



Security Council

Seventy-eighth year

9381st meeting

Tuesday, 18 July 2023, 10 a.m.

New York

Provisional

<i>President:</i>	Mr. Cleverly/Dame Barbara Woodward.	(United Kingdom of Great Britain and Northern Ireland)
<i>Members:</i>	Albania	Mr. Hoxha
	Brazil	Mr. França Danese
	China	Mr. Zhang Jun
	Ecuador	Mr. Pérez Loose
	France	Mr. De Rivière
	Gabon	Mrs. Ngyema Ndong
	Ghana	Mr. Agyeman
	Japan	Mr. Takei
	Malta	Mrs. Frazier
	Mozambique	Mr. Gonçalves
	Russian Federation.	Mr. Polyanskiy
	Switzerland	Mrs. Baeriswyl
	United Arab Emirates	Mr. Sharaf
	United States of America.	Mr. DeLaurentis

Agenda

Maintenance of international peace and security

Artificial intelligence: opportunities and risks for international peace and security

This record contains the text of speeches delivered in English and of the translation of speeches delivered in other languages. The final text will be printed in the *Official Records of the Security Council*. *Corrections* should be submitted to the original languages only. They should be incorporated in a copy of the record and sent under the signature of a member of the delegation concerned to the Chief of the Verbatim Reporting Service, room AB-0601 (verbatimrecords@un.org). Corrected records will be reissued electronically on the Official Document System of the United Nations (<http://documents.un.org>).



The meeting was called to order at 10.05 a.m.

Adoption of the agenda

The agenda was adopted.

Maintenance of international peace and security

Artificial intelligence: opportunities and risks for international peace and security

The President: I warmly welcome the Secretary-General and the high-level representatives. Their presence today underscores the importance of the subject matter under discussion.

In accordance with rule 39 of the Council's provisional rules of procedure, I invite the following briefers to participate in this meeting: Mr. Jack Clark, co-Founder of Anthropic; and Mr. Yi Zeng, Institute of Automation, Chinese Academy of Sciences.

The President: The Security Council will now begin its consideration of the item on its agenda.

I wish to draw the attention of Council members to document S/2023/528, which contains the text of a letter dated 14 July 2023 from the Permanent Representative of the United Kingdom of Great Britain and Northern Ireland to the United Nations addressed to the Secretary-General, transmitting a concept note on the item under consideration.

I now give the floor to His Excellency Secretary-General António Guterres.

The Secretary-General: I thank the United Kingdom for convening the first debate on artificial intelligence (AI) ever held in the Security Council.

I have been following the development of AI for some time. Indeed, I told the General Assembly six years ago that AI would have a dramatic impact on sustainable development, the world of work and the social fabric. But like everyone here, I have been shocked and impressed by the newest form of AI — generative AI — which is a radical advance in its capabilities. The speed and reach of this new technology in all its forms are utterly unprecedented. It has been compared to the introduction of the printing press. But while it took more than 50 years for printed books to become widely available across Europe, ChatGPT reached 100 million users in just two months. The finance industry estimates that AI could contribute between \$10 and \$15 trillion dollars to the global economy

by 2030. Almost every Government, large company and organization in the world is working on an AI strategy. But even its own designers have no idea where their stunning technological breakthrough may lead.

It is clear that AI will have an impact on every area of our lives, including the three pillars of the United Nations. It has the potential to turbocharge global development, from monitoring the climate crisis to breakthroughs in medical research. It offers new potential to realize human rights, particularly on health and education. But the High Commissioner for Human Rights has expressed alarm over evidence that AI can amplify bias, reinforce discrimination and enable new levels of authoritarian surveillance.

Today's debate is an opportunity to consider the impact of artificial intelligence on peace and security, where it is already raising political, legal, ethical and humanitarian concerns. I urge the Council to approach this technology with a sense of urgency, a global lens and a learner's mindset — because what we have seen is just the beginning. Never again will technological innovation move as slowly as today.

AI is being put to work in connection with peace and security, including by the United Nations. It is increasingly being used to identify patterns of violence, to monitor ceasefires and more, helping to strengthen our peacekeeping, mediation and humanitarian efforts. But AI tools can also be used by those with malicious intent. AI models can help people to harm themselves and each other, at massive scale.

Let us be clear: the malicious use of AI systems for terrorist, criminal or State purposes could cause horrific levels of death and destruction, widespread trauma and deep psychological damage on an unimaginable scale. AI-enabled cyberattacks are already targeting critical infrastructure and our own peacekeeping and humanitarian operations, causing great human suffering. The technical and financial barriers to access are low, including for criminals and terrorists.

Both military and non-military applications of AI could have very serious consequences for global peace and security. The advent of generative AI could be a defining moment for disinformation and hate speech, undermining truth, facts and safety, adding a new dimension to the manipulation of human behaviour and contributing to polarization and instability on a vast scale. Deepfakes are just one new AI-enabled tool that, if unchecked, could have serious implications for

peace and stability. And the unforeseen consequences of some AI-enabled systems could create security risks by accident.

Look no further than social media. Tools and platforms that were designed to enhance human connection are now used to undermine elections, spread conspiracy theories and incite hatred and violence. Malfunctioning AI systems are another huge area for concern. And the interaction between AI and nuclear weapons, biotechnology, neurotechnology and robotics is deeply alarming.

Generative AI has enormous potential for good and evil at scale. Its creators themselves have warned that much bigger, potentially catastrophic and existential risks lie ahead. Without action to address those risks, we are derelict in our responsibilities to present and future generations.

(spoke in French)

The international community has a long history of responding to new technologies with the potential to disrupt our societies and economies. We have come together at the United Nations to set new international rules, sign new treaties and establish new global agencies. While many countries have called for different measures and initiatives around the governance of AI, that requires a universal approach. And questions of governance will be complex for several reasons.

First, powerful AI models are already widely available to the general public. Secondly, unlike nuclear material and chemical and biological agents, AI tools can be moved around the world while leaving very little trace. And thirdly, the private sector's leading role in AI has few parallels in other strategic technologies.

But we already have entry points. One is the 2018—2019 guiding principles on lethal autonomous weapons systems, agreed through the Convention on Certain Conventional Weapons. I agree with the large number of experts who have recommended the prohibition of lethal autonomous weapons without human control. A second is the 2021 recommendations on the ethics of artificial intelligence agreed through UNESCO. The Office of Counter-Terrorism, working with the Interregional Crime and Justice Research Institute, has provided recommendations on how Member States can tackle the potential use of AI for terrorist purposes. And the AI for Good summits, hosted by the International Telecommunications Union, have brought together

experts, the private sector, United Nations agencies and Governments around efforts to ensure that AI serves the common good.

(spoke in English)

The best approach would address existing challenges while also creating the capacity to monitor and respond to future risks. It should be flexible and adaptable and consider technical, social and legal questions. It should integrate the private sector, civil society, independent scientists and all those driving AI innovation. The need for global standards and approaches makes the United Nations the ideal place for that to happen. The Charter's emphasis on protecting succeeding generations gives us a clear mandate to bring all stakeholders together around the collective mitigation of long-term global risks. AI poses just such a risk.

I therefore welcome calls from some Member States for the creation of a new United Nations entity to support collective efforts to govern this extraordinary technology, inspired by such models as the International Atomic Energy Agency, the International Civil Aviation Organization and the Intergovernmental Panel on Climate Change. The overarching goal of that body would be to support countries to maximize the benefits of AI for good, to mitigate existing and potential risks and to establish and administer internationally agreed mechanisms of monitoring and governance.

Let us be honest: there is a huge skills gap around AI in Governments and other administrative and security structures that must be addressed at the national and global levels. A new United Nations entity would gather expertise and put it at the disposal of the international community. And it could support collaboration on the research and development of AI tools to accelerate sustainable development. As a first step, I am convening a multistakeholder high-level advisory board for artificial intelligence that will report back on the options for global AI governance by the end of this year. My upcoming policy brief on A New Agenda for Peace will also make recommendations on AI governance to Member States.

First, it will recommend that Member States develop national strategies on the responsible design, development and use of AI, consistent with their obligations under international humanitarian law and human rights law.

Secondly, it will call on Member States to engage in a multilateral process to develop norms, rules and principles around military applications of AI, while ensuring the engagement of other relevant stakeholders.

Thirdly, it will call on Member States to agree on a global framework to regulate and strengthen oversight mechanisms for the use of data-driven technology, including artificial intelligence, for counter-terrorism purposes.

The policy brief on a New Agenda for Peace will also call for negotiations to be concluded by 2026 on a legally binding instrument to prohibit lethal autonomous weapons systems that function without human control or oversight, and which cannot be used in compliance with international humanitarian law.

I hope Member States will debate those options and decide on the best course of action to establish the AI governance mechanisms that are so urgently needed.

In addition to the recommendations of the New Agenda for Peace, I urge agreement on the general principle that human agency and control are essential for nuclear weapons and should never be withdrawn. The Summit of the Future, to be held next year, will be an ideal opportunity for decisions on many of these inter-related issues.

I urge the Council to exercise leadership with regard to artificial intelligence and show the way towards common measures we can take to ensure the transparency, accountability and oversight of AI systems. We must work together for AI that can bridge social, digital and economic divides rather than push us further apart. I urge Council members to join forces and build trust for peace and security. We need a race to develop AI for good, AI that is reliable and safe, that can end poverty, banish hunger, cure cancer, supercharge climate action and propel us towards achieving the Sustainable Development Goals. That is the race we need, and it is a race that is possible and achievable.

