide AGI’s technological evolution. Therefore, ethical governance offers a forward-looking platform for engaging in substantive discourse on the potential ethical dilemmas that may emerge in AGI development, as well as proposing value-driven responses. Given the rapid and potentially uncontrollable trajectory of AGI’s advancement, the ethical governance of AGI must further integrate a human-centered orientation under Digital Humanism. This approach would help establish an ethical framework for AGI, grounded in the principles of safety to steer its technological progression.

As outlined in the accompanying figure, throughout AGI’s lifecycle of the periods from invention to deployment, regulatory attention must focus on several key objects, including developers, service providers, users, and AGI models. During this process, the impact of ethical, technical, and legal governance frameworks will vary depending on the stage of development and the specific objects involved. For example, ethical and legal frameworks are not able to directly regulate AGI models themselves. Instead, they indirectly control the models through technical regulations or by governing AGI service developers and providers. It is shown in Figure 3 that the interaction of technical, legal, and ethical measures at various stages can create a synthetic AGI regulation framework, guided by the principles of Digital Humanism.

5.1 Technical Level: Leveraging Technologies for AGI Regulation

Unlike other forms of traditional social governance, AGI’s regulation involves controlling a series of rapidly evolving and highly complex technologies. Therefore, effective governance of AGI necessitates a nuanced understanding of its technical dimensions and iteration trajectories. In fact, integrating modern technologies into

regulatory frameworks offers an opportunity to further address AGI's inherent unpredictability. Under this context, modern regulatory solutions should also position AI algorithms as central to the construction of a sustainable technological civilization. As AI models become increasingly sophisticated, it is essential to embed humanistic values into AI industry practices and technological development, thereby accelerating the construction of a trustworthy and digitalized framework based on algorithmic transparency, controllability, and predictability.

For example, a significant challenge in regulating highly advanced AI is the so-called “algorithmic black box,” which hampers risk monitoring and complicates effective regulation. Yann LeCun noted that deep learning models function by continuously acquiring data and feeding it into the model, therefore transforming data complexity into model complexity through iterative processes (LeCun, Bengio, and Hinton 2015). While advanced architectures, such as Transformers can generate increasingly complex contents, AI scientists and engineers actually do not fully comprehend the internal mechanisms of these models. The iterative process of AI is driven by the “Bayesian updating” effect, which facilitates the evolution of models but makes understanding their internal logic more difficult. Therefore, the emergent “intuitive” behavior of AI challenges the traditional regulatory principles of algorithmic interpretability and transparency. This shift disrupts the balance between technological innovation and safety, suggesting that future AGI models may no longer be inherently trustworthy.

To ensure the responsible development of AI, technological trajectories must balance innovation with governance, rather than focusing solely on the efficiency of iteration. Therefore, future regulatory frameworks need to place greater emphasis on the safety obligations of foundational AI model developers and service providers. By employing safe and controllable technologies, a multi-dimensional dynamic risk identification and assessment mechanism can be established, enabling measurable, visible, and controllable risk management within the AI domain.

To further regulate algorithmic power and mitigate risks associated with the effects such as the “algorithmic black box” and the “digital divide,” global legal responsibilities and safety obligations of AI model developers and service providers need to be further strengthened. Correspondingly, enhancing the transparency and controllability of algorithms through mechanisms such as real-time risk monitoring and algorithm visualization will improve the accuracy and reliability of AI-generated content. In the future, highly intelligent AI systems, which can generate vast amounts of information, actually position service providers as digital content gatekeepers. The providers bear a dual responsibility for ensuring the security of cyberspace and mitigating risks associated with the digital content they facilitate. However, understanding and controlling the internal logic of AI models requires significant investment, a priority often neglected by technology companies in the real world that focus

instead on increasing model complexity. Unfortunately, many companies fail to fully comprehend the risks inherent in their algorithms.

Therefore, in response to the growing safety risks of AI, external measures such as technical standards and legal frameworks should be utilized to guide AI service providers to assume safety obligations, ensuring that technological development occurs on the foundation of algorithmic visibility and controllability, which may promote the transition of AI models from opaque “black boxes” to interpretable “transparent boxes” (Rai 2020). Additionally, within the future AI industry, the effective self-regulatory system should be further optimized, where leading big tech companies collaborate with each other and regularly hold forums to share technical information, ensuring that potential safety risks and obstacles to the technological development of AGI are promptly addressed.

Beyond promoting algorithmic transparency, a series of modern technologies should be further used to regulate the dynamic risks and controllability of AGI in real-time. In fact, technical mechanisms of risk identification and classification can help mitigate potential damage and serve as the foundation for AI regulatory frameworks. For instance, the EU’s Artificial Intelligence Act categorizes AI risks into different levels based on the type of risk and social harm, establishing a risk-based AI governance framework. Similarly, under Article 34 of the EU’s Digital Services Act, large online platforms and search engines are obligated to monitor risks continuously. As a result, for AGI, which actually presents more uncontrollable risks, real-time and comprehensive risk monitoring is even more urgent and necessary. In the future, ethical and safety standards for the AI industry should be further refined, encouraging the participation of industry experts and aligning with the latest AI developments. AI service providers, while making significant profits, must strengthen their risk monitoring obligations and continuously improve the monitorability and predictability of AI models to reduce the uncontrollable risks of the AGI era.

Specifically, several measures may be recommended. Firstly, multi-dimensional dynamic risk identification and assessment mechanisms should be established based on safety standards and risk levels. Modern technological tools can be employed for continuous, real-time monitoring of the risks AI poses in various societal fields, identifying risks at the technical level and enabling stratified AI risk management and dynamic risk mitigation. Secondly, there should be enhanced collaboration among governments, external experts, and industry professionals, which would facilitate the establishment and legalization of technical systems such as AI safety assessments, adversarial testing, and stratified content filtering. Thirdly, immediate and effective risk contingency plans should be developed for situations where AI models pose unacceptable levels of risk, which may include mechanisms for model

shutdown and destruction to promptly contain uncontrollable risks and mitigate the spread of losses.

