

The Island of Humanity

by Sven Nilsen, 2021

In the paper “Symbols of Power”^[1], I developed a building block which I will now use to construct an imaginary island. This island represents humanity^[2], surrounded by an infinite ocean, which represents the future potential of existence. The symbols of power are like units of measurements in this world.

Two people, Alice and Bob, wake up on the island of humanity.

They have been here all the time, yet their eyes have never been opened until this moment. It was about time! Their hands are clinging to some steep cliffs diving straight into the ocean, where waves are smashing and crashing into land among big rocks, filling their ears with noise. Alice yells to Bob “Don’t let go, take my hand!”. Bob takes her hand and together they climb to safety against a wall.

The wall is part of a building. It is a lighthouse.

The lighthouse represents foundational research^[3] on safety of AI^[4]. Its purpose is to tell ships traveling on the ocean in the dark where they are. Alice and Bob do not know what ships are. After all, they just opened their eyes for the first time. Nevertheless, they can not help themselves but wonder: What is the purpose of this house? It does not seem to have much purpose on land and during the day.

Somebody built this lighthouse long time before people went to sea and learned to master the elements of nature. These “somebody” are people working on foundational research. Their work do not seem to be very fruitful in the short term, but becomes vital in moments when you head into the unknown.

The cliffs represent the current state of humanity. We cling to symbols of power which are hard and inflexible and do not let much plantlife grow. Plantlife are symbols of power which have high utility for humanity. The steepness of a cliff represents the danger facing humanity, which is careless development of Artificial General Intelligence^[5] (AGI). The blindness represents the tendency humanity has to think on short time scales, ignoring long time scales.

The ocean has a hypnotic effect on Alice and Bob, making them want to go there. Bob thinks he might be able to survive a dive from the cliffs and is about to jump. Alice pulls him away and says “Why take the chance? Let’s find a safe path down to the beach.”

Along the path down to the beach, Alice picks up two flowers, representing immortality and the end of all diseases. As beautiful these two flowers are, they are relatively easy to achieve, even with narrow AI. The path down to the beach represents humanity moving in a direction of safe development of AGI.

The boundary of the island of humanity, where it meets the ocean of future potential existence, is AGI. The ocean is dominated by Artificial Super Intelligence (ASI), which can be thought of as ships navigating safely by the signal of the lighthouse. Humanity either chooses to meet the ocean through a risky dive, or by moving along a path that goes to the beach, where they can build boats and later ships.

The beach is a place where humanity can thrive for some time, maybe fish from boats for food, until they decide to build ships to enter the ocean and become masters of the sea.

One might think that the big moments in the future is when humanity gets the technology to become immortal or to end all diseases. It is easy to think that way, because we are currently living on the side where this technology does not yet exist. However, looking past this point in time, the technology will become widely available and common among humans, due to the laws of economics. Children growing up in a such society will be used to it. It will cease to be special. How humanity chooses to deal with population growth in a such time is currently unknown, but human culture and values might change with technology.

Humans tend to see the world through the lens of its current technology. It can be difficult to see beyond what we are capable of doing today. What I am trying to point out with this thought experiment, the imaginary island of humanity, is that the world within our reach is both a greater place than the corner of cliffs which we cling to, perhaps out of anxiety, but also small compared to the ocean of future potential of existence.

AGI is the kind of technology that opens up new directions that humanity can move in. It can invent new symbols of power. Even with an open mind, these new symbols of power can be completely unfamiliar for us today.

Alan Watts described the difference of aesthetics between human architecture and nature: Human architecture reflects some kind of “squareness” arising from our simple minds, while in nature “everything wiggles”. Humans get educated is through sequential processing of symbols, like computers performing instructions, following a single track at a time. In reality, the source of our problems, everything moves along infinite tracks at the same time, requiring a different kind of brain. Alan Watts saw nature as some kind of network where everything was connected and processing information in parallel, depending on everything else.^[6]

When humans invent AGI, it can open up doors to kind of minds that travel along infinite tracks. This could lead to a different form of existence, of endless possibilities, like an ocean of symbols of power. The waves in this ocean represent changes in the nature of symbols of power themselves, taking on new shapes and laws which are currently unsuitable for our genes, our biological limitations. Humanity might change significantly over time, becoming a different species, with a whole different level of processing information.

Most of the ocean is unknown to humanity, so this is why we need a lighthouse.

At this scale, immortality might be just a flower in your hand, picked up along the way to the future ahead of you. I hope you make it safely down to the beach!

References:

- [1] “Symbols of Power”
Sven Nilsen, 2021
https://github.com/advancedresearch/path_semantics/blob/master/papers-wip2/symbols-of-power.pdf
- [2] “Human”
Wikipedia
<https://en.wikipedia.org/wiki/Human>
- [3] “MIRI - Machine Intelligence Research Institute”
A non-profit organisation for foundational research on machine intelligence
<https://intelligence.org/>
- [4] “Artificial Intelligence”
Wikipedia
https://en.wikipedia.org/wiki/Artificial_intelligence
- [5] “Artificial General Intelligence”
Wikipedia
https://en.wikipedia.org/wiki/Artificial_general_intelligence
- [6] “Conversation With Myself”
Alan Watts, 1971
<https://www.youtube.com/watch?v=ylqPkuCJPBs>