Review of: "[Essay] Not Quite Like Us? — Can Cyborgs and Intelligent Machines Be Natural Persons as a Matter of Law?"

Santo Di Nuovo[1]

1 University of Catania

Potential competing interests: No potential competing interests to declare.

How much a human-like artificial agent can be ethically and legally responsible?

A review of Gervais' essay "Not quite like us? — Can cyborgs and intelligent machines be natural persons as a matter of law?"

Starting from the advances of Artificial Intelligence (AI) and its cybernetic applications, which can simulate the higher cognitive and social functions typical of humans, Gervais (2023) wonders if it is possible to identify legally relevant differences between artificial and human agents.

The arguments to answer this question are advanced following a transdisciplinary strategy, including neuroscience, linguistics, philosophy of mind, and moral philosophy.

While agreeing with the basic structure of the essay, some integrations are needed to focus attention on the psychological-social and forensic aspects of the complex issue concerning the recognition of a human-like identity to artificial and cybernetic agents, which may have legal value.

What intelligence is artificially simulated?

To ask whether the cyborg-producing "intelligence" has legal relevance, we need preliminary to agree on which intelligence is reproduced in the artificial agent.

The intelligence used in cognitive problem solving, activating the neocortex according to the old MacLean's "triune brain model", is not enough - as Gervais admits - because emotional and social intelligence plays a large part in the practice of learning, communications, and relationships, which are the essence of real life.

The practical intelligence used in everyday life does not always evaluate all the possibilities and heuristics. Still, it selects some of them on the basis of different motivations (subjective time, accessibility of the elements for the solution, interests, attitudes, and emotions). This should be fully simulated in the artificial mind to give it "sentience" and "sapience" - not just "intelligence" - as studies on simulations of problem-solving and human decision-making have shown (for a review: Joksimovic et al., 2023). Also a decision not only rational, as trusting on others, could be simulated in artificial agents (Falcone & Castelfranchi, 2010).

To have ecological validity, artificial intelligence should represent not just theoretical intelligence but a practical one, activated by the dynamic neural "network", continuously connected, in response to contingent stimuli inside and outside the mind.

To what extent can we simulate aspects of mental life that are complex and unpredictable in terms of linear causality, such as desire or anhedonia, striving for fulfillment or apathy, empathy or selfishness: that is, all motivations linked to the “chaotic” complexity of emotions, which constitute the human intelligence actually used, and which is subject to ethical and juridical evaluations? These components should be simulated not only in an algorithmic way but also with respect to the multiple and variable adaptive or maladaptive effects dependent on the interaction between subjective motivation and a variable and often unpredictable context.

Artificial life - of which AI is the foundation - cannot consist only of an appropriately organized set of intelligent neural circuits, perhaps applied to a robotic body. This reduction in complexity would not represent the whole existential reality, that makes us perceive well-being or discontent, love and hate, enjoy and suffer, be in groups and isolate ourselves, create societies, and destroy them. The capabilities of “intuition, insight, and inspiration”, according to the constructivist view quoted by Gervais, require the natural language typical of mankind, symbolic as well as syntactic, ambiguous, and fuzzy as well as formal logic.

AI and its products must not dictate, with a top-down procedure, what life is or should be; rather they should "learn" from it, just like children learn: a concept that evolutionary robotics has clearly affirmed and empirically demonstrated (Cangelosi & Schlesinger, 2015).

The problem, therefore, does not lie only in creating Cyborg organisms that are between human and machine, or Cobots built by implanting intelligent elements in humans or portions of human DNA in the artificial, and then evaluating to what extent the agent produced is human-like and therefore ethically and legally capable of rights and duties.

AI products can have applications in organisms that have educational, economic and social, assistance and rehabilitation, or even military functions - which could therefore have “legally relevant” personality - provided that they are able to reproduce the evolution of a real organism that learns how to behave in these fields, based on the specific situations and conditions of real life, and has specific awareness of them. The perspective of developmental intelligence must be accompanied by the ecological one, i.e. constant interaction with the environment which in turn varies, and to which the recurrent neural network which is the basis of the organism must continuously choose the adjustment methods going further (or even modifying) the initial programming algorithm. Can the incapacity to do this, have ethical and legal relevance, compared to the incapacity to “understand and want” that attenuates the intentionality and the responsibility in criminal judicial procedures towards humans? In this sense, some of Gervais' conclusions should be repositioned starting precisely from the legal systems that pertain to the different human societies.

Like humans in their ontogenetic development, the artificial agent - whatever its proportion of artificial and human - must learn to relate its own intentionality with the ethical and juridical system of the system in which it acts, and therefore its behavior could be assessed in legal proceedings in terms of liability.

In many legal systems the attribution of *responsibility* is associated with that of *dangerousness*, which can be assessed and sanctioned together with, or even as an alternative, to the charge for the action. A person can be legally judged *dangerous* even if incapable of understanding and therefore *not responsible*, and in any case obliged to undergo treatments or restrictions on personal freedom.

The same goes for animals: if one of them attacks and kills humans, it can be put in a position to do no harm (in some cases, even killed). The same should happen with the artificial agent.

If a child commits a crime, not being aware of what he/she is doing, whoever should control his behavior (i.e. the parents) is held responsible.

