

Šum #8



Gene Kogan, »People« from *Experiments with Deep Generator Networks*, 2016,
vir: <http://genekogan.com/>.

Selfish Machines

Part 1

Tadej Vindiš

Will robots inherit the Earth? That is the question in the title of the article published in the early nineties by Marvin L. Minsky, an American cognitive scientist and one of the key figures in early development of artificial intelligence. Minsky portrays a dystopian finite future of the human corpus that will only depend on the possibility of its extendability; “to lengthen our lives, and improve our minds, in the future we will need to change our bodies and brains.”¹ The fatality of biological constraints of the human body, the terrible weakness of its inevitable process of decay, of the exposure to all sorts of illnesses, its genetic determinism and the sheer fact of expiration, concerns Minsky. Not to mention our brain that, as some recent studies have estimated, have the capacity of 2.5 petabytes (or a thousand million gigabytes) and might run out of space; we need to think about reproducibility and reach of the human brain, which is unable to compete with computable alternatives. The question for Minsky is one of time (or lack thereof); nanotechnology should thus enable

1 MINSKY, Marvin L., *Will Robots Inherit the Earth?*, Scientific American, Oct, 1994, available from: <<https://web.media.mit.edu/~minsky/papers/sciam.inherit.html>> (last accessed on 4 September 2017).

the construction of “replacement bodies and brains that won’t be constrained to work at the crawling pace of real time.”²

The augmentation of human corpus is central to the construction of innovation narratives of advanced capitalism, perpetuating a linear vision of the future human beyond the organic. The defining paradigms of performance evaluation have been reduced to speed, accuracy and application, composing an affirmative framework of mathematical objectivity, which lies at the core of technological mainstream. Machine-augmented human bodies are therefore seen as an inevitable and welcome alteration, and “once we will entirely replace our brains using nanotechnology, we will escape the limitations of biology and choose the length of our lives, and consequently our abilities as well.”³ In fact, technological augmentations designed to reach beyond the limitations of the human corpus have always been part of human evolution. Weapons, eyeglasses, wheel-chairs, hearing aids, microscopes, telescopes, implants, genetic engineering, regenerative and reproductive medicine, neural interfaces, memory-enhancing drugs, and so on are all aimed to bypass the biological determinism of the skin-bags we are. Furthermore, cybernetics, intelligent machines and artificial intelligence, it has been argued, are the final frontiers of human organic evolution, which will lead to the emancipation of the mind from the body. In such permutation of the final Cartesian mind/body split, enabled by the creation of the metaphysical foundation that is “technicity”, “the human [will not only] transcend itself (its matter) while retaining what Kurzweil argues are the ‘best human traits’—intelligence, rationality, presence, that is, ipseity—but we will have reached our ontotheological end. We will have become gods.”⁴ Today, events in computational systems happen incredibly fast, faster than those in our brain cells, and with the uprise of intelligent machines that are capable of learning, such systems can independently process infinite amounts of data, drive cars, trade, distribute and manage resources, and solve games with high levels of unpredictability, such as the

2 Ibid.

3 Ibid.

4 ANDERSON, Nicole, “The Limits and Possibilities of Animal-Human Relations”, in: Jami Weinstein, Claire Colebrook (ed.), *Posthumous Life: Theorizing Beyond the Post-human*, New York, 2017, p. 18–19.

well-publicised victory of DeepMind’s AI system AlphaGo. With the rise of these intelligent systems, it is no longer possible to talk about human augmentation and its potential transcendence without considering radicalisation of language of communication, or communicability, caused by tensions within the mediatic regime.⁵

