

Gnirut: The Trouble With Being Born Human In An Autonomous World

Luca Viganò
King's College London
London, UK
luca.vigano@kcl.ac.uk

Diego Sempredoni
King's College London
London, UK
diego.sempredoni@kcl.ac.uk

ABSTRACT

What if we delegated so much to autonomous AI and intelligent machines that They passed a law that forbids humans to carry out a number of professions? We conceive the plot of a new episode of Black Mirror to reflect on what might await us and how we can deal with such a future.

CCS CONCEPTS

• **Computing methodologies** → **Philosophical/theoretical foundations of artificial intelligence**; **Theory of mind**; • **Security and privacy** → *Social aspects of security and privacy*; • **Human-centered computing** → *HCI theory, concepts and models*;

KEYWORDS

Autonomous AI; Reverse Turing Test; Theory of Mind; Deception

ACM Reference Format:

Luca Viganò and Diego Sempredoni. 2018. Gnirut: The Trouble With Being Born Human In An Autonomous World. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3184558.3191612>

1 PROLOGUE: “WHAT DO YOU WANT TO BE WHEN YOU GROW UP?”

EXT. MAIN STREET – DAY.

An eight-year old child (C), one of the child's parents (P) and a woman they just met (W).

W: And what do you want to be when you grow up?

C: Well, I...

W: A philosopher like your father?

C: I... I...

P: Come on, don't be shy. Answer the nice lady.

W: Or perhaps a painter like your mother?

C: I...

P: Come on, speak up! Please, forgive him. He's so shy. Come on!

C: I...

P: This is not ok. Not ok, you understand? The nice lady asked you a question and you should answer.

W: No worries, no worries.

P: Wait till we get home...

W: You'll tell me another time, ok?

C: ... a pilot! I want to be a pilot!

W: Oh, but that's impossible, dear.

C: Why?

W: Because you can't.

C: I can't? Why?

W: It's not just you. We. We can't.

C: Why?

P: That's enough. Don't be a nuisance.

W: Don't be silly, no nuisance at all.

C: Why can't I be a pilot?

P: That's enough, I said!

W: Oh, no, no, please allow me to explain. There is a law.

C: A law?

W: A law that says we are not allowed to be pilots.

C: Why?

W: Because people were afraid.

P: People were afraid of flying, so They made a law.

C: How come people were afraid of flying? If at all, people should be afraid of falling.

W: Ha! That's clever.

P: People were too afraid that their time had come, you understand?

C: But if it had come, why worry? It had come anyway. On a plane or on the ground.

W: A-ha, and what if it's the pilot's time to go?

P: What if the pilot got sick, or decided to commit suicide?

W: What if the pilot made a mistake?

C: I wouldn't make any mistakes.

W: But of course you would. Of course. We all do.

C: I wouldn't get sick or kill myself. They could test me.

W: Oh, They did. They did test us. That's why They passed the law.

P: That's why They passed all those laws.

W: For our good.

P: Only for our good. That's why we have to obey.

C: But...

P: No but! You must learn to know your place.

C: But I want to be a pilot.

W: But you can't dear.

P: That's why They made a law. To protect us.

W: Pilots, doctors, surgeons, judges, construction builders, ... that's not for us anymore.

P: To protect us.

C: Yes, but why? Why can't I be a pilot?

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191612>

We wrote this scene as the set-up of a future Black Mirror episode, in which “They” have passed laws that forbid humans to carry out a number of professions. How could this have happened?

2 FROM THE HUMAN-TO-AI RATIO ...

The current recommendations (and, in some cases, regulations) require a *many-to-1 human-to-robot ratio* (typically between *2-to-1* and *5-to-1*) for tasks that are particularly security-sensitive and/or safety-sensitive as they involve the protection of diverse critical assets such as data, systems, infrastructures and, ultimately, human life. Examples of such tasks are those carried out in military operations, such as drone strikes or rescue, demining or bomb-disposal operations — studies carried out in the early 2000’s in the context of military operations indicated the then current state of practice for aerial systems to be many-to-1 (as witnessed by the Global Hawk and Predator crews) and that ground robots are more effective with a 2-to-1 ratio [4, 5, 19]. Other, more “civilian”, examples are tasks to be carried out in critical infrastructures or in specific automation environments and situations where human life or expensive assets might be in danger, such as remote surgeries or tasks carried out in power plants, in autonomous and assisted transportation systems, underwater [22], in space [21], or in harsh conditions, say due to excessive cold, heat or radiation [12].

