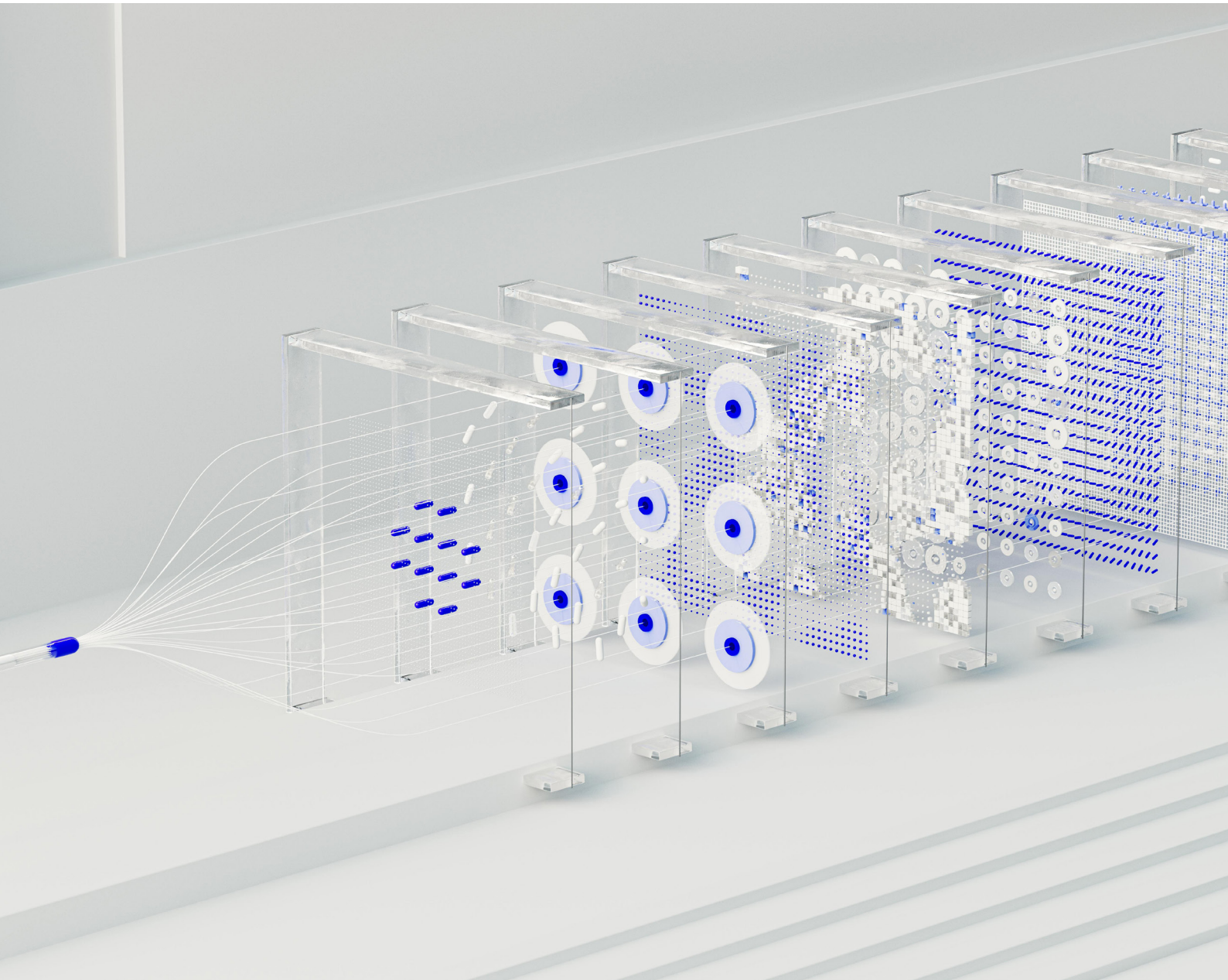
UNITED ARAB EMIRATES
MINISTER OF STATE FOR ARTIFICIAL INTELLIGENCE,
DIGITAL ECONOMY & REMOTE WORK APPLICATIONS OFFICE

Towards a Future of Responsible AI

WHITE PAPER



Foreword
1

The dawn of the artificial intelligence (AI) age is upon us, ushering in a dynamic new era of technological capabilities and possibilities and paving the way for an AI-enabled future for the benefit of humanity.

The pervasiveness of the technology is evident, and its impact is remarkable. Its use is likely to bring forth a new age of human civilization, and humanity aspires for a future where the use of AI is responsible. But what does responsible AI mean? Who decides on the definition? How do we envision humanity's future in the age of AI?

This was the focus of a Round Table Assembly hosted by the United Arab Emirates at the World Governments Summit (WGS), where experts from government, academia, industry, and NGOs came together to discuss issues such as open-sourcing foundational technologies, the explainability of AI, and technological equity.

The thought-provoking discussions from this meeting and the deep conversations that followed show the importance of collaborative efforts by leaders and experts in shaping the future of responsible AI. As we navigate this rapidly evolving landscape, it is essential that we continue to have meaningful conversations. Only through collaborative efforts can we ensure our future is a future of responsible AI.

Omar Sultan Al Olama
Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications

Introduction to the White Paper

2

The World Governments Summit represents an annual opportunity to bring together thought leaders from government, policy, private sector, technology, and industry to engage in high-level discussion and debate about future trends, issues, and opportunities facing humanity. This annual event has been taking place in the United Arab Emirates since 2013 and has consistently brought thought leaders together on critical issues.

This White Paper, entitled Towards a Responsible Future of AI is designed to represent the proceedings of a Round Table Assembly attended by global thought leaders in the field of Artificial Intelligence. This forum provided a multinational platform for dispassionate debate relating to the opportunities and challenges for a global regulatory framework in support of a human-enabling, ethical future for AI.

