Bachelor thesis

Anthropomorphization of Artificial Intelligence

Tomáš Kopeček

Brno 2015

Masaryk University
Department of Sociology
Supervisor: doc. PhDr. Csaba Szaló, Ph. D.

Čestné prohlášení

Prohlašuji, že jsem tuto bakalářskou/diplomovou práci vypracoval samostatně s použitím pramenů a literatury uvedené v bibliografii.

V Brně dne

Anotace

Téma umělé inteligence se objevuje ve vědeckém a čím dál více i ve veřejném prostoru. Od počátku je provázeno otázkami, které překračují běžnou vědeckou debatu ve smyslu Kuhnovy normální vědy. Je stvoření člověkupodobné umělé inteligence morální? Nahradí lidstvo? Je existence takové bytosti vůbec možná, když nemá duši? Bude mít volební právo? Většina této diskuze je ovlivňována mnoha jevy, ať již je to pozice vědy, náboženská stanoviska, či politické cíle a v neposlední řadě materialita inženýrské práce. Na druhé straně dopady nejsou jen uvnitř vědeckého světa, ale ovlivňují i své okolí a celou společnost. A to například ve formě zákonů či formulaci společenských obav nebo „mocenského“ souboje vědy a náboženství.

Tématu se okrajově věnují jak práce sociologické, tak filozofické či právnické. V této práci se pokouším podat ucelený přehled sociologicky zajímavých aspektů tohoto pole. Naopak se nevěnuji tématům ryze technicistním. Metodologicky se jedná o horizontální přehledovou studii, která se opírá o různé filozoficko-sociologické přístupy od Bourdieho polí, přes Latourovo pojetí materiality po Habermasovu komunikační teorii. Práce identifikuje čtyři základní narativy diskurzu umělé inteligence (vědecký, náboženský, sekuritizační, umělecký) pomocí nichž budou ilustrována jednotlivá témata.

Klíčová slova

Umělá inteligence, diskurz, filozofie vědy, náboženství, materialita

Počet znaků

87 836

Annotation

Artificial intelligence is becoming topic in scientific and more and more in public space as well. From the beginning it is followed by questions which outreach ordinary scientific discussion in terms of Kuhn's normal science. Is creating of human-like artificial intelligence moral? Will it replace mankind? Is existence of such being even possible if it has no soul? Should it have right to vote? Most of this discussion is influenced by many effects from position of science, religious stances or political goals and not last in the line materiality of engineering work. On the second side it has no impact only in scientific world, but also in its neighbourhood and whole society in forms of law or in formulation of public fears or "power" struggle of science and religion.

Topic is slightly researched by sociological, philosophical and juridical literature. I will try to show overview of sociologically relevant aspects of the field. In opposite I will not deal with topics which are clearly technical. Methodologically it will be horizontal survey which is rooted in different philosophical-sociological approaches from Bourdieu's theory of fields through Latour's view on materiality to Habermas' theory of communication. Text identifies four basic narratives of AI discourse (scientific, religious, securitization and artistic) with whose will be illustrating particular topics.

Keywords

Artificial intelligence, discourse, philosophy of science, religion, materiality

Character count

87 836

Contents

1 Introduction

> Everyone takes the limits of his own vision for the limits of the world.
>
> *(Arthur Schopenhauer[1])*

> "You may read *Principia Mathematica* without finding discourse of souls, spirits, cogitation, or what-have-you," said Isaac. "It is about planets, forces, gravity, and geometry. I do not address, and certainly do not pretend to solve, the riddles that so confounded Monsieur Descartes. Why should we attempt to frame hypotheses about such matters?"
> "Because if you do not, Sir Isaac, others, of less brilliance, will; and they will frame the wrong ones,"
>
> *(Neal Stephenson[2])*

1.1 Why

Three ugly words in title of something what should not be a research work but a bachelor thesis. That needs some kind of explanation. Let's cast naïve question: "Why study something like Artificial Intelligence (AI) (something which does not exist) from sociological stance?" and similarly naïve answer follows: "Because it matters (and sociology deals with things which 'don't exist'." Now we have to see it more broadly.