As stated by Murphy and Burke in 2010, the human-robot-interaction literature has (often) ignored safety, making the assumption that low human-to-robot ratios are desirable per se, and

often pursuing an arbitrary goal of 1-to-many based on expected advances in vehicle autonomy. [19]

Nowadays, this 1-to-many goal is not so arbitrary anymore: thanks to advances in robot autonomy, we appear to be not too far removed from a ratio of 1 human to 5 robots (*1-to-5*) for all kinds of tasks, while ensuring safety and security (and privacy) of team, bystanders, and robots, as well as logical efficiency at the same time.

But why stop here? Solutions are currently being sought for full automation of systems by removing the human and transforming the operating space into a highly instrumented and controlled area for robots, as has been common in car manufacturing or even in places like Amazon warehouses (with carefully segregated spaces). The present reality is that in many dynamic and challenging environments we still cannot remove the human as there will be times an automated system is operating at the boundaries of its competence and might require human intervention. However, it is not utopian to imagine a near future reality in which the human-to-robot ratio has been reduced to *0-to-5*, fulfilling the vision of a fully autonomous world in which robots carry out unmanned tasks. Actually, many tasks don’t necessarily need a physical entity such as an arm or wheels to be carried out, but could be accomplished directly by some form of artificial intelligence, so we can speak more generally of a *0-to-5 human-to-AI ratio*.

But why stop here? Why not invert the ratio, and consider the case in which 1 human is supervised by one or more AIs? In 2017, Rachel Botsman carried out a small but enlightening experiment that involved her 3-year old daughter Grace and Amazon’s Alexa, which she summarized in a New York Times Sunday Review Opinion by observing that in the last few years, we have been (often passive) witnesses to

a profound shift in our relationship with technology. For generations, our trust in it has gone no further than feeling confident the machine or mechanism will do what it’s supposed or expected to do, nothing more, nothing less. We trust a washing machine to clean our clothes or an A.T.M. to dispense money, but we don’t expect to form a relationship with them or call them by name. [1]

Now we can call them by name: Amazon’s Alexa, Apple’s Siri, Google’s Google Assistant and Microsoft’s Cortana are all intelligent personal assistants that recognize natural voice (in many different languages) without the requirement for keyboard input and support a wide range of user commands, ranging from answering questions to providing real-time information (such as news, weather forecast or traffic information), from making phone calls to compiling to-do lists, from setting alarms to playing music or audiobooks or streaming podcasts. Alexa can also control several smart devices using itself as a home automation system. It’s the two-faced, sirenlike, beauty of the *Internet of Things*, in which we silently enter an agreement to hand over the keys to our houses, to our appliances, to our data, to large parts of our lives:

Today, we’re no longer trusting machines just to do something, but to decide what to do and when to do it. The next generation will grow up in an age where it’s normal to be surrounded by autonomous agents, with or without cute names. The Alexas of the world will make a raft of decisions for my kids and others like them as they proceed through life — everything from whether to have mac and cheese or a green bowl for dinner to the perfect gift for a friend’s birthday to what to do to improve their mood or energy and even advice on whom they should date. In time, the question for them won’t be, “Should we trust robots?” but “Do we trust them too much?” [1]

3 ... TO THE AI-TO-HUMAN RATIO

Our thesis in this paper is that we will not stop here. We will silently but willingly go for the full inversion of the ratio, from human-to-AI to AI-to-human, and hand over the vast majority of our choices and decisions, of our data and personal information, of our privacy, the vast majority of the different facets and aspects of our lives. We won’t stop at the 1-to-1 AI-to-human ratio exemplified by little Grace’s naive, and somewhat cute, trust in Alexa:

With some trepidation, I watched my daughter gaily hand her decisions over. “Alexa, what should I do today?” Grace asked in her singsong voice on Day 3. It wasn’t long before she was trusting her with the big choices. “Alexa, what should I wear today? My pink or my sparkly dress?” [1]

We will go “all in”. The AI-to-human ratio will soon be *1-to-some* (for instance, what if all of Botsman’s family members, including the adults, had developed such a close relationship with Alexa?) and, ultimately, *some-to-many*.