Round Table Delegates 3

Government

H.E. Omar AlOlama
Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications, UAE

H.E. Faisal AlBannai
Secretary General of ATRC and Chairman, AI7

H.E. Majed AlMesmar
Director General Telecommunications and Digital Government Regulatory Authority, UAE

Talal AlKaissi
Executive Vice President and Chief Product and Global Partnerships Officer, G42

H.E. Dr. Amr Talaat
Minister of Communications and Information Technology, Arab Republic of Egypt

Andres Sutt
Member of Parliament Parliament of Estonia

H.E. Giedre Balcytyte
Chancellor, Government of the Republic of Lithuania

Jungwoo Ha
The Presidential Committee on the Digital Platform Government - South Korea

H.E. Kathi Vidal
Under Secretary of Commerce for Intellectual Property and Director, U.S. Patent and Trademark Office, U.S. Department of Commerce

H.E. Mauricio Lizcano Arango
Minister of Information Technologies and Communications, Republic of Colombia

H.E. Mohamed Louly
Ministry Of Digital Transformation Mauritania

H.E. Ulvi Mehdiyev
State Agency for Public Service and Social Innovations Under the President – Azerbaijan

Private Sector

AJ Abdallat
President and CEO
Beyond Limits

Dr. Alexander Karp
Co-founder and CEO
Palantir Technologies

Alexander Sukharevsky
Managing Partner
QuantumBlack

Arya Bolurfrushan
Chief Executive Officer
Applied AI Company

Çağlayan Çetin
President
Trendyol Group, Republic of Türkiye

Prof. Eric Xing
MBZUAI
President and University Professor

Gary Kazantsev
Head of Quant Technology Strategy, Bloomberg

Greg Wilson
Worldwide Public Sector Government CTO, Microsoft

Jensen Huang
Founder and CEO
NVIDIA

Jonathan Ross
Founder and CEO
Groq

Lucas Joppa
Senior Managing Director, Haveli Investments and Conservation Philanthropist

Luis Videgaray
Director
MIT Management Sloan School

Miguel Correa
General, Affinity Partners

Naim Yazbeck
General Manager
Microsoft Corporation

Dr. Ray O. Johnson
Chief Executive Officer
Technology Innovation Institute

Sachin Duggal
Chief Wizard
Builder.ai

Seth Gerson
Chief Executive Officer,
Survios

Shawn Edwards
Chief Technology Officer
Bloomberg

Suad Khawaja
Parsons International Ltd

Dr. Werner Vogels
CTO
Amazon

Dr. Yann LeCun
Turing Award Laureate,
Vice President and Chief AI Scientist, Meta

Glossary 4

Terminology in AI is a fast-moving topic, and the same term can have multiple meanings. The glossary below should be viewed as a snapshot of contemporary definitions.

Artificial Intelligence

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

Artificial Intelligence Model

An AI model is a key part of an AI system. It is a program trained to recognize patterns in data and make specific predictions or decisions. For example, an AI model might learn to detect objects in images or transcribe human speech. While the AI model does the learning and predicting, it needs to be part of an AI system to be used effectively for real-world applications.

Artificial Intelligence System

An artificial intelligence (AI) system is a computer program designed to achieve specific goals. It learns from the data it gets to make predictions, create content, suggest options, or make decisions that can affect the real or online world. Some AI systems, if designed, can potentially improve their performance over time after training.

Automation

The use or introduction of automatic equipment in a manufacturing or other process or facility.

Autonomous Cars

An autonomous car is a vehicle that can guide itself without human conduction.

Bias

Inclination or prejudice for or against one person or group, especially in a way considered to be unfair.

Deep Learning

Deep learning is a subset of machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data.

Ethics

Moral principles that govern a person's behavior or the conducting of an activity.

Federated Learning

Federated Learning is a machine learning setting where the goal is to train a high-quality centralized model with training data distributed over a large number of clients, each with unreliable and relatively slow network connections.

Fine-Tuning

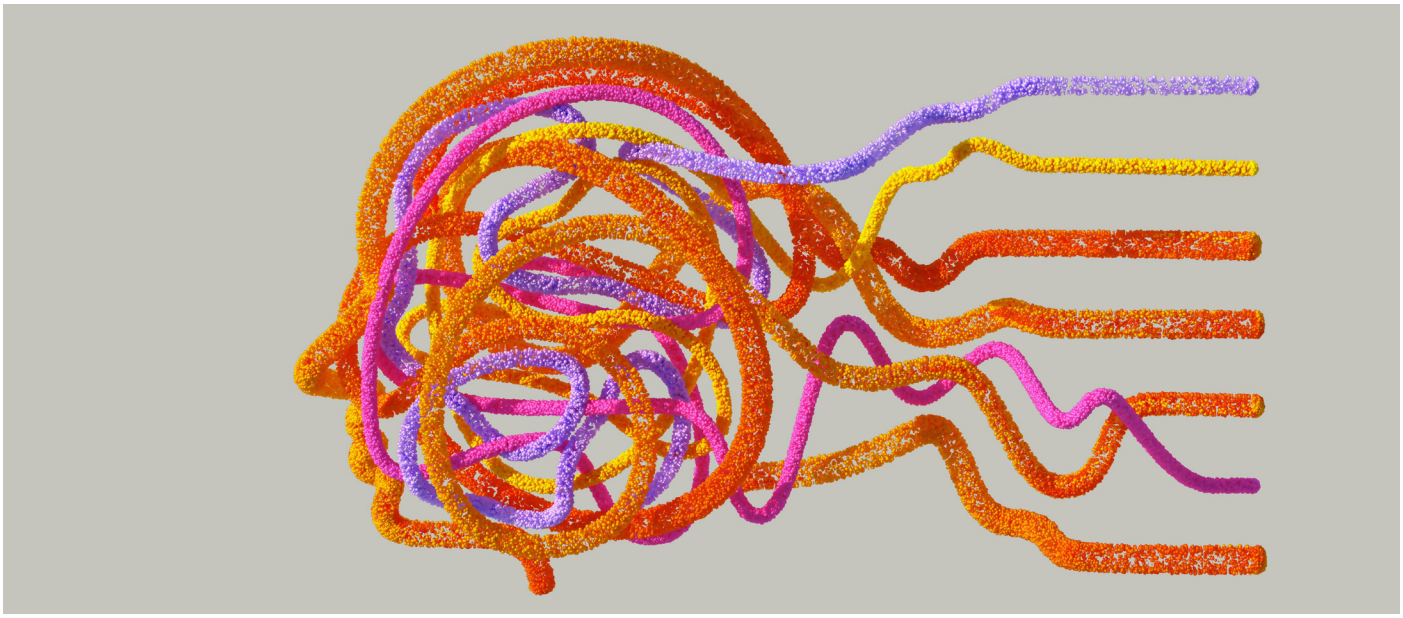
The process of adapting a pre-trained model to perform a specific task by conducting additional training relevant to that task, leading to updating the pre-trained model's parameters to better suit them to the task in question.

