In March 2024, **Open AI**, the Artificial Intelligence research company that produced, among others, ChatGPT, released a video showing a humanoid robot following the requests of a human operator, handing him an apple, arranging dishes and glasses and carefully putting them back in the washing machine basket. It speaks. It chooses objects and handles them carefully. All this, says the manufacturer, without any external control: a robot with ChatGPT functions built in.

Prof. **Bruno Siciliano** comments here on what he calls a change of perspective for robotics.

Bruno Siciliano is an Italian engineer, academic and populariser of science. He is a full professor of robotics at the University of Naples Federico II and President of the Scientific Committee of the ICAROS Centre, the Interdepartmental Centre for Robotic Surgery, which aims to create synergies between clinical and surgical practice and research into new technologies for computer/robot-assisted surgery.

Open AI enters robotics. A step that opens up important perspectives
The relationship between robotics and artificial intelligence (AI) began several decades ago, is complex and has gone through various phases. Recently, there was an experiment that changed the relationship between these two disciplines. Figure 01, the humanoid robot resulting from the collaboration between the Californian start-up Figure AI and Open AI, showed that it is able to understand sentences spoken by a human and perform the required tasks.
The robot implements ChatGPT and is equipped with cameras to analyse the context. It can interpret voice commands, speak and move objects.
This kind of integration is actually not new. The company Boston Dynamics, famous for the acrobatic performance of its robots, even earlier integrated AI, a text and video recogniser, on Spot, a quadruped robot that responds and performs actions requested by the operator.
The novelty of Figure 01 is Open AI's active collaboration, from the very beginning of the project, with Figure AI, a young company whose mission is to implement autonomous humanoid robots on a global scale that perform unsafe and undesirable tasks. This is a radical change of perspective within a society that until a few years ago only saw Artificial Intelligence as the future, implicitly downgrading robotics to one of the possible applications of generative AI.
Today, we are generally seeing a change in the attitude of the big AI companies towards robotics.

The element of the Figure 01 robot that makes the difference is not movement — there are in fact robots such as Atlas and Unitree that are much better in this respect — but the integration of ChatGPT into the control system.

A physical generative Artificial Intelligence?

With the start of these closer collaborations between AI and robotics companies, a vision of robots endowed with significant interpretative capabilities, and thus able to operate more easily in human environments, even in non-specialised ones, could be established.

euROBIN, a new European network of excellence bringing together the main centres for AI and robotics research in Europe, comes at a crucial time in the development of robotics in Europe due to the spread of interaction technologies (IAT), which are fostering the transition from digital to physical twin, thus integrating artificial intelligence into robotic systems.

Along these lines, at a meeting of a few months ago in Brussels, we proposed a kind of challenge: to try to develop a physical generative AI: Action GPT, an *intelligent interaction technology.*

Generative AI could control the physical system and perform certain operations comparable to those developed with a control algorithm using a mathematical model of the physical system. Not simple movements of the humanoid's arms as in the case of Figure 01, but more complex movements such as a walk of a bipedal robot.

As an expert in control systems, I remain a little skeptical that robot's performance can be achieved solely on the basis of data, which ChatGPT does by accessing various datasets.

In robotics, the AI should also take data describing the system's behaviour and transmit, for example, to a quadruped, a type of robot that is certainly better suited for intelligent actions, all the information concerning the movement of its four legs, its speed, its path, and planning decisions. The complexity is great but the challenge is set.

These experiments certainly do not exclude the role of those who develop mathematical models of physical systems. From a research point of view, one could try to improve the performance of a model-based system by introducing a considerable amount of real-time data, with a view to a hybrid model.

Deep learning capabilities associated with model-based control could be a starting point to be refined by means of reinforcement learning techniques, which go through the feedback of human experience, of the user of the system itself.

Any control system is always organised on the basis of a functional architecture, with hierarchical levels, from the low level of control up to the levels of social interaction and recognition and understanding of human behaviour. Hence, we could have a low-level

control system that takes into account the physics, mechatronic constraints of the system and, at a higher level, a cognitive system powered by generative AI.

One could entrust generative AI with tasks of recognition and adaptation to the human to be served, or whose commands to obey. The generative AI could, for instance, try to recognise — on the basis of data collected on the network — the characteristics of the human, his profile, gender, age, and so on, and also, if possible, trace a more complete profile of the user so as to be able to modulate the required actions. The next step would be whether the generative AI could also communicate the user's physical actions to the robot, not just text or voice commands.

Generative AI could also have the function of detecting failures, errors, which could then be corrected. Just as it could be equipped with the ability to detect impossible tasks, or incorrect commands and select the correct answer or admit that it does not have one. Quite a challenge!

Generative AI could also have the function of detecting failures, errors, which could then be corrected. Just as it could be equipped with the ability to detect impossible tasks, or incorrect commands and select the correct answer or admit that it does not have one.

Quite a challenge!

There is always the problem of quantifying, of introducing metrics; for example, the intervention of a generative AI could be modular — and so applicable to different robots.