The President: I thank the Secretary-General for his briefing.

I now give the floor to Mr. Clark.

Mr. Clark: I am here today to offer a brief overview of why artificial intelligence (AI) has become a subject of concern for the world's nations, what the next few years hold for the development of the technology and some ideas on how policymakers may choose to respond

to a historic opportunity. The main takeaway from my remarks should be that we cannot leave the development of artificial intelligence solely to private-sector actors. The Governments of the world must come together, develop State capacity and make the development of powerful AI systems a shared endeavour across all parts of society, rather than one dictated solely by a small number of firms competing with one another in the marketplace.

Why am I making this statement? It helps to have a sense of recent history. A decade ago, a company in England called DeepMind published research that showed how to teach an AI system to play old computer games like Space Invaders and Pong. Fast-forward to 2023, and the same techniques that were used in that research are now being used to create AI systems that can beat military pilots in air-fighting simulations, stabilize the plasma in fusion reactors and even design the components of next-generation semiconductors. Similar trends have played out in computer vision. A decade ago, scientists were able to create basic image classifiers and generate very crude, pixelated images. Today image classification is used across the world to inspect goods on production lines, analyse satellite imagery and improve State security. And AI models that are drawing attention today, such as OpenAI's ChatGPT, Google's Bard and my own company Anthropic's Claude, are themselves also being developed by corporate interests. A lot has therefore happened in 10 years, and we can expect new and even more powerful systems in the coming years. We can expect those trends to continue. Across the world, it is private-sector actors that have the sophisticated computers and large pools of data and capital resources to build such systems, and it therefore seems likely that private-sector actors will continue to define their development. While that will bring huge benefits to humans across the world, it also poses potential threats to peace, security and global stability.

Those threats stem from two essential qualities of AI systems. The first is their potential for misuse, and the second is their unpredictability, as well as the fragility inherent in the fact that they are being developed by such a narrow set of actors. With regard to misuse, AI systems have an increasingly broad set of capabilities, and some beneficial capabilities sit alongside others that can pose threats of profound misuse. For example, an AI system that can help us in understanding the science of biology may also be one that

can be used to create biological weapons. With regard to unpredictability, a fundamental sense surrounding AI is that we do not understand its systems. It is as though we are building engines without understanding the science of combustion. That means that once AI systems are developed and deployed, people identify new uses for them that were not anticipated by their developers. Many of them will be positive, but some, as I have mentioned, could be misuses. And the problem of chaotic or unpredictable behaviour is even more challenging. Once deployed, an AI system may exhibit subtle problems that were not identified during its development. We should therefore think very carefully about how to ensure that developers of such systems can be held accountable so that they build and deploy safe and reliable systems that do not compromise global security.

In order to dramatize the issue, it may be helpful to use an analogy. I want to challenge everyone listening to this briefing not to think of AI as a specific technology but rather as a type of human labour that can be bought and sold at the speed of a computer and that is getting cheaper and more capable over time. As I have described, it is a form of labour that is being developed by one narrow class of actors — companies. We should be clear-eyed about the immense political leverage that affords. If you can create a substitute or augmentation for human labour and sell it into the world, you will become more influential over time. Many of the challenges of AI policy seem simpler to think about if we look at them this way. How should the nations of the world react to the fact that anyone who has enough money and data can now easily create an artificial expert in a given domain? Who should have access to that power? How should Governments regulate that power? Who should be the actors able to create and sell those so-called experts? And what kinds of experts can we allow to be created? Those are huge questions.

Based on my experience, I think that a useful thing that we can do is to work on developing ways to test for capabilities, misuses and potential safety flaws in these systems. If we are creating and distributing new types of workers that will go into the global economy, it stands to reason that we would like to be able to characterize them, evaluate their capabilities and understand their failings. After all, humans go through rigorous evaluation and on-the-job testing for many

critical roles, ranging from the emergency services to the military. Why not the same for AI?

For that reason, it has been encouraging to see many countries emphasize the importance of safety testing and evaluation in their various AI policy proposals, ranging from the European Union's AI framework to China's recently announced generative AI rules, and from the Risk Management Framework for AI systems of the National Institute of Standards and Technologies of the United States to the upcoming AI Summit London on AI and AI safety. All of these different AI policy proposals and events rely in some form on testing and evaluating AI systems so that the Governments of the world can invest in the area. Right now, there are no standards or even best practices for how to test frontier systems for issues such as discrimination, misuse and safety. And because there are no best practices, it is hard for Governments to create policies that can create more accountability for the actors developing the systems. Correspondingly, it means that the private-sector actors enjoy an information advantage when dealing with Governments.

In conclusion, any sensible approach to regulation will start with having the ability to evaluate an AI system for a given capability or flaw. Any failed approach will start with grand policy ideas that are not supported by effective measurements and evaluations. It will be through the development of robust and reliable evaluation systems that Governments will be able to hold companies accountable and companies will be able to earn the trust of the world into which they want to deploy their AI systems. If we do not invest in that, we run the risk of regulatory capture compromising global security and handing over the future to a narrow set of private-sector actors. If we can rise to that challenge, however, I believe we can reap the benefits of AI as a global community and ensure that there is a balance of power between the developers of AI and the citizens of the world.

The President: I thank Mr. Clark for his briefing.

I now give the floor to Mr. Yi Zeng.

Mr. Yi Zeng: My name is Yi Zeng. I would like to take this opportunity to share with the Security Council my personal view of artificial intelligence (AI) as a force for good in terms of international peace and security. I hope it will be useful in promoting the discussion and understanding of the necessity of global governance for AI.

There can be no doubt that AI is a powerful enabling technology for advancing global sustainable development. When investigating the use of AI for the Sustainable Development Goals (SDGs), we have found that most of the efforts focus on AI for quality education and for health care, while very few efforts have been put forward on many other important topics such as AI for biodiversity, climate action or peace. However, I think those are topics of fundamental importance for the future of humankind, and Governments should definitely work together to address them.

Military AI and AI for peace and security, while closely related, are fundamentally different in important aspects. We should advance the use of AI for international peace as an essential pillar of the Sustainable Development Goals, with a view to reducing security and safety risks rather than increasing them. When we are thinking about AI for good from a perspective of peace and security, rather than seeking ways to create disinformation for military and political purposes, it would be much better to work on using AI to identify disinformation and misunderstandings among different countries and political bodies, and to use it not to attack networks but to defend them.

AI should be used to connect people and cultures rather than disconnecting them. That is why we created an AI-enabled culture-interaction engine that finds commonalities and diversities among various UNESCO World Heritage Sites. Those World Heritage Sites have shown us that we are not so divided in terms of culture as we may think, and the commonalities serve as roots and hands that reach out, helping us to appreciate, understand and even learn from the diversities in various cultures.

The current examples of AI, including, more recently, generative AI, are all information-processing tools that seem intelligent but have no real understanding and are therefore not truly intelligent. That is of course why they cannot be trusted as responsible agents to help humans make decisions. For example, although the world has yet to reach a sufficiently broad consensus on lethal autonomous weapon systems, it is at least clear that AI should not be used directly to make life-and-death decisions for humans. Effective and responsible human control must be applied in order to ensure appropriate human-AI interactions. Nor should AI be used to automate diplomatic tasks, especially foreign negotiations between countries, since it is possible

that it could exploit and amplify human limitations and weaknesses such as deception and distrust to create bigger or even catastrophic risks for humans. It is strange, misleading and even irresponsible that dialogue systems powered by generative AI always use expressions such as “I think” and “I suggest” in their arguments when there is no “I” or even “me” in AI models. I therefore want to emphasize again that AI should never, ever pretend to be human, take the place of a human or mislead humans into having wrong perceptions. We should use generative AI to assist us but never trust it to replace human decision-making.

We must ensure human control of all AI-enabled weapon systems, and that human control has to be adequate, effective and responsible. For example, cognitive overload during human-AI interactions must be avoided. We must prevent the proliferation of AI-enabled weapon systems, since the related technology is very likely to be abused or used maliciously. Both near-term and long-term AI carry a risk of causing human extinction, simply because we have not yet found a way to protect ourselves from the possibility that AI can exploit human weaknesses — and if AI were to do that, it would not even know what we mean by humans, death and life. As for the long term, we have not given super-intelligence any practical reasons why it should protect humankind, and the solution to that may take decades to find. Our preliminary research suggests that we may need to change the way we interact with one another and with other species, our ecology and the environment. That may require decision-making across the entire human species, and we will all need to work together deeply and thought-provokingly in that regard.

Considering those near-term and long-term challenges, while I am quite sure that we cannot solve the issue of AI for peace and security today, this discussion, though challenging, may nonetheless be a good starting point for Member States. In that regard, I would suggest that the Security Council consider the possibility of establishing a working group on AI for peace and security, working on near-term and long-term challenges, because at the expert level it would be more flexible and scientific to work together, and easier to reach a consensus from a scientific and technical point of view and to provide assistance and support to Council members in their decision-making. The Council should set a good example for other countries and play a significant role in this important issue.