This will happen gradually, with the change differentials initially too small to be noticed, or without us really realizing that the sequence of differentials over a longer period of time have caused a momentous change. This is somewhat reminiscent of the legendary

social experiment involving 5 monkeys, a ladder and a banana in which, after a number of cold showers whenever one of the monkeys tried to reach for the banana, and after all 5 original monkeys have been stepwise replaced with 5 new monkeys, what is left is a group of 5 monkeys that, without ever having received a cold shower, continue to beat up any monkey who attempts to climb the ladder. The change has occurred gradually, one monkey at a time, but in the end a new status quo has been reached whose justification has long been forgotten: if it was possible to ask the monkeys why they beat up on all those who attempted to climb the ladder, their most likely answer would be “because that’s the way it’s always been done around here.”¹ Will we end up being the monkeys in our own experiment towards a fully autonomous world?

There are already plenty of articles, news feeds and books explaining the future of employment and how intelligent machines will replace humans in many jobs in the context of the “industrial revolution 4.0” [3, 13, 15, 16, 29, 30, 33, 34]. It is not just that many boring or strenuous jobs (such as taxi driver [32], factory worker [9], etc.) are under threat of automation and will result in technological unemployment — this is often referred to as the *economic singularity* [7] that will bring us one step closer to the *technological singularity* in which ordinary humans will someday be overtaken by artificially intelligent machines or cognitively enhanced biological intelligence [25]. In a laudable attempt to safeguard human safety, we will also likely soon legislate that some jobs should be carried out only by intelligent machines. This won’t be limited to what is already happening (as we remarked above, robots have, at least partially, taken over dangerous operations such as demining, bomb-disposal, or nuclear meltdown inspections). With the advances in the automation of decision making, planning and autonomy, this will also encompass professions in which human nature and subjectivity might slow down reaction time or adversely affect the end-result.

Think, for example, about the story of the “Miracle on the Hudson”: the US Airways Flight 1549, in the climbout after takeoff from New York City’s LaGuardia Airport on January 15, 2009, struck a flock of Canada geese and consequently lost all engine power. Unable to reach any airport, the pilot Chesley “Sully” Sullenberger and his co-pilot Jeffrey Skiles glided the plane to a ditching in the Hudson River off Midtown Manhattan. All 155 people aboard were rescued by nearby boats and there were only few serious injuries. As masterfully portrayed in the movie *Sully* directed by Clint Eastwood from a screenplay by Todd Komarnicki (based on the book *Highest Duty* by Sullenberger and Jeffrey Zaslow), Sullenberger and Skiles were subject to an investigation by the National Transportation Safety Board, which initially seemed to conclude that they could have safely landed in one of the nearby airports instead of attempting a one-in-a-million landing in the river. A number of pilots were asked to carry out human-piloted simulations of the accident which all showed that it was actually possible to land

safely in one of the nearby airports. This is Tom Hank’s/Sully’s reply

...you’re still not taking into account the human factor. These pilots are not reacting like human beings. Like people who are experiencing this for the first time. ... You have allowed no time for analysis and decision making. And with these sims, you have taken all the humanity out of the cockpit. [11]

Reaction-decision time is then set at thirty-five seconds and the simulations are run again... and all result in a crash. With that delay, attempting to land in the Hudson was indeed the only option. An intelligent machine, on the other hand, might have required a much lower reaction-decision time and safely made it to a runway.