Foundation Model

A foundation model is an AI model that can be adapted to a wide range of downstream tasks. Foundation models are typically large-scale (e.g., billions of parameters) generative models trained on a vast array of data, encompassing both labeled and unlabeled datasets.

Generative AI

AI models specifically intended to produce new digital material as an output (e.g., text, images, audio, video, and software code), including when such AI models are used in applications and their user interfaces. These are typically constructed as machine learning systems that have been trained on massive amounts of data.



Glossary 4

General AI
General AI refers to a machine that can perform any intellectual, physical, and even emotional task that a human being could.

Hallucinations
Hallucinations occur when models produce factually inaccurate or untruthful information. Often, hallucinatory output is presented in a plausible or convincing manner, making detection by end users challenging.

Large Language Model (LLM)
A Large Language Model (LLM) is a type of generative AI model designed to understand and produce human-like text. Initially trained on vast amounts of diverse text data as a foundation model, an LLM can generate coherent and contextually relevant text based on the input it receives. After the initial training phase, an LLM can be used as is or fine-tuned for specific tasks, such as translation, summarization, question answering, or powering conversational agents.

Machine Learning
Machine learning is a branch of artificial intelligence that involves training algorithms to learn from and make predictions or decisions based on data. It uses statistical techniques to give computers the ability to "learn" from data without being explicitly programmed for specific tasks. Through this learning process, the algorithms can improve their performance over time as they are exposed to more data.

Model Bias
Model bias refers to systematic errors in an AI model's predictions or decisions that result from prejudices in the training data or the model's design. Bias can lead to unfair or discriminatory outcomes, particularly against certain groups of people based on attributes such as race, gender, or age.

Model Explainability
Model explainability refers to the ability to understand and interpret the decisions and predictions made by an AI model. It involves providing clear insights into how the model processes input data to produce outputs, making the model's behavior transparent and understandable to users.

Model Parameters
Model parameters are the internal settings of an AI model adjusted during training to help it make accurate predictions. These parameters determine how the model processes input data and generates output. During training, these parameters are tweaked to improve the model's performance, resulting in better quality output or more accurate predictions.

Model Training
Model training is the process of teaching an AI model to perform a specific task. This involves feeding the model large amounts of data so it can learn patterns and relationships within that data. For example, training a model to recognize cats in photos would involve showing it many images of cats until it becomes adept at accurately identifying new cat images on its own.

Narrow AI
Narrow AI is artificial intelligence that is focused on a single narrow task.

Neural Network
A computer system modeled to simulate the human brain and nervous system.

Pre-Training
Pre-training is a preliminary phase of training where a model learns from a very large general dataset before being further trained for a specific task. This preliminary training process helps the model develop a broad understanding of data, which can then be specialized with additional training. For instance, a language model might be pre-trained on a massive collection of text from the internet before being fine-tuned for tasks like answering questions or translating languages.

Reinforcement Learning
Reinforcement learning is a field of machine learning concerned with how software agents ought to take actions in an environment so as to maximise some notion of cumulative reward.

Robotics
Branch of technology that deals with the design, construction, operation, and application of robots.

Supervised Learning
Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

Unsupervised Learning
Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.