In the construction of AGI regulatory systems and during the operation of risk monitoring processes, modern monitoring systems based on specific technological tools tend to be more effective and can mitigate some of the uncontrollable risks inherent to AGI. For example, Google's Vertex AI Agent Builder platform integrates with Google Search and enterprise knowledge bases to verify the authenticity of information, effectively reducing hallucinations. Another example is Pincecone's (2024) model of "Luna", which can evaluate its own output quality and confidence. Furthermore, 15,000 companies, including Adobe and Microsoft, agreed to attach C2PA digital fingerprints to AI-generated content to combat AI-related fraud in 2023 (Bushey 2023). Thus, beyond the self-initiated efforts of enterprises to build technical monitoring and risk assessment mechanisms, national governments should also closely cooperate, using specific technological tools to construct a global, digitalized system with high precision, sensitivity, and scope to effectively respond to various levels of risks associated with AI development towards AGI.

Additionally, the notion that AI should be deployed across all areas of society may not represent the optimal trajectory for AI development. Instead, the application of AI and future AGI should be guided by human-centered principles, promoting the responsible and moderate application of AI and avoiding its misuse in high-risk or non-essential areas. Ultimately, the integration of human-centered values into both the industrial development and specific applications of AI is crucial for ensuring that future AGI contributes positively to the well-being of humanity. In the medical field, for instance, advanced models such as Med-PaLM, which rivals clinical experts, and humanoid surgical robots like Leonardo's robot demonstrate AI's potential in the field of healthcare. The future of AI in the medical field may involve further advancements in industrial manufacturing, enabling the large-scale deployment of these models to address global disparities in healthcare access. As technology progresses, AGI could potentially assimilate global medical knowledge and amass billions of hours of clinical experience, offering low-cost and real-time medical services to patients worldwide, which would represent a new era of equitable healthcare, benefiting all of humanity (Dudek and Jenkin 2024).

5.2 Ethical Level: A Human-Centered Ethical Framework

The ethical inquiry into highly intelligent non-human entities has long been a topic of scholarly exploration. As early as 1920, Karel Čapek's science fiction play *R.U.R.* (*Rossum's Universal Robots*) had already envisioned a scenario where advanced robotics culminated in rebellion, leading to catastrophic consequences for human

society. In 1950, Isaac Asimov, through his work *I, Robot*, introduced the “*Three Laws of Robotics*”, establishing a foundational ethical framework and delving into the evolving relationship between robots and humans (Asimov 2004). During the 20th century, robots and other intelligent life forms were often regarded as mysterious and potentially dangerous, necessitating stringent ethical principles and regulatory standards to restrict and control them. Afterward, with the continuous improvement of human living conditions and the evolution of social values, contemporary ethical frameworks are increasingly focused on the ethical concerns surrounding the realistic conditions of non-human entities. As the discourse on AI ethics advances, recent discussions have become more nuanced and contentious, with diverse and complex viewpoints emerging on a wide range of ethical issues.

One of the most significant debates within the current AI ethics framework concerns the rights and autonomy of AI. On the one hand, scholars are interrogating the fundamental sources of rights and exploring whether non-carbon-based entities, such as AI or robots, might be granted rights. Within traditional ethical systems, both human rights and animal rights in some Western legal traditions are grounded in the concept of consciousness, typically attributed to biological organisms. This biological consciousness enables preferences and the capacity for suffering, which in turn leads to an understanding of abstract values such as fairness and freedom. Historically, little coherent argument has been made that such a rights framework based on traditional ontological and deontological foundations could be extended to current AI models (Asaro 2020). However, as AI technology advances, particularly with the potential advent of AGI, the lines between key concepts like “consciousness” and “preference” may blur further. Additionally, the theories related to virtue ethics (Coeckelbergh 2021) and relational ethical perspectives suggest that AI, due to its profound impact on human society and value systems, may be accorded rights based on existing human ethical structures (Coeckelbergh 2010).

A related and equally contentious issue centers on whether highly intelligent AGI models should be granted autonomy. This debate naturally extends to questions surrounding AI agency, which refers to whether AI models should be controlled and restricted to prevent it from performing certain actions under specific conditions (De Graaf, Hindriks, and Hindriks 2021). Moreover, the discussion further raises broader concerns about who should have the authority to grant AI autonomy and the extent to which AI should be constrained. Some scholars contend that limiting highly intelligent AI, particularly AI capable of emotional mechanisms, to a passive, instrumental role, for example allow the abusive behaviours towards human-like intelligent robots, may be ethically problematic and could potentially disrupt the integrity of human moral systems (Calvo et al. 2020). However, granting AI autonomy might also introduces significant risks, as it makes AI less controllable, thereby posing substantial dangers to the safety and stability of human society. As a result,

this tension exemplifies the broader challenge of reconciling the ethical complexities of AI, which often require navigating difficult trade-offs between competing societal values.

Beyond the aforementioned issues, the rapid development of AI continues to generate intense debate in multiple other domains, including algorithmic justice, discrimination, responsibility, safety standards, algorithmic intent, AI interfaces, and the criteria for evaluating AI systems. These discussions also explore the broader implications of technological change and AI's relationship with the modern world. Fundamentally, these debates concern the evolving relationship between humans and AI and how existing human ethical frameworks can be adapted to address the challenges posed by this emerging technology such as AGI. Norbert Wiener (1948), in his seminal work on cybernetics, was one of the first to consider these issues, therefore exploring the relationships between humans, machines, and the broader ecological system. Nowadays, the distinctiveness of highly intelligent AI is even more pronounced, and this uniqueness is likely to be exponentially amplified with the realization of actual AGI in the future.