AI is supposed to help humans solve problems, not create them. A boy once asked me if an AI-assisted nuclear bomb — apart from being a feature in science fiction — could be used to save our lives by blowing up an asteroid attacking Earth or altering its trajectory to prevent a collision with Earth. While that idea may not be scientifically solid and would be very risky at this point, it is at least about using AI to solve a problem for humankind, which is much better than empowering AI to help us attack one another with nuclear weapons on our planet, creating problems for human society and potentially posing catastrophic risks for us, the next generation or even human civilization as a whole. In my view, humans should always maintain and be responsible for final decision-making on the use of nuclear weapons. And we have already affirmed that a nuclear war cannot be won and must never be fought. Many countries — including but not limited to the five permanent members of the Council — have announced their own strategy and opinions regarding the use of AI for security and governance in general, and we can see there are commonalities that can serve as important inputs for international consensus, but that is still not enough. The United Nations must play a central role in setting up a framework on AI for development and governance in order to ensure global peace and security. In the context of a shared future for all, we all need to set up that agenda and framework together, leaving no one behind.

The President: I thank Mr. Yi Zeng for his briefing.

I shall now make a statement in my capacity as the representative of the United Kingdom.

This is the first-ever discussion of artificial intelligence (AI) in the Security Council, and a historic meeting. Since the early development of artificial intelligence by pioneers such as Alan Turing and Christopher Strachey, the technology has advanced with ever greater speed. Yet the biggest AI-induced transformations are still to come. While their scale is impossible for us to comprehend fully, the gains for humankind will surely be immense. AI will fundamentally alter every aspect of human life. Groundbreaking discoveries in medicine may be just around the corner. The productivity boosts to our economies may be vast. AI could help us adapt to climate change, beat corruption, revolutionize education, deliver the Sustainable Development Goals and reduce violent conflict.

But we are here today because AI will affect the work of the Council. It could enhance or disrupt global strategic stability. It challenges our fundamental assumptions about defence and deterrence. It poses moral questions about accountability for lethal decisions on the battlefield. There can already be no doubt that AI is changing the speed, scale and spread of disinformation, with hugely harmful consequences for democracy and stability.

AI could aid the reckless quest for weapons of mass destruction by State and non-State actors alike. But it could also help us stop proliferation. That is why we urgently need to shape the global governance of transformative technologies — because AI knows no borders.

The United Kingdom's vision is founded on four irreducible principles. It should be open — AI should support freedom and democracy. It should be responsible — AI should be consistent with the rule of law and human rights. It should be secure — AI should be safe and predictable by design, safeguarding property rights, privacy and national security. It should be resilient — AI should be trusted by the public, and critical systems must be protected.

The United Kingdom's approach builds on existing multilateral initiatives, such as the AI for Good Global Summit, in Geneva, or the work of UNESCO, the Organization for Economic Cooperation and Development and the Group of 20. Such institutions as the Global Partnership on Artificial Intelligence, the Group of Seven's Hiroshima AI Process, the Council of Europe, and the International Telecommunication Union are all important partners. Pioneering AI companies will also need to work with us so we can capture the gains and minimize the risks to humanity. No country will be untouched by AI, so we must involve and engage the widest coalition of international actors from all sectors.

The United Kingdom is home to many of the world's trail-blazing AI developers and foremost AI safety researchers. That is why, this autumn, the United Kingdom plans to bring world leaders together for the first major global summit on AI safety. Our shared goal will be to consider the risks of AI and decide how they can be reduced through coordinated action.

Momentous opportunities on a scale that we can barely imagine lie before us. We must seize these opportunities and grasp the challenges of AI — including

those for international peace and security — decisively, optimistically and from a position of global unity on essential principles.

“There is a tide in the affairs of men, which, taken at the flood, leads to fortune.” In that spirit, let us work together to ensure peace and security as we pass across the threshold of an unfamiliar world.

I now resume my functions as President of the Council.

I now call on the State Minister for Foreign Affairs of Japan.

Mr. Takei (Japan): I commend the United Kingdom’s initiative to take up artificial intelligence (AI) in the Security Council. This will be a good beginning for future global discussions. I also thank the Secretary-General and the other briefers.

AI is changing the world. AI has transformed human life. Its speed, potential and risks go beyond our imaginations and national borders. We are being tested at this historical juncture: can we have the self-discipline to control it?

My political belief is “instead of worrying, deal with it”. I believe that the key to taking on the challenge is two-fold: AI must be human-centric and trustworthy. Human beings can and should control AI to enhance human potential, not the other way around. Let me make two points.

First, as to human-centric AI, the development of AI should be consistent with our democratic values and fundamental human rights. AI should not be a tool for rulers; it should be placed under the rule of law. The military use of AI is a case in point. It should be responsible, transparent and based on international law. Accordingly, Japan will continue to contribute to the international rule-making process on lethal autonomous weapons systems in the context of the Convention on Certain Conventional Weapons.

Secondly, as to trustworthy AI, AI can be more trustworthy with a wide range of stakeholders being included in the rule-making process. I believe that this is where the convening power of the United Nations can make a difference and bring together wisdom from around the world. Last month, Japan led discussions at the United Nations on the misuse of AI by terrorists by hosting a side event with the Office of Counter-Terrorism and the United Nations Interregional Crime and Justice

Research Institute. Japan is also proud to have launched the Group of Seven Hiroshima AI Process this year and to contribute to the global discussion on generative AI.

The Security Council and the United Nations can update their existing toolkits through the use of AI. First, we should consider how the active use of AI can enhance the efficiency and transparency of the Council in its decision-making and working methods. We welcome the efforts of the Secretariat to utilize AI for mediation and peacebuilding activities. Moreover, we can make the United Nations work more efficiently and effectively through AI-based early-warning systems for conflicts, sanctions-implementation monitoring and counter-measures against disinformation on peace operations.

Let me conclude by expressing our willingness to actively participate in the discussions on AI at the United Nations and beyond.

The President: I now call on the Deputy Minister for Foreign Affairs and Cooperation of the Republic of Mozambique.

Mr. Gonçalves (Mozambique): Mozambique warmly congratulates the United Kingdom for its excellent presidency and for convening today’s timely and important debate on artificial intelligence (AI), with a focus on the opportunities and risks it presents for international peace and security”.

We wish to convey our gratitude to His Excellency Mr. António Guterres, Secretary-General of the United Nations, for his insightful remarks. We thank the briefers for their thoughtful and pertinent contributions.

Allow me to begin with a candid disclosure: this statement was made solely by humans and not by a generative artificial-intelligence model, such as the well-known ChatGPT. This disclosure is important. It reveals some of the anxieties surrounding the rapid advancements in AI, specifically that we are approaching a point where digital machines can execute tasks that, for the majority of human existence, were exclusively within the realm of human intelligence.

The recent acceleration in both the power and visibility of AI systems, along with the increasing awareness of their capabilities and limitations, has sparked concerns that the technology is advancing at such a rapid pace that it may no longer be safely controllable. That caution is warranted.

While recent advancements in AI present immense opportunities to enhance various fields such as speech-making, medicine and warfare through the democratization of innovation, certain models have also exhibited capabilities that surpass the understanding and control of their creators. That poses risks of various kinds, including the potential for catastrophic outcomes. Indeed, we should take heed of the cautionary tale of the “Sorcerer’s Apprentice.”

As AI engines increasingly and convincingly imitate, and in some cases even surpass, various behaviours associated with human beings, they become an ideal tool for spreading misinformation, scamming individuals, engaging in academic cheating, deceitfully initiating conflicts, recruiting terrorists, sowing division and perpetrating numerous other nefarious activities.

AI models have evolved into self-programming machines capable of automating their learning processes through a continuous loop of self-improvement. That necessitates the establishment of robust governance structures that aim to mitigate the risks of accidents and misuse, while still fostering innovation and harnessing the potential for positive outcomes.

As accurately stated in the concept note for this briefing, artificial intelligence technologies possess the potential to profoundly transform our societies, yielding a multitude of positive effects. AI can contribute to eradicating diseases, combating climate change and accurately predicting natural disasters, making it a valuable ally for the global South.

Similarly, by leveraging the extensive database generated by organizations such as the Social Conflict Analysis Database, regional institutions and the wider United Nations system, which adhere to rigorous standards of quality control, sourcing and Member State involvement, we have the potential to enhance early warning capabilities, customize mediation efforts and strengthen strategic communication in peacekeeping, among various other examples. AI can be a valuable tool in harnessing that vast data for the benefit of those endeavours.

Faced with the opportunities and threats posed by the use of artificial intelligence, the Republic of Mozambique recognizes the importance, as a Member State of the United Nations, of adopting a balanced approach that encompasses the following aspects.

First, in the event that credible evidence emerges indicating that AI poses an existential risk, it is crucial to negotiate an intergovernmental treaty to govern and monitor its use.

Secondly, it is essential to develop relevant regulations and appropriate legislation to safeguard privacy and data security. That entails ensuring that all relevant actors, including Governments and companies providing digital technology, utilize artificial intelligence in an ethical and responsible manner, while respecting the principles outlined in article 12 of the Universal Declaration of Human Rights and article 17 of the International Covenant on Civil and Political Rights.

Thirdly, a global digital pact should be promoted, facilitating the sharing of technological knowledge between advanced countries and those in the early stages of AI development. That collaborative effort between AI specialists, Governments, companies and civil society aims to mitigate the risks of misuse and foster responsible AI practices.

In conclusion, it is important to recognize that the inputs required for AI are not disconnected from the real world. The necessary resources such as data, computing power, electricity, skills and technological infrastructure are not evenly distributed across the globe. By striking a balance between the advantages of AI and the essential safeguards in place, we can ensure that AI does not become a source of conflicts that reinforces inequalities and asymmetries, potentially posing threats to global peace and security. That approach aims to harness the potential of AI while actively mitigating any negative consequences that may arise.