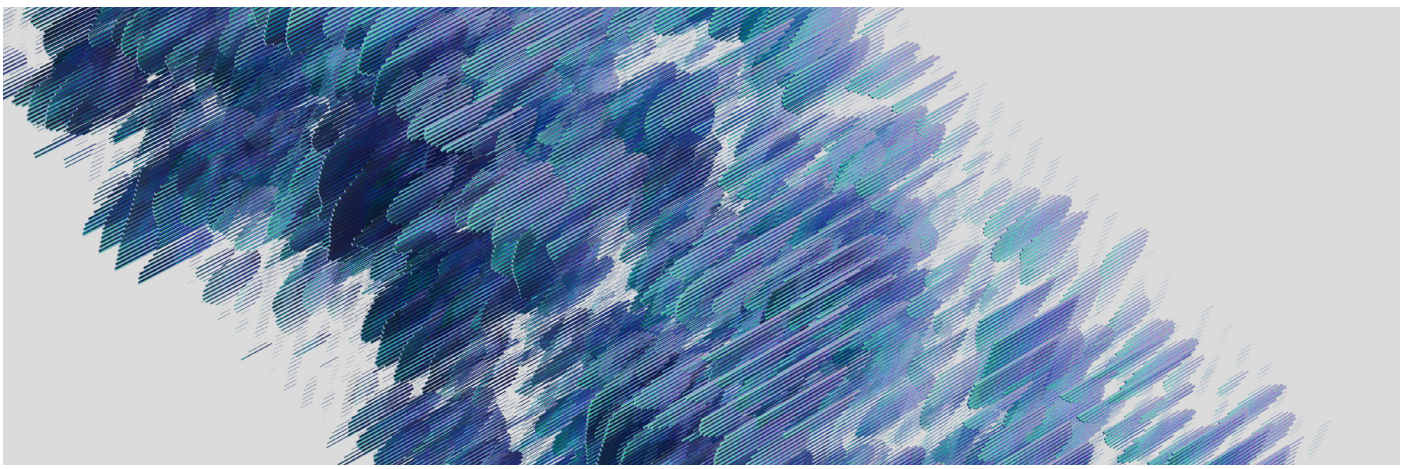
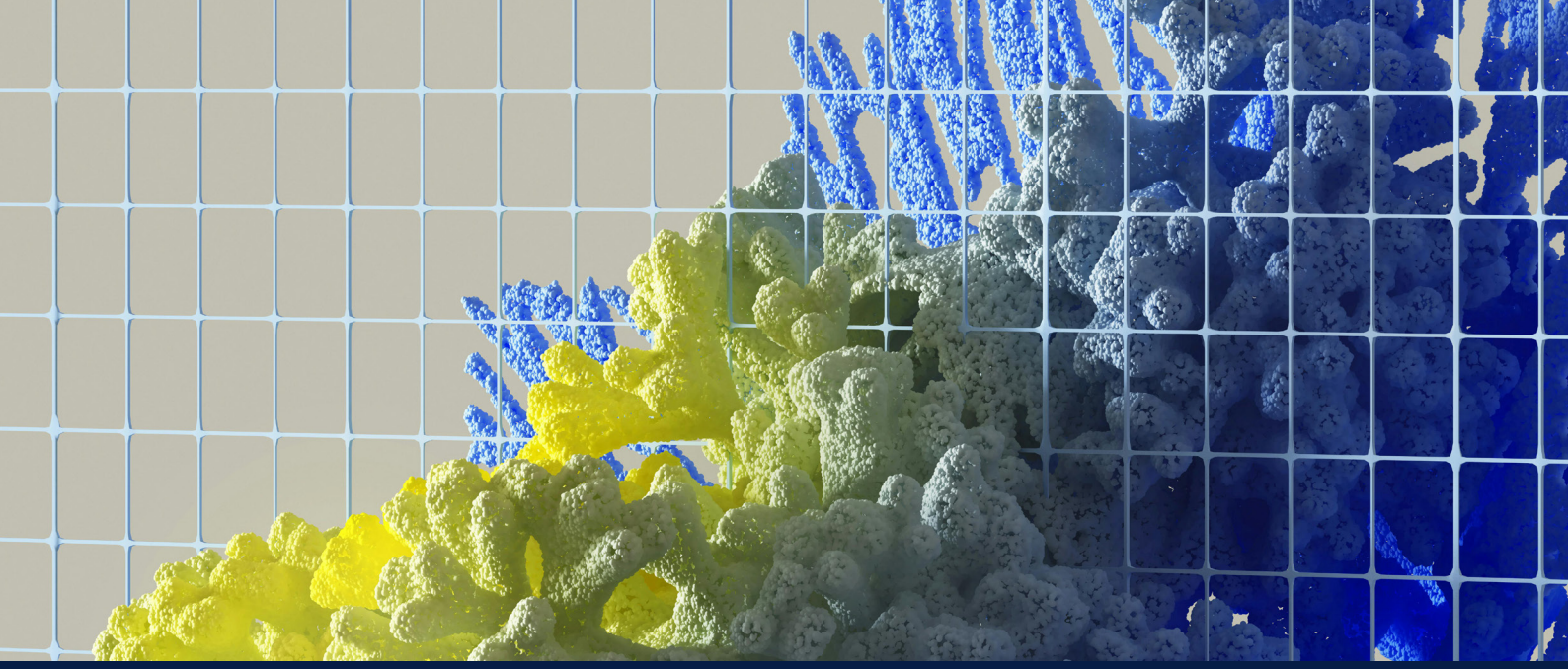