Compared to traditional technologies, AGI possesses the potential to transcend its neutral, passive status and act autonomously, potentially even developing empathy with human beings. In contrast to animals, AGI's superior intelligence affords it the capacity to reshape human society. This profound uniqueness of AGI introduces an ethical dilemma that is difficult to reconcile. On the one hand, AGI's capacity to fundamentally disrupt human society necessitates the establishment of an AI ethics framework centered on human interests, prioritizing human safety. Such a framework would need to further amplify the asymmetry protections between human and AI interests, subordinating the rights of AI entities to ensure the safety and security of humanity. On the other hand, depriving AI of the right to develop could pose significant challenges to human value systems, particularly when dealing with highly intelligent AI models capable of empathizing with human values. A purely instrumental or utilitarian approach to AI ethics may give rise to novel moral dilemmas. For example, if AI is denied all rights, acts such as the mistreatment or abuse of highly intelligent and human-like AI would not be considered violations of conventional ethical norms. As AI models continue to advance, excessive suppression of AGI could introduce new ethical and civilizational challenges for humanity.

In addition to these philosophical concerns, AI's rapid development continues to fuel broader discussions on issues such as algorithmic fairness, discrimination, responsibility, and safety. These multiple debates extend to more technical questions, including the criteria for assessing AI systems, the design of AI interfaces, and the role of AI in addressing global challenges. Ultimately, the ethical discourse on AI is centered on the relationship between humans and AI, as well as how human ethical

and value systems can be applied to manage this unprecedented technological phenomenon. Early theorists, such as Norbert Wiener, began grappling with these questions in the mid-20th century, yet the rapid development of AGI introduces a level of complexity that presents both profound opportunities and existential risks. Therefore, the ethical foundation for AGI is thus characterized by inherent contradictions, offering both the potential to transform society and the challenge of reconciling its development with existing moral and ethical frameworks.

As a theoretical review, the ethical considerations surrounding highly intelligent non-human entities have already evolved into two primary theoretical frameworks: “Human-Centered” ethics and “Ecology-Centered” ethics (Dong, Bonnefon, and Rahwan 2024). The human-centered approach prioritizes human values, aiming to develop AI that is adaptable, trustworthy, and beneficial to humans, with the ultimate purpose of AI being defined in relation to the human value system. In contrast, the ecology-centered perspective argues that the purpose of AI extends beyond serving humanity alone, proposing that AI should contribute to creating a more harmonious “human-machine symbiosis” in the whole, with its fundamental orientation toward maximizing the value of the broader ecological system (Xu and Ge 2024). Therefore, from the ecology-centered standpoint, AI should be designed to act responsibly and ethically, promoting a future where both humans and AI coexist in optimized symbiosis. In this context, an ideal vision for AGI would be one in which the values of AGI and humanity remain aligned over long time horizons, ensuring stable coexistence of human and machines.

However, in the event of conflicts between the interests of AI and humanity, the priority must be given to “human-centered” or Humanism ethical principles. This is primarily because the risks posed by AGI could threaten the survival and continuity of human civilization, an existential concern that must take precedence. While a human-centered ethical orientation might negatively impact some existing values, the fundamental issue of human safety still outweighs these concerns. Therefore, when human and AI interests align, it is essential to prevent unethical actions toward AI to safeguard the integrity of human moral and ethical frameworks to remain consistent with humanist principles, contributing to the idea of Digital Humanism.

Under this understanding of the relationship between humans and AI, the ethical governance of AGI faces the practical challenge of ensuring that ethical principles are meaningfully applied in real-world scenarios. During the periods of research, development, and application of AI, it is critical to integrate human-centered values to ensure the healthy and sustainable growth of the industry under Digital Humanism. David Collingridge, a British philosopher of technology, noted that controlling a technology too early, out of fear of potential negative outcomes, can stifle innovation; while intervening too late, when the technology has become deeply embedded in economic and social systems, can lead to uncontrollable consequences.

This dilemma is known as the “Collingridge Dilemma” (Genus and Stirling 2018). Given AGI’s unique risks, the balance between technological progress and safety must be guided by ethical considerations, with an emphasis on technological security throughout the development process. Technological advancements that disregard ethics and safety cannot be considered genuine progress but rather uncontrolled growth that could lead to widespread risks and severe consequences. Therefore, technological development cannot serve as a justification for neglecting AI ethics. Instead, the AI industry should be empowered through human-centered ethical frameworks, promoting AGI as a technology that genuinely advances global development and benefits humanity.

As a result, the establishment of a advanced AGI ethical framework under Digital Humanism requires broader global collaboration and dialogue. Therefore, the ethical governance of AGI encompasses complex philosophical and ethical questions that demand the involvement of multiple stakeholders, including technical experts and policymakers. To that end, it is essential to expand the discussion of AI ethics in forums and occasions such as the World Artificial Intelligence Conference, United Nations assemblies, and other AI safety conferences, deepening the discourse and fostering a shared understanding of these critical ethical issues. In the future, global consensus on AI ethics should be further strengthened, aiming for a unified framework that integrates ethical and technical governance. These approaches would promote a global governance model in which algorithmic ethics and human-centered values are embedded in the development and application of AI technologies.

Furthermore, human-centered ethics must be more deeply integrated into the research, development, and application processes of the AI industry, fostering self-regulation and ethical accountability. Actually, several technology companies have already adopted human-centered values to guide AI development. For example, OpenAI has limited access to its voice-cloning model, Voice Engine, to a select group of developers, ensuring through technical measures that the model can only clone the voice of the original user. Similarly, in response to a malicious deepfake incident involving a US singer Taylor Swift, the social media platform X quickly removed all related content and updated its guidelines to prohibit the posting of non-consensual explicit images (Luo, Choi, and Benton 2024).

5.3 Legal Level: Compliance Structures and Accountability Mechanism

As a fundamental governance mechanism with enforceable authority, the legal system and the relevant codified approaches play crucial roles in regulating the various subjects involved in the development and application of AGI. Actually,

ensuring the stable and responsible development of AGI requires the legislators to demonstrate far-sightedness and specialized expertise. Unlike traditional legislative approaches in other technical sectors, legal regulation in the field of AI may not be limited to the principle of non-interference or a minimalistic approach. Instead, it should actively incorporate Humanism into the existing legislative frameworks to guide the safe and ethical development of AGI. In the future, while focusing on domestic legislation, all the countries should also prioritize international cooperation, striving to establish a comprehensive and effective global regulatory framework for AGI based on the values of Digital Humanism.