At first comes AI concept. It is a thin stream which springs in twentieth century (but it has quite ancient roots as we will see later) and is slowly gaining momentum to these days. Now AI concept is something which almost everyone is more or less familiar with.

From early fifties when Alan Turing comes with his famous Imitation game [Turing, 1950] through all buzz of excited sixties when Marvin Minsky becomes a new apostle of AI religion until these days when there is no month without some 'progressive' news from the field.

[1] [Schopenhauer, 2008]

[2] [Stephenson, 2005]

Story of AI is not only about technical expertise, but also encapsulates many others and a lot of them are sociologically relevant. Permanent struggle of science and religion stages another battlefield here which ironically shows more religious zeal on side of (positivist) science. Politics also become involved as there is a few forces which has to emancipate themselves through permeating law systems with AI-related concepts ending even with *Charter of Fundamental Rights and Freedoms* AI amendment proposals.

There is still one word left — anthropomorphization. There is a reason, in fact more of them, why most work related to AI is thinking about it as about a human-compatible creature. We will see that it is quite problematic — from point of 'creature' to 'human-compatible' concepts. There are some reasons why experts expects it to be alike, but there are also not a few problems which such thinking raises.

1.2 What

Let's dive into main lines of thinking about AI and look to which social and cultural aspects played and still plays important roles here. Whole AI sector even if it looks like ultra-rationalist and firmly rooted in modern positivism is full of contradictions. We will go so far that in extreme position we will describe it as a religious project which builds on Judeo-Christian roots touched by hermetism and gnosticism. AI story is becoming part of a much larger cultural stream from Enkidu through hermetic homunculi, golems and entering modern era via Mary Shelley's 'Modern Prometheus' — Frankenstein's monster [Shelley, 2012]. We will see how is AI discursively constructed, how this construct backfires to its authors and how still non-existent (strong) AI is changing existing world already.

Whole discourse on AI is quite complex and for purposes of this work I have to tear out just few stripes. To keep (not so false) complex system idea I am going to not choose only one line but few narratives whose should serve as a base for limited horizontal overview of the field. Thus, we will go through scientific and securitization stories.[3] In the end it should draw picture how such detached topic (as AI for most of population is) can change and changes our everyday world.

1.3 From Where

Let's call this section 'Section of clarifications'. First clarification is to say, from where I am coming — I am not born-and-raised sociologist. My epistemo-onto-

[3] Stories of Religion, Alchemy and Art were omitted, as I have not enough space to go through them in meaningful way.

logical origins lies near to the enemy. In fact, I am coming directly from the eye of the storm — theoretical informatics, more specifically from mathematical linguistics branch directly building on Wittgenstein, Chomsky and mathematical intuitionism[4] speaking with 'Leibniz-like' philosophical language. On the other side I have to deal with quite different languages like theological philosophy of Teilhard de Chardin [de Chardin, 1966].

We are both speaking about areas somehow related to AI problematic, but for many people each of these languages is incomprehensible. I wanted quickly illustrate how wide is the stream of thinking which I have to grasp. But instead of it I see that I am falling to one of the modern traps — building dichotomy. Mathematics vs. philosophy, 2008 vs. 1949 So I will try to fix it a bit. Let's add some new axes and let these two extremes slowly melt into the amalgam of (post)modernity.

1.4 Through What

From the point of philosophy I will be mainly operating on one side with Leibniz, Kant and Nietzsche — in topics related to ethics and epistemology. Reasons for that will be given in section on materiality. On the other side there will be Wittgenstein, Habermas [Habermas and Cronin, 2008, Habermas, 2003, Habermas and Outhwaite, 1996], Alexander [Alexander, 1989] useful in area of meaning construction. To the mix will be added some authors from the science of philosophy arena which will be touching the topic from more angles as it is the only reflective apparatus which could be used to reflect on scientific dimension of AI.