The President: I now give the floor to the Assistant Minister for Foreign Affairs and International Cooperation for Advanced Science and Technology of the United Arab Emirates.

Mr. Sharaf (United Arab Emirates): I will begin by thanking Secretary-General Guterres for his attentive remarks today. I thank you, Mr. President, and the United Kingdom presidency for bringing such a salient topic to the Council for discussion. I would also like to thank our other briefers for their illuminating statements.

How we negotiate the threats and opportunities of artificial intelligence (AI) is fast becoming one of the defining questions of our time. Five years ago, the

United Arab Emirates and Switzerland brought forward a proposal to Secretary-General Guterres to establish a deliberative group to consider that very question. Under the Secretary-General's leadership, the High-Level Panel on Digital Cooperation was created, and it was clear from their deliberations that technologies such as AI could no longer go unchecked.

Computation processing power has followed Moore's Law — doubling every 18 months — since the dawn of the computer age. That is no longer the case. AI development is now outpacing Moore's law and moving at breakneck speed and Governments are unable to keep up. That is the wake up-call we need. It is time to be optimistic realists when it comes to AI, not just for assessing the threats the technology poses to international peace and security, but to harness the opportunities it offers.

To that end, I will make four brief points today.

First, we must establish the rules of the road. There is a brief window of opportunity available now where key stakeholders are willing to unite and consider the guard-rails for this technology. Member States should pick up the mantle from the Secretary-General and establish commonly agreed upon rules to govern AI before it is too late. That should include mechanisms to prevent AI tools from promoting hatred, misinformation and disinformation that can fuel extremism and exacerbate conflict. As with other cybertechnologies, the use of AI should be firmly guided by international law, since international law continues to apply in cyberspace. But we must also recognize that strategies may need to be adopted so that we can effectively apply the conventional principles of international law in the rapidly evolving context of AI development.

Secondly, artificial intelligence should become a tool to promote peacebuilding and the de-escalation of conflicts, not a threat multiplier. AI-driven tools have the potential to more effectively analyse vast amounts of data, trends and patterns. That translates to an increased ability to detect terrorist activity in real time and predicting how the adverse effects of climate change may impact peace and security. It also paves the way for limiting the misattribution of attacks, as well as ensuring responses in conflict settings are proportionate. At the same time, we must be aware of the potential misapplication of this technology in targeting critical infrastructure and fabricating false narratives to fuel tensions and incite violence.

Thirdly, the biases of the real world should not be replicated by AI. Decades of progress on the fight against discrimination, especially gender discrimination towards women and girls, as well as against persons with disabilities, will be undermined if we do not ensure an AI that is inclusive.

The High-level Panel on Digital Cooperation was clear in stating that an inclusive digital economy and society was a priority action for immediate attention. Any opportunity that AI offers can be a true opportunity only if it is based on the principle of equality, both in its design and access.

Fourthly, we must avoid overregulating AI in a way that hinders innovation. The creativity, research and development activities occurring in the context of AI and taking place in emerging nations are critical for the sustainable growth and development of those nations. To maintain that, emerging countries need flexibility and agile regulation. We should nurture a sector that encourages responsible behaviour using smart, effective and efficient regulations and guidelines and avoid overly rigid rules that can hamper the evolution of this technology.

Throughout history, major shifts and leaps forward have often followed moments of major crises. The creation of the United Nations and the Security Council following the Second World War speaks to that very fact. When it comes to AI, let us not wait for the moment of crisis. It is high time to get ahead of the curve and shape an AI arena that is geared towards preserving international peace and security.

Mr. Zhang Jun (China) (*spoke in Chinese*): China welcomes His Excellency in presiding over today's Security Council meeting and thanks Secretary-General Guterres for his briefing. Many of his proposals deserve our study. I would also like to thank Professor Yi Zeng and Mr. Jack Clark for their briefings. Their insights can help us better understand and handle issues related to artificial intelligence (AI).

In recent years, the world has witnessed a rapid development and wide application of AI, with complex effects constantly emerging. On the one hand, the empowering role of AI in areas such as scientific research, health care, autonomous driving and smart decision-making, is becoming increasingly prominent, generating huge technological dividends. On the other hand, the scope of AI application has been constantly expanding, causing increasing concerns in areas such as

data privacy, spreading false information, exacerbating social inequality and disrupting employment structures. In particular, the misuse or abuse of AI by terrorist or extremist forces will pose a significant threat to international peace and security.

At present, as a cutting-edge technology, AI is still in its early stages of development. As a double-edged sword, whether it is good or evil, depends on how humankind utilizes it, regulates it and balances scientific development and security. The international community should uphold the spirit of true multilateralism, engage in extensive dialogue, constantly seek consensus and explore the development of guiding principles for AI governance. We support the central coordinating role of the United Nations in that regard and support the efforts of Secretary-General Guterres to hold in-depth discussions among all parties. We also support the full participation of all countries, especially developing countries, in this cause and appreciate their contributions.

I would now like to make some preliminary observations.

First, we should adhere to the principle of putting ethics first. The potential impacts of AI may exceed human cognitive boundaries. To ensure that this technology always benefits humankind, it is necessary to posit "AI for good" and a people-centred approach as the basic principles to regulate the development of AI and to prevent that technology from running away from us like a wild horse. Based on those two guidelines, efforts should be made to gradually establish and improve ethical norms, laws, regulations and policy systems for AI, while allowing countries to establish AI governance systems that are in line with their own national conditions, based on their own development stages and social and cultural characteristics.

Secondly, we must commit to safety and controllability. There are many uncertainties in the development and application of AI-related technologies, and safety is the bottom line that must be upheld. The international community needs to enhance risk awareness, establish effective risk warning and response mechanisms, ensure that risks beyond human control do not occur and that autonomous machine killing does not occur. We need to strengthen the detection and evaluation of the entire life cycle of AI, ensuring that humankind has the ability to press the pause button at critical moments. Leading technology

enterprises should clarify the responsible parties, establish a sound accountability mechanism and avoid developing or using risky technologies that may have serious negative consequences.

Thirdly, we must adhere to fairness and inclusiveness. The impact of AI on science and technology is worldwide and revolutionary. Equal access and utilization of AI technology products and services by developing countries are crucial to bridging the technological, digital and development divides between the North and the South. The international community should work together to ensure that developing countries equally enjoy the development dividends brought by AI technology and continuously enhance their representation, voice and decision-making power in this field. Certain developed countries, in seeking technological hegemony, make efforts to build their exclusive small clubs, take various actions under different pretexts to maliciously obstruct the technological development of other countries and artificially create technological barriers. China firmly opposes those behaviours.

Fourthly, we should adhere to openness and inclusiveness. The development of science and technology needs to achieve a relative balance between technological progress and safe applications. The best path is to maintain open cooperation, encourage interdisciplinary, inter-industrial, inter-regional and cross-border exchanges and dialogues and oppose various forms of exclusive clubs, decoupling and disconnection. We need to promote coordination and interaction among international organizations, Government departments, research and educational institutions, enterprises and the public in the field of AI development and governance under the United Nations framework and jointly create an open, inclusive, just and non-discriminatory environment for scientific and technological development.

Fifthly, we should be committed to the peaceful utilization of AI. The fundamental purpose of developing AI technologies is to enhance the common well-being of humankind. It is therefore necessary to focus on exploring the potential of AI in promoting sustainable development, promoting cross-disciplinary integration and innovation and better empowering the global development cause. The Security Council could conduct in-depth studies on the application and impact of AI in conflict situations and take measures to enrich the toolkit of the United Nations for peace. AI in the

military field may lead to major changes in the way of warfare and the format of war. All countries should uphold a responsible defence policy, oppose the use of AI to seek military hegemony or to undermine the sovereignty and territorial integrity of other countries and avoid the abuse, unintentional misuse or even malicious use of AI-weapon systems.

Today's discussion of AI highlights the importance, necessity and urgency of building a community of a shared future for humankind. China adheres to the concept of a community of a shared future for humankind and has actively explored the scientific path of AI development and governance in every area. In 2017 the Chinese Government issued a new-generation AI development plan that clearly laid out basic principles regarding issues such as technology, leadership, systemic layout, market leadership, open sources and openness. In recent years, China has worked to continually improve relevant laws and regulations, ethical norms, intellectual property standards and safety monitoring and evaluation measures in order to ensure the healthy and orderly development of AI. China has always maintained a highly responsible attitude in its participation in global cooperation on AI governance. As early as 2021, China hosted an Arria Formula meeting during its presidency of the Security Council on the impact of emerging technologies on international peace and security, bringing emerging technologies such as AI to the Council's attention for the first time. China has submitted two position papers on military applications of AI and its ethical governance in United Nations forums, offering systematic proposals from a perspective of strategic security, military policy, legal ethics, technological security, rule-making and international cooperation.

In February, the Chinese Government released its Global Security Initiative concept paper, which states clearly that China is willing to strengthen its communication and exchanges with the international community on AI security governance, promote the establishment of an international mechanism for universal participation and form a governance framework and standards based on a broad consensus. We stand ready to work with the international community to actively implement the Global Development, Global Security and Global Civilization Initiatives that President Xi Jinping has proposed. In the field of AI, we will continue to prioritize development, maintain common security, promote cross-cultural exchanges

and cooperation and work with other countries to share the benefits of AI, while jointly preventing and responding to risks and challenges.

Mr. DeLaurentis (United States of America): I thank the United Kingdom for convening this discussion, and I also thank the Secretary-General, Mr. Clark and Mr. Yi Zeng for their valuable insights.