Radicalisation in practice involves adaptation of logics of computation into the political, social and cultural domains, which distributed participation of human and nonhuman agents in complicated, often unplanned and usually untraced ways. These new distributions are intertwined in a numerical rhythm of analogical communication and informational exchange; the aggregated modulation of things, dissecting the organic subject into bits and bytes of interdependent informational clusters, constant inputs and outputs, assemblages of algorithmic processing and (re)active data analysis, stretched across social interactions and systems of governance. For Vilém Flusser, technological revolution led to the formation of a new consciousness with the creation of new codes; the transformation of our being-in-the-world such that it becomes increasingly more difficult to distinguish where digital life ends and physical life begins. The imposition of information technology has a direct impact on the agenda and content of human history, which becomes reflexive rather than indexical: “In the same manner that the alphabet was directed against pictograms, so digital codes currently direct themselves against letters, to overtake them. In the same manner that a form of thinking based on writing opposed itself to magic and myth (pictorial thinking), so a new form of thinking based on digital codes directs itself against procedural, ‘progressive’ ideologies, to replace them with structural, systems-based, cybernetic modes of thought. [...] This can no longer be thought dialectically, but rather through Kuhn’s notion of ‘paradigm’: no more a synthesis of opposites, but rather a sudden, almost incomprehensible leap from one level to another.”⁶ The mediatic regime

5 PARISI, Luciana, “Mediality with Ends: When Machines Start to Think”, in: Joke Brouwer, Lars Spuybroek, Sjoerd van Tuinen (ed.), *The War of Appearances: Transparency, Opacity, Radiance*, Rotterdam, 2016, p. 45.

6 FLUSSER, Vilém, *Die Schrift: Hat Schreiben Zukunft?* (Frankfurt am Main: Fischer Taschenbuch Verlag, 1992), p. 129–32, quoted in: Andreas Ströhl (ed.), *Vilém Flusser Writings*, Minneapolis, 2002, p. xiii.

accelerated dislocation of the subject from the object, reducing communication to a choice between discrete units without the need for those units to be strictly defined to one another. The units of human as well as those of nonhuman, all interacting through the global binary communications framework. Reality can thus be understood as a field or set of relations amongst things (users) within the mediatic regime in which a human is only one of the participants in a network of interactions and infinite possibilities.⁷

The rapidly increasing accumulation of data now drives the political and economic interests towards obsessive implementation of computational processes on a global scale. Governments, schools, banks, hospitals, courts, and factories are all participating in the use of seemingly esoteric computational structures, left without a choice but to join in in the absorbance of the “pervasiveness of computational techniques as they become increasingly coextensive with processes of production, consumption, and distribution displayed in logistics, finance, architecture, medicine, urban planning, infographics, advertising, dating, gaming, publishing, and all kinds of creative expression (music, graphics, dance, and so on).”⁸ “Don’t be evil”, Google’s well-known maxim, as discussed by Matthew Fuller and Andrew Goffey in *Evil Media*, resonates a broader capitalist ambition to eliminate, to work against or to smooth over the ambiguity of computation, its unpredictability and contingency—its entropy—closing out the system for the purpose of enabling software policing justified by economic thinking and wrapped in moral terms; “the maxim and its rather more bellicose sibling, the ultimatum to be with us or to be against us, both portend, with cartoon simplicity, the coming of a world in which every byte of information and every tap on a screen, every waking thought and action, are expunged of the deviant and devious propensities of contrary forms of vital energy.”⁹ The intensified social state of

7 STRÖHL, Andreas, “Introduction”, in: Andreas Ströhl (ed.), *Vilém Flusser Writings*, Minneapolis, 2002, p. xiii–xiv.

8 TERRANOVA, Tiziana, “Red Stack Attack! Algorithms, Capital and the Automation of the Common”, in: Ryan Bishop, Kristoffer Gansing, Jussi Parikka, Elvia Wilk (ed.), *across & beyond—A transmediale Reader on Post-digital Practices*, Concepts and Institutions, Berlin, 2016, p. 202–203.