The question then is: Was the dangerous artificial agent capable of self-control, and did it not? Or the dangerousness depends on its programmer or external controller, and then this should be held responsible?

The ethics of the artificial, and the problem of "control"

Gervais' analysis of the artificial from the point of view of moral philosophy should also be extended to the psychological and juridical aspects of the dilemma between 'external' and 'internal' control. The morality - and the compliance with the norms - of an agent, human or artificial, can depend on another controller agent (in the case of AI, the programmer, or another monitoring agent), or on the same organism capable to plan and perform its own actions.

Developmental psychology tells us that in the early stages of childhood external control prevails, and then moves on to forms of self-regulation based on principles, rules, and norms that one freely chooses to respect.

A correct consideration of the ethical and control aspects of the artificial life of the cyborg agents produced by it is considered essential today, and has been repeatedly reaffirmed by international bodies, e.g. the 2017 Asilomar handbook and the 2019 EU code of ethics, aimed at "Building Trust in Human-Centric Artificial Intelligence".

For these aspects, a close transdisciplinary collaboration between cognitive sciences and social sciences such as psychology, sociology, and law is necessary.

Artificially simulated models can create organisms that develop and adapt to specific environments; or evolve individual awareness towards collective, cybernetic, and even 'trans-human' forms of consciousness. These developments in the relationship between the natural and the artificial re-propose the old questions relating to ethics and social policy, including legal aspects.

The ethical problem arises from the possible discrepancy between the values and purposes of the programmers of artificial minds and organisms, and the values and purposes present in the contexts of the use of the applications produced by the simulation. If ethical behavior is based on internal self-control, the subject of ethics is the agent himself (in the autonomous artificial, as in humans), if instead the control is external, ethics is attributed to the programmer or supervisor.

The dilemma that consequently, and further, arises concerns social policies and juridical-forensic aspects.

Will the evolution of simulation and the artificial always need a human 'controller', who monitors the process as already happens for educators, therapists, rehabilitators, financial analysts, and meteorologists? In this case, the human who controls it is responsible for the simulated agent.

Or the management of the cyborgs can in turn be entrusted to other artificial 'controller' systems which, making use of the increasingly powerful data processing capabilities of intelligent supercomputers, are able to monitor the evolution of the 'controlled' organism up to the goal?

This second possibility is currently the subject of ambitious scientific projects, military applications, and literary or cinematographic fiction; if it were to be realized, it will pose, in addition to problems of economic feasibility, problems relating to project management:

- Who will be able to finance and organize these 'controllers' of artificial life systems and therefore manage their power?

- Who (or what) will "control the controllers" if these are artificial agents capable of self-regulation and adaptation even with respect to modifications of the final objectives (what in philosophical terms is called "heterogenesis of ends")?

These essential questions would still arise even if the controller remains human, even if he makes use of sophisticated cybernetic systems.

The solution could be the control by a democratically organized community, which establishes a direct dialogue between those who assume responsibility for programming and the reference communities that will be involved in application uses. Maintaining responsibility for action programs should be shared with users or their associations, rather than asking them for a technology mandate.

As Bostrom (2014) underlined in his book on the 'intelligent' use of Artificial Intelligence, it must be ensured that those who program artificial agents, and those who monitor them, share goals compatible with the survival and well-being of humanity, purposes that cannot be modified autonomously by the artificial agent itself.

Unfortunately, the solution of a democracy of control has not always succeeded even with the relationship between human technologies, as history has amply demonstrated, and perhaps borders on utopia.

Several authors have discussed it, starting with Kevin Kelly's pioneering 1992 essay: the list would be very long. But shared conclusions have not been reached, and perhaps will never be reached: difficult to reach when the questions are

- if a world of artificial intelligence will improve the quality of life of humans, or cause addiction;

- if it is really possible to simulate/replicate nature and the human psyche, including consciousness;

- and, still upstream, what is the conscious life that one intends to reproduce or recreate.

Let us limit ourselves to concluding that the problems of ethics and control - with their legal implications - are among the

great challenges that humans, indeed the 'post-humans', will have to face in the coming decades (Di Nuovo & Conti, 2017). A transdisciplinary approach like the one suggested by Gervais, extended also to social sciences, is essential to meet the challenge.

References.

Bostrom N. (2014). *Superintelligence: paths, dangers, strategies.* Oxford: University Press.

Cangelosi A., Schlesinger M. (2015). *Developmental robotics: from babies to robots,* Cambridge: MIT Press - Bradford Books.

Di Nuovo S., Conti D. (2017). Robots as post-humans: some issues about artificial life. In: M. Maldonato & P.A. Masullo (eds) *Posthuman. Consciousness and Pathic Engagement* (pp. 7-20), Brighton, UK, and Portland, US: Sussex Academic Press.

Gervais D.J. (2023). Not quite like us? — Can cyborgs and intelligent machines be natural persons as a matter of law? *QEIOS*, https://www.qeios.com/read/9WPMG4

Kelly K., (1992). *'Out of control' The New Biology of Machines, Social Systems, and the Economic World* New York: Basic Books.

Joksimovic, S., Ifenthaler D., Marrone R., De Laat M., Siemens G. (2023) Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence, 4*, 100138. https://doi.org/10.1016/j.caeai.2023.100138