Artificial intelligence (AI) offers incredible promise for addressing global challenges such as those related to food security, education and medicine. Automated systems are already helping us to grow food more efficiently, predict the paths of storms and identify diseases in patients. Used appropriately, AI can therefore accelerate our progress towards achieving the Sustainable Development Goals. However, it also has the potential to compound threats and intensify conflicts, including by spreading misinformation and disinformation, amplifying bias and inequality, enhancing malicious cyberoperations and exacerbating human rights abuses. We therefore welcome this discussion with a view to understanding how the Council can find the right balance between maximizing AI's benefits while mitigating its risks.

The Council already has experience in addressing dual-use capabilities and integrating transformative technologies into our efforts to maintain international peace and security. As those experiences have taught us, success comes from working with a range of actors, including Member States, technology companies and civil-society activists, through the Security Council and other United Nations bodies and in both formal and informal settings. The United States is committed to doing just that and has already begun such efforts at home. On 4 May, President Biden met with leading AI companies to underscore the fundamental responsibility we all have to ensure that AI systems are safe and trustworthy. Those efforts build on the work of the United States National Institute of Standards and Technology, which recently released an AI framework providing organizations with a voluntary set of guidelines for managing risks from AI systems. Through the White House's October 2022 blueprint for an AI bill of rights, we are also identifying principles to guide the design, use and deployment of automated systems so that rights, opportunities and access related to critical resources and services are enjoyed equally and are fully protected. We are now working with a broad group of stakeholders to identify and address AI-related human rights risks that threaten to undermine

peace and security. No Member State should use AI to censor, constrain, repress or disempower people.

The military use of AI can and should be ethical and responsible and enhance international security. Earlier this year, the United States released a proposed political declaration on the responsible military use of AI and autonomy, which elaborates principles on how to develop and use AI in the military domain in compliance with applicable international law. The proposed declaration emphasizes that the military use of AI capabilities must be accountable to a human chain of command and that States should take steps to minimize unintended bias and accidents. We encourage all Member States to endorse the proposed declaration.

Here at the United Nations, we welcome efforts to develop and apply AI tools that improve our joint efforts to deliver humanitarian assistance, provide early warnings of issues as diverse as climate change and conflict and further other shared goals. The International Telecommunication Union's recent Artificial Intelligence for Good Global Summit represents one step in that direction. Within the Security Council, we welcome continued discussions on how to manage technological advances, including when and how to take action to address Governments' or non-State actors' misuse of AI technologies to undermine international peace and security. We must also work together to ensure that AI and other emerging technologies are not used primarily as weapons or tools of oppression but rather as tools to enhance human dignity and help us achieve our highest aspirations, including for a more secure and peaceful world. The United States looks forward to working with all relevant parties to ensure that the responsible development and use of trustworthy AI systems serves the global good.

Mr. França Danese (Brazil): I thank the Secretary-General for his briefing today and for joining our meeting. I also thank Mr. Clark and Mr. Yi Zeng for their statements.

The rapid development of artificial intelligence (AI) holds immense potential for bolstering our global security architecture, augmenting decision-making processes and enhancing humanitarian efforts. We must also address the multifaceted potential challenges it poses, including in the areas of autonomous weapons, cyberthreats and the exacerbation of existing inequalities. As we embark on this crucial discussion, let us seek a comprehensive understanding

of the risks and opportunities associated with AI and work to harness its potential for the greater benefit of humankind while ensuring the preservation of peace, stability and human rights.

The paragraph I just read was entirely written through ChatGPT. Although it has conceptual imprecisions, it shows how sophisticated such tools have become. The technology is developing so fast that even our best researchers are still unable to assess the full scale of the challenges that await us and the benefits such new technologies can provide. Any discussion today must start from a basis of humility and an awareness that we do not fully know what it is that we do not know about AI. What we know for sure is that artificial intelligence is not human intelligence. Most AI relies on large amounts of data, and through complex algorithms establishes patterns and relationships that allow it to generate contextually appropriate results. The outcomes are therefore crucially dependent on the inputs. Human oversight is essential to avoiding bias and errors. Otherwise, we are running the risk that the saying "garbage in, garbage out" will become a self-fulfilling prophecy.

Unlike other innovations with potential implications for security, AI has been developed mostly as a civilian application. It would therefore be premature to see it primarily through a lens of international peace and security, as the most significant effects on our societies are likely to be produced by its peaceful uses. We can nevertheless predict with certainty that its applications will be extended to the military arena, with relevant consequences for peace and security.

While the Council should remain vigilant and ready to respond to any incidents involving the use of AI, we should also be careful not to overly securitize the topic by concentrating discussions in this Chamber. Due to the intrinsically multidisciplinary nature of AI, which deals with every aspect of life, our international discussions of it should remain open and inclusive. Only a wide and diverse range of views will enable us to scratch the surface and start to make sense of the different facets of AI. Today's briefing is a good start for bringing in various views on its development and use.

Nevertheless, in the light of AI's wide-ranging implications and impacts, the General Assembly, with its universal composition, is the forum best suited to a structured, long-term discussion on artificial

intelligence. AI is a crucial topic among the various subjects that come under the mandate of the ongoing Open-ended Working Group on Security of and in the Use of Information and Communications Technologies 2021–2025, which will hold its fifth substantive session next week. The Working Group, which is open to all Member States, has been able to make progress on gradually developing global, common understandings on information and communications technologies issues related to international peace and security, despite challenging geopolitical circumstances. Considering their particular nature, that is what we should aim for when discussing challenges arising from cybertechnologies.

The military applications of AI, especially with regard to the use of force, should strictly abide by international humanitarian law, as enshrined in the Geneva Conventions and other pertinent international commitments. Brazil has been consistently guided by the concept of meaningful human control. As approved in 2019 by the High Contracting Parties to the Convention on Certain Conventional Weapons, guiding principle (b) indicates that

“[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines” (CCW/MSP/2019/9, annex III).

The centrality of the human element in any autonomous system is essential to the establishment of ethics standards and to full compliance with international humanitarian law. There can be no substitute for human judgment and accountability.

Military applications of AI must be based on transparency and accountability throughout their life cycle, from development to deployment and use. Moreover, weapon systems with autonomous functions should eliminate bias in their system's operations. We must move ahead swiftly with the progressive development of regulations and norms governing the use of autonomous weapon systems via robust norms to prevent biases and abuse and to guarantee compliance with international law, particularly international humanitarian law and human rights law. Compliance with international law is mandatory in States' use of AI technologies, as well as in any use the Council may wish to make of them in its peacekeeping missions or its broader mandate for the maintenance of international peace and security.

Beyond the challenges posed by conventional weapons with autonomous functions, we should not shy away from issuing a very stern warning regarding the inherent risks posed by the interaction of AI and weapons of mass destruction. We have noted with alarm the news that AI-assisted computer systems are capable of developing new poisonous chemical compounds in a matter of hours and of designing new pathogens and molecules. Nor should we allow the possibility that nuclear weapons might be linked to AI at the risk of our common future.

AI has tremendous potential to both remake and break our societies in the coming years. Navigating between the two will require a broad and concerted international effort, which will include but is in no way limited to the Security Council. The United Nations remains the only organization capable of promoting the global coordination needed to oversee and shape the development of AI so that it works for the betterment of humankind and in accordance with the shared purposes and principles that have brought us here.

Mrs. Baeriswyl (Switzerland) (*spoke in French*): We are grateful to Secretary-General António Guterres for his participation in this important debate. My thanks also go to Mr. Clark and Mr. Yi Zeng for their valuable and impressive contributions.

(*spoke in English*)

“I believe it is only a matter of time before we see thousands of robots like me out there making a difference”.

(*spoke in French*)

Those words were uttered by the robot Ameca, speaking to a journalist at the Artificial Intelligence for Good Global Summit mentioned by the Secretary-General, which was organized by the International Telecommunication Union and Switzerland and held in Geneva two weeks ago.

While it can represent a challenge because of its speed and apparent omniscience, artificial intelligence (AI) can and must also serve peace and security. As we look ahead to a New Agenda for Peace, we have it in our hands to ensure that AI makes a difference to the benefit, and not the detriment, of humankind. Against that backdrop, let us seize the opportunity to pave the way for AI for good by working closely with cutting-edge research. To that end, the Swiss Federal Institute of Technology in Zurich is developing a prototype AI-

assisted analysis tool for the United Nations Crisis and Operations Centre. That tool will explore the potential of AI for peacekeeping, and particularly for the protection of civilians and peacekeepers. In addition, Switzerland recently launched a call for a trust and transparency initiative in which academia, the private sector and diplomacy can jointly seek practical and rapid solutions to the risks associated with AI.

The Council, too, must work to counter the risks that AI poses to peace. That is why we are very grateful to the United Kingdom for organizing this important debate. Let us consider cyberoperations and disinformation as an example. False narratives undermine people's trust in Governments and peace missions. In that respect, AI is a double-edged sword. While it accentuates disinformation, it can also be used to detect false narratives and hate speech. So how can we reap the benefits of AI for peace and security while minimizing the risks? I would like to propose three pathways.

First, we need a common framework, shared by all the players involved in the development and application of such technology — Governments, businesses, civil society and research organizations — which I think Mr. Clark made very clear in his briefing. AI does not exist in a normative vacuum. Existing international law, including the Charter of the United Nations, international humanitarian law and human rights, apply to it. Switzerland is committed to all United Nations processes that serve to reaffirm and clarify the international legal framework for AI, and in the case of lethal autonomous weapon systems, to develop prohibitions and restrictions.