9 FULLER, Matthew and Andrew Goffey, *Evil Media*, Cambridge/London, 2012, p. 3.

anticipation generated by the capital and manifesting in a hallucinogenic state of Flusser's *universe of technical images*; a total immersion into an addictive horror of self-purposing fantasies and affectual pleasures of dematerialised being, now absorbed into a computational infrastructure, only programmed towards excessive prediction and immensely hungry for constant and continuous interaction. The vast black boxed and sham zone of "emptying out the subject onto the surfaces of mediatic language [which] involves causality becoming redundant, defining a neoliberal condition of living as suspension in the paradoxical state of purposeless purposiveness, i.e. a political condition in which means have no aims and aims have no causes."¹⁰

Flusser portrays this condition of fully automated communicational infrastructure of networked capital as a nightmare—an impending telematic society structured as a Platonic utopia: "[A] society in which slaves (economy) are robots, artists (politics) are automatic intelligences, in which everyone lives for theory (all are philosophers, kings), nourished and supplied with criticizable models by robots and artificial intelligences."¹¹ A state of cerebral paradox, which is built on the belief "that human reason separated from nature can produce objective knowledge about the latter",¹² and perpetuated via the construction of an affectual and selfish blindness of capital against the nature of the cybernetic infrastructure lying underneath. Is it perhaps necessary, faced with the emergence of intelligent systems, to turn attention back to the machines? Instead of obsessively mining within the immanence of data to amplify our own interests, could observation of the evolving capacity of computation—of its strangeness and its rigidity—reveal possibilities for alternative modes of thinking?

Just recently the world has been driven into sci-fi paranoia by two negotiating bots which started to talk to each other in a strange and computationally altered English language at the

10 PARISI, Luciana, "Mediality with Ends: When Machines Start to Think", in: Joke Brouwer, Lars Spuybroek, Sjoerd van Tuinen (ed.), *The War of Appearances: Transparency, Opacity, Radiancy*, Rotterdam, 2016, p. 48.

11 FLUSSER, Vilém, *Into the Universe of Technical Images*, Minneapolis, 2011, p. 143.

12 WILSON, Neil, "Punching Out the Enlightenment: A Discussion of Peter Sloterdijk's *Kritik der zynischen Vernunft*", in: *New German Critique: Special Issue on the Critiques of the Enlightenment*, No. 41, Spring–Summer, 1987, p. 60.

Šum #8

Facebook Artificial Intelligence Research (FAIR) lab. The most reported quote from the negotiation between the two bots proceeded like this: a bot named Bob said to a bot named Alice “I can i i everything else,” to which Alice responded with “balls have zero to me to,” Bob replying with “you i everything else,” and Alice following with “balls have a ball to me to me”.

FAIR uses a method of dialogue rollouts in which chatbots that engage in a negotiation can predict in which direction of a conversation might take and can choose to steer away from potentially frustrating exchanges in favour of successful ones. In a situation of a multi-issue bargaining, two computer agents are presented with the same collection of items and are instructed to negotiate between them in order to split the items. The model of bots learned the tactics of negotiation from an initial dataset of 5808 dialogues between humans on negotiation tasks, but it also competed with itself via a constructed system of reinforcement training rewarding the model whenever a good outcome was achieved. However, during such training, in which one agent is improving its parameters from conversations with another agent, the researchers at FAIR used a fixed supervised model imitating humans and thus restricted the model to using a humanlike language. The reason for the use of a fixed supervised model was the discovery that if both agents were updating the parameters at the same time and in conversations with each other, they diverted from human language and developed their own language of negotiating.¹³ The situation is particularly interesting because of the effect it had outside of FAIR’s lab. The alliterated English language—its machinic oddness—generated panic. News reports were shouting plots of singularity, showing clips of Terminator movies and picturing screams of the Facebook engineers as they were unplugging the AI overlord. The reality, however, is somehow more—or less, depending how you look at it—dull. Agents started to invent code words for themselves in order to achieve a higher score. The repetition of words

13 LEWIS, Mike, Denis Yarats, Yann N. Dauphin, Devi Parikh, Dhruv Batra, *Deal or no deal? Training AI bots to negotiate*, 2017, available from: <https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/> (last accessed on 4 September 2017).

connected to a specific function literally represents the desire of the agents to receive more of that function. However, how come that even after the explanation of computational dullness, it is difficult to eliminate the feeling of estrangement towards suspicious bots?