In sociology I am quite influenced by Latour's notion of networks [Latour, 2007, Latour and Porter, 1991], nevertheless it will be used in somehow *naïve* way, as this is not research work and there are only tiny resources dedicated to AI topic. It will be set in contrast with Alexander's critic expanded by Hodder [Hodder, 2012] mostly concerning the materiality issues. Of course, we can't evade classical sociologists as Marx, Durkheim and Weber mainly through their critics or followers. Guide to the religious aspects is Geraci [Geraci, 2008, Geraci, 2014, Geraci, 2007, Geraci, 2011]

Of course there will be a lot of references to insider scholars and practitioners dealing with AI and related fields such as Alan Turing [Turing, 1950], Marvin

[4] I am not going to dive into the deepness of math theory. It should be sufficient to know, that there *is* some reflection in math about what truth means. Intuitionism deals with concept of truth as with something what can be constructed and proved. Simply, the Truth is something what can be proved — Plato was wrong. If you are interested in the topic, start with [Brouwer and Heyting, 1980].

Minsky [Minsky and Papert, 1972], Hans Moravec [Moravec, 1988, Moravec, 1991] or Ray Kurzweil [Kurzweil, 2014, Kurzweil, 2001, Kurzweil, 2005].

There is no definitive methodology I will be using. It could be some loss and on the other side some highlight of the work. So, if I have noted that I will add few more axes to distorted image I will be adding also sociological and philosophical approaches which should allow us to see more holistic picture.

Such mix of schools will provide, I believe, some insight into the field with wider scope than choosing one limited approach which would require some speculations and would be more appropriate for empiric research such is not intended part of this bachelor's thesis. From the same reason is not any specific method of inquiry used.

2 Story of Science

> The chance of the quantum theoretician is not the ethical freedom of the Augustinian.
>
> *(Ray Kurzweil[1])*

> For lack of better terms, we call ourselves sociologists, historians, economists, political scientists, philosophers or anthropologists. But to these venerable disciplinary labels we always add a qualifier: 'of science and technology'. 'Science studies', as Angloamericans call it, or 'science, technology and society'.
>
> *(Bruno Latour[2])*

Long story short. Science and in this case hard science is minefield of epistomological and ontological misunderstandings. Similar dangers also lies on the sociology side. We will slowly navigate through it and also probably hit some hidden mines.

Narrative is probably most-known in public discourse. It is quite logic (even if word 'logic' will be soon problematized) as it is related to origin of actors which are involved in AI development. In first place this narrative is 'natural' for them and not in last it provides benefits. From the legitimization by science or possibility to use scientist status to utter new 'facts'.

2.1 Scientific Version of History

Let's dive into history of AI as initial framing. As this is scientific story, I ignore more mystical roots and to demarcate some starting point I deliberately choose Gottfried Wilhelm Leibniz. Some authors [Pratt, 1987, Churchland, 2001] puts him as grandfather of AI even if it sounds quite exaggerating. Umberto Eco [Eco, 1995] takes different approach and stays more with what Leibniz really formulated and thrived for — which is universal philosophical language *characteristica universalis* designed to describe everything in the universe. This idea is coming

[1] [Kurzweil, 2005]

[2] [Latour and Porter, 1991]

in the second half of seventeenth century. What is new with Leibniz approach is 'scientific' mindset. He is leaving searching mode of original God language and moving to 'modern' way of constructive science. His vague design of philosophical language has heavily influenced even Kurt Gödel [Rescher, 2011]. Gödel himself is second big foundation stone of modern computation[3]. On the technical side there is Blaise Pascal, Charles Babbage and Ada Lovelace (famous *Enchantress of Numbers*) on their road to mechanical computer.[4]

Alan Turing is finally the person which has brought this philosophical, mathematical and engineering streams of thinking to its climax [Muggleton, 2014] and started the AI revolution. In year 1950 he publishes fundamental text with which contemporary scientific world still has not settled in satisfactory way. In *Computing Machinery and Intelligence* [Turing, 1950] he creates famous *Imitation Game* where he proposes 'test' (which is immediately named after him) which should verify if computer (machine, more generally) can think.

