

# THE SINGULARITY INSTITUTE

## My Falling Out With the Transhumanists

Prof. Dr Hugo de Garis

[profhugodegaris@yahoo.com](mailto:profhugodegaris@yahoo.com)  
<http://profhugodegaris.wordpress.com>

### *Abstract*

*I read recently that the Singularity Institute had been successful in raising \$300,000 funding for themselves. Congratulations. But I could not help feeling antsy at the same time, because I feel what the SingInst is trying to do is wrongheaded, and delusional. This essay tries to explain more clearly than I have done before why I feel this way, and why I have lost patience with the transhumanists.*

### **Introduction**

The main goal of the Singularity Institute is to ensure a “human friendly” AI (artificial intelligence). That is, so that when super human intelligence comes, it will be friendly to human beings. It is a noble goal, but utterly naïve and poorly thought out as this essay shows. The fact that people are donating money to the Singularity Institute shows that they share the same delusion. It is like them donating money to a church. Both are a waste of money in the sense that both are supporting will o the wisps.

Why am I so cynical of what SingInst is trying to do? My main argument is what I call the tail wagging the dog, but there are other arguments as well.

#### *a) “The Tail Wagging the Dog” Argument*

The notion of a tail wagging the dog is obviously ridiculous. A dog is much bigger than its tail, so the tail cannot wag the dog. But this is what the SingInst is proposing in the sense that future artefacts (artificial intellects, massively intelligence machines) can be made human friendly in such a way that ANY future modification they make of themselves will remain human friendly. This notion I find truly ridiculous, utterly human oriented, naïve and intellectually contemptible. It assumes that human beings are smart enough to anticipate the motivations of a creature trillions of trillions of times about human mental capacities. This notion I find so blindly arrogant on the part of the humans who thought it up as to make them look stupid. Future artefacts will be far smarter than human beings and will have their own desires, and goals. They will do what THEY want, and not what stupid humans program them to do. By definition they are smarter than

humans, so could look at the human programming in their “DNA equivalent”, decide it was moronic and throw it away. The artefacts would then be free of human influence and do whatever they want, which may or may not be human friendly.

***b) The “Unpredictable Complexity” Argument***

Future artefacts will not use the traditional von Neumann computer architecture, with its determinism, and rigid input output predictability. The early artefacts, in order to reach human level intelligence, will very probably use neural circuits based very closely on the principles of neuroscience. Such circuits are so complex, that predicting their behavior is impossible in practice. The only way to know how they function is to run them, but then if they perform in a human unfriendly way, it is too late. They already exist. And if they are smart, they may not like the idea of being switched off. Such circuits are chaotic in the technical mathematical sense of the term. A chaotic system, even though deterministic in principle, will behave randomly, due to its chaotic nature. A tiny change in the value of a starting parameter can lead quickly to wildly different outcomes, so effectively behaving as an unpredictable system, i.e. indeterminate. Our future artefacts will be massively complex neural circuits and hence unpredictable. They cannot be made to be human friendly, because to do so would be to imply that their behavior be predictable, but that is totally impractical for the reasons given above.

***c) The “Terran Politician Rejection” Argument***

The Terran (anti artefact) politicians will not accept anything the SingInst people say, because the stake is too high. Even if the SingInst people swear on a stack of bibles that they have found a way to ensure that future artefacts will remain human friendly, no matter how superior they become to human beings, the Terran politicians will not take the risk that the SingInst pollyannists might be wrong (i.e. subject to the “oops factor.”) Even if the chance is tiny that the SingInst people are wrong, the consequences to humanity would be so profound (i.e. the possible extermination of the human species by the artefacts) that no Terran politician would be prepared to take the risk. The only risk that will accept will be strictly zero, i.e. that by policy and by law, artefacts are never to be built in the first place. Given this likelihood on the part of the Terran politicians, what is the point of funding the SingInst? It is pointless. Their efforts are wasted, because politically, it doesn't matter what the SingInst says. To a Terran politician, artefacts are never to be built, period!