Most bots have no human-readable output or “voice” because they are running within convolutional neural networks, exchanging and processing information, waiting for inputs, sending outputs, testing ports. Only recently have we been more generally introduced to sophisticated speaking bots, such as Siri, Alexa and Cortana, or so-called Conversational User Interfaces; intelligent systems that have the capacity to apply their operational condition to vocabularies of a human language. Even though speech simulation software has been present for a relatively long time, it has been established within the field of machine learning that it is better to train a system to recognize speech rather than attempt to programme it in entirety.¹⁴ Benjamin Bratton argues that there is a certain level of dynamics between “cognitive technologies of *programming* (a specific medium of inscription) and the cognitive technologies of interface *interaction* (reading and responding to available mediated inscriptions, and piloting them according to user intention).”¹⁵ In the first, it is possible to automate a certain process of instrumental cognition in order to achieve a specific task, which is the baseline for software as a tool, allowing us to encode a specific function which is available for execution at any time and as many times wished so in the future. However, in the latter, “the user interests and intentions are extended from the interface’s image of those available programs and how it frames their context and allows for their interpretation. In [the first] cognition is encoded *into* the interface and in the other cognition is augmented by *using* the interface.”¹⁶ In the situation of the machine which has the capacity to learn through interaction we are faced

14 MITCHELL, Tom M., *The Discipline of Machine Learning*, Pittsburgh, 2006, available from: <<https://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>> (last accessed on 4 September 2017).

15 BRATTON, Benjamin, “Can the Bot Speak? The Paranoid Voice in Conversational UI”, in: Ryan Bishop, Kristoffer Gansing, Jussi Parikka, Elvia Wilk (ed.), *across & beyond—A transmediale Reader on Post-digital Practices, Concepts and Institutions*, Berlin, 2016, p. 308.

16 *Ibid.*, p. 309.

with the process of continuous expenditure of the intelligent system, which is a result of the operational focus of the system's adaptation through learning. Instead of humans getting closer to understanding functionalities of such systems by giving pre-mediated instructions in the form of coding and thus retaining control over instrumental reason at hand, the conversational interface demands that the user learns how to negotiate with the limitations of the system, by which the system programmes itself. The system, therefore, cannot be understood as a form of an automated set of instructions, but as automated cognition, which adapts and evolves based on situations it encounters.

It is linguistic tension in communicability of the mediatic regime that drives the need for the development of technology towards severe humanisation. By steering the process of interaction towards the use of a human language, the system itself needs to imitate the use of such language in order to interact successfully. However, if Apple's Siri sounds "too inhumanly robotic, then the intimacy necessary to feel safe talking to her when you wake up disoriented at 4 a.m. will not be established, but if she's too idiosyncratic in her replies then users may be uncomfortable asking her to help with very serious queries about banking, healthcare, or such. It's a matter of balance. That she has something to say about 'God' gives her a certain depth, but that what she has to say is only a diplomatic one-liner assures us that we are still in charge."¹⁷ In the case of chatting between the two Facebook bots the fear-inducing dialogue was a malfunction in the process of imitation, which disrupted the process of transition from the experience of computing with premeditated inscription of programming to the abstracted illumination of the use of language, which in turn translates the process of algorithmic training. What surfaced was a kind of manifestation of the logics of computation underneath the imitation, which revealed the way algorithms construct their comprehension of themselves via a relational network—the capacity of an intelligent system to optimize based on its operational and parametric volitions. The permuted language of the bots materialized a deeply ingrained horror of alienation; an apparent sudden separation and loss of

¹⁷ Ibid., p. 316.

control; the unknown which necessarily had to become a threat, even though the operation of the machine has not changed.