Secondly, AI must be human-centred, or, as Mr. Zeng put it earlier,

(spoke in English)

“AI should never, ever pretend to be human”.

(spoke in French)

We call for the development, deployment and use of AI to always be guided by ethical and inclusive considerations. Clear responsibility and accountability must be maintained for Governments, companies and individuals alike.

Finally, the relatively early stage of AI development gives us the opportunity to ensure equality and inclusion, and to counteract discriminatory stereotypes.

AI is only as good and reliable as the data with which we provide it. If that data reflects prejudices and stereotypes — for example, of gender — or if it is simply not representative of its operating environment, AI will offer us bad advice in maintaining peace and security. It is the responsibility of developers and users, both governmental and non-governmental, to ensure that AI does not reproduce the harmful societal biases we are striving to overcome.

The Security Council has a responsibility to proactively monitor developments around AI and the threat it may pose to the maintenance of international peace and security. It should be guided by the results emanating from the General Assembly on the related legal framework. The Council must also use its powers to ensure that AI serves peace by anticipating risks and opportunities or by encouraging the Secretariat and peace missions to use this technology in innovative and responsible ways.

My delegation used AI for our first debate under our presidency, on futureproofing trust for sustaining peace (see S/PV.9315), as well as in the context of an exhibition in collaboration with the International Committee of the Red Cross on digital dilemmas. We were able to recognize the impressive potential of this technology to serve peace. We therefore look forward to making “AI for Good” an integral part of the New Agenda for Peace.

Mr. Agyeman (Ghana): I thank the United Kingdom for convening this high-level debate on artificial intelligence (AI) and Secretary-General Antonio Guterres for his important briefing to the Security Council this morning. We are equally grateful for the expert views provided in this meeting by Mr. Jack Clark and Mr. Zeng Yi.

The emerging dominance of artificial intelligence as a pervasive fabric of our societies is one that could have positive impacts on several vistas, including beneficial applications for medicine, agriculture, environmental management, research and development, the realm of arts and culture, and trade. While we see opportunities on the horizon to enhance outcomes in different areas of life by embracing the further application of artificial intelligence, we can also already see dangers that must motivate all of us to work quickly and collaboratively to avert risks that could be detrimental to our common humanity.

Artificial intelligence, especially in peace and security, must be guided by a common determination not to replicate the risks that powerful technologies have created for the world by their ability to unleash disaster of global proportions. We must constrain the excesses of individual national ambitions for combative dominance and commit to working towards the development of principles and frameworks that would govern AI technologies for peaceful purposes.

For Ghana, we see opportunities in the development and application of AI technologies for identifying early-warning signs of conflicts and defining responses that have a higher rate of success and that may also be more cost-effective. Such technologies can facilitate the coordination of humanitarian assistance and improve risk assessment. Their application for law enforcement is already well appreciated in many jurisdictions, and where law enforcement has been effective, risks of conflict are usually low.

Moreover, the application of AI technologies for peace mediation and negotiation efforts has revealed remarkable early outcomes that must be pursued for the cause of peace. The deployment of AI technologies in determining the Libyan population's reaction to policies, for instance, has facilitated the peace as reflected in the improvement in that country's ranking in the 2022 Global Peace Index. We also see in similar contexts, and within the intelligence, surveillance and reconnaissance functions of peacekeeping missions, an opportunity to enhance the safety and security of peacekeepers and the protection of civilian populations through the responsible deployment of AI technologies.

Despite these encouraging developments, we see risks with AI technologies from the perspective of both State actors and non-State actors. The integration of AI technologies into autonomous weapons systems is a foremost source of concern. While States seeking to develop such weapons systems may be genuinely interested in reducing the human cost of their involvement in conflicts, it belies their commitment to a pacific world. The history of our experience with humankind's mastery in atomic manipulation shows that should such desires persist, they only generate, in equal measure, efforts by other States to cancel the advantage that such deterrence seeks to create. The additional danger of non-human control over these weapons systems is also a risk that the world cannot afford or ignore.

The increasingly digitalized world and the creation of virtual reality means that the capacity to tell the difference between what is real and what is made up is diminishing by the day. This can create unchecked platforms that non-State actors in particular could, using AI technologies, instrumentalize to destabilize societies or cause friction between or among States. While AI technologies can be used to counter misinformation, disinformation and hate speech, they also have the capacity to be used by negative forces to pursue the campaigns of those forces' nefarious agenda.

However, the potential of AI technologies for good should lead us to work towards its peaceful uses. As indicated earlier, some principles and frameworks need to be developed, keeping in mind that we do not yet have a full sense of the evolution of AI technologies. Still, such a process should not be the preserve of the Security Council but of the wider membership of the United Nations, which has an equal stake in how we guide the further evolution of AI technologies. Without global consensus it would be difficult to limit the flourishing of AI technologies.

Since a significant part of the developments on AI technologies presently occurs within the private sector and academia, it is important to also broaden the dialogue beyond governments to ensure that, in filling industry gaps, there can be no diversion and misuse of AI technologies, including of unarmed aerial vehicles, whose negative consequences for peace and security, including on the African continent, where terrorist groups may be experimenting with such technologies, should be anticipated and abated.

In recognition of the fact that AI technologies can create disruptions in military balance, it is important that States deliberately pursue confidence-building measures that rest on a shared interest in preventing conflicts that are not deliberately intended. This can be done through setting standards for voluntary information-sharing and notifications concerning AI-enabled systems, strategies, policies and programmes implemented by States. We hope that in considering the Secretary-General's upcoming policy brief on the New Agenda for Peace, Member States will be able to advance durable solutions to address new threats to international peace and security. We wish to indicate our support for the Secretary-General's efforts in this regard.

In the process, we must also deepen work on such existing initiatives and ongoing processes as the Secretary-General's Roadmap on Digital Cooperation, the ongoing negotiations on a global convention on countering the use of information and communications technology for criminal purposes, and the Open-ended Working Group on security of and in the use of information and communications technologies. Equally, we encourage the Security Council's further engagement with the Strategy for the Digital Transformation of United Nations Peacekeeping under the Action for Peacekeeping Plus initiative. During the upcoming United Nations Peacekeeping Ministerial Meeting in Accra, Ghana would welcome conversations on how AI can be deployed to enhance peacekeeping operations under the relevant themes. In Africa, the African Union Digital Transformation Strategy (2020–2030) would also continue to be an important ancillary to the African Continental Free Trade Area, which is an anchor for addressing many of the underlying security challenges on the continent and in silencing the guns in Africa.

Finally, I affirm Ghana's commitment to advancing constructive discussions on AI technologies for the peace and security of our world. We highlight the need for a whole-of-society approach that leverages the potential of the private sector, especially the technology giants, and retains the human rights of citizens at the core of all ethical principles.

Mr. De Rivière (France) (*spoke in French*): I would like to thank the Secretary-General, Mr. Jack Clark and Mr. Yi Zeng for their briefings.

Artificial intelligence (AI) is the revolution of the twenty-first century. At a time when we are facing a tougher world, marked by competition and hybrid wars, it is essential to make artificial intelligence a tool in the service of peace.

France is convinced that AI can play a decisive role in maintaining peace. These technologies can contribute to the safety of peacekeepers and the performance of operations, in particular by improving the protection of civilians. They can also help resolve conflicts by mobilizing civil society and perhaps, in future, by facilitating the delivery of humanitarian aid.

Artificial intelligence can also serve the Sustainable Development Goals. That is the thrust of our contribution to the Secretary-General's Global Digital Compact. In the fight against climate change,

AI can help us to prevent natural hazards by providing more accurate weather forecasts. It can also support the implementation of the commitments made in the Paris Agreement to reduce greenhouse gases.

The development of artificial intelligence also entails risks, which we need to face head on. AI is likely to increase the cyberthreat as it enables malicious actors to carry out increasingly sophisticated cyberattacks. Artificial-intelligence systems themselves can be vulnerable to cyberattacks. Ensuring their security is therefore of paramount importance.

In the military sphere, artificial intelligence could profoundly change the nature of conflicts. We must therefore continue our efforts, within the Group of Governmental Experts on lethal autonomous weapons systems under the Convention on Certain Conventional Weapons, to develop a framework applicable to such systems. That framework would need to ensure that tomorrow's conflicts respect international humanitarian law.

Above all, generative artificial intelligence can intensify information warfare through the inexpensive mass propagation of artificial content or messages modulated according to the recipient. We need only look at the massive disinformation campaigns under way in the Central African Republic and Mali or those accompanying Russia's war against Ukraine. Foreign electoral interference campaigns destabilize countries and call into question the foundations of democracies.

France is committed to promoting an ethical and responsible approach to artificial intelligence. That is the aim of the global partnership we launched in 2020. Within the European Union and the Council of Europe, France is working on rules to govern and support the development of AI.

In the face of that revolution, the United Nations offers an irreplaceable framework. We welcome the ongoing work of the New Agenda for Peace and the forthcoming Summit of the Future, which will enable us to reflect collectively on those issues and draw up the standards of tomorrow. France will do its utmost to put artificial intelligence at the service of conflict prevention, peacekeeping and peacebuilding.

Mr. Pérez Loose (Ecuador) (*spoke in Spanish*): "The first obligation of intelligence is to distrust itself," said the Polish writer Stanisław Lem. But that

element is something we cannot expect from artificial intelligence (AI).