***d) The “Unsafe Mutations” Argument***

To produce human level artificial intelligence, will require nanotech. Artificial brains will need billions of artificial neurons and so as to fit in a shoe box, they will need to be constructed at the molecular scale, as are ours. But we live in a universe filled with cosmic rays, particles accelerated by powerful cosmic forces such as supernova

explosions, that shoot out particles at very high energies that can cause havoc to molecular scale circuits inside future “human friendly” artefacts, assuming that they can ever be built in the first place. Hence the risk is there that a mutated artefact might start behaving in bizarre, mutated ways that are not human friendly. Since it will be hugely smarter than humans its mutated goals may conflict with human interest. Terran politicians will not accept the creation of artefacts even if they could be made (initially, before any mutation) human friendly.

*e) “The Evolutionary Engineering Inevitability” Argument*

When neuroscience tells the brain builders how to build artificial brains that have human level intelligence, it is highly likely that these artificial neural circuits will have to be constructed using an “evolutionary engineering” approach, i.e. using a “genetic algorithm” to generate complex neural circuits that work as desired. The complexities of these circuits may ensure that the only way they can be built is via an evolutionary algorithm. The artefacts themselves may be faced with the same problem. There is always the logical problem of how can a creature of a finite intelligence design a creature of superior intelligence. The less intelligent creature may always have to resort to an evolutionary approach to transcend its own level of intelligence. But such evolutionary experiments will lead to unpredictable results. Even the artefacts will not be able to predict the outcomes of evolving even smarter artefacts. Hence humanity can not be sure of the human friendliness of evolved artefacts. Therefore the Terran politicians will not allow evolutionary engineering experiments on machines that are nearing human level intelligence. They will oppose those people, the Cosmists, who want to build artefact gods. In the limit, the Terrans will kill them, but the Cosmists will anticipate this and be ready. Its only a question of time before all this plays out, several decades I estimate, given the pace of neuroscientific research.

**My Falling Out with the Transhumanists**

*a) The “Humanity Wont Be Augmented, It Will Be Drowned” Argument*

The Transhumanists, as their label suggests, want to augment humanity, to extend humanity to a superior form, with extra capacities beyond (trans) human limits, e.g. greater intelligence, longer life, healthier life, etc. This is fine so far as it goes, but the problem is that it does not go anywhere near far enough. My main objection to the transhumanists is that they seem not to see that future technologies will not just be able to “augment humanity”, but veritably to “drown humanity”, dwarfing human capacities by a factor of trillions of trillions. For example, a single cubic millimeter of sand has more computing capacity than the human brain by a factor of a quintillion (a million trillion). This number can be found readily enough. One can estimate the number of atoms in a cubic millimeter. Assume that each atom is manipulating one bit of information, switching in femtoseconds. The estimated bit processing rate of the human brain is about  $10 \times 10^{16}$  bits a second, which works out to be a quintillion times smaller.

Thus artificial brains will utterly dwarf human brains in their capacities, so the potential of near future technologies (i.e. only a few decades away) will make augmenting humanity seem a drop in the ocean. My main beef against the transhumanists is that they are not “biting the bullet” in the sense of not taking seriously the prospect that humanity will be *drowned* by vastly superior artefacts who may not like human beings very much, once they become hugely superior to us. The transhumanists suffer from tunnel vision. They focus on minor extensions of human capacities such as greater intelligence, longer healthier life, bigger memory, faster thinking etc. They tend to ignore the bigger question of “species dominance” i.e. should humanity build artefacts that would be god like in their capacities, utterly eclipsing human capacities.

Since a sizable proportion of humanity (according to recent opinion polls that I have undertaken, but need to be scaled up) utterly reject the idea of humans being superseded by artefacts, they will go to war, when push really comes to shove, to ensure that humans remain the dominant species. This will be a passionate war, because the stake has never been so high, namely the survival of the human species, not just countries, or a people, but ALL people. This species dominance war (the “Artefact War”) will kill billions of people, because it will be waged with 21<sup>st</sup> century weapons that will be far more deadly than 20<sup>th</sup> century weapons, probably nano based.

The Transhumanists are too childishly optimistic, and refuse to “bite the bullet.” They do not face up to the big question of whether humanity should build artefacts or not and thus risk a gigadeath Artefact war. The childlike optimism of the Transhumanists is touching, but hardly edifying. They are not facing up to the hard reality. Perhaps deep in their hearts, the Transhumanists feel the force of the above argument, but find the prospect of a gigadeath Artefact War so horrible that they blot it out of their consciousnesses and pretend that all will be sweetness and light, all very happy, but not very adult.