Table of contents



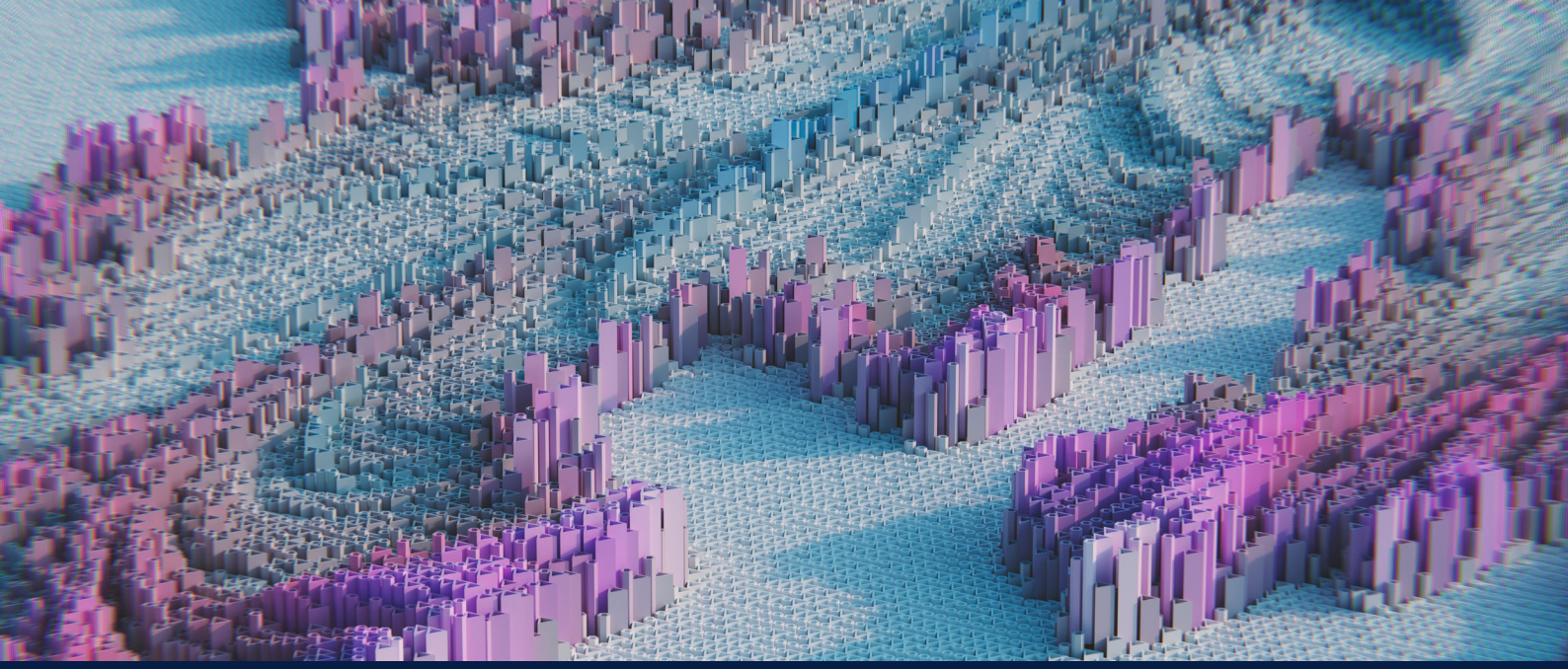
Executive Summary

5

In recent years we have seen artificial intelligence transition from the realms of scientific research and speculative fiction to becoming a cornerstone of everyday life. Its applications, ranging from simple automation to complex decision-making systems, have profound implications not only for economic growth and efficiency but also for societal norms, individual rights, and ethical standards. As we stand on the precipice of what many call the "AI era", the need for responsible, ethical, and well governed frameworks for AI development and deployment has never been more critical.

The burgeoning capabilities of AI systems, characterized by their ability to learn, adapt, and make autonomous decisions, pose unique challenges to existing regulatory and ethical frame works. Issues such as data privacy, algorithmic bias, and the accountability of AI systems have sparked intense debate among policymakers, technologists, and to some extent the general public. These concerns underscore the necessity of developing robust policy guidelines that not only foster innovation, but also ensure the protection of rights and values in a domain that, by definition, requires multi-national, multi-jurisdictional cooperation.

Acknowledging existing policy frameworks, academic literature, and legislation, this white paper seeks to describe and thematize the proceedings of the World Government Summit Round Table Assembly "Towards a Future of Responsible AI" to chart a course towards a future where AI contributes to the betterment of society while mitigating risks and dispelling fear, which is one of the biggest issues. AI-enabled thriving for human-kind recognizes that AI is simply a tool, and like any tool it reflects the values of those who design, develop, and deploy it. Therefore, a multidisciplinary approach, incorporating insights from computer science, law, ethics, sociology, human psychology, economics,



and politics, is paramount in crafting policies that are effective and equitable.

Several policy frameworks already provide a foundation for such an endeavor. The EU General Data Protection Regulation (GDPR) and the Ethics Guidelines for Trustworthy AI set early efforts for privacy and ethical considerations. Similarly, the academic discourse on algorithmic bias and human-value alignment offers critical insight into the complexities of AI governance. Furthermore, national and supra-national legislation, such as the EU AI Act, illustrates ongoing efforts to address the multifaceted challenges posed by AI technologies. The pioneering foresight demonstrated by the United Arab Emirates in their appointment of the world's first Minister for Artificial Intelligence in 2017, foregrounds this as an issue which requires prioritization from the height of political strategy.

This white paper aims to build upon these precedents to summarize, describe, and thematize internal discussions relating to a comprehensive framework for the future of responsible AI. By highlighting multidisciplinary discussions examining key principles such as transparency, accountability, representativeness, fairness, and respect for privacy—all in the pursuit of human-enabling innovation. In doing so, we not only address the immediate challenges but also summarize the groundwork for a future where AI serves as a force for good, enhancing human capabilities and enriching the fabric of society.

The World Government Summit (WGS) Round Table Assembly "Towards a Future of Responsible AI" brought together diverse perspectives on the ethical, transparent, and socially responsible development of artificial intelligence (AI). This white paper synthesizes key insights and recommendations from the discussion, emphasizing the need for a global ethical code for AI and output-driven regulation shaped by collaborative and human-orientated efforts.

Introduction to the Round Table

6

The Round Table Assembly initiated discussions on the challenges and opportunities in regulating AI. The prevailing sentiment was that the "cat is out of the bag," and as such its regulation becomes challenging. However where regulation is needed, it must be grounded in a human-centric approach. It was noted that AI is not only becoming more prevalent but rather it is already ubiquitous, so its use must be both ethical and transparent.

In this context, sustainability can be thought of as multi-faceted, from environmental considerations to the employability of blue-and white-collar members of the labor force. The discussion emphasized the need to move a more positive narrative, helping people to no longer view AI as a latter-day evil that is being imposed but rather a mechanism through which greater quality of life can be achieved. That is the challenge. The example was given of personal digital maps, and how their absence could harm the quality of life. An active dialogue with world-leading experts was encouraged to help determine what we technically can implement and what we cannot, given an often political and ideological view of the possible priorities. Debates relating to the necessity of an AI regulatory framework shared a common agreement that "bad actors" often move more quickly than good actors, with less regard for legislation, the rule of law, ethical frameworks, and governance. However, an ethical code that is universal, human-enabling, trust-building, and grounded in bottom-up as well as top-down inputs, would be not only a valuable but a necessary asset in the pursuit of responsible AI.

