Matrix re-viewed: challenges from AI as seen in the light of Paul Tillich’s thought

Eduardo R. Cruz

RESUMO
O recente lançamento de Matrix Reloaded trouxe de novo à tona aspectos religiosos do filme original, Matrix, objetos de muitas controvérsias. Aqui não fazemos uma análise comparativa destas possíveis leituras do filme, mas procuramos avançar na discussão apontando para além das mensagens dele, do ponto de vista da tradição Cristã. Começamos a análise salientando um aspecto do filme, em relação a outras obras que seguem na esteira do anti-tecnológico Frankenstein de Mary Shelley. De fato Matrix, ao invés de apresentar andróides e outros artefatos físicos (hardware), coloca como protagonista um “robô” da mente, um software: o próprio “Matrix” e os agentes, programas sentientes. A diferença maior está no grau de perfeição atingido por estes “softbots”, assim como as infinitas possibilidades de logro e auto-ilusão que os softwares apresentam. Esta situação leva a duas respostas por parte da tradição Cristã (para usar a terminologia do método da correlação de Tillich). Primeiro discute-se a ambiguidade de toda perfeição humana, da qual nem Neo, o “redentor” no filme, escapará. Alguns paralelos são também traçados com o filme AI, de Spielberg. Sobre o logro (deception, utilizando um termo do estudo Darwiniano do comportamento), é importante ressaltar-se que nenhuma das respostas religiosas derivadas do filme (versões do gnosticismo, do Budismo e do Cristianismo) dele escapa. Como bem apontou Tillich, todo esforço religioso humano ainda é feito no âmbito da esfera da finitude e da ambigüidade, mesmo que nenhuma falta moral esteja envolvida. A sombra da idolatria sempre se projeta sobre toda salvação obtida na história, por mais perfeita que seja. Por isso Tillich aponta o símbolo da Cruz como o símbolo redentor por excelência, que atinge o cerne da tendência da elevação do finito ao infinito. Por mais profundas que tenham sido as implicações religiosas de Matrix, suas propostas de redenção ainda não escapam deste destino.
From Frankenstein to Sentient Programs: The Person and Its Creature in a Foreseeable Future

In two previous papers of mine (Cruz 1998 and 2001), I have had the opportunity to develop some considerations on the meaning of person, expanding on and analyzing the notion offered by Paul Tillich, drawing specially on his analyses of the ambiguities of life and of the ambiguity of perfection. It is still worth to pursue this trend of thought, to the extent that the notion of personhood is at stake with stronger passions than just a few years ago. This is particularly the case now when gurus of the Artificial Intelligence (hereafter AI) such as Mark Pesce, Hans Moravec, and Ray Kurzweil, as well as some outstanding movies (such as Matrix and AI), point to a foreseeable future full of androids competing with our biological species for a human status.

Not long ago creatures made by humans in their image and likeness were material beings. Frankenstein, the first in a series from Mary Shelley onwards, challenges his creator about his identity (ambiguity maintained: the one of the creature and above all the one of the creator). This mixture of sci-fiction and art reaches a recognizable climax with Blade Runner, where revolt against the creator and the quest for being are superseded in the affirmation of life. Matrix introduces two new elements in this trend. First, the battle for superiority is over, and was won by the artificial creatures. Second, these machines recede to a background in the plot, and a computer program (and add-ons, such as the “agents”) places for humans the main obstacle to the recovery of their freedom: deception.

“Infobots,” “softbots,” and “agents” are not new to the community of computer experts (see, e.g., the site http://agents.umbc.edu. But the general public never thought of softwares as something so autonomous vis-a-vis hardware, to the point of being “Frankensteins” on their own right. What is so scary about this? Androids are projected to carry something that is part of the best of us, to the point of perfection. That is the case, e.g., of “Gigolo Joe” in AI. As Anne Foerst has put it, “The main vision behind that is to solve the last mystery which is still remaining: the full and complete understanding of ourselves—and the overcoming of all problems (psychological and somatic) which keep human life away from perfection.” (Foerst 1996a, 2). So the task seems...
to be simple: just look for the imperfections that we are conscious of, and correct them in the creatures that we are building.

Two problems arise: on the one hand, we are subject to the limits of our knowledge of the world and, even when some sort of perfection is achieved, this perfection reveals itself to be ambiguous. [1] On the other, “perfection” entails all those biological traits that helped to make us humans, including those that we usually (and naively) regard as nasty: that is the case of deception and self-deception.

Evolutionary biologists studied this trait for long, common as it is to all primates (Alexander 1979; 1987). As such, it is essential to our survival. The machines in Matrix seem to have learned this lesson quite well, having developed sophisticated software specially devised to maintain human beings in a perpetuated state of self-deception. Perfection here is not at stake: the software has no flaws, and has to be cracked out from within, following strictly its own rules. Another level of ambiguity has to be sought for, one that goes beyond our usual notions of “defective” and “bad.”

Indeed, the agents in the movie act perfectly: the rebels, smart as they are, cannot face the agents head on—every once a while they have to run. Mastery of the software is not enough, and the rebels have to yearn for the “chosen one.” At the end of the movie, the agents finally meet their turn to run away—they realize that Neo, as it were, “has become like one of us, knowing good and evil” (Gn. 3:22). Neo, however, is already more than they are. The Genesis saga thus continues: “what if he now reaches out his hand and takes fruit from the tree of life also, eats it and lives for ever? ” (3:23). Neo is free because, as opposed to the agents, he chose his own destiny (the oracle did not give him any warranty that in fact he was the chosen one) and, in his freedom, he is superior to both the rebels and the agents.