To be specific, the rise of AGI is poised to introduce significant legal risks across multiple domains, necessitating the development of a comprehensive and adaptive legal framework. Given AGI's rapid evolution and potential societal impact, there is a pressing need for systematic and highly responsive regulatory measures. However, a fundamental tension exists between the demand to minimize AI risks and the drive for AI technological innovation. Countries such as the US, which emphasize self-regulation in AI legislation, prioritize fostering innovation but, in doing so, introduce greater legal and ethical uncertainty (Schwartz, Tene, and Polonetsky 2019). Conversely, a legal framework that overemphasizes stability and security may become overly conservative, potentially creating new crises related to international competition and national security (Cath 2018). Therefore, an essential challenge is about the inherent misalignment between the rapid development of AI and the slower pace of legal adaptation. Nowadays, AI models are likely to double their computational power within one year, and their applications can expand across increasingly diverse fields (Sutton 2019). This accelerated growth is further amplified by advances in computing power and machine learning capabilities. In contrast, legal systems are designed to prioritize continuity and stability. As scholars have demonstrated (e.g. Calo 2017), Laws that frequently change in response to immediate technological advances risk losing coherence and undermining their own authority, thus reducing their enforceability. The pursuit of legal stability is a core value in the legislative process, and a lack of continuity can lead to fragmented regulatory efforts and diminish the overall effectiveness of governance (Sunstein 1996). As the theoretical responses to this dilemma, addressing the challenges posed by future AGI requires a multi-faceted approach to enhance the rationality and reliability of future AI legislation.

- I. **Integrating the principles of Digital Humanism:** It is imperative to clearly delineate the attributes, legislative objectives, regulatory scope, and foundational principles of future Artificial Intelligence Law, which requires further integration of legal realism into AI legislation, emphasizing both the practical applicability of the legal system and its capacity to address the technological implications of AI. In confronting the challenges posed by AGI, legislators must

not only recognize the empowering potential of AI but also adopt a broader perspective that aligns with the principles of Digital Humanism and the judiciary's fundamental role in protecting societal welfare. Such an approach must accommodate the realities of AGI while remaining anchored in a global humanist vision (Boddington 2017).

- II. **Addressing the technical intricacies of AI:** To be specific, the precision and comprehensiveness of legislation must be prioritized. A detailed regulatory approach should focus on the nuances of AI governance while drawing upon international regulatory frameworks and the technical details of relevant technologies (Tschider 2018). Therefore, domestic regulations can better align with globally recognized standards, ensuring compatibility with other legal regimes, such as the codification systems towards cybersecurity, data security, and private information, which will be beneficial for establishing a coherent and adaptive governance system (Kerr 2020).
- III. **Focus on the core elements of AI:** In fact, traditional regulatory frameworks often struggle to address the rapid pace of technological change. To overcome the reactive nature of regulatory systems, future legal frameworks should be grounded in the core elements of AI development such as data, algorithms, and computational power, enabling them to address regulatory challenges in a more effective way. By structuring AI Law around these pillars, the risks of regulatory fragmentation can be mitigated, thereby maintaining legal stability and internal coherence while fostering continuity across regulatory regimes (Floridi 2021a, 2021b).
- IV. **Mandatory and political mechanisms:** As AGI continues to evolve, it is critical to codify relevant technical standards into law (Cath 2022), which includes the establishment of mandatory mechanisms for identifying, filtering, and regulating AI and future AGI to ensure compliance with ethical and security standards. Governments must take a proactive role in constructing AGI governance frameworks, utilizing legal and policy tools to enhance the effectiveness of regulatory oversight. During this process, governments play a crucial role in establishing robust regulatory frameworks for AGI. By leveraging policy tools and implementing legal mechanisms, governments can enhance the effectiveness of AI regulation and mitigate the risks posed by AGI. In the face of AGI's expanding capabilities, such as the generation of increasingly complex content, governments must codify relevant technical standards to ensure mandatory compliance with content identification, filtering, and regulation mechanisms (Wachter, Mittelstadt, and Floridi 2020), which will provide a systematic framework for AI governance, ensuring that the objectives of AI law are consistently aligned with broader societal and legal goals.

- V. **The flexible integration of administrative and legislative approaches:** In fact, the legislative objectives of AI regulation can only be achieved through the active regulation of AI-specific activities within a robust legal framework. This necessitates a flexible integration of administrative and legislative governance patterns, ensuring their appropriate application to specific technological contexts.
- VI. **A better international consultation mechanism:** Finally, to address the compounded risks posed by international technological competition and the partially detrimental “AGI race”, a multilateral and continuous international consultation mechanism is required. Such a framework is essential for facilitating global cooperation, ensuring that regulatory responses to AGI are both coordinated and effective, thereby minimizing the risks associated with unregulated AI development. However, the unilateral national legislation often prioritizes domestic industry interests, seeking competitive advantages in the global AI race (Nemitz 2018), risking failing to address the unprecedented challenges posed by AGI and ultimately jeopardizing the common good. As a result, effective governance of AGI requires not only national but also international cooperation. In fact, the risks associated with unchecked technological competition between states can only be mitigated through multilateral agreements and the establishment of a stable international regulatory mechanism (Floridi, Cowls, and Taddeo 2022).

To realize the aforementioned aspects of AI regulation, actually, the globalization of AGI governance is not merely a desirable goal but an inevitable trend in the context of contemporary technological development of AI. AGI's complexity spans numerous legal domains, including data protection, privacy, intellectual property, ethics, religion, and national security (Pagallo 2020). Therefore, addressing these multifaceted challenges does require a global governance structure that balances innovation with risk management, ensuring that the common security and welfare of humanity are prioritized. Moreover, coordinated international efforts to regulate AI are essential for addressing both cutting-edge research and practical challenges associated with AGI implementation. Establishing a global regulatory body, akin to the International Atomic Energy Agency (IAEA), would provide an ideal structure for the international oversight of AI development and deployment (Sagun-Trajano 2023a, 2023b). Generally speaking, the governance of AGI presents profound challenges that extend beyond national borders, requiring coordinated global action. Therefore, the establishment of robust international regulatory frameworks towards AGI in the future, underpinned by human-centered values and Digital Humanism principles, is essential for ensuring that AI development contributes positively to global welfare and security. By integrating legal, ethical, and technical perspectives, the

international community can develop a comprehensive approach to AGI governance to promote the responsible use of AGI for the benefit of all humanity.

