European Commission - Speech

Remarks of President von der Leyen at the Bletchley Park AI Safety Summit

Brussels, 2 November 2023

"As prepared"

Session I: AI safety priorities for 2024 and beyond

Dear Prime Minister Sunak, Rishi,

Here in Bletchley Park, we honour the great mind of Alan Turing, the father of modern computing. I do not know how close we are to machines that can reason. Some say they will exist in 5 years, some say not. This was very much discussed yesterday.

We are entering a completely different era. We are now at the dawn of an era where machines can act intelligently. My wish for the next five years is that we learn from the past, and act fast!

Given the complexity of these intelligent machines, AI safety is correspondingly much more complex. Experience from other technologies can therefore be our guide. Take the history of atomic energy and the nuclear bomb. Scientists discovered quantum physics which led to nuclear energy – good – but also with societal risks, and also to the atomic bomb.

This teaches a first important lesson: The independence of the scientific community is essential. We need a system of objective scientific checks and balances. We need to nurture a community of outstanding, independent scientists. Scientists, with access to resources to evaluate the risks of AI, and free to call out those risks.

Second, we need to establish AI safety standards that are accepted worldwide.

Aviation is a good source of inspiration. Air travel has become extremely safe because we learned systematically from mistakes. Any error can lead to a catastrophic result. The aviation community made it a standard practice, that any incident or error is made public and followed-up. It is not seen as a failure but as responsible and appropriate to flag an error. Any error is investigated and the results are publicly available and recommendations are always followed up. This approach shows the value of shared standards and procedures.

My third point, we can also learn from more recent experience in developing a culture of cybersecurity. Organisations are better prepared against cyberattacks when there is effective information sharing. Widely shared security alerts prevent the worst viral spread. It is a matter of time. AI systems also evolve and learn.

Complex algorithms can never be exhaustively tested. So, above all else, we must make sure that developers act swiftly when problems occur, both before and after their models are put on the market. In short, I hope that in 5 years' time, we will all have systems in place that implement those lessons. Doing all of this is the key to unlock the huge benefits of AI.

Session II: Concrete steps to make frontier AI safe

Last summer, I said we should join forces for a global approach to understanding the impact of AI. At that time, I thought about something like the IPCC for climate. Since then, the debate has been so intense, for example in our G7 Hiroshima process, that we see even more clearly what is needed.

I believe that a framework to understand and mitigate the risks of very complex AI systems should be built on 4 pillars.

And we discussed this in the first session:

First, we need a thriving and independent scientific community, equipped with the means to evaluate AI systems.They need public funding and access to the best supercomputers. In the past 5 years, the EU has built the largest public network of supercomputers in the world.And we already

give access to Lumi in Finland and Leonardo in Italy to start-ups and testers.

Second, we need to develop internationally accepted procedures and standards for testing AI safety.

Third, it has to be standard procedure,that every significant incident caused by errors or misuse of AI is reported and followed-up.

And fourth, we need an international system of alerts fed by trusted flaggers.

These 4 pillars should constitute an effective system of governance.

Now, at the heart of it all, there needs to be a culture of responsibility.

For private actors, this means one general principle: The greater the AI model's capability and attendant risks, the greater the responsibility. That means solid, traceable corporate responsibility, embedded within your business model. But this goes beyond purely corporate responsibility.

Public authorities are ultimately responsible for the safety and security of citizens. So we must put in place binding principled rules. And public authorities must have powers of intervention, as a complement and backstop to self regulation. These are the guardrails.

And just as on a motorway: guardrails are not barriers – they allow the traffic to keep to the road and proceed safely.

That is why we proposed an AI act. Its basic principles are to support innovation, harvest the benefits of AI and to focus regulation only on the high risks.

The AI act is in the final stages of the legislative process. In that process, we are discussing the foundation of a European AI Office. This Office could deal with the most advanced AI models, with responsibility for oversight, in the logic of the 4-pillar framework I have outlined. The European AI Office should work with the scientific community at large. It could contribute to fostering standards and testing practices for frontier AI systems. It could complement the private sector in investigation and testing. It should be able to act on the basis of alerts and make sure that developers take responsibility.

And finally, a European AI Office would enforce the common rules in all 27 Member States for those most advanced models.

That is good news for those businesses and for safety in Europe. But the European AI Office should also have a global vocation. It should be open to cooperate with similar entities around the world. Including, of course, dear Rishi, dear Kamala with your newly formed AI Safety Institutes.

Dear colleagues, history is watching. As Mark Twain famously remarked: “The secret of getting ahead, is getting started”.

SPEECH/23/5502

Press contacts:

Arianna PODESTA (+32 2 298 70 24)

General public inquiries: Europe Direct by phone 00 800 67 89 10 11 or by email