DANIEL ANDLER

Professor Emeritus at Sorbonne University, Member of the Académie des sciences morales et politiques, Philosopher

Welcome to the second session on AI today. Our session is sandwiched between the very fascinating previous session in French and before another session, which will be on generative AI for businesses, so in a way, they complement one another. Clearly, we have all hit the road of artificial intelligence and we all know that it will take us far, that is about the extent of our certainty. We do not know how far and where we are heading is unclear, but we have a say. As far as we know, AI does not drive itself and we cannot let the Magnificent Seven and their counterparts in China decide for us. When we look ahead, the most general question is how good AI is going to get but there are two ways of understanding this question. The first and most frequent construal is of what cognitive feats will it eventually be capable of. The second and most significant construal is how good it will be for us. We cannot simply assume that what is good for AI is good for humanity as some people seem to think, nor of course, should we assume the opposite, as others believe. We can try to strike a balance, we can recommend that we reap the benefits of AI while preventing or limiting the damage it can inflict but that does not help, it does not provide any hint of how to approach such a goal. We must do better, and we can. AI's achievements and effects so far are a rich body of evidence, of case studies so to speak, that we can use to assess whether we are moving in the right direction or should correct the present trajectory in small or big ways. Of course, there is an important complication, the directions that AI can in fact take are not entirely pre-ordained, they depend on what scientists and engineers will discover by intent or serendipity, in the short-term and in the fullness of time. As always in science, we sometimes find what we have been looking for and sometime not, and what we find is sometimes that we were looking for and sometimes not. This should not dissuade us from trying to gain as clear a perspective as possible on where AI can take us and where we want it to take us.

The panelists in the present session, as in the next one and the previous one, will provide a wide spectrum of perspectives on this issue. By way of introduction, let me offer a simple, indeed simplistic conceptual map of the kind of concrete questions that need to be raised. There are, on the sunny side of the AI road, achievements and promises that are promoted and on the other, shady side, pitfalls or downsides that are flagged. Regarding achievements and promises, I suggest a 2x2 table, on one dimension some are real or realistic, some are unsupported by present achievements and available ideas. On the other dimension, some are *prima facie* beneficial, some dangerous, risky or outright toxic. Examples that combine realism and *a priori* beneficence might be smart search engines, scientific, medical, legal and driving assistance, which was brought out very well in the previous session by Mr. Ezzat's discussion. Examples that combine realism and risk of toxicity are near-perfect chatbots, especially of the AI companion kind, as well as AI poets, novelists, composers, etc. On the side of beneficial but unsupported promises and exaggerated achievements, explainable artificial general

intelligence, AGI, safe Level 5 self-driving vehicles, and I know some people are going to disagree with me. On the side of dangerous but mercifully, unsupported systems, black box AGI, full-autonomous AI physicians, teachers, judges, policemen, therapists, politicians, and CEOs. Concerning pitfalls or downsides, I propose to distinguish between predicaments that have no solution but need to be faced as lucidly as possible so as to be mitigated as best we can, and challenges that can and therefore should be actively taken up. Examples of predicaments might be autonomous lethal weapons and the new battlefield, cybersecurity, increased inequalities within and between nations, and to use a word that is not admitted in polite, AI-friendly society, dehumanization. Examples of challenges that must be met head-on are respect of privacy, of intellectual property, of simple factual truth, some degree of transparency as far as it does not contradict the very idea of AI, environmental impact, deskilling and last but not least, fabrication of false persons. This last example is a case where we could have imposed from day one a total ban on all technologies that would contribute to this deadly attack on trust, something analogous in the social domain to nuclear weapons to biological survival. It was perfectly clear that these technologies would lead to the present hell of fakery and AI companions and perfectly clear, at least outside Silicon Valley, that it would be disastrous, but we can, and we must claw our way back to sanity and trust. Some of these examples are controversial and each of them, and countless others, calls for careful consideration. Two things stand out, though, it seems to me. One is that while debates on AI in general are necessary, they tend to be inconclusive, progress is more likely when the discussion is restricted to specific areas such as medicine, education, warfare, health, justice, political decision-making, etc. The other is that innovation *per se* is not a sovereign good and therefore, it need not be protected regardless of cost. Regulation may slow it down, but good regulation does it for a reason. When innovation and regulation work hand in hand, humanity benefits.

With those introductory remarks, we will now move on to the panel. I will not introduce the speakers, you have their bios in the brochure, and we will go in the order in which people are seated, and they each have six or seven minutes, and I will be a terribly rigorous chairman.