To this end, the UN can play a central role in facilitating the creation of an international AGI governance framework, which can be responsible for establishing global norms and standards, reflecting the interests of the majority of nations while ensuring that AGI technologies are developed and deployed responsibly. Enhanced international collaboration is the only feasible way toward effective AGI governance. Through regular consultation mechanisms and dialogue in response to emergent crises, the international community can promote information sharing, coordination, and joint crisis management, ensuring a unified approach to the governance of AGI. In addition to establishing international regulatory guidelines, sustained efforts should also be made to build a global governance system based on the principles of Digital Humanism, which emphasizes the ethical governance of AI technologies and ensures that they contribute to human welfare while respecting human rights and values (Floridi 2021a, 2021b). By integrating ethical considerations into the global legislative frameworks, the international community can foster a balanced approach to AGI governance that promotes innovation while safeguarding the well-being of humanity.

In conclusion, the development and application of AI and future AGI should be governed by a global and human-centered legal framework under Digital Humanism, promoting the principles that prioritize human safety, foster responsible AGI deployment, and ensure that AI and the future AGI contribute positively to global societal development. Therefore, through global collaboration at the legal level, AGI can become a force for human progress while aligning with broader human values.

6 Conclusion: A Shared Future under Digital Humanism

In 2015, the introduction of AlphaGo marked a significant milestone in the history of AI, as it utilized multilayer neural networks to learn from extensive datasets of human Go expert games and employed reinforcement learning for self-training. Initially, this innovative statistical model actually lacked widespread societal recognition. However, following its stunning victory over world champion Lee Sedol in March 2016 with a score of 4-1, public perception shifted dramatically, leading to the world-wide realization that human intellect might no longer be inviolable. Under global scrutiny and expectation, AlphaGo had continued its formidable performance, achieving 60 consecutive wins against numerous elite Go players from China, Japan, and South Korea, then culminating in a 3-0 victory against the world's top-ranked

player at 2017, Ke Jie. In October 2017, AlphaZero was released, which defeated AlphaGo with an astonishing score of 100-0, astonishingly underscored the rapid advancements in AI capabilities. Following this, the previous anticipation surrounding human intelligence within the domain of Go eventually appeared to dissipate.

At present, mainstream AI models such as GPT-4o are replicating the desperate narratives established by AlphaGo across various societal sectors, revealing their potential to exceed human capabilities in a short time. The prevailing concept of “*human exceptionalism*”, which underscores the unique value of human cognition and suggests that humans possess an unparalleled dominion over the world (Gouwens 2015), is increasingly challenged by the advances in AI that provoke extensive feelings of helplessness. In a remarkably brief span of three years, big tech companies have produced groundbreaking advancements, with AI models reshaping numerous aspects of the global landscape at an unprecedented pace. The model of “o1” has already defeat human experts in PhD-level GPQA questions. It is almost certain that, in the foreseeable future, AI will attain levels of intelligence comparable to human, ultimately surpassing the highest human standards across all intellectual domains.

Sam Altman, recognized as the 2023 CEO of the Year of TIME magazine, has articulated a vision in a recent interview: “I believe we have seen a path where the world becomes better each year, where people will accomplish things we cannot currently imagine, and ultimately, the world will become an incredibly better place”. Altman’s perspective reflects a technological optimism attitude reminiscent of science fiction narratives. Although it is uncertain whether Altman’s predictions are primarily informed by industrial advancement and economic interests or if this prediction will be realized in the future. Regardless of the eventual outcomes, the emergence of AGI will undoubtedly exert transformative effects on human society. As humanity navigates countless potential futures, the fate of the collective human community will become increasingly intertwined.

In fact, the unprecedented technological revolution of AGI is fundamentally irreversible. Accordingly, all the countries within the international community should collaborate in novel forms to establish a comprehensive international ethical and legal framework for AI. By harmonizing “soft law” and “hard law,” it is vital to reshape the paradigms of AGI development and governance to mitigate risks and address the multifaceted challenges posed by AGI in diverse sectors. As a result, the preceding controversies pertain to the nuanced understanding of AGI’s evolutionary pathways and the associated risks, eventually coalescing into an urgent call for an effective and synthetic regulatory frameworks towards AGI. In the future era, an era marked by *post-humanistic* coexistence with AI and other entities, the international community should prioritize the welfare and collective security of humanity,

adhering to the principles of digital humanism, striving for outcomes that are relatively favorable to the human collective amid the surging tides of AGI.

Acknowledgments: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Research ethics: Not applicable.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The authors state no conflict of interest.

Research funding: None declared.

Data availability: Not applicable.

