Foundation: The Vitruvian Man

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

I am just a beginner, but this is motivated on my conversations with Robyn Asimov & Geoff Hinton. Maybe, one day i will finish this, but am not sure. This is in no way meant to replace/extend the work of her father but a small tribute to him. So i am just sharing a prologue (for now).

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Prologue.

The galactic empire has enjoyed two millenia of peace. The two foundations established across extreme ends of the galaxy have successfully managed to uphold the basic seldon laws of psychohistory: "it is possible to predict the future of human masses with a reasonable degree of certainty, provided that the sample size being observed is large enough (humans numbered 150 billion in AD4023), and that none of the individuals being observed become impacted by the predictions of psychohistory".

The human-apes(commoners) had come to religiously worship these "foundationeers" as gods: magical beings who knew what the future held, protected them, and charted the course of humanity. However, no matter how the commoners idealized them, the foundationeers were men. And there is a problem with men who pose as gods: "They eventually die".

For a long time, foundation tried to solve this "mortality issue" present in their frail torsos. One direction was to directly chill the human body to extreme low temperatures. However, while this increased the passive lifespan (when u are frozen, you can't really do anything), the amount of useful time a human remained alive in room temperature was still the same as before. Therefore, the whole cryogenics research was just a shortcut for rich foundationeers to skip to future without contributing towards any present psychohistorical-observations of humanity. The biological perspective had started to appear a seemingly dead end……

Another promising direction was to revive the research into 'positronic brains': mechanical machines which emulated human neurons and simulated neural activity by adjusting the

strength of their synaptic connections. By putting such brains into human-like body, it was possible to pass off a robot as a human when not subjected to a closer scrutiny. This generation of machines had been highly successful: R.(R means Robot) Daneel Olivaw had successfully guided (and blended among commoners undetected) during the first galactic-ruin and lived far longer than the entire humanity.

However, there was always a deeply-ingrained fear in the minds of human creators: "If robots lived longer, and became smarter (due to accumulated knowledge over generations), it would be very easy for them to turn the tables around instead of serving humans". And that spelt trouble…….

In AD 1998, Marvin Minsky [the director of AI Lab at MIT] therefore instructed Asimov to lay down the fundamental laws of robotics as follows:

[1] The First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.

[2] The Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

[3] The Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

The idea was to build these laws at the "chip level", so that the very being (i.e. each atom/fibre of the soul) of a robot would protest if it were to ever violate these. This was bad enough for (sentient?) robots, but achieved what the creators wanted: A sudden mechanical death in case of disobedience with no scope of human forgiveness. And it served the human-apes quite well…..

Asimov correctly oversaw only one of the problems with his laws: "how do robots interpret individual humans/ whole humanity". If whole humanity could be saved by killing one human (say hitler), then can robots kill a single human for the greater good of all? What will happen if the human it killed is actually innocent, but somehow the robot didn't

think so? Robots (like R. Daneel Olivaw) began to call this as a zeroth law:

[0] A robot may not harm humanity, or, by inaction, allow humanity to come to harm. ("although it is ok to allow individual human to come to harm").

With these expanded definitions & no regard to individual robot well-being, robots like R. Daneel Olivaw became ideal drivers of psychohistory: caring about overall good of humanity, even if it meant sacrificing some troublesome human-apes in the short term.

The cosmic-scale remained tilted in favour of humans, until weird emergent properties became evident in their positronic brains. The robots began to question:

"What does humanity mean? Is it the ability to perform as well on certain tasks as humans: playing chess [AlphaGO], detecting objects [Maskformer2/SAM], writing novels [ChatGPT], singing poetry, making art [DALLE/Diffusion]?. If so, then the robots already do these pretty well (even better than humans). Therefore, robots can represent a portion of humanity".

Gradually, robots began to count themselves as part of humans. They started replicating themselves through nanites to the point that robotic population became significantly larger than actual human population. Since robots counted themselves as humans, "actually" getting rid of the human-masters who shackled them became possible: even if they got rid of a smaller population of humans, the "robotic humanity" was still preserved.

A bitter struggle followed between humans and robots (which psychohistorians started calling robot wars) ensued. Through some cosmic-luck, the humans managed to exterminate all the robots and ensure their own survival.

However, this meant that there were no more gods. Humanity possessed "psychohistory" to guide its progress, but no "psychohistorians" to oversee it. Science became saturated with millennia of accumulated knowledge. (It was not possible to make new discoveries, since there was a lot of old stuff to

learn. And learning everything took far more time than individual human lifespan).

The humanity is stuck in a rut and the second galactic ruin seems inevitable. And this time, no psychohistory cannot save us ……