To overcome, and prevent, similar problems, the rise in AI will alter also legal frameworks. As observed by Bowcott in his analysis of a report by the International Bar Association [3],

Among the professions deemed most likely to disappear are accountants, court clerks and ‘desk officers at fiscal authorities’. ... Even some lawyers risk becoming unemployed. “An intelligent algorithm went through the European Court of Human Rights’ decisions and found patterns in the text,” the report records. “Having learned from these cases, the algorithm was able to predict the outcome of other cases with 79% accuracy ... According to a study conducted by [the auditing firm] Deloitte, 100,000 jobs in the English legal sector will be automated in the next 20 years.”

But why stop here? Why not hand over the actual legislative power to intelligent machines, which, after all, have the ability to take better and more consistent decisions? “They” might then legislate that the list of professions that should not be accessible to humans ought be extended to include also all kinds of drivers and pilots (no more Sully, then), doctors, surgeons, accountants, lawyers, judges, lawmakers, soldiers, ... you get the drift.

What will be left for us humans to do? Well, first of all, we will hope that the system doesn’t go *Skynet* on us², but then we will also be faced with the problem of what to do with our lives. Will we end up fat and lazy, although potentially happy, like the humans in *Wall-E* [27]? Will we all be rich and bored, and have lives of leisure while the machines are taking the sweat [34]? Will we succumb to pessimism like Cioran, whose best-known work [8] inspired the subtitle of this paper?

- *What do you do from morning to night?*
- *I endure myself.*

So, what will be left for us to do? Actually, Cioran himself (unknowingly) suggests an answer:

A zoologist who observed gorillas in their native habitat was amazed by the uniformity of their life and their vast idleness. Hours and hours without doing anything.

¹We wrote “legendary” as, apparently, the experiment was never really carried out but was rather inspired in part by the experiments of G.R. Stephenson, found in “Cultural acquisition of a specific learned response among rhesus monkeys”, as well as experiments with chimpanzees conducted by Wolfgang Kohler in the 1920s. Over the years, it was pieced together to form the urban legend as it now stands, see <http://www.wisdompill.com/2014/05/28/the-famous-social-experiment-5-monkeys-a-ladder/> as well as Jeff Bridges’/President Jackson Evans’ excellent rendition in the movie *The Contender*, written and directed by Rod Lurie [18].

²*Skynet* is the fictional AI that serves as the main antagonist in the movies of the *Terminator* franchise [6]. Skynet came to the logical conclusion that all of humanity would attempt to destroy it. In order to continue fulfilling its programming mandates of “safeguarding the world” and to defend itself against humanity, Skynet launched a series of nuclear attacks killing over three billion people and gathered a slave labor force from surviving humans.

Was boredom unknown to them? This is indeed a question raised by a human, a busy ape. ... Man alone, in nature, is incapable of enduring monotony, man alone wants something to happen at all costs — something, anything... [8]

A common distinguishing trait of us human beings with respect to animals, monkeys, apes and primates, and possibly with respect to present and future AI, is our desire for more, our insatiable curiosity.³ We desire what we don't have, and in some cases what we can't have, and we strategize and make long-term plans to achieve what we desire. In the scene at the beginning of this paper, the child C wants to become a pilot. C doesn't care that the laws forbid it, C wants to fly. C asks "why?", C is the monkey that challenges the status quo, the way things are done around here.

What can C do then? If it really wishes to become a pilot, C will muster up a plan. C will be prepared to lie to achieve what it wants. Perhaps inspired by Yentl, the Jewish girl who disguises herself as a boy to enter religious training in Isaac Bashevis Singer's short story and play *Yentl, the Yeshiva Boy* [26] as well as in Barbra Streisand's eponymous movie [28] (but see also similar plotlines in *Tootsie* [23], *Albert Nobbs* [14] and many more movies, books and plays going back to Greek mythology and actually even earlier than that).

4 CAN LIFE IMITATE AI?

C could attempt to disguise itself as an AI to enter the "pilot academy" of this futuristic, but not utopian, fully autonomous world. Will C be able to pass the (to-be-developed) *Gnirut Test*, a fully reverse Turing test that will involve an AI judge and a human subject which attempts to appear artificial? The underlying presumption here is that an AI subject will always be judged artificial, and a human is said to "pass the Gnirut test" if it is also judged artificial.