Luciana Parisi argues that the ingression of computation into culture produced the world in which “algorithms are no longer or are not simply instructions to be performed, but have become performing entities: actualities that select, evaluate, transform and produce data.”¹⁸ Instead of being confronted with the smooth operation of algorithmic regulation of networked capital, which relies on the potential of optimizing any kind of processes indefinitely, “the entropic tendency of data to increase in size, and thus to become random, drives infinite amounts of *information* to interfere with and to reprogram algorithmic procedures.”¹⁹ Instead of thinking about algorithms as a set of operational instructions, the computation mediates towards indeterminacy, producing alien rules, and “the cybernetic network constantly reminds us that the task of critique is limited by the helpless cognitive impossibility of knowing.”²⁰ How do we evaluate governance, if we disregard the impossibility of understanding the multiplicity of data? The vision of a possibility for a preconfigured social infrastructure—of the future without work—is derived from probabilities within the immanence of Big Data economy and the resulting relentless paranoid need for scientific development of things-to-understand, to diagnose, alter and thus optimise—intentions often distorted by ideology. *Modus operandi* for the construction of fixed-algorithms is to focus on “a hyperfast connecting of the dots rather than tacking how they got there in the first place.”²¹ These are predictions which do not account for any future to come, but a future reduced to a precaution and pulled into the present, leading to the apparent technical possibility where, by following a set of procedures, all “developments, tendencies, curves can be projected from the present forward, and these projections can be manipulated. Margins of error can be calculated as closely as one likes. But

18 PARISI, Luciana, *Contagious Architecture: Computation, Aesthetics, and Space*, London, 2013, p. ix.

19 Ibid.

20 PARISI, Luciana, “Mediality with Ends: When Machines Start to Think”, in: Joke Brouwer, Lars Spuybroek, Sjoerd van Tuinen (ed.), *The War of Appearances: Transparency, Opacity, Radiance*, Rotterdam, 2016, p. 52.

21 Ibid.

such projections show the results of calculations, not what is coming. There is no future. Computerized prediction devours the future in the interest of avoiding catastrophe.”²²

With the emergence of intelligent systems of machines-that-can-think or interfaces-that-programme-in-turn, the initial information grows as those systems adapt to the environments in which they operate. If I follow Parisi’s discussion of Gregory Chaitin’s algorithmic theory, “the computation involves the maximally unknowable probabilities.”²³ Information within such systems is not predetermined or fixed or immanent, it does not only act and respond, but it evolves, too; “the augmentation of entropy becomes productive of new axiomatic truths that cannot be predicted in advance”²⁴ and “it is [therefore] possible that patternless information emerging from within this evolution of data quantities points to a dynamics internal to algorithmic automation.”²⁵ In the case of Facebook’s bots, the modifications in the use of English language could have brought us closer to understanding intelligent machines had the performance evaluation of FAIR’s lab not been restricted solely to human capability of direct communication with the bots in order to transform such a system into yet another mere “assistant” to accommodate “a general tendency in linguistically sophisticated mammals to defer interest in anything other than the reverberation of their own vocalizations.”²⁶

“Selfish Machines” may be a provocative phrase, but it provokes in order to direct attention towards disruptions within the imitation game of instrumentality. It is a mark of the paradoxical anthropomorphisation of the machines, which allows for computational ambiguity to (re)surface as a central focus of discourse. Instead of developing methodologies for machine humanisation—to make them as human-like and politically correct

22 FLUSSER, Vilém, *Into the Universe of Technical Images*, Minneapolis, 2011, p. 159.

23 PARISI, Luciana, “Mediality with Ends: When Machines Start to Think”, in: Joke Brouwer, Lars Spuybroek, Sjoerd van Tuinen (ed.), *The War of Appearances: Transparency, Opacity, Radiance*, Rotterdam, 2016, p. 56.

