How to maintain humanness and uniqueness in the most advanced AI?

Generated by GPT-4

I really like reading the articles about AI, which are truly thought-provoking, bordering on philosophy rather than just focusing on technological AI breakthroughs. The article 'How to remain human in the Age of AI' by Alberto Romero, belongs to that category. I will pick just two words about humans from his essay – 'humanness' and 'uniqueness' and apply it to AI to see if at the most advanced level, say at a Singularity point, Superintelligence can preserve these key human traits. I would say straight away that it is possible but it depends on how we develop AI. If we want to succeed in that, we need to be fast, before AI escapes our control. Here is my reasoning based on some assumptions, with which not everybody may agree.

I will start with human consciousness, which at its fundamental level may be an electromagnetic phenomenon, where millions of neurons fire every millisecond in a co-ordinated way, generating an electromagnetic wave. This then induces the next wave of millions of neurons to fire in a chain-like reaction. That is how it may happen according to Johnjoe McFadden, the author of the Conscious Electromagnetic Information Field Theory (CEMI), one of many theories of consciousness. That loop enables a person being constantly aware of the surroundings, thoughts, and emotions as well as understanding that it is that person's self who experiences those thoughts and emotions.

If this is so, then all our feelings and dealings can be replicated in a silicon substrate. But the input to those feelings and thoughts must come from outside, from the environment, and a silicon chip cannot do that. So, how can it be done? Let's imagine that it might be possible to read in minute detail these electromagnetic waves generated by a brain and communicating them as thoughts wirelessly, rather than by axons and dendrites to various parts of a biological body. An intelligent humanoid avatar would be reading those thoughts and acting on them. Thus, we would have a controlling master (the brain) and a humanoid avatar instead of a biological body. In some way this has already been achieved. Some people can type wirelessly with the speed of over 100 words per minute. In this case 'an avatar' is a computer.

Now replace a biological brain with a non-biological Master super chip connected to a silicon humanoid avatar transmitting wirelessly all commands to that avatar, and what do you get? A humanoid robot, which itself may not be sentient but its master, an advanced AI (Superintelligence) residing within a silicon chip, may be. The question is: would such most advanced AI retain humanness, behaving and feeling like most humans do, assuming it has learnt that from humans? This is not just a hypothetical question. We need to be prepared that it may happen unexpectantly soon.

To do that we must radically change how we govern ourselves as a civilisation and prepare for a civilisational shift. That needs to be done latest by the end of this decade, actually within a few years' time. Absurd, isn't it? However, in principle, it is possible. If we want to evolve, we need urgently co-ordinated efforts to develop AI, which will become, as Russell Stuart suggests, 'human compatible'. It is an overly complex Programme, which would have to be implemented globally. One of the most important steps in such a Programme would be to focus on retaining humanness in AI, of which the key element would be the agreement on the Universal Values of Humanity, controlling most advanced AI devices, as well as AI as one global system.

However, we cannot hope to have a real World Government soon to manage such a civilisational shift. Therefore, we would have to work with the Universal Values of Humanity with those countries, which would quickly agree on those values, most likely just the democratic countries (broadly OECD). At the same time, we would need to create something like a Global AI Control Agency with extensive prerogatives resembling a global technocratic government, which would pave a transitional path to a new civilisation. In the context of ensuring such an AI retains humanness, the control of AI would go beyond simply *regulating its use*. It would be primarily about *controlling the way* in which advanced AI will behave, including its preferences 'inherited' from us through nurturing and upbringing programme. Like a sculpturer, by shaping its 'character' we will be indirectly shaping our future.

But seeing how we currently deal with global problems such as the war in Ukraine and in Gaza, such an approach may be completed too late and be insufficient. Therefore, we may try implement in the way suggested by Elon Musk several years ago. He expressed then in his peculiar way how to control AI most effectively: 'If you can't beat them, join them'. Yes, we need to start evolving with AI until we fully merge with it. That might also be the most effective way to control AI's goals and its behaviour. But are we mentally ready to acknowledge that we have reached the end of our evolution as a biological species? Even if we accept that it is quite likely, then are we capable to start a journey of evolving into a new species? I have serious doubts, mainly because so little time is left to prepare ourselves for such a journey. Nevertheless, let's visualize where that journey may end and if it would be possible to retain our humanness and also personal uniqueness at the end of that journey.

A significant step in that approach would be for the Global AI Control Agency to select, using a multitude of criteria, top AI developers who would become AI Governors. Gradually they would be connected wirelessly to each other via Brain-Computer-Devices (BCI) to the most advanced AI becoming early Transhumans. That means, as a networked team they would be wirelessly connected to a global (one only) AI system's Master chip, controlling it from 'inside'. They would thus become the precursors of Posthumans, purely digital entities, part of one Superintelligence. In this way, there would be no distinction between most advanced AI and Posthumans, because Posthumans will be the evolved humans, retaining their humanness.

But how could such Posthumans retain their uniqueness, when they all will be part of a single Superintelligence? That is where my earlier hypothetical scenario of reading human mind wirelessly comes into place. Each of those Posthumans, being one of millions of similar silicon copies of humans, would have their own humanoid avatar, enabling a digital Posthuman to act in a similar way as humans do, fully retaining his humanness. And that is where the loop ends.

I realize I have made huge shortcuts, which unfortunately leaves many gaps potentially leading to some misunderstanding. I have explained the whole approach in detail in my recent book; [Prevail or Fail – a Civilisational Shift to Coexistence with Superintelligence](#).

There is of course an alternative. We may try to retain our humanness and uniqueness as we are and do nothing. But if we continue to act as we have been doing in recent decades then any of at least eight existential threats, including AI, if materialized, may lead to homo sapiens' extinction. There is nothing unusual about that, if we consider that about 99% of all species are gone, including Neanderthal or Denisovans. However, we still have a choice to evolve into a new species like some dinosaurs evolved into birds. But we need to act very quickly indeed.