The idea of a deceptive machine is fundamental to AI research and has been present in the AI literature since Turing introduced the *Imitation Game*, so why not consider a deceptive human in our plot? Will C be able to deceive the AI judge? Difficult, probably impossible, but we wish to believe that at least in our Black Mirror episode C will stand a chance.⁴

Advances in *Deception Theory* might provide C with a range of techniques to fool the judge, especially when coupled with advances in the *Theory of Mind*, which investigates the creation of intelligent machines that are able to model other agents' minds. Maybe then C will be able not only to pass as an AI but also to lower its reaction-decision time to that of an intelligent machine in case of an accident to the plane it is piloting. C will hopefully be able to leverage on the (to-be-developed) *Theory of Artificial Mind* to plot its disguise in what could be an entertaining and, as usual, thought-provoking new episode of Black Mirror aptly titled *Gnirut*.

5 EPILOGUE

Like most episodes of Black Mirror, our own Gnirut would also be discomfiting and scary. It would ask a lot of "What ifs?" and act

as a cautionary tale, putting us in front of that black mirror that we can find

on every wall, on every desk, in the palm of every hand: the cold, shiny screen of a TV, a monitor, a smartphone. [31]

Should we be scared by the reflection that we see? Of course, we should, at least a little. Should we retire Alexa to the closet like Botsman tells us that she did after the experiment with her daughter [1]? Our answer is no, we should not.

Rather than halting technological progress, we should make sure to accompany it, bringing together, in a multi-disciplinary supervision effort, teams of experts including AI experts, informaticians, mathematical and social scientists, psychologists, lawmakers and more. We should be careful who we trust [2] and, above all, we should not lose our faith in expertise [20]. Only then will the "magnificent and progressive fate of the human race" [17] stand a chance to materialize in its full glory.

ACKNOWLEDGMENTS

We thank Daniele Magazzeni for many interesting discussions (and for allowing us to paraphrase a couple of his still unpublished sentences).

REFERENCES

- [1] Rachel Botsman. 2017. Co-Parenting with Alexa. *The New York Times* Oct. 10 (2017).
- [2] Rachel Botsman. 2017. *Who Can You Trust? (How Technology Brought Us Together — and Why It Could Drive Us Apart)*. Penguin Random House.
- [3] Owen Bowcott. 2017. Rise of robotics will upend laws and lead to human job quotas, study says. *The Guardian* Apr. 4 (2017). <https://www.theguardian.com/technology/2017/apr/04/innovation-in-ai-could-see-governments-introduce-human-quotas-study-says>
- [4] Jennifer L. Burke, Robin R. Murphy, Michael D. Coovert, and Dawn L. Riddle. 2004. Moonlight in Miami: Field Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise. *Human-Computer Interaction* 19, 1-2 (2004), 85–116.
- [5] Jennifer L. Burke, Robin R. Murphy, Erika Rogers, Vladimir J. Lumelsky, and Jean Scholtz. 2004. Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Trans. Systems, Man, and Cybernetics, Part C* 34, 2 (2004), 103–112.
- [6] James Cameron (created by). 1984–. The Terminator Franchise. (1984–). <http://www.imdb.com/title/tt0088247/>.
- [7] Calum Chace. 2016. *The Economic Singularity: Artificial intelligence and the death of capitalism*. Three Cs.
- [8] Emil M. Cioran. 1973. *De l'inconvenient d'être né* ("The Trouble With Being Born"). Gallimard.
- [9] Steve Davis. 2018. Soft robots could be the factory workers of the future. *The Conversation* Jan. 10 (2018). <http://theconversation.com/soft-robots-could-be-the-factory-workers-of-the-future-89885>
- [10] Philip K. Dick. 1968. *Do Androids Dream of Electric Sheep?* Doubleday.
- [11] Clint Eastwood (directed by). 2016. *Sully*. (2016). <http://www.imdb.com/title/tt3263904/>.
- [12] Martin Fackler. 2017. Six Years After Fukushima, Robots Finally Find Reactors' Melted Uranium Fuel. *The New York Times* Nov. 19 (2017).
- [13] Martin Ford. 2016. *The Rise of the Robots: Technology and the Threat of Mass Unemployment*. Oneworld Publications.
- [14] Rodrigo Garcia (directed by). 2011. *Albert Nobbs*. (2011). http://www.imdb.com/title/tt1602098/?ref=nm_flmg_act_21.
- [15] Jerry Kaplan. 2015. *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*. Yale University Press.
- [16] Gerd Leonhard. 2016. *Technology vs. Humanity: The coming clash between man and machine*. Fast Future Publishing.
- [17] Giacomo Leopardi. 1845. Wild Broom (XXXIV). In *The Canti*. OUP Usa.
- [18] Rod Lurie (written & directed by). 2000. *The Contender*. (2000). <http://www.imdb.com/title/tt0208874/>.
- [19] Robin R. Murphy and Jennifer L. Burke. 2010. The Safe Human-Robot Ratio. (2010), 31–49 pages.