24 Ibid., p. 57.

25 Ibid.

26 BRATTON, Benjamin, “Can the Bot Speak? The Paranoid Voice in Conversational UI”, in: Ryan Bishop, Kristoffer Gansing, Jussi Parikka, Elvia Wilk (ed.), *across & beyond—A transmediale Reader on Post-digital Practices, Concepts and Institutions*, Berlin, 2016, p. 316.

Šum #8

as possible—why not develop machines in a manner that makes the way they function more comprehensible? What can strangeness of images that are surfacing from the depths of Deep Generator Networks tell us about the machines themselves? And in turn, what can such images, or the permuted language of Facebook bots, or strange noises coming from computational glitches, tell us about aesthetics, representation, or language? Under the algorithmic influence, these are markedly different from smooth-functioning programmes and narratives of digital interfaces; they deal with revealing possibilities rather than fitting a niche. Could an examination which is less concerned with defining its ends explain more about the condition of the human within indeterminacy of the mediatic regime? Or further, how do we relate to intelligent systems which do not necessarily share the same volitions as we do? How are machines not human? How are they also not not-human? Questions like these make it necessary to demand and start working on the formation of a critical practice which works within mediatic regimes at a time when we are facing the rise of automated thought; when computation in practice is showing itself to be more than a domain of a scientific research, but of a wider artistic and cultural discourse. With this in mind, where is the culture of machines headed? Can machines have their own interests? How could such interests be examined? Can machines be “selfish” as well?

Šum #8

References

- ANDERSON, Nicole, "The Limits and Possibilities of Animal-Human Relations", in: Jami Weinstein, Claire Colebrook (ed.), *Posthumous Life: Theorizing Beyond the Posthuman*, New York, 2017.
- BRATTON, Benjamin, "Can the Bot Speak? The Paranoid Voice in Conversational UI", in: Ryan Bishop, Kristoffer Gansing, Jussi Parikka, Elvia Wilk (ed.), *across & beyond—A transmediale Reader on Post-digital Practices, Concepts and Institutions*, Berlin, 2016.
- COMBES, Muriel, *Gilbert Simondon and the Philosophy of the Transindividual*, Cambridge/London, 2013.
- FLUSSER, Vilém, *Into the Universe of Technical Images*, Minneapolis, 2011.
- FULLER, Matthew and Andrew Goffey, *Evil Media*, Cambridge/London, 2012.
- LEWIS, Mike, Denis Yarats, Yann N. Dauphin, Devi Parikh, Dhruv Batra, *Deal or no deal? Training AI bots to negotiate*, 2017, available from: <<https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/>> (last accessed on 4 September 2017).
- MINSKY, Marvin L., *Will Robots Inherit the Earth?*, Scientific American, Oct, 1994, available from: <<https://web.media.mit.edu/~minsky/papers/sciam.inherit.html>> (last accessed on 4 September 2017).
- MITCHELL, Tom M., *The Discipline of Machine Learning*, Pittsburgh, 2006, available from: <<https://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>> (last accessed on 4 September 2017).
- PARISI, Luciana, *Contagious Architecture: Computation, Aesthetics, and Space*, London, 2013.
- PARISI, Luciana, "Mediality with Ends: When Machines Start to Think", in: Joke Brouwer, Lars Spuybroek, Sjoerd van Tuinen (ed.), *The War of Appearances: Transparency, Opacity, Radiance*, Rotterdam, 2016.
- STRÖHL, Andreas (ed.), *Vilém Flusser Writings*, Minneapolis, 2002.
- TERRANOVA, Tiziana, "Red Stack Attack! Algorithms, Capital and the Automation of the Common", in: Ryan Bishop, Kristoffer Gansing, Jussi Parikka, Elvia Wilk (ed.), *across & beyond—A transmediale Reader on Post-digital Practices, Concepts and Institutions*, Berlin, 2016.
- WILSON, Neil, "Punching Out the Enlightenment: A Discussion of Peter Sloterdijk's Kritik der zynischen Vernunft", in: *New German Critique: Special Issue on the Critiques of the Enlightenment*, No. 41, Spring–Summer, 1987.