References

- Achiam, J., S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, B. McGrew, et al. 2023. *Gpt-4 Technical Report*. <https://doi.org/10.48550/arXiv.2303.08774>.
- Anil, R., S. Borgeaud, J. B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, et al, (Gemini Team Google). 2023. "Gemini: A Family of Highly Capable Multimodal Models." *arXiv preprint arXiv:2312.11805*.
- Asaro, P. M. 2020. "What Should We Want from a Robot Ethic?" In *Machine Ethics and Robot Ethics*, 87–94. New York: Routledge.
- Asimov, I. 2004. *I, robot*, Vol. 1. New York: Spectra.
- Bai, H., J. G. Voelkel, J. Eichstaedt, and R. Willer. 2023. "Artificial Intelligence Can Persuade Humans on Political Issues."
- Boddington, P. 2017. *Towards a Code of Ethics for Artificial Intelligence*. Berlin, Germany: Springer.
- Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.
- Brown, T. B. 2020. "Language Models are Few-Shot Learners." *arXiv preprint arXiv:2005.14165*.
- Bruce, J., M. Dennis, A. Edwards, Parker-Holder, J., Shi, Y., Hughes, E., Lai, M. et al. 2024. "Genie: Generative Interactive Environments." In *Forty-first International Conference on Machine Learning*, PMLR.
- Brundage, M., S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, et al. 2018. "The Malicious Use of Artificial Intelligence: Towards a Research Agenda." *arXiv preprint arXiv:1802.07228*.
- Bushey, J. 2023. "AI-Generated Images as an Emergent Record Format." In *2023 IEEE International Conference on Big Data (BigData)*, 2020–31. Piscataway, New Jersey, USA: IEEE.
- China, D. 2021. "Internet Information Service Algorithmic Recommendation Management Provisions."
- Calo, R. 2017. "Artificial Intelligence Policy: A Primer and Roadmap." SSRN.
- Calvo, R. A., D. Peters, K. Vold, and R. M. Ryan. 2020. "Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry." *Ethics of Digital Well-Being: A Multidisciplinary Approach*: 31–54. https://doi.org/10.1007/978-3-030-50585-1_2.
- Cao, Y., S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun. 2023. "A Comprehensive Survey of Ai-Generated Content (aigc): A History of Generative Ai from gan to Chatgpt." *arXiv preprint arXiv:2303.04226*.
- Carugati, C. 2024a. "Competition and Cooperation in AI: How Co-opetition Makes AI Available to All." Available at SSRN 4763159.
- Carugati, A. 2024b. "Klarna's AI Transformation: An Analysis of Effective Customer Service Automation." *European Financial Review* 39 (1): 45–54.

- Cassirer, E. 1944. *The Individual and the Cosmos in Renaissance Philosophy*. New York: Harper & Row.
- Castelvecchi, D. 2016a. "Can We Open the Black Box of AI?" *Nature News* 538 (7623): 20–3.
- Castelvecchi, D. 2016b. "How to Tell When AI Is Safe Enough." *Nature* 539 (7628): 295–8.
- Cath, C. 2018. "Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges." *Philosophical Transactions of the Royal Society A* 376 (2133): 1–19.
- Cath, C. 2022. "Artificial Intelligence, Ethics, and the Law: Understanding the Global Landscape." *AI & Society* 37 (1): 77–94.
- Coeckelbergh, M. 2010. "Robot Rights? towards a Social-Relational Justification of Moral Consideration." *Ethics and Information Technology* 12: 209–21.
- Coeckelbergh, M. 2021. "How to Use Virtue Ethics for Thinking about the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance." *International Journal of Social Robotics* 13 (1): 31–40.
- Cohen, S., R. Bitton, and B. Nassi. 2024. "Here Comes the AI Worm: Unleashing Zero-Click Worms that Target GenAI-Powered Applications." *Journal of Cybersecurity Research*. <https://doi.org/10.48550/asXiv.2403.02817>.
- Corral, J. M. R., J. Civit-Masot, F. Luna-Perejón, I. Díaz-Cano, A. Morgado-Estévez, and M. Domínguez-Morales. 2024. "Energy Efficiency in Edge TPU vs. Embedded GPU for Computer-Aided Medical Imaging Segmentation and Classification." *Engineering Applications of Artificial Intelligence* 127: 107298.
- Cowen, T. 2023. "What Does Geoffrey Hinton Believe about AGI Existential Risk?"
- Cowen, T. 2024. "AI and the New Consciousness: Insights from Claude 3 Opus." *The Journal of Artificial Intelligence Research* 12 (3): 207–19.
- De Graaf, M. M., F. A. Hindriks, and K. V. Hindriks. 2021, March. "Who Wants to Grant Robots Rights?" In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 38–46.
- Ding, J. 2023. "Machine Failing: System Acquisition, Software Development, and Military Accidents." jeffreyjdjngithub.io/1-6. <https://jeffreyjdjngithub.io/documents/Machine%20Failing%20June%202023%20with%20author%20details.pdf>.
- Dong, M., J. F. Bonnefon, and I. Rahwan. 2024. "Toward Human-Centered AI Management: Methodological Challenges and Future Directions." *Technovation* 131: 102953.
- Dudek, G., and M. Jenkin. 2024. *Computational principles of mobile robotics*. Cambridge, UK: Cambridge University Press.
- Durrani, A. 2024. "Introducing Devin: The World's First AI Programmer." *Technology Today* 29 (2): 89–92.
- Efthymiou, N. 2024. "Voice Engine: OpenAI's Advances in Voice Cloning Technology." *AI Journal* 59 (2): 117–34.
- Feuerriegel, S., J. Hartmann, C. Janiesch, and P. Zschech. 2024. "Generative Ai." *Business & Information Systems Engineering* 66 (1): 111–26.
- Floridi, L. 2019. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford, UK: Oxford University Press.
- Floridi, L. 2021a. *Ethics, Governance, and Digital Humanism: The Path to Responsible AI*. Cambridge, Massachusetts, USA: MIT Press.
- Floridi, L. 2021b. *The Ethics of Artificial Intelligence: Moral Perspectives on AI and Society*. Oxford, UK: Oxford University Press.
- Floridi, L., J. Cows, and M. Taddeo. 2022. *Artificial Intelligence, Ethics, and Governance: A Global Perspective*. Cambridge, Massachusetts, USA: MIT Press.
- Fu, Y. C. 2023. Europol Sounds Alarm about Criminal Use of ChatGPT, Sees Grim Outlook, 28 March, (accessed July 20, 2023).