³As Albert Einstein said, "I have no special talents, I am only passionately curious."

⁴Or maybe the AI judge will fall in love with C and help it to go unnoticed like Rick Deckard does with Rachel in *Blade Runner* [24] (although the original novella by Philip K. Dick [10] actually followed a slightly different storyline).

- [20] Tom Nichols. 2017. *The Death of Expertise: The Campaign Against Established Knowledge and Why it Matters*. OUP Usa.
- [21] Laura Niles. 2015. First Humanoid Robot In Space Receives NASA Government Invention of the Year. https://www.nasa.gov/mission_pages/station/research/news/invention_of_the_year Jun. 17 (2015).
- [22] Narcís Palomeras, Arnau Carrera, Natàlia Hurtós, George C. Karras, Charalampos P. Bechlioulis, Michael Cashmore, Daniele Magazzeni, Derek Long, Maria Fox, Kostas J. Kyriakopoulos, Petar Kormushev, Joaquim Salvi, and Marc Carreras. 2016. Toward persistent autonomous intervention in a subsea panel. *Auton. Robots* 40, 7 (2016), 1279–1306.
- [23] Sydney Pollack (directed by). 1982. *Tootsie*. (1982). http://www.imdb.com/title/tt0084805/?ref_=fn_al_tt_1.
- [24] Ridley Scott (directed by). 1982. *Blade Runner*. (1982). <http://www.imdb.com/title/tt0083658/>.
- [25] Murray Shanahan. 2015. *The Technological Singularity*. MIT Press.
- [26] Isaac Bashevis Singer. 1983. *Yentl, the Yeshiva Boy*.
- [27] Andrew Stanton (directed by). 2008. *Wall-E*. (2008). <http://www.imdb.com/title/tt0910970/>.
- [28] Barbra Streisand (directed by). 1983. *Yentl*. (1983). <http://www.imdb.com/title/tt0086619/>.
- [29] Richard Susskind and Daniel Susskind. 2017. *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford University Press.
- [30] Max Tegmark. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Allen Lane.
- [31] The Guardian — Gadgets. 2011. Charlie Brooker: the dark side of our gadget addiction. *The Guardian* Dec. 1 (2011). <https://www.theguardian.com/technology/2011/dec/01/charlie-brooker-dark-side-gadget-addiction-black-mirror>
- [32] The UK Autodrive project 2018. (2018). <http://www.ukautodrive.com>
- [33] Jane Wakefield. 2015. Intelligent Machines: The jobs robots will steal first. *BBC News — Technology* Sep. 14 (2015). <http://www.bbc.co.uk/news/technology-33327659>
- [34] Toby Walsh. 2017. Will robots bring about the end of work? *The Guardian* Oct. 1 (2017). <https://www.theguardian.com/science/political-science/2017/oct/01/will-robots-bring-about-the-end-of-work>