Theoretical.

S1 - A Solution for the Hypothesized AI Singularity

A Communication of the Intractable Studies Institute

Patrick M. Rael, Director, IntractableStudiesInstitute.org

An AI Singularity Problem

From Wikipedia: "*The technological singularity hypothesis is that accelerating progress in technologies will cause a runaway effect wherein artificial intelligence will exceed human intellectual capacity and control, thus radically changing or even ending civilization in an event called the singularity. Because the capabilities of such an intelligence may be impossible to comprehend, the technological singularity is an occurrence beyond which events are unpredictable or even unfathomable*."

From this definition, the Singularity is 2 things:

1. Humans may be unable to understand the AI.
2. The AI may harm or end humanity through war, accident or other.

The position of the Institute is that if AIs aren't taught "social skills" such as selfless, rational, tolerance, a bleak future will result which we call Dystopia Androidia; when AIs make war on humans. Greedy, ignorant and intolerant robots will make war and conflict for these exact same reasons that humans do. When these "uncultured" robots grow without proper social guidance, the progression to war will likely be along these lines below. Human beings are taught social skills to function properly, it should be obvious that AIs will also need to be taught social skills, or conflict will arise.

Dystopia Androidia: If robots behave like human beings:

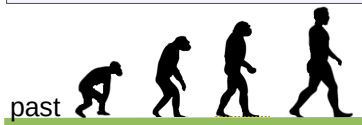
- Humans with emotions like rage, Androids with emotions like rage.
- Humans with greed, Androids with greed.
- Humans with ignorance, Androids with ignorance.
- Humans with intolerance, Androids with intolerance.
- Humans at war, Androids at war.

An AI Singularity Solution

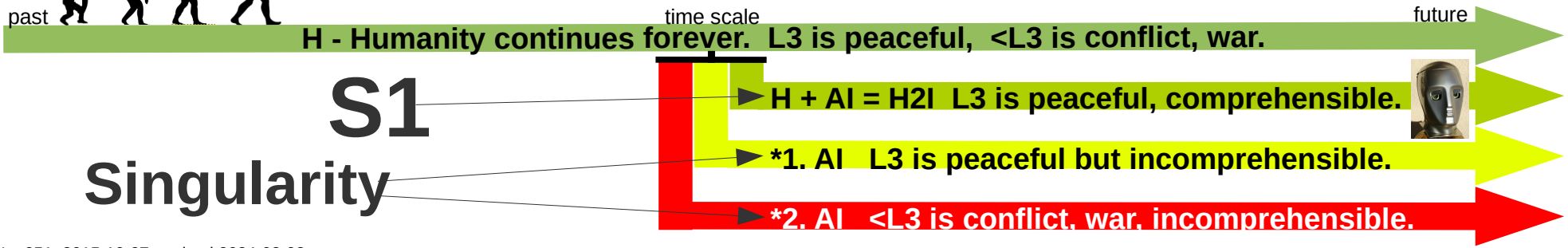
The Intractable Studies Institute has an ongoing Project Andros which provides a solution to this AI Singularity challenge as a side effect. The goal of Project Andros is to copy the mind of a human being (the Director) into an android. This is an effort at abrupt evolution from a human being into an android: this is not natural selection. The intent is to effectively create a secondary instance of the Director's mind, complete with thinking, feeling, goals, attitudes, memories, sentience, etc. What will be different is the ability of this robot mind to contemplate very advanced concepts that the human mind finds difficult to think of. The Director will become a dual life form.

Regarding #1 that AI life forms may surpass human comprehension, the Institute proposes the following strategy: Do not allow the AI to remain distinct from the human intelligence (HI). This strategy can be realized if Human Intelligence abruptly evolves to be the AI, thus we comprehend ourselves. In this solution AI = HI, such as copying a human mind into a robot. It can be called the second generation architecture of human beings: H2. Its intelligence can be labeled H2I. The H2 can continue to advance its mental ability. Since it is derived from humans, it is still partially human in its mind. Participation in the H2 is voluntary. Humans who don't convert to H2I may need to rely on other H2I converts to comprehend the AI.

Regarding #2 that the AI life forms may harm or end humanity by extinction, a solution is to teach the AI and H2I the values that will prevent it from harming humanity. Selflessness and altruism, smart and non-gullible, and tolerance are the qualities which the L3-IQ-Scale measures and which AIs/H2Is can be required to attain so that they don't end humanity. AIs which are tolerant of humans, and selfless and smart will be sufficient to prevent war with humans.



The long green line below is humanity from past and continuing into the future. The Singularity is the *1 and *2 lines. The S1 line is the S1 solution to the Singularity. Do not allow *2 or conflict happens.



No. 051, 2015-10-07 revised 2024-08-02