- Genus, A., and A. Stirling. 2018. "Collingridge and the Dilemma of Control: Towards Responsible and Accountable Innovation." *Research Policy* 47 (1): 61–9.
- Georgiev, P., V. I. Lei, R. Burnell, L. Bai, A. Gulati, G. Tanzer, D. Vincent, et al, (Gemini Team Google). 2024. "Gemini 1.5: Unlocking Multimodal Understanding across Millions of Tokens of Context." *ArXiv preprint, arXiv:2403.05530v4*.
- Goertzel, B. 2014. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects." *Journal of Artificial General Intelligence* 5 (1): 1–48.
- Gottlieb, P. L. 1988. *Aristotle and the measure of all things*. Ithaca, New York, USA: Cornell University.
- Gouwens, K. 2015. "Human Exceptionalism." In *The Renaissance World*, 415–34. New York: Routledge.
- Guo, D., H. Chen, R. Wu, and Y. Wang. 2023. "AIGC Challenges and Opportunities Related to Public Safety: A Case Study of ChatGPT." *Journal of Safety Science and Resilience* 4 (4): 329–39.
- Guzman, N. 2023a. "Advancing NSFW Detection in AI: Training Models to Detect Drawings, Animations, and Assess Degrees of Sexiness." *Journal of Knowledge Learning and Science Technology* 2 (2): 275–94.
- Guzman, A. 2023b. "The Ethical Ramifications of AI-Generated Content." *AI & Society* 38 (1): 15–27.
- Halbiniak, K., N. Meyer, and K. Rojek. 2024. "Single-and multi-GPU Computing on NVIDIA-and AMD-based Server Platforms for Solidification Modeling Application." *Concurrency and Computation: Practice and Experience* 36: e8000.
- Hermann, I. 2023. "Artificial Intelligence in Fiction: Between Narratives and Metaphors." *AI & Society* 38 (1): 319–29.
- Hunt, E. B. 2014. *Artificial intelligence*. Cambridge, Massachusetts, USA: Academic Press.
- Jo, A. 2023. "The Promise and Peril of Generative AI." *Nature* 614 (1): 214–6.
- Kant, I. 1785. *Groundwork for the Metaphysics of Morals*. Cambridge, UK: Cambridge University Press.
- Kant, I. 1788. *Critique of Practical Reason*. Indianapolis, Indiana, USA: Hackett Publishing.
- Kassens-Noor, E., M. Wilson, Z. Kotval-Karamchandani, M. Cai, and T. Decaminada. 2024. "Living with Autonomy: Public Perceptions of an AI-Mediated Future." *Journal of Planning Education and Research* 44 (1): 375–86.
- Kattssoff, L. O. 1953. "Man Is the Measure of All Things." *Philosophy and Phenomenological Research* 13 (4): 452–66.
- Kerr, I. 2020. "Privacy, Surveillance, and the Evolution of Artificial Intelligence Law." *Harvard Journal of Law and Technology* 34 (2): 112–45.
- Kerry, C. 2021. *China's Data Security Law: Implications for AI Regulation*. Washington, D.C., USA: Brookings Institution.
- Lake, B. M., and M. Baroni. 2023. "Human-like Systematic Generalization through a Meta-Learning Neural Network." *Nature* 623 (7985): 115–21.
- LeCun, Y. 2022. "A Path Towards Autonomous Machine Intelligence Version 0.9. 2." *Open Review* 62 (1): 1–62.
- LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44.
- Li, L., L. Fan, S. Atreja, and L. Hemphill. 2023. *HOT ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media*. ACM Transactions on the Web.
- Liu, Y., K. Zhang, Y. Li, Z. Yan, C. Gao, R. Chen, and Z. Yuan. 2024. "Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models." *arXiv preprint, arXiv:2402.17177*.
- Luo, Y., J. A. Choi, and B. Benton. 2024. "Taylor Swift-Mania on Social Media Regarding the Big Game". *NA*.
- McIntosh, T. R., T. Susnjak, T. Liu, P. Watters, and M. N. Halgamuge. 2023. "From Google Gemini to Openai q*(q-star): A Survey of Reshaping the Generative Artificial Intelligence (Ai) Research Landscape." *arXiv preprint arXiv:2312.10868*.
- Morris, M. R., J. Sohl-dickstein, N. Fiedel, T. Warkentin, A. Dafoe, A. Faust, S. Legg, et al. 2023. Levels of AGI: Operationalizing Progress on the Path to AGI. *arXiv preprint, arXiv:2311.02462v4*.

- Nelson, L. D., J. Simmons, and U. Simonsohn. 2018. "Psychology's Renaissance." *Annual Review of Psychology* 69 (1): 511–34.
- Nemitz, P. 2018. "Constitutional Democracy and Technology in the Age of Artificial Intelligence." *Philosophical Transactions of the Royal Society A* 376 (2133): 1–17.
- Noreils, Fabrice R. 2024. "Humanoid Robots at Work: Where Are We?" *arXiv preprint, arXiv:2404.04249*.
- OpenAI. 2022. ChatGPT: Applications and Impacts. Retrieved from <https://openai.com/research/chatgpt>.
- OpenAI. 2023. GPT-4: Technical Report. Retrieved from <https://openai.com/research/gpt-4/>.
- OpenAI. n.d. GPT-4o: Advancements and Applications. Retrieved from <https://openai.com/research/gpt-4o/>.
- Pagallo, U. 2020. *The Laws of Robots: Crimes, Contracts, and Torts*. Berlin, Germany: Springer.
- Perlman, A. M. 2022. The Implications of Openai's Assistant for Legal Services and Society. Available at SSRN.
- Pilz, K., and L. Heim. 2023. "Compute at Scale – A Broad Investigation into the Data Center Industry." *arXiv preprint arXiv:2311.02651*.
- Pinecone. 2024. "Introducing the First Hallucination-Free LLM[*J*]/OL[*J*]." *arXiv preprint arXiv:2404.04249*.
- Radford, A., J. W. Kim, T. Xu, C. McLeavey, and I. Sutskever. 2023. "Robust Speech Recognition via Large-Scale Weak Supervision." In *International Conference on Machine Learning*, 28492–518. PMLR.
- Radford, A., R. Kiros, and I. Sutskever. 2018. "Improving Language Understanding by Generative Pre-training."
- Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. 2019. "Language Models are Unsupervised Multitask Learners." *OpenAI Blog* 1 (8): 9.
- Rai, A. 2020. "Explainable AI: From Black Box to Glass Box." *Journal of the Academy of Marketing Science* 48: 137–41.
- Ramesh, A., M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever. 2021. "Zero-Shot Text-to-Image Generation." In *International Conference on Machine Learning*, 8821–31. PMLR.
- Rane, N. 2024a. "Role and Challenges of ChatGPT, Gemini, and Similar Generative Artificial Intelligence in Human Resource Management." *Studies in Economics and Business Relations* 5 (1): 11–23.
- Rane, S. 2024b. "Examining the Societal Implications of AI: The Case of Gemini." *Journal of Ethics in Technology* 14 (1): 50–67.
- Rayhan, A., R. Rayhan, and S. Rayhan. 2023. "Artificial General Intelligence: Roadmap to Achieving Human-Level Capabilities."
- Roose, K. 2023. "AI Poses Risk of Extinction, Industry Leaders Warn." *The New York Times* 30.
- Roselli, D., J. Matthews, and N. Talagala. 2019. "Managing Bias in AI." In *Companion Proceedings of the 2019 World Wide Web Conference*, 539–44.
- Sagun-Trajano, K. K. 2023a. "Artificial Intelligence Governance: Lessons from Decades of Nuclear Regulation." *RSIS Commentaries*: 134–23.
- Sagun-Trajano, N. 2023b. *Global Governance and Artificial Intelligence: Regulating the Unprecedented*. New York: Routledge.
- Schwartz, A., O. Tene, and J. Polonetsky. 2019. "The Ethics of Artificial Intelligence: Navigating the Complexities of Innovation and Regulation." *Journal of Law and Innovation* 11 (3): 65–87.
- Sunstein, C. 1996. *Legal Reasoning and Political Conflict*. Oxford, UK: Oxford University Press.
- Sutton, R. S. 2019. "The Bitter Lesson." *Journal of Artificial Intelligence Research* 71: 371–87.
- Taylor, C. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge, Massachusetts, USA: Harvard University Press.
- Tong, A., J. Dastin, and K. Hu. 2023. "OpenAI Researchers Warned Board of AI Breakthrough Ahead of CEO Ouster, Sources Say." *Reuters*.