We know that ChatGPT or other generative AIs can be asked for text and content refinement based on our other requests, so that the generated product is closer to what we want. If we could apply this to a refinement of a robot's actions in relation to the requested tasks, the environment, the user, it would be a big step that could improve the movement and performance of the physical machine — also based on failure analysis.

The robot could, thanks to an advanced Human-Machine Interface (HMI) correct errors and perfect movements with feedback from the human.

An Internet of Skills

With 5G or 6G in perspective, robots will be able to be dynamically controlled in real time and connected with people and machines both locally and globally. One can see, then, how the Internet of Things (IoT) will be overtaken by the Internet of Skills (IoS) a haptic Internet to enable a remote physical experience through haptic devices that match the skills, the abilities of, for instance, the drone operator or the surgeon dealing with an operation

performed through a remote robotic system. What we describe belongs to a future dimension towards which research is heading.

I believe that the next step will be when these generative techniques are able to support physical actions that also involve the handling of real-time data involving contact. In the case of the demo in Figure 01, the big difference would be if the robot could detect the user's ability to lift the apple, if he is, for example, disabled and cannot move his arms.

The challenge to generative AI is to understand the physical interaction with machines and humans: this is the novelty, over and beyond performance, of the Open AI and Future AI collaboration that, I believe, will not remain isolated.

We are at a turning point. Before, there were two communities, the ones who develop Language Models, which are agents that operate on machines; and the robot developers and manufacturers, who equip them with controls based on mathematical models. The collaboration between these two communities could produce what I call InterAction Technology (IAT) or physical AI.

It could result in a human–robot interaction in which the robot's intelligence acquires some of the human's judgement skills, because it is refined in terms of its ability to understand the human's statements, its continuously improved prompts. Not a sentient system, no doubt, but symbiotic with the human. In some cases, the machine could recognise wrong, dangerous or illegal commands and avoid them; and the human could both trick the machine, it is possible, and improve its performance by interacting with it in natural language.

With generative interventions or other metrics, it might be possible to equip the machine with the ability to know its own functioning and operate on that basis so that it can respond to commands 'being aware' of its own limitations and capabilities. Knowing that it was built with that hierarchical control system with those mathematical models, which allows it to perform those given movements required by the human. That it is moved by electrical instead of pneumatic energy and that this causes a difference in performance and behaviour.

In this way we could be sure that the class of actions in the *ActGPT model* are possible for that class of robot.

Similarly, the physical machine should be able to recognise whether a command from generative AI is incorrect, dangerous or illegal. This is information that the machine should have, as well as the ability to act or not act according to this, because only on the basis of datasets and not mathematical models could the robot not act correctly.

This is indeed a challenge. Because recognising the apple and offering it to the human who says he wants to eat it is one thing, but managing, for example, the physical contact between robot and human requires the robot to have sensory information, and that is not that of the generational models but to be picked up on the spot, at that moment. This would be the case even if the interaction was remote via a dedicated 5G network or the 6G that would allow the haptic, sensory Internet to take place via two haptic devices.

How much to trust machines?

As humans, we tend to trust people and entities that are accredited.

But how do we determine the credibility of a physical generative AI system? It could be the case that generative AI would intimidate the user by being able to know his profile and play on weaknesses or emotions.

This would particularly affect young people, the Z generation, who move rapidly from the real to the virtual and transfer the virtual into the real and vice versa, and who, should they work or interact with machines equipped with generative AI, may no longer be able to distinguish the reality of the physical machine from an entity that appears sentient and conscious.

Should the human feel inferior to the machine, the latter could take control of the situation, while one would have to go so far as to design a machine with generative AI so sophisticated as to be endowed with parameters for identifying similar situations, and thus withdraw from actions that are not clear or well defined by the human, self-adapting to the operator's capabilities, degree of acceptability and intentions.

This is a challenge within a challenge. Not only should we be able to design intelligent machines that can generate physical actions requested by the user, as in the demo in Figure 01, but also that can recognise, through its generative capabilities, the situation, and especially the profile of the user. For instance, the machine should refrain from actions that could harm the human, even if such actions have been requested by him or her.

Another fundamental aspect would be to be able to equip the machine with ethnic, cultural, religious profile recognition capabilities. As also, to make the machine's behaviour vary according to the nations in which it operates, depending on the norms and laws that the machine would have to respect, a condition that would be possible to fulfil thanks to the geo-localisation of the system.

This in part already happens, if we think of autonomous vehicles, where the machine system already recognises the passenger's level of experience, his or her waking/sleeping state, so

there is an adaptation of the system to the user's profile, both in terms of gender, physical structure, and previous driving experience. It is therefore a question of understanding how, in the machine, the generation of an intelligent action can be modulated according to the human that this action requires.

A problem arises here: the more we live in a wired environment, the greater the time and power of the network connection we operate on, the greater the intrusion of technology into our private lives: will we gain in security and control, and lose in privacy?