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

Stephen DeCanio[1]

1 University of California, Santa Barbara

Potential competing interests: No potential competing interests to declare.

Gervais' Essay "aims to situates [sic] the difference in law between human and machine in a way that a court of law could operationalize." He acknowledges early on that the law can define almost anything to be a "person" (e.g., lakes, rivers, corporations) but he seeks to go beyond this truism to provide criteria for what constitutes a "human being."

This is no easy task, especially when considering the possibility of adding mechanical or electronic enhancements to human bodies. Leaving those hybrid cases aside, the crucial distinction to be made is between ordinary humans as they exist today and potential machine intelligences (AIs) that may come into existence in the future.

Gervais' Essay is grounded in a fundamentally materialistic metaphysics. With that starting point, it is indeed difficult to come up with a clear definition of what a "human being" is, because AIs potentially can exhibit any or all of the observable behaviors normally associated with humans – agency, creativity, interactivity, humor, and so forth. We already have built autonomous vehicles and weapons systems, and AIs surpass human capabilities in numerous cognitive and physical tasks.

Gervais' solution (or "working hypothesis") is "that biology and the unique biologically-embodied mental abilities of humans are key pieces of the definitional line separating human and machine." This may be adequate for courtroom work, but it would seem to beg the question of establishing the ontological status of AIs that can do most of the things human beings can do.

Drawing a distinction between human beings and AIs entails metaphysics in a fundamental way. I would argue that the key difference cannot be framed in purely materialistic terms. *Human beings have non-material* ***souls****.*

This assertion reaches beyond the task Gervais set for his Essay. Yet it has been lurking since the start of the modern literature on the possibility of intelligent machines. The founder of this literature, Alan Turing, argued in "Computing Machinery and Intelligence" (1950) for a kind of functional equivalence between AIs and humans based on the observational indistinguishability of their cognitive capabilities. He sharpened the question, "Can machines think?" to a thoroughly operational matter of whether a computer could successfully play an "imitation game." Turing's answer was positive. But in discussing a number of objections to this position he brought up the "theological objection" to the notion that computers can "think." (Turing here used "think" as shorthand for the type of equivalence he was positing,) Thus, the

theologian might say, and Turing would answer:

> *Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think…..*
>
> *I [Turing] am unable to accept any part of this, but will attempt to reply in theological terms. …It appears to me that the argument quoted above implies a serious restriction of the omnipotence of the Almighty. It is admitted that there are certain things that He cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this soul. An argument of exactly similar form may be made for the case of machines. It may seem different because it is more difficult to "swallow". But this really only means that we think it would be less likely that He would consider the circumstances suitable for conferring a soul. …In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates (p. 443).*

Thus, Turing-equivalence of a human and an AI neither requires nor rules out the possession of a soul by the AI. This must be the case because Turing-equivalence is purely phenomenological, and the soul is not a material or observable thing. Of course, a purely materialist metaphysics rules out any kind of transcendent God who alone can create souls. But it must always be kept in mind that materialism (and specifically its concomitant principle of the causal closure of physics) is a metaphysical *assumption*, "an unargued and undefended assumption—a kind of intellectual prejudice, in the literal meaning of the word" (BonJour, 2010).

Gervais acknowledges that his Essay does not address the ontological question of the existence of souls when he states (in footnote 16)

> *A related question that this Essay does not explore is this: even if one believes in the existence of soul, consciousness, or sentience as something more than just a biological phenomenon, since we're not sure exactly where or how it is created or otherwise emerges, how do we know we're not going to create it ourselves in an AI system, perhaps especially in AI systems that are partly biologically based.*

It may not be germane for legal practice to know whether AIs have souls. Courts could find AIs (and/or their builders) liable for damages, and willful destruction of an AI could be a serious crime like the destruction of other valuable artifacts. Yet the ultimate moral standing of an entity depends on whether or not it has a soul, and "soul" is of a category different from anything we humans might create. There may be extraterrestrial creatures that have souls, but exploring the implications of that possibility falls under the purview of theology rather than ordinary science (see, for example, Thigpen, 2022).

Regardless of these issues, Gervais' Essay contains a wealth of information and thought-provoking arguments, surveys a broad swath of several distinct literatures, and asks important questions. It makes timely reading as we struggle with the consequences and implications of the proliferation and expansion of AI systems today.

**References**

BonJour, Laurence, 2010. "Against Materialism." In *The Waning of Materialism*, ed. Robert C. Koons and George Bealer. Oxford: Oxford University Press.

Thigpen, Paul, 2022. *Extraterrestrial Intelligence and the Catholic Faith*. TAN Books: Gastonia, North Carolina.

Turing, Alan M., 1950. "Computing Machinery and Intelligence." *Mind* **59**, 236: 433-460.