Noam Chomsky's *Syntactic Structures* [Chomsky, 2002] are going to print in 1957. Following Wittgenstein's linguistic turn and merging it with formal logic, Chomsky once more opens door for computers to metaphysics. His essential concept of language classes is inherently connected to problems of computability and computation theory of brain modelling. Engineers got one more bullet to their crucial argument that brain *is* machine. When it got linked with Turing machine limits [Turing, 1936] and Church-Turing thesis [Kleene, 1952] it based almost unbreakable foundations[5]. Research in physiology and biology in the same time adds to the theory of AI. It is probably not an accident that year before publication is first *Dartmouth Project*[6] run, where the term 'Artificial Intelligence' is coined (To what AI is and how it became term of itself is covered in *Discursive Construction of AI (2.2)*). This is the point where AI is definitively connected with computers and computational theory of mind.

[3] It would be worth to mention at least famous Gödel's first incompleteness theorem which has immense impact on computing and related philosophies. Theorem states: "that in any consistent formal system F within which a certain amount of arithmetic can be carried out, there are statements of the language of F which can neither be proved nor disproved in F" [Raatikainen, 2015].

[4] While Babbage was probably firm on his claim, that his machine can't think, we can find in texts related to *Analytical Engine* pattern of 'starting' anthropomorphization. Even stating that "machine is not a thinking being" [Menabrea, 1842] is the point which have to be taken seriously in our quest.

[5] Oversimplified these all different concepts from mathematical constructivism says that what is computable is identical on some level — languages, mathematical functions and what is ideal computer able to compute. It is often used to argue that human brain is on exactly same level and thus there is some limit of epistemology inherently included (I strip the embodiment/cognition problem here).

[6] Dartmouth Summer Research Project on Artifical Intelligence

Pragmatism, as one of the post-positivist ways out, allows opening of full scale of differently educated theories, which further allows building model of human behaviour, brain and human (or 'human-like') manifests. Idea of so-called 'Rule-based reasoning' as an inference model for human behaviour is born. It builds on same thesis as *homo economicus*. Man is driven by rational rules. It is emerging in the time when Skinner's radical behaviourism [Skinner, 1976] is going to be one of the leading schools of USA's psychology science. This paradigm evolves with psychology through Dennett's *Computational Theory of Mind* [Dennett, 1992] until contemporary trends in *Computational Ethics* [Allen et al., 2000, Torrance, 2008, Leuenberger, 2014] and finally *Model-based Reasoning* as convolution with neuroscience [Thagard, 2010, Blank, 2013, Jones, 2003]. Long-time paradigm shift is from constructing AIs from scratch to design of learning machines which can evolve on their own [Cristianini, 2014, Velik, 2010].

2.1.1 Philosophy

The philosophy approaches which are utilized by AI scientists, could be helpful to understand what drives and what also limits them. Philosophy in AI is divided to few quite distinct areas. First one is approach to science itself. What is the science, what is its purpose and how reflectively they look to it. Second (and definitely being most explored) field is ethics. And the last one is relationship between science and lived world.

In population and especially in academics there is present idea, that technology is not a field where theory of science is reflected well. If it is true, false or irrelevant is not of our concern now. I will look to points where sociologists and philosophers looks upon AI science and how science builds its own philosophical foundations.

There is a tradition from Plato and Aristotle, through Leibniz and Newton to Russell which is slowly taking mathematics to its knees and forcing it to reflect upon itself. Especially Whitehead's and Russell's *Principia Mathematica* [Whitehead and Russell, 1963] (of course pointing to Newton's *Philosophiæ Naturalis Principia Mathematica* [Newton, 1723]) has shown problems which are hidden in the base of math in the modern light. In the early twentieth century schools divided according to epistemological and ontological explanations to three basic branches: formalism, intuitionism and logicism. Mainly discussions of epistemological qualities directly influenced AI research. Floridi mentions that in year 1964 'philosophy of AI had already produced more than a thousand articles" [Floridi, 2004, p. 566] which is stunning compared to the fact that in that time computers were able to solve limited set of word mathematical problems [Bobrow, 1964] which is quite far from what we imagine under word AI.