Tragic irony: it was in the attempt to live for ever that human beings extended themselves through machines, and now Neo is exactly at the same point his predecessors were in the first place! The religious overtones of Matrix have been discussed elsewhere, and there is no reason to present them here once again. [2] The important point to make is that mastery of knowledge and technology is not enough: the religious dimension points to the overcoming of human finitude and, if
death is the ultimate limit of the human, to life everlasting.

Nevertheless, as Paul Tillich has pointed out, religion itself is ambiguous (Tillich 1963, 86-106). Notions such as idolatry and symbols such as the Anti-Christ remind us that religion itself is the great deceiver of all. Marx and Freud could help us to analyze this in terms that are more scientific. Not only self-deception is a given for humans (for persons as well, even if we differentiate one from the other), but it is also something that we long for. While training Neo, Morpheus explains to him how appearances are deceiving, and how all humans were connected to this surrogate of nature and culture which is the Matrix: “You have to understand, most of these people are not ready to be unplugged. And many of them are so inert, so hopelessly dependent on the system that they will fight to protect it.” Church services are not shown in the movie but, oh, the machines would love them in the Matrix!

The way out offered by the movie is somewhat elitist, to the point that some have suggested a Gnostic bent to it. It resumes for the same reason ideals from the Enlightenment—have complete knowledge, and you will be saved. Consequently, the movie assumes that self-deception can be done away, with proper means. This goal is defeated by what was said in the preceding paragraph. Biology, moreover, tells us that self-deception is here to stay. Curiously, there are some in the technological community who even propose a future when we will have the “right to live in a simulation”—that is scary! [3]

The parallels between the Judeo-Christian tradition and characters in the movie are explicit. Do they make Neo some sort of Jesus Christ, or “New Being,” to use Paul Tillich’s terminology? I think exactly the opposite is true. Neo is more of an Übermensch than the Ecce Homo of John 19:6. The Christian Savior saves us by his kenosis, his emptying himself out, delivering himself to the System, and ironically defeating it from within. This sort of reversal is not present in the movie, no irony, no paradox—that is why the sole irony is the one pointed out above: Neo arrives exactly at the same point where human beings were starting from in the first place. By not breaking the tragic consequences of self-elevation in religion, Neo is yielding a new cycle for humans where deception will happen again.
Let us tie together some of these loose ends. Human beings are by nature self-deceivers, entailing that deception is an unavoidable part of our personhood. We may fight against self-deception, but we also rush toward it, using all sorts of narcotics to numb our senses, the most sophisticated of them being simulations enabled by softwares. It is a matter of survival—better stay in a state of “blissful ignorance” (or “dreaming innocence,” as Tillich puts it) than to face constantly a harsh and dreadful reality. Our artifacts, even the most noble of them (painting, music, and so on) do not escape this fate. [4] Robots and infobots seem particularly suitable to perform this stupefying task, to the extent that they carry our dreams and wishes to perfection.

Kant’s hope for the use of reason in the public domain, fighting deception (superstition, ignorance, and so on)—a moral task—and avoiding self-deception—an epistemological task—seems to be an unreasonable one for today’s person. Any person is simultaneously a member of a community and part of a social contract, on the one hand, and part of mass society (Cruz 1998) on the other. In mass society, deception is part of the game.

As Matrix has rightly implied, religion stages the entire drama of salvation: we are finite beings, engulfed ourselves in a state of complete alienation by tinkering with AI, and only a savior can effectively rescue us from this state. But religion is also a human craft, although a very unconscious one, and does not escape the ambiguities of life by resorting only to cult and belief. Even the best intentioned of human saviors can lead us astray again. According to Paul Tillich, within history we can hope for reconciliation (understood here as the opposite of alienation) only in fragment and anticipation (Tillich 1963, 138-141). The same is true for full personhood as we enter into a new century.

What kind of person can we expect from the interaction of virtual creatures, simulations and infobots, that are also in the business of deception? Persons that are enmeshed with their own creations, potential prisoners of their own artifacts; [5] at the same time, in the midst of their deception, persons that have an intuition of their destiny in freedom, love, and full awareness of reality. Affirmation of life does not shun away from death, and the expectation of life everlasting is not the same as planning an indefinite span of time to our own existence. A
full and sound account of this disposition is the best contribution that
teology could give to AI.

REFERENCES

of Washington Press.


Bostrom, Nick. 2001. “Are You Living in a Simulation?” http://simulation-
argument.com.


NOTAS


[4] Playing with “Fate” and “Destiny” is not without a purpose in this argument. Fate points to the insertion of a person in a chain of causality—even contingency is out of one’s control. Depending on artifacts is an easy route to alienate our own responsibility over our
acts—“The system is down” is a common excuse heard nowadays. Destiny points rather to the outcome of free decisions, engaging a vocation and our free will. Without suggesting a duality between both situations, we would like to indicate that “destiny” is the redemption of both technology and religion.

[5] Some people today put too much emphasis on relationality, a word that has acquired a soteriological character almost just by being professed. However, when faced with something of their own make, human beings move too quickly from Narcissism to hate. That is perhaps one of the most striking messages of the movie AI.