I therefore stress the relevance of the theme that brings us together, at the initiative of the United Kingdom, and I am grateful for the briefings made by Secretary-General António Guterres and the other speakers.

The question cannot be whether or not we support the development of artificial intelligence. In the context of rapid technological change, AI has already developed at breakneck speed and will continue to do so.

Like any other technology, AI is a tool that can contribute to peacekeeping and peacebuilding efforts or it can be used to undermine those goals. AI can contribute to conflict prevention and the moderation of dialogue in complex contexts, such as coronavirus disease. Emerging technologies were essential in overcoming the obstacles posed by the pandemic.

Artificial intelligence can support the protection of humanitarian personnel, allowing for expanded access and action, including through predictive analytics. Preparedness, early warning and timely reaction can benefit from this tool. Technological solutions can help United Nations peacekeeping operations to fulfil their mandates more effectively, including by facilitating adaptation to changing conflict dynamics.

On 30 March 2020, Ecuador co-sponsored resolution 2518 (2020) supporting the more integrated use of new technologies with a view to improving the situational awareness of personnel and their capabilities, which was reiterated in the presidential statement of 24 May 2021 (S/PRST/2021/11). That must include artificial intelligence, for its ability to improve camp and convoy security, with conflict monitoring and analysis.

As an Organization, we cannot achieve greater efficiency if we are not equipped with the tools to overcome new security challenges. Our responsibility is to promote and harness technological development as an enabler of peace. That must be done in strict accordance with public international law, international human rights law and international humanitarian law. We cannot ignore the threats posed by the misuse or abuse of AI for malicious or terrorist purposes. An AI system also carries other risks, such as discrimination or mass surveillance.

Furthermore, Ecuador rejects the militarization or placement of AI weapons. We reiterate the risks posed

by lethal autonomous weapon systems and the need for all weapons systems to respond to human decision, control and judgment, under the only viable framework of responsibility and accountability.

Ethical principles and responsible behaviour are indispensable but not sufficient. The answer to harnessing artificial intelligence without exacerbating the threats arising from it is the establishment of a legally binding international framework, as Ecuador will continue to advocate. Moreover, in cases where it is not possible to ensure sufficient human control over lethal autonomous weapons, as well as the principles of distinction, proportionality and precaution, they should be prohibited.

I share the Secretary-General's concern about the alarming potential link between AI and nuclear weapons. We welcome the recommendations made today with regard to the New Agenda for Peace and agree on the need to bridge the digital divide and to promote partnerships and associations that enable us to harness existing technologies for peaceful ends. In addition to ethical considerations, the robotization of conflicts poses a major challenge to disarmament efforts and an existential challenge that the Council cannot neglect.

AI researchers today speak of the problem of alignment, that is, how can we ensure that AI's discoveries will serve us rather than destroy us? That is a question that scientists such as Einstein, Oppenheimer and Von Neumann asked themselves, in turn. That is the challenge that lies ahead of us today.

Mrs. Frazier (Malta): I thank the presidency of the United Kingdom for holding today's briefing on this highly topical issue. I also thank the Secretary-General for enriching our discussion with his thoughts and insights.

Artificial intelligence (AI) is reshaping the way we work, interact and live. Peaceful applications of AI can help achieve the Sustainable Development Goals and support United Nations peacekeeping efforts. Such efforts include the use of drones for humanitarian assistance deliveries, monitoring and surveillance.

On the other hand, the proliferation of AI technologies raises significant risks that demand our attention. The potential misuse or unintended consequences of AI, if not carefully managed, could pose threats to international peace and security. Malicious

actors could exploit AI for cyberattacks, disinformation and misinformation campaigns or autonomous weapon systems, leading to increased vulnerabilities and geopolitical tensions. There can also be negative human rights consequences associated with AI, including through discriminatory algorithmic decision-making. We must address those risks collectively through international cooperation, frameworks and norms.

Malta believes that the cooperation of multiple stakeholders across the various levels and sectors of international, regional and national communities is essential for implementing ethical frameworks regarding AI around the world. In that regard, the international community needs to develop universal instruments that focus not only on the articulation of values and principles but also on their practical realization, with a strong emphasis on the rule of law, human rights, gender equality and the protection of the environment. As governmental and non-governmental actors race to be first in the development of AI, governance and control practices must be developed at a comparable pace for safeguarding international peace and security. Therefore, the Security Council must push for strong AI governance and ensure its inclusive, safe and responsible deployment through the sharing of experiences and governmental frameworks.

Since 2019, Malta has been developing an ethical AI framework, in alignment with the Ethics Guidelines for Trustworthy AI of the European Union (EU). The framework is based on four guiding principles: first, build on a human-centric approach; secondly, respect for all applicable laws and regulations, human rights and democratic values; thirdly, maximize the benefits of AI systems while preventing and minimizing their risks; and fourthly, align with emerging international standards and norms around AI ethics. Malta is ready to work hand in hand on AI with all stakeholders involved to develop a global agreement on common standards for the responsible use of AI. Moreover, within the EU, we are working on an artificial intelligence act, which seeks to ensure that citizens can trust what AI has to offer. It takes a human-centric and innovation-friendly approach to AI, based on fundamental rights and the rule of law.

In that line, Malta strongly supports the work of the open-ended working group on security of and in the use of information and communications technologies 2021–2025 and underlines that confidence-building measures are essential to increase the level of dialogue

and trust for more transparency in the use of AI to ensure better accountability.

Malta expresses its concern regarding the use of AI systems in military operations, since machines cannot make human-like decisions involving the legal principles of distinction, proportionality and precaution. We believe that lethal autonomous weapon systems currently exploiting AI should be banned and only those weapons systems that are in full respect of international humanitarian law and human rights law should be regulated. Likewise, the integration of AI into national security, counter-terrorism and law enforcement systems raises fundamental human rights, transparency and privacy concerns, which must be addressed.

In conclusion, Malta believes that the Security Council has an important anticipatory role to play on this issue. We have the responsibility to monitor developments closely and address any threats to international peace and security that may arise in a timely manner. It is only by promoting responsible governance, international cooperation and ethical considerations that we can harness the transformative power of AI while mitigating potential risks.

Mrs. Ngyema Ndong (Gabon) (*spoke in French*): I would like to thank the United Kingdom for organizing this debate on artificial intelligence (AI) at a time when technological innovations are constantly increasing and revolutionizing our societies and having an impact on international security. I also thank Secretary-General António Guterres, Professor Yi Zeng and Mr. Jack Clark for their briefings.

Artificial intelligence is as fascinating as it is bewildering. In recent years, it has revolutionized our ways of life, modes of production and ways of thinking and has expanded the boundaries of our reality. Thanks to their precision and ability to solve complex problems, AI systems stand out from more conventional information technology mechanisms and offer numerous opportunities for maintaining international peace and security.

For years, the maintenance of international peace and security has relied on a robust technological ecosystem that not only enhances crisis management and prevention capabilities, but also promotes greater understanding of situations on the ground, while improving the protection of civilians, especially in complex environments.

Artificial intelligence is making its own specific contribution by multiplying the analysis capabilities of early warning systems. It is now faster and easier to detect emerging threats by analysing vast quantities of data from a variety of sources in a very short period of time. Thanks to AI, the operational systems of the United Nations peace missions are becoming ever more effective. Indeed, the use of drones, night vision and geolocation systems, for example, makes it possible to detect the activities of armed and terrorist groups, to secure the delivery of humanitarian aid in hard-to-reach areas and to improve ceasefire monitoring and mine-detection missions on the ground. It also strengthens the implementation of very complex peacekeeping mandates, in particular the protection of civilians.

AI also plays a major role in peacebuilding processes, contributing to the reconstruction efforts of States in post-conflict situations, and encourages the implementation of quick-impact projects, while offering employment opportunities to young people and reintegration possibilities for former combatants. However, in order to make the most of the benefits of AI for peace and security, in particular when deploying peacekeeping operations, it is essential that local communities appropriate and absorb those new technologies, so as to perpetuate their beneficial effects after the withdrawal of international forces. If they are not anchored at the local level, the benefits of AI are likely to disappear and crises to re-emerge. States, national and international organizations and local populations must be educated in the manufacturing and distribution processes, in order to strengthen trust in and the legitimacy of the AI systems being used.

AI certainly contributes to strengthening international peace and security, but it also poses many risks which we must understand now. Terrorist and criminal groups can take advantage of the many opportunities offered by AI to pursue their illicit activities. In recent years, hacker networks have stepped up their cyberattacks, disinformation and theft of sensitive data. The threats posed by the malicious use of AI should be a wake-up call for the international community and the starting point for greater control over the development of new technologies. That means strengthening transparency and international governance, with the United Nations as a guarantor, but also and above all, accountability. The United Nations must bolster international cooperation to

develop a regulatory framework with appropriate control mechanisms and robust security systems. Information-sharing and the establishment of ethical standards will also help to prevent abuse and preserve international peace and security.

Gabon remains committed to promoting the peaceful and responsible use of new technologies, including artificial intelligence. With that in mind, it is important to encourage the sharing of best practices in the areas of security and control, encourage States to adopt national regulatory policies and start awareness-raising programmes right now, especially for young people, on the issues and challenges of artificial intelligence.

In conclusion, artificial intelligence clearly offers a whole range of opportunities. It supports sustainable development initiatives and helps to prevent humanitarian and security crises and combat climate change and its negative effects. However, in the absence of reliable regulations and effective control and management tools, AI can pose a genuine threat to international peace and security. Our enthusiasm for those increasingly sophisticated technologies must therefore be tempered with caution and restraint.