Mission

If we look to almost random paper, which is not purely technical, we will see some similar features. Typically, it is self-confidence that what science is doing here is important and even 'the mission' of mankind. Creation of AI is inevitable. "The Singularity denotes an event that will take place in the material world, the inevitable next step in the evolutionary process that started with biological evolution and has extended through human-directed technological evolution." [Kurzweil, 2005][7] Background could be different — here Kurzweil reasons with evolution, otherwise it could be simple eternal value of knowledge like here: "Artificial life is foremost a scientific rather than an engineering endeavor." [Bedau et al., 2000] — ignore engineers, do science no matter what.

Ethics

There is a lot of open questions regarding ethics. We investigate only few here and some more later in *Story of Securitization (3)* which is definitely more concerned with it.

Questions being asked here are about what makes AI living, creature, human or superhuman. They are important in ethical part simply because they define how we will behave to and with AIs. Typical question which haunts the field is question if AIs are genuine or if they just 'simulate' [Bickhard, 2014]. Searle's Chinese room dilemma redefined this typical western dichotomy, where we can see Alexander's binary cultural codes. It is not important here if AI is compatible or indistinguishable from human. Important is what's its ontology [Levine, 2014]. Another dichotomy is put if there is some personhood or not [Laukyte, 2014]. Binary division is also seen if AI needs to have emotions [Megill, 2014]. These few examples shows that general tendency is to grasp some specific concepts where almost all of them are taken from (philosophical) human traits and as such are not definable and try to find if they are applicable to AIs. These concepts can be defined relationally as human-only and as such is futile to try to map them to AIs. It mostly creates only false contradictions as there is no objective position from which it can be evaluated.

Final comment about this topic is reflection of these anhropomorphic terms via Coeckelbergh: "Verbeek's turn to what things *do* remedies this problem, but his conclusion that we therefore should talk about 'moral' artefacts or the 'morality of things' goes at the cost of diluting the anthropocentric meaning we like to give

[7] *Technological Singularity* is crucial term in strong-AI and transhumanism program. The point-of-no-return when humankind can no longer estimate future as it is defined by self-improving AIs themselves.

to the terms 'moral' and 'morality'." [Coeckelbergh, 2009, p. 183][8] We see here, that whole topic of ethical things is quite problematic in binary view. There are two ways out — to stop thinking in dichotomy and adopt non-discriminating approaches, which is sometimes proposed, mostly inspired by Latour [Adam, 2008] or throw away this type of morality at all and stays on human-only level.

Philosophy strikes back

"The new technologies make a public discourse on the right understanding of cultural forms of life in general an urgent matter. And philosophers no longer have any good reasons for leaving such a dispute to biologists and engineers intoxicated by science fiction." [Habermas, 2003, p. 15] As we see here, Habermas has quite clear view on the problem, that ethical and philosophical decisions are being made by 'unqualified' engineers. Many philosophers are disturbed that tunnelled view of engineers should be contemplating about future of mankind. It is disputable if these reactions are relevant and if philosophy is reflective enough upon itself. There is a lot of opinions which says that opposite is true. When we are talking about Habermas we can point e. g. Latour's reaction to him. In [Latour and Porter, 1991, p. 60] he makes an attack based on idea, that Habermas intention is to defend modernity. According to Latour he denies postmodernity's right to existence (he categorizes him as 'pre-postmodern'). In such situation is hard to believe that philosophy as a field has some 'superior' rights to speak about AI ethics. We also see examples (especially Searle) which more defends their place in the postmodern science than to use rational arguments. In case of Habermas it is quite ridiculous compared with his theory of communicative action. Of course, only at first sight. We see, what he tries to communicate here 'philosophy is more special than engineering'.

