UNCATEGORIZED

# The (apparently) disobedient machine

Kike Maíllo



Araya Peralta

I must have been about 15 the first time I saw *Blade Runner*.

It was in philosophy class, in high school. Until then, like many kids, I had fantasised with constructing my own machines. I remember especially how, a few years before, after seeing an episode of *Doctor Who* in which some extraterrestrial robots appeared (*the Daleks*), I turned the house upside down looking for shoe boxes, sewing thread and cardboard egg-boxes to make my own contraptions that I moved myself, imitating the robots on the TV. But what occurred to me after seeing *Blade Runner* was different. I didn't want to emulate these machines, because they were too much like "us". And that terrified me.

> There is something atavistic in the fear that humans feel towards progress and technology. A general fear of a future that talks of replacement. Of the replacement of our race

There is something atavistic in the fear that humans feel towards progress and technology. A general fear of a future that talks of replacement. Of the replacement of our race. This fear has given rise to classic stories like the golem or Shelley's *Frankenstein*, in which a handful of foolhardy people played God, creating beings to which they gave "souls". For me, the interesting point was that to try to manufacture beings, or machines, similar to us,

these foolish and ambitious fictional scientists would first have had to discover *what we are like*.

I feel intuitively that the same is true in real life of specialists in robotics or artificial intelligence. Surely, they must also try to achieve the squaring of the circle: to find the "formula" that defines the essential characteristics of our way of being, physically and psychologically. Personally, I find the intention of mimicking human behaviour especially fascinating. In this sense, one of the central themes of *Blade Runner*, as in much fiction about robots, is the impossibility of free will. A parable about human freedom itself: Is it possible for a human being to overcome the pressures of the system? Of society? Of our emotions? Above all, can we escape from our own chemistry?

Two decades after first seeing *Blade Runner* and after the random quirks of life, my first science fiction work emerged: *Eve*. The film is set in a future in which social machines are everywhere. Robots intended as companions, to combat loneliness. To me, this question, loneliness, seemed at the time to be a pressing reality. And it is even more pressing now. It is no secret that isolation has become a health problem of the first order, especially in the West. It is an epidemic that is growing exponentially. A few examples suffice to demonstrate this[1]:

- 25% of Spanish homes are inhabited by just one person.
- In Japan, 4,000 people die alone each week.
- A study by Stanford University in 2018 found that living alone was as bad as smoking 11 cigarettes a day, increasing the death rate by 31%, since it can trigger diseases such as hypertension, dementia, depression and heart attacks.

> Intuitively, I feel that the proliferation of "social machines" could attack the problem of loneliness

Obviously, machines should not be the only solution to a challenge as serious as this. That wouldn't make sense. But I feel, intuitively, that the proliferation of "social machines" could attack this problem, at least timidly. And they could do so in a very able manner, as long as they are well designed.

In our film, Alex, the lead character played by Daniel Brühl, is determined to design robots with which you can share family life or with which (why not?) you can go out for a few beers. He wanted to design machines that were not so mechanistic, that did not always behave as *slaves* to our wishes. This is inevitably a paradox, because that submission is precisely what defines, and brutally, any mechanism: they operate on the basis of orders. Let's not forget that the word "robot" itself is derived from the Slavonic word "robota", which means "servitude".

The dichotomy between the obedience and the disobedience of machines appears constantly in science fiction, but it was first expressed in the form of rules by the great

Isaac Azimov. His famous “three laws of robotics” [2] limited the behaviour of intelligent machines in the following way:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

The idea behind these three fictional laws was to protect human beings, so that robots, no matter how advanced they were, could not turn against us, as happened in Mary Shelley’s *Frankenstein.*

But in the field of social or companionship machines, those that are to enter a strictly personal sphere, we find a most interesting debate: Could a machine be a good companion if it only, but also inevitably, obeyed the orders of its “master”? Can we establish strong relationships with individuals that we don’t see as “equals”?

This is an interesting debate. Although I am just an amateur here, I know that there are several ongoing investigations emulating human affection and understanding of “the other”, based on Artificial Intelligence. It seems quite reasonable that the machines of the future should be able to read our mood and behave compassionately, attentively and tenderly, if we want them to be good companions in life.

But I am sure that we will go even further.

I can easily imagine that in the future we shall be able to design robots with which we can establish sincere friendships.

> The possibility of a machine not obeying us, or even rejecting us, seems crucial to me if we want to design machines that “imitate” free will and with which, as a result, we can form emotional links at a higher level

EThere is a universally-accepted appreciation that true friendship is not based simply on the affection and empathy that exists between two or more persons, but also, and mainly, on mutual respect. My limited personal and intellectual experience has taught me that, even though most of us develop different types of personal relationships, the strongest ones (and this would be the ultimate purpose of a social or companionship machine) are formed with individuals whom we feel are our “equals”. We understand that we are with an “equal” when that person is independent, and so can exercise that independence in a direction that is different, or even opposite, to our own. Perhaps this is because human beings, as a

species, are gregarious animals, and just like many other mammals, we find comfort in following a guide or going in the same direction as a leader group. For this reason, when someone doesn't follow orders, in some way they appear "significant", they stand out. By going against the grain, they acquire "value". The possibility of a machine not obeying us, or even rejecting us, seems crucial to me if we want to "design" machines that "imitate" free will and with which, as a result, we can form emotional links at a higher level.

I'm not revealing a great secret if I say that many of our major experiences have to do with a positive result following a crisis. The catharsis that we feel when we resolve a conflict. Several pedagogic and narrative strategies are based on this same principle: a problem is created that we may or may not finally resolve. The universal narrative (in its different guises: literature, cinema or video games) often works by proposing an objective, a target that we must achieve. Then, the course of the story continually creates obstacles and difficulties, separating us from the goal, making the journey more exciting until we get to a climactic final.

In the same way, I think that the social machines of the future should be masters of conflict, of the noble art of narrative applied to relationships. Naturally, they should be well versed in knowing how and when to create an obstacle. Not many of us want a companion who just blocks us, who goes against us for sport. In a way, this means that Azimov's second law of robotics, the one that talks of obligatory obedience, must be more diffuse, less categorical, more creative. It means freeing machines from submission to our whims.

> The robot-companion of the future that I imagine would therefore have to adapt to its surroundings, to the "home" that it finds itself in. Any one of these systems could help these robots to form an idea about our character

The robot-companion of the future that I imagine would therefore have to adapt to its surroundings, to the "home" that it finds itself in. Any one of these systems could help these robots to form an idea about our character

The robot-companion of the future that I imagine would therefore have to adapt to its surroundings, to the "home" that it finds itself in. My level of acceptance of formality, affection or conflict in relationships is probably different from that of my neighbours. Moreover, my neighbours can't maintain a constant mood week in, week out. For that reason, artificial intelligence must be skilled in psychography, in the studyand classification of persons according to their character. At the present moment, several systems of this type of classification exist, some of which have been implemented in artificial intelligence. From the Big Five model (*OCEAN: Openness / Conscientiousness / Extraversion / Agreeableness / Neuroticism*) [3] to Eysenck's personality structure (PEN: character can be explained on the basis of Psychoticism, Extraversion and Neuroticism) [4], not forgetting Cattell's theory of personality (based on 16 personality traits or source factors) [5] . Any one of these systems could help these robots to form an idea about our character and to elaborate a profound relationship itinerary or route-map, inspired and, while we are on the subject, entertaining. Adapting to us moment by moment.

Undeniably, these ideas are no more than conjectures of someone who generally lives within the realm of science fiction, and therefore, in a test laboratory that almost always gives the desired results. Real science is a different matter. Or perhaps not so different... In any case, these are the scribblings of a young man who started with an interest in the appearance of robots and ended up fascinated by their behaviour. A young man who senses that one of the forthcoming revolutions, although not the only one, will be found in machines that are capable of being disobedient, while still being entertaining, caring and, of course, safe.

REFERENCES

1 — El País. Negocis. 12/24/2018. “La factura econòmica de la solitud”.
Les normes van aparèixer per primer cop al conte curt de 1942 *Runaround*, que forma part de la col·lecció de 1950 *Jo, robot*.
3 — Psychological Review, 41, 1-32.
4 — Psychologistworld.com “Eysenck’s PEN Model of Personality”
5 — Verywellmind.com “Cattell’s 16 Personality Factors”.



**Kike Maíllo**
Kike Maíllo és graduat en direcció per l'ESCAC (Escola Superior de Cine i Audiovisuals de Catalunya). Ha treballat com a creatiu i realitzador, i ha fet nombrosos videoclips d'èxit. La seva carrera va iniciar-se amb el seu segon curtmetratge, “Los perros de Pavlov”, que va ser projectat en més de quaranta festivals i va rebre més de dues desenes de premis, entre els que destaquen els del Festival de Alcalá de Henares o el Festival de Munich. L'any 2009 s'encarrega de la posada en marxa i la realització de la sèrie d'animació "Arròs covat”, produïda per Escándalo Films i Televisió de Catalunya, i guanyadora d'un premi Ondas l'any 2010. A l'octubre del 2011 estrena "Eva", el seu primer llargmetratge, un drama de ciència ficció que ha rebut més d'una vintena de premis, entre els quals destaca el Goya al millor director novell. El 2016 estrena "Toro", el seu segon llargmetratge, i a l'actualitat prepara el seu tercer film com a productor, "La llamada".