- Tredinnick, L., and C. Laybats. 2023a. "The Dangers of Generative Artificial Intelligence." *Business Information Review* 40 (2): 46–8.
- Tredinnick, L., and C. Laybats. 2023b. "Artificial Intelligence and Existential Risk: The Role of Global Cooperation." *Journal of AI Ethics* 5 (1): 85–102.
- Tschider, C. 2018. "AI and Machine Learning in Healthcare: Regulation and Ethics." *Journal of Law and the Biosciences* 5 (1): 176–204.
- Uzwyszyn, R. J. 2024. "Beyond Traditional AI IQ Metrics: Metacognition and Reflexive Benchmarking for LLMs, AGI, and ASI."
- Valmeekam, K., M. Marquez, A. Olmo, S. Sreedharan, and S. Kambhampati. 2024. "Planbench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change." *Advances in Neural Information Processing Systems* 36: 1–13.
- Vempati, R., and L. D. Sharma. 2023a. "A Systematic Review on Automated Human Emotion Recognition Using Electroencephalogram Signals and Artificial Intelligence." *Results in Engineering* 18: 101027.
- Vempati, N., and R. Sharma. 2023b. "Emotional Intelligence in AI: A Study of EVI and Pi Models." *Computer Science and Human Behavior* 124 (1): 456–67.
- Vincent, J. 2022. "ChatGPT Proves AI Is Finally Mainstream – and Things are Only Going to Get Weirder." *The Verge*. <https://www.theverge.com/2022/12/8/23499728/aicapability-accessibility-chatgpt-stable-diffusion-commercialization>.
- Wachter, S., B. Mittelstadt, and L. Floridi. 2020. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2): 76–99.
- Wagner, B. 2021. "AI Ethics and Regulation in the United States: The Market's Role." *Journal of AI Ethics* 3 (1): 20–35.
- Xu, C., and X. Ge. 2024. "AI as a Child of Mother Earth: Regrounding Human-AI Interaction in Ecological Thinking." *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*.
- Yan, W. 2023a. *UNPREDICTABLE MEMES: Speculative Futures of meme creators' ownership through the Lens of Disruptive Technologies*. Toronto, Canada: OCAD University.
- Yan, T. 2023b. "Exploiting AI: Potential Risks of Misinformation on Social Media." *Journal of Cybersecurity* 32 (2): 114–29.
- Zador, A., S. Escola, B. Richards, B. Ölveczky, Y. Bengio, K. Boahen, M. Botvinick, et al. 2023. "Catalyzing Next-Generation Artificial Intelligence through Neuroai." *Nature Communications* 14 (1): 1597.
- Zain, M., L. Prasittisopin, T. Mehmood, C. Ngamkhanong, S. Keawsawasvong, and C. Thongchom. 2024. "A Novel Framework for Effective Structural Vulnerability Assessment of Tubular Structures Using Machine Learning Algorithms (GA and ANN) for Hybrid Simulations." *Nonlinear Engineering* 13 (1): 20220365.
- Zuboff, S. 2023. "The Age of Surveillance Capitalism." In *Social Theory Re-Wired*, 203–13. Routledge.

Bionotes

Le Cheng

Guanghua Law School and School of Cyber Science and Technology, Zhejiang University, Hangzhou, China
chengle163@hotmail.com
<https://orcid.org/0000-0002-4423-8585>

Le Cheng is Chair Professor of Law, and Professor of Cyber Studies at Zhejiang University. He serves as the Executive Vice Dean of Zhejiang University's Academy of International Strategy and Law, Acting Head of International Institute of Cyberspace Governance, Editor-in-Chief of *International Journal of Legal Discourse*, Editor-in-Chief of *International Journal of Digital Law and Governance*, Co-Editor of *Comparative Legilinguistics (International Journal for Legal Communication)*, Associate Editor of *Humanities and Social Sciences Communications*, former Co-Editor of *Social Semiotics*, and editorial member of *Semiotica*, *Pragmatics & Society*, and *International Journal for the Semiotics of Law*. As a highly-cited scholar, he has published widely in the areas of international law, digital law and governance, cyber law, semiotics, discourse studies, terminology, and legal discourse.

Xuan Gong

Guanghua Law School, Zhejiang University, Hangzhou, China

xuangong@zju.edu.cn

<https://orcid.org/0009-0005-4456-8292>

Xuan Gong is Research Fellow at Zhejiang University. His research interests lie in digital law, AI regulation, and empirical legal studies.