Nevertheless, Habermas slowly leads us to religiosity, with which philosophy is very closely related. "The scientistic belief in a science which will one day not only supplement, but replace the self-understanding of actors as persons by an objectivating self-description is not science, but bad philosophy." [Habermas, 2003, p. 108] It is obviously another attack on incompetent engineers. If we ignore the tone and get base message, we've to agree. It is belief in science. It is not rationally defensible position. Habermas further explains that science is mostly performed as a religion. It is aligned with late Berger's diversion from secularization thesis. Society is no more secular. In this postsecular society still stays a lot of endemic places which are not aware of it. One of these is (western) science itself. Ironically, science which has proposed concept of postsecularism is disturbed by another postmodern attack on its primacy. Most of science (in this case hard-science) tend to ignore this interpretation and stays in secular phase, which is from outside-view

[8] italics in original

interpreted as a religious institution [Evans, 2014, Knight and Murphy, 2010]. In such paradigm makes sense to communicate with science but definitely not to adhere to its ontological meanings.

Another example of fear from mathematical-naturalistic hegemony is shown in Floridi's paper [Floridi, 2004]. Informational theory of cognition and human behaviour and its concepts are in his view so powerful, that they ultimately dodge almost every criticism. "Informational concepts are so powerful that, given the right level of abstraction (LoA) (Floridi and Sanders 2004), anything can be presented as an information system, from a building to a volcano, from a forest to a dinner, from a brain to a company, and any process can be simulated informationally — heating, flying, and knitting." [Floridi, 2004, p. 566] He concludes that in such case this theory lies in the realm of dogma. Such science is no more Popper's science for Floridi. It is something what is pragmatic engineering but not contestable science. They have to leave that paradigm which on the one side proclaims Gödel's paradox and in the same moment does not want to measure itself by it.

Why I am elaborating on these issues is my attempt to show that independently on real topic of discussion, AI creates arena for fight of two scientific fields. In Bourdieu's [Bourdieu, 1984] sense we see on one side hard-science with its questionable 'objective truth'[9] and on the other side philosophy which pulls itself out of science to higher-ground of patronizing referee position 'above science'. Fight is staged via standard 'scientific' means such as papers or conferences and also out of the field, legitimizing itself and fight for public.

2.1.2 Materiality

"The political task starts up again, at a new cost. It has been necessary to modify the fabric of our collectives from top to bottom in order to absorb the citizen of the eighteenth century and the worker of the nineteenth. We shall have to transform ourselves just as thoroughly in order to make room, today, for the nonhumans created by science and technology." [Latour and Porter, 1991, p. 136] This Latour's idea is one of the central problems of previously mentioned *Computation Ethics.* Ideas on which this concept builds are not uninteresting. We can see how process of theory choosing is driven by a) hegemonic (informatics) discourse (more in *Discursive Construction of AI (2.2)*) and b) technical possibilities. Especially

[9] There is whole field dealing with how this 'truth' is established and how the field itself is defending its position against sociological introspection. One for all — Rosental [Rosental, 2003] shows how even concept of 'logic' is not only methodological tool, but also "privileged object that enables exploration of the material and social forms of intellectual work, including the building of credibility". Its elaborated uses (which are studied only with big investment [Latour, 2007]) can also be viewed as active tool of defence.

second aspect is interesting from the point-of-view of sociology and even more from latourian perspective. We will see how these (basic) questions are practically without any freedom determined by technology which dictates its own rules. View on if what means ‘ethical agent’ is basal for such discussion.

To speak with language of the field, scientists distinguishes between *ethical-impact agents*, *implicit ethical agents*, *explicit ethical agents* and *full ethical agents* (like in [Tonkens, 2009, DeBaets, 2014]).[10] Agents would lead us to Latour’s concept of agency. In fact a lot of AI ethicists are acquainted with it and are using it in some way [Coeckelbergh, 2009]. The problem of agency concept is similar to one which is once and again hit by Latour’s disciples and critics — it sounds too anthropomorphic. It no more matters that it is not[11], the problem is that it is often perceived as such. In this concrete case the first two types of agency (ethical-impact and implicit) are connected more to things and latter two with humans (AIs, creatures, . . .).

While ethical-impact agent