Governments and Data

7

Governments are increasingly compelled to manage vast quantities of real-time data to effectively meet the needs of their citizens and deliver services. The quality of AI data is an integral part of this challenge.

As governments accumulate more data on virtually every aspect of society, the necessity for sophisticated tools to handle this growing pool of information becomes evident. Artificial Intelligence emerges as an indispensable solution in this context. As AI's ability to analyze and interpret complex datasets in meaningful ways positions it as a critical asset for governments seeking to leverage data for decision-making, policy formulation, and service improvement. The integration of AI technologies will therefore play a pivotal role in transforming how governments operate and interact with their citizens, making it an essential component of modern governance strategies and delivery for citizens.

Establishing an Ethical code for AI

8

Establishing an ethical code for AI

The round-table discussion emphasized the necessity of establishing a comprehensive ethical code for AI. This code should be shaped by global collaboration, drawing on the insights of various stakeholders and organizations, such as the United Nations (UN), UNESCO, the G20 and the World Economic Forum (WEF), who are all working on AI governance. The goal is to avoid regulatory fragmentation and potential "arms races" based on competing interests, temporal obsolescence (non-future-proof), and lacking diversity and inclusivity. It was noted that approximately 160 guidelines on ethics for AI exist, with UNESCO leading efforts in this direction, and that any governance ideally needed to be grounded in a respectable global body, such as the United Nations. A global concensus at that level is clearly a challenge; however, the UAE was noted as an ideal convening entity by the delegates, having led the way globally in fostering AI at the heart of its own government. Drawing on existing positive exemplars, such as the AI. Safety Summit hosted by the UK in November 2023, the UAE is advocating for open-source AI and will have a head-of-state discussion on the matter with other nations in the future.

Explainability, transparency, and traceability

The importance of explainability in AI was highlighted, with a call to promote tools that provide accessible answers to users. This includes open-source initiatives that facilitate transparency in AI decision-making processes and were key to supporting trust at a societal level. The sense of a series of systems which are thought of as inaccessible to the public, which operate on a "black box" basis, is one of several key sources of fear among the public at large. This is further enabled by both foregrounding of negative use cases (e.g., deepfake images, the reallocation of labor to machines), and a lack of understanding of positive everyday use cases, such as mapping software with live traffic data, as well as decades of Science Fiction writing.

The matter of explainability was first discussed with delegates arguing that tools should be put in place that allow users to directly ask why an AI model came to a certain prediction and what facts were used to build this conclusion. They should allow for a more open understanding of how the model works. However, delegates explained that this question was nearly always answered with a rather unsatisfatory "because a specific training dataset was used." Using different training datasets could lead to different answers. This concept of traceability was consistently brought up as a consideration from a policy-making perspective, albeit it somewhat at odds with what is possible given current architecture (e.g., LLMs).

This settled discussions on the notion that legislation for use is more practicable than legislation at the input level. However, this heightened the case for open-source infrastructure, as a foundational assumption on the input/build side creates an axiomatic transparency, which was noted as "as much as we can hope for" given current architecture. As renewed architectures further enable AI this may render a more compelling and conducive case for legislation.

The Assembly addressed the matter of explainability, mentioning that current systems, with autoregressive LLMs, are inaccurate. In their current form, they hallucinate and are unable to reason and plan. They lack a persistent memory, an understanding of the physical world, and common sense. The reliable systems in terms of explainability and traceability that people want

are not possible with the current system, despite their usefulness. There will first need to be a major change in architecture. The point was made that there is no specific way to make an AI system more explainable; however, a better AI that has a better understanding of our world, a persistent memory, and the ability to reason will end up being more accurate. The delegates stated that there will be significant progress in building systems that exhibit such abilities within five years.

Some delegates wanted to explain how these models work to make the challenge of explainability clearer. The models work in a similar way to how one would play a game of chess. There is a sequence of moves that can be selected. In chess, you have around 30 moves: LLMs, however, have 30,000. They rank the moves and choose the next best one,, but instead of moving a piece across the board, they generate a word or a token. This was described as the model's intuition. If asked to explain why it made a decision, the intuition would again rank what is the next best word, rather than truly answering the question. This is why LLMs sometimes confabulate or hallucinate answers.

The Assembly had a debate around the matter of the model's fundamental understanding of the "real world". One point of view was that the models have no true understanding. Another view was that the models have a very limited basic understanding through intuition. However, there was no clear concensus.

Throughout the debate, delegates repeatedly emphasized the importance of traceability. Despite the complexity of underlying data, there is a need to ensure that decisions taken are traceable to raw pieces of data. This is absolutely necessarily for the technology to be practically usable, because only traceable decisions can be debugged and efficiently disputed to increase trust in AI.

Audit trails

The concept of audit trails was highlighted when discussing responsible artificial intelligence (AI). This concept of trust emerges as a fundamental cornerstone. Establishing trust is crucial for encouraging the widespread adoption and acceptance of AI in society. The essence of fostering this trust lies in the ability of users to understand and rely on the decisions made by AI systems. In the current landscape, dominated by machine learning and deep learning technologies, the process of arriving at decisions often resembles a "black box"—answers are provided without clear explanations of the underlying rationale. This opacity challenges the establishment of trust. One delegate brought up the example of the use of machine learning in the criminal justice system in the Midwest of the United States, where it is being trialed for various use cases. It was suggested that one can only be comfortable with machine learning evaluating a case if it includes an audit trail, honoring explainability and giving explainable answers to questions of probity.

The importance of explainability extends beyond mere transparency; it also addresses the quality of the training data. A recent example highlighted the limitations of AI in pediatric medical diagnoses, where the accuracy

was significantly low (15% was suggested). This discrepancy could be attributed to the training data, which may not have adequately represented the diverse patient demographics. This situation underscores the critical role of training data in the performance of AI systems and reinforces the argument that explainability ultimately hinges on the quality of the data used.