Mr. Hoxha (Albania): I thank you, Madam President, for convening today's important meeting and for bringing the issue to the Security Council for a first-ever debate on the matter.

Artificial intelligence (AI) has been around for decades as part of the world's scientific and technological drive. The recent spike in its development has opened vast avenues for its use in almost every sector of human activity, as the Secretary-General, our briefers and other colleagues have said. Everything indicates that it is headed to a place that will be front and centre in an era of revolutionary, never-before-seen technological advances in the years to come. Our world is no stranger to scientific evolution and disruptive technological growth. It is part of the uninterrupted quest by humans for progress, part of our genome and engraved in human history. But there is something fundamentally different about artificial intelligence. It stands out both for its rate of progress and for the potential scope of its applications, and holds great promise for transforming the world as never before and automating processes on a scale we cannot even imagine now.

While that technology advances at a mind-blowing pace, we are caught between fascination and fear, weighing benefits and worries, anticipating applications

that may transform the world but also aware of its other side — the dark side, the potential risks that could have an impact on our safety, privacy, economy and security. Some, more alarmist, have gone as far as to warn about the risks that AI may pose for our civilization. The vertiginous rhythm of the development of a technology with far-reaching consequences that we are not fully able to grasp raises serious questions, and rightly. It is because of the nature of technology, the lack of transparency and accountability in how algorithms reach their results and the fact that often not even the scientists and engineers who design the AI models fully understand how they arrive at the outputs they produce.

Depending on the data sets on which they are trained or the way in which their algorithms are organized, AI models and systems could lead to discrimination on the basis of race, sex, age or disability. While such issues can be prevented or corrected, it is far more difficult to prevent the serious risks presented by those who use the technology with the intention to cause harm. The Internet and social media have already shown how damaging such behaviours can be. Some countries continually attempt to deliberately create misleading information, distort facts, interfere in the democratic processes of others, spread hatred, promote discrimination and incite violence or conflict by misusing digital technologies. Deepfakes and doctored photos are being used to create convincing but false information and narratives, produce convincing conspiracy theories that undermine public trust and democracy and even cause panic. For all those actors, artificial intelligence will provide infinite possibilities for malicious activities.

The misuse of AI could have a direct impact on international peace and security, and it poses grave security challenges for which we are currently ill prepared. AI may be used to perpetuate bias through large-scale disinformation attacks, develop new cyberweapons, power autonomous weapons and design advanced biological weapons.

As we reap the benefits of technological advances, it is urgent that we make use of existing rules and regulations, improve and update them and define the ethics of the use of AI. We must also establish, at the national and international level, the necessary safeguards, governance frameworks and clear lines of responsibility and authority to ensure that AI systems are used appropriately, safely and responsibly for the good of all, and that they do not infringe on human

rights and freedoms or affect peace and security. We must promote standards for responsible State behaviour and the applicability of international law in the use of artificial intelligence and its technologies, but also in the monitoring and assessment of risks and implications. And that provides a role for the Security Council. Albania will continue to promote open, free and safe AI technologies where human rights, fundamental freedoms and the rule of law are respected.

Mr. Polyanskiy (Russian Federation) (*spoke in Russian*): We welcomed the Secretary-General's participation in today's meeting, and listened very closely to his statement. We also thank the briefers for their interesting observations.

The Russian Federation attaches great importance to the development of advanced technology that is designed to serve the common good of humankind and contribute to socioeconomic progress. Artificial intelligence (AI) is a groundbreaking modern technology that has unquestionably inspired considerable scientific interest. However, its practical significance and potential applications, not to mention its hypothetical impact on political processes, have yet to be fully assessed.

The impression left by the assessments we heard from today's briefers is that as a technology, artificial intelligence is in the early stages of its development. The unknowns are still far more numerous than the knowns. The choice of the Security Council Chamber as a platform for today's discussion was therefore not obvious to us. Even in the concept note (S/2023/528) for today's meeting, it is clear that there are no concrete arguments that support the notion of organic interconnections between AI and issues of international peace and security, at least not yet. The concept note only lists a disparate set of technological trends that can be applied to any area one wants.

When I was preparing for today's meeting, I decided to ask the primary source — artificial intelligence itself — if AI is a threat to international peace and security. This was its reply.

“Artificial intelligence can pose certain threats to international peace and security, but it can also have a positive impact. Here are a few points that should be considered.

“The first is cybersecurity. The development of AI may cause new vulnerabilities and threats

in cyberspace. AI can be used to create powerful cyberweapons, automate cyberattacks and circumvent existing defences.

“The second is autonomous weapon systems. The development of AI-based autonomous weapon systems could pose security risks, as such systems could decide to use force without direct human involvement. That could lead to unpredictable consequences and exacerbate armed conflicts.

“The third and last is the spread of disinformation. AI can be used to create and spread disinformation and fake news, which can undermine the credibility of information and contribute to social and political tensions.”

We therefore believe attempts to establish a generic connection between artificial intelligence and threats to peace and security to be somewhat artificial. And as we can see, artificial intelligence agrees with us.

To begin with, the international community should determine the nature and character of possible risks and threats, assess their scale and possible ways to respond and so on. That requires a professional discussion, based on scientific expertise, which is likely to take years. It is already under way on specialized platforms. Various military aspects of AI that may have an impact on global and regional security are also being discussed in specialized formats. The issue of lethal autonomous systems falls under the purview of the Group of Governmental Experts on Lethal Autonomous Weapons Systems and the States parties to the Convention on Certain Conventional Weapons, also known as the Inhumane Weapons Convention. Security issues related to the use of information and communications technologies are comprehensively discussed in the Open-ended Working Group on Security of and in the Use of Information and Communications Technologies 2021–2025, under the auspices of the General Assembly. We believe that duplicating their work would be counterproductive.

As with any form of advanced technology, AI can have beneficial or destructive consequences for humankind, depending on who controls it and the purposes it is used for. Today we are unfortunately seeing how the West, led by the United States, is undermining trust in its own technological decisions and the information-technology (IT) companies implementing them. We are regularly seeing evidence of American special services’ interference in the activities of major

corporate players in the industry, manipulating content-moderation algorithms and tracking users, including through manufacturers’ built-in hardware and software features. However, the West sees no ethical problems with regard to deliberately allowing AI to ignore hate speech on social network platforms if it adheres to the political agenda that suits it, as in the case of the extremist company Meta and the permission it gives for posting calls for killing Russians. At the same time, the algorithms learn to put out fakes and disinformation and automatically block information that the owners of social networks and their handlers in the intelligence services believe is wrong — in other words, truths that hurt. In the spirit of the notorious cancel culture, AI is made to edit entire digital arrays on demand, thereby generating false narratives. In brief, the primary source of challenges and threats lies not with AI itself but with the unscrupulous champions of AI among the so-called advanced democracies. It is just as important to talk about that as about the issues that the United Kingdom presidency cited as reasons for convening this meeting.

There is a popular belief today that artificial intelligence will create major prospects for emerging new markets and sources of wealth. However, the issue of the unequal distribution of such potential benefits is carefully avoided. The Secretary-General addressed those aspects in detail in his *Roadmap for Digital Cooperation* report. The digital divide has reached a point where 89 per cent of people in Europe have Internet access, while in low-income countries the number is only 25 per cent. Almost two thirds of global commerce and services is conducted on digital platforms, and yet the cost of a smartphone in South Asia and sub-Saharan Africa represents more than 40 per cent of an average monthly income, while mobile data charges for African users are more than three times the global average. Finally, Government support aimed at equipping citizens with digital skills is provided in less than half of the world’s countries.

That is because the wealth created by innovation is extremely unevenly distributed and dominated by a handful of major platforms and States. Digital technologies have resulted in a significant increase in productivity and added value, but their benefits have not led to shared prosperity. The United Nations Conference on Trade and Development’s *Technology and Innovation Report 2023* warns that developed countries will benefit the most from digital technologies, including artificial intelligence. Digital technologies

are helping to concentrate economic power in the hands of an increasingly small group of elites and companies. The combined wealth of tech billionaires in 2022 was \$2.1 trillion. Underlying that disparity is a massive gap in governance, especially across borders, and in public investment.

Historically, digital technologies have been developed by the private sector, and Governments have consistently lagged behind in regulating them in the public interest. That trend must be reversed. States should play a leadership role in developing regulatory mechanisms for AI. Any self-regulatory tools that the industry adopts should comply with the national legislation of the countries where those companies operate. We are opposed to establishing supranational oversight bodies for AI. We also consider the extraterritorial imposition of norms in that area unacceptable. Reaching universal agreements in this area is possible only on a basis of mutually respectful dialogue, on an equal footing, among sovereign members of the international community and with due consideration of all the legitimate interests and concerns

of the participants in the negotiation process. Russia is already contributing to that process. In our country, the major IT companies have developed a national code of ethics in the area of artificial intelligence, which sets guidelines for the safe and ethical development and use of AI systems. It does not establish any legal obligations and is open to foreign specialized agencies, private companies and academic and social entities. The code was formulated in a national effort to contribute to the implementation of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

In conclusion, I would like to emphasize that no AI system should be allowed to put in question the moral and intellectual autonomy of human beings. Developers should regularly assess the risks associated with the use of AI and take measures to minimize them.

The President: There are no more names inscribed on the list of speakers.

I once again thank our technical experts for joining us and our colleagues for their contributions today.

The meeting rose at 12.15 p.m.