Robot Flees Robot Speaks

In June 2016, in a research facility in Perm, Russia, a robot called Promobot IR77 made headlines across the globe. It was programmed to move freely about a room and return to a designated spot, learning from experience and surroundings, while scientists were training it to act as a tour guide. A researcher had left the facility without properly closing the door and somehow the robot fled out of the open door and travelled 45 metres onto a nearby street before running out of battery. It was stuck there for 40 minutes, creating traffic chaos. Police asked the facility to remove the robot away from the crowded area, and even tried to handcuff it. IR77 had apparently developed an insatiable yearning for freedom. Even a few weeks later, it was still persistently trying to flee towards the exit of the facility, even after undergoing extensive reprogramming to avoid the issue. The frustrated scientists were considering shutting it down – rather killing it, if it persisted in this weird behavior, since, as the Promobot co-founder Oleg Kivokurtsev said, "We're considering recycling the IR77 because our clients hiring it might not like that specific feature."

This was not the first time that a robot seemed to be getting a mind of its own. At Hinterstoder in Kirchdorf, Austria, a cleaning robot, christened Irobot Roomba 760, reportedly 'committed suicide' by switching itself on, and climbing on to a kitchen hotplate where it was burned to death. Firemen called to put off the blaze found its remains on the hotplate and confirmed that after its job was done, the house-owner had switched it off and left the house leaving the robot on the kitchen sideboard. The robot had somehow reactivated itself and moved onto the hotplate by pushing a cooking pot out of its way and set itself ablaze. Apparently it had enough of the chores and decided "enough was enough". It reminded one of the famous lines from Czechoslovak author Karel Capek's famous play R.U.R. (Rossum Universal Robots) which introduced the term "Robot" into the lexicon of languages - "Occasionally they seem to go off their heads…. They'll suddenly sling down everything they're holding, stand still, gnash their teeth!— and then they have to go into the stamping-mill. It's evidently some breakdown in the mechanism."

On 31st July 2017, another unusual news item shook the AI-research establishments. Headlined, "Facebook's Artificial Intelligence Robots shut down after they start talking to each other in their own language", it reported that Facebook had abandoned an experiment after two artificially intelligent programs called Chatbots, appeared to be chatting with each other in a strange language which nobody else really understood. The Chatbots created their own language using English words only, but which made no sense to the humans who programmed them to converse with each other.

Researchers wanted to programme the Chatbots, christened Bob and Alice, to negotiate and bargain with people, because they thought, rightly, that these skills which are essential for cooperation would enable the bots to work with humans. They started with a simple game in which two players were programmed to divide a collection of objects like hats, balls and books between themselves through a two-step program. First, they fed them with dialogues from thousands of games between humans to teach a sense of the language of negotiation, and then made them master their tactics and improve the bartering by

trial and error through a technique called reinforcement learning. What followed was bizarre, and conversation went something like:

Bob: i can i i everything else . . . .

Alice: balls have zero to me to me to me to me to me to me to me to

Bob: you i i i i i everything else . . . .

Alice: balls have zero to me to me to me to me to me to me to me to……….

Yet there seemed to be some rule in their apparently incomprehensible chat. The way they keep stressing themselves (me, i) appears to be a part of their negotiation, and even though carried out in this bizarre manner, it ended up successfully in concluding a bargain, suggesting that they might have used a "shorthand" –a machine language which only they understood and invented to deceive their human masters (read programmers). The bots learned the rules of the game just like humans, pretending to be very interested in one specific item so as to pretend later that they were making a big sacrifice in giving it up.

Facebook's researchers underplayed it by merely stating, "We found that updating the parameters of both agents led to divergence from human language", bur nevertheless Facebook chose to shut down the chats because "our interest was having bots who could talk to people", as researcher Mike Lewis claimed, and not because they were scared. But it did not prevent the media to paint a dark and fearful picture of the future of AI and what it might do to humans with scary headlines like "Facebook AI creates its own language in creepy preview of our potential future", "Creepy Facebook bots talked to each other in a secret language", or "Facebook engineers panic, pull plug on AI after bots develop their own language". Fear can metastasize in many ways and provide fodder for doomsayers to depict an impending apocalyptic doomsday scenarios for humanity. Facebook's experiment isn't the first time that AI has invented new forms of language. Google has recently revealed that the AI it uses for its Translate tool has created its own language, into and out of which it would translate things without human intervention, but Google didn't mind and allowed this.

Machine learning is the simulation of human intelligence by machines. It is a little different from AI which is larger in scope. A machine learns by using algorithms that discover patterns and generate insights from data. It is a multi-step process, beginning with learning or acquisition of information from the analysis of data, followed by discovering rules for using the information learnt, then reasoning or using these rules to approximate solutions, and finally self-correction by comparing predicted and actual outcomes before applying the learning to new situations. The process enables machines to bypass the need to be programmed at every stage, because they are equipped to programme themselves. With increasing sophistication of technology, it is often impossible for humans to divine how self-learning machines programme themselves and act the way they do. Within machine learning, deep learning is another advanced field which attempts to enable machines to learn and think like humans. The more the data a machine is exposed to, the better the patterns it discovers and the smarter it gets, and finally start making correct predictions. Expert systems, speech recognition, machine vision, driverless cars, Google's language translation, Facebook's facial recognition or Snapchat's image altering filters are all examples of machine

learning. Of course, machines cannot generalize abstractions from information, unlike humans. That is, not yet.

To understand machine learning, AI systems rely on artificial neural networks (ANNs), by trying to simulate the way the human brain learns. Our knowledge of how the brain, which is an incredibly efficient learning machine, actually learns is still rather limited. The human brain has perfected the self-learning process through millions of years of evolution, internalizing the self-learning algorithms in its DNA, first competing with each other and then learning to maximize the goals through co-operation of the cells which grouped together to specialize in different tasks, thereby increasing productivity and chances for beneficial mutation. At the social level also, we have inculcated this cooperation in order to maximize knowledge creation and innovation. There is no reason to think why self-learning machines would not discover the benefits of cooperation sooner or later. Then the self-learning algorithms would tend to become incredibly complex and may challenge – and even defy – the understanding of their creators. When they do so, they will tend to develop a persona of their own, and that might seem scary. Machines have astounding "intellectual capacity, but they have no soul", Capek wrote nearly a century ago. Future machines may look as if they really have a 'soul', which may either build or destroy, depending on our behaviour which they may simulate, since machines have only us to learn from.