Despite the potential for inaccuracies, acknowledging and understanding the limitations and biases of AI systems can strengthen trust. Providing insights into the decision-making process, including the data and training methodologies used, will enable individuals to better understand the context of the outcomes. Recognizing that errors can occur, the key is to ensure that these errors are transparent, allowing for rectification. This approach not only builds trust, but also fosters a collaborative relationship between humans and AI In essence, the path to achieving broad acceptance of AI lies in our ability to imbue these systems with transparency and explainability, thereby establishing a foundation of trust that encourages engagement and integration into society.

Fostering trust through governance

It was emphasized that fostering trust is crucial for AI adoption. Delegates stressed the need for AI systems to be explainable, especially in critical areas like healthcare, where inaccuracies can have severe consequences. An example was given that even if AI gives bad answers, with sufficient explainability, we could still determine where and why it got it wrong. This speaks to ongoing discussions about the stratification of risk levels at the use end of AI, where the requirement for regulation in finance, healthcare, education, and sustainability makes an explicit regulation against harm imperative. On this topic, other delegates mentioned that AI has a role in governance worldwide that could centralize public services and thereby better serve the interests of all segments of society.

Human-in-the-loop

Incorporating human insight into the operation of AI is crucial. This is because solely relying on data might lean AI towards easy solutions. For more uncommon but perhaps creative solutions, having a human-in-the-loop might be a prerequisite. One way to prevent this gravitation towards simpler solutions is by using multi-agent LLMs trained on different, diverse datasets and have them communicate with each other before a solution is provided.

The Assembly addressed the need for AI sovereignty and a common open-source infra-

guages and dialects, allowing for easier customization of the platform and ensuring more cultural alignment with the place where an AI platform is to be used. The Assembly cautioned against regulatory blueprints that might hinder the development of AI Delegates emphasized the importance of changes in planning and architecture for a more reliable AI within a five-year time horizon, predicated upon infrastructure that can meet thresholds for persistent memory, understanding of the world, etc. This kind of reliable AI is good AI. The consistent message was that there is nothing to fear from AI as it has a limited ability to deal with the "real world," and that the best defense against current AI issues is to develop the next architecture of AI.

More broadly, in the context of governance objectives, the discussion underscored the importance of setting clear objectives within an agreed framework. However, the current challenge of explainability within the architecture of models was acknowledged, where choices are often limited and only grounded in intuition.

The role of open-source AI

The importance of open-source AI was echoed by delegates. The consensus in the Assembly stressed the importance of adhering to the wider concept of "open-source", especially the foundational models that can serve as a basic common building blocks. "We do not need a hundred different foundational models," delegates emphasized. This is a situation that might occur if the foundational models were not open-sourced. Regardless, a delegate added that whatever regulations come to be in effect, there will be pressure to open-source AI platforms, in a similar vein to what happened to the internet and its infrastructure. Delegates highlighted that Linux is not regulated, and is a collection of open-source operating systems. The Linux Foundation and Linus Torvalds, the creator of Linux, have not been sued to date for bugs or malfunctions of the system, nor can they be held liable for the system failing. Despite this, the use of Linux is pervasive and a wide range of consumer technology relies on it to keep running. Linux runs embedded systems in cars and automobiles. Meta (Facebook's parent company) makes use of Linux in many of its services. The internet runs on Linux. Cell phone towers rely on Linux. All Android phones are built on Linux. Open-source has allowed Linux to be safe and reliable without stringent regulation. Artificial Intelligence, if it is to become basic infrastructure, could - and perhaps should - operate on the same basis.

Safety, oversight, and privacy

9

The Assembly emphasized the importance of understanding inputs and adopting a pragmatic, application-focused approach. Delegates also highlighted key aspects of AI such as safety, oversight, privacy, fairness, security, transparency, and explainability, urging the establishment of standards and risk acceptance levels.

Jurisdictional challenges and innovation

Delegates from the Global South raised particular concerns about jurisdictional misalignment, especially in data-driven AI development. Licensing, delegates argued, might hinder innovation and create biases in AI models. It was also noted, as a special case of such bias, that only 2% of LLM inputs are established in Spanish. Delegates from the global south challenged the more technical delegates in this regard. Their response was to indicate the trend towards smaller AI applications to embrace a low volume but high-quality data environment in many global south countries.

Regulation against harm and responsible AI

The Assembly articulated that a process driven by standards is crucial for gaining trust, particularly in the patenting sphere. This further suggests that regulating the human element is key here, where the logical trajectory is that human transactional inputs will likely decrease over time.

Bridging the gap with public and policy makers

The Assembly agreed the importance of researchers bridging the gap between AI developments and public policy, suggesting that regulators themselves will need AI training for effective policy-making. This was seen by the panel as something where the UAE had shown clear leadership.

Cybersecurity as a foundational element

Delegates highlighted the foundational role of cybersecurity in the AI governance discussion, underscoring its importance in ensuring the safety and integrity of AI systems. A delegate also emphasized the issue with languages that are little used and the inherent bias in current LLMs that results from this. Another delegate further illustrated the language and culture weaknesses of current LLMs by informing the panel that one million people speak Estonian, and as such this issue is not limited to the Global South, but is a worldwide issue.

Addressing bad actors and ensuring digital trust

Delegates raised concerns about bad actors outpacing good actors in AI development. It was suggested that creating a parallel "digital trust" industry to verify models and quantify risks. The point was that heavy-handed regulation simply will not work due to the complexities and consistents previously discussed. It will also create "regulatory arbitrage" in which individuals will locate their work in regions with the least regulation, whilst at the same time empower bad actors who have little or no concern about regulations or ethical considerations.

Ensuring enforcement of regulation

One AI startup delegate highlighted the challenges of enforcing regulations and the need for regulations on engineers. The ethics of job displacement were discussed, with suggested mitigation mechanisms such as whistleblower protection. These issues are further considered in the following section.

Sovereign AI and Decentralized Cloud Services

The Assembly discussed the concept of sovereign AI and the importance of independent, decentralized cloud services in fostering responsible AI development. Concern was expressed that the major providers are focused on a small number of countries, meaning that regulation by those countries could be detrimental to the wider world. This could create a two-tier "have and have not" world of AI.

Information asymmetry and regulatory bodies

Delegates pointed out the issue of information asymmetry and stressed the importance of investing in regulatory bodies. As an example, there was concensus for the need to train and educate judges for effective execution of legislation. In a similar vein, the matter of copyright in the age of AI was discussed.

Human and systems-level challenges

10

Global internet accessibility and AI impact

Delegates highlighted the global inequalities of internet inaccessibility and the growing impact of AI on a daily basis. The impact of this may further marginalize some groups, as inequitable access to technology may limit the capacity for the digital preservation of some cultures. Delegates discussed the need to prioritize the digitalization of cultural heritage, so that AI could make use of it, ensuring more inclusive generation.

However, the divide in financial resources in different countries was acknowledged as adding to the challenge of ensuring equitable representation. The discussion touched on the importance of diverse voices in AI development to reduce bias. Affirmative action on data was suggested to enhance inclusivity, with a focus on digitized cultural awareness and heritage. To exemplify the significance of this scenario, delegates discussed examples of biases based on divergent inputs and training data. e.g., pediatric models trained on data of adult men but then used in scenarios involving children.

Acknowledging and incorporating differing perspectives

At a more local level, delegates noted that it is essential for the leaders of organizations to engage with a representative quorum of employees because the perspective obtained from a top-down approach is markedly different than that from the bottom-up, allowing for a more holistic view of the situations where AI will be used and deployed. This should include entrepreneurs and other key stakeholders to ensure a comprehensive understanding.

Necessary breakthroughs

Delegates noted the importance of two emerging breakthroughs that may address some of the technological concerns: first, the acquisition of small but high-quality datasets, followed by the strategic feeding of this curated data into models.

Human interaction with the digital world

The Assembly discussed a future where we will not interact with the digital world through search engines, but rather through AI-powered assistants. This change in our "digital die" will be significant. Judges and regulators will also need to be educated to make knowledgeable decisions on AI. However, it is not only judges and regulators that will need to be brought up to speed on AI matters. The Assembly agreed that development of critical skills was particularly key across a multitude of scenarios in AI (e.g., in order not to fall victim to potential hallucinations of generative AI) as vital in this context. The nature of our changing rela-

tionship with the digital world will change: there might be a need for some standard of digital detox in place and to study how a regular break from the technology could be an essential human resource.

Educating users of AI

The Assembly discussed the idea that there is more to AI than generative AI. Although Gen AI has been garnering attention in recent times. The delegates noted that other technologies, such as graph theory or variant of a neural networks, might be more appropriate than generative AI. In other words, generative AI is not the only AI technology and ought not be used solely for the sake of using it. Other more traditional methods, such as trees, are still useful and may even represent the better fit for certain scenarios. Education should therefore extend to letting people know what AI methods (or other kinds of algorithms) are best used in certain situations, rather than simply assuming that one option fits all. It is up to business owners and other users of this technology to learn and then apply appropriate technology.

AI-induced job loss

Addressing the potential uncertainty around unemployment was discussed at length, along with the recognition that AI is different from past technological upheavals, and now white-and blue-collar jobs are going to be impacted. However the further context is that of global population. We may have a peak in 15-20 years, with a resulting shortage of labor; hence, AI could be a positive force in this context. The Assembly's consensus was that the better AI becomes, the more we need to understand the consequences, and regulators must model and consider AI in this context.

Computer power

Delegates agreed that we do not need 100 different foundation models for AI, we need only a few. An example was provided that pertained to the regional languages of India. Rather than having separate models, a foundational framework could be fine-tuned for each individual language. From this, we can infer that not having open-source foundational models will lead to a number of closed-source variations, which will increase the need for ever greater societal cost associated with computing power.

Conclusion 11

The Assembly's discussions highlighted the complexity of responsible AI development, requiring a delicate balance between innovation and regulation. The need for a global ethical code, transparency, explainability, and inclusivity emerged as key pillars for building trust and ensuring the responsible deployment of AI technologies.

There was a clear theme relating to learning from the regulation of previous technologies by recognizing that ethical considerations are often driven by technology and the nascent nature of current AI technology. The focus should be on developing universal ethical principles at the global level but using existing regulatory frameworks to provide the necessary AI "guardrails" without legislating for low or non-existing risk. It was clearly recognized that the anthropomorphic characteristics of AI have raised issues amongst the general population, and generated responses from legislators who typically do not have appropriate insight into the new technology. Legislating to respond to the anthropomorphic emulation concerns will inhibit AI research from breakthroughs in non-emulation methodologies and we should focus on enabling research that will take us to the next generation of AI architectural platforms. In the interim, we should educate our governments and global population in AI; we should open-source AI wherever and whenever possible; we should use existing regulatory frameworks to provide the necessary guardrails, and focus on expanding the global dataset to encompass all the world's languages and cultures. We should fundamentally embrace a human-led and global AI future. The insights shared during this Round Table Assembly provide a foundation for ongoing discussions and collaborative efforts towards a future of responsible AI.

UNITED ARAB EMIRATES
MINISTER OF STATE FOR ARTIFICIAL INTELLIGENCE,
DIGITAL ECONOMY & REMOTE WORK APPLICATIONS OFFICE

Saqr Binghalib
Executive Director

Hasher Bin Dalmook
Head of Policy

David Toman
Senior Advisor

جامعة خليفة
Khalifa University

Dr Mohammad Alsharid
Advisor

UNIVERSITY OF BIRMINGHAM

Dr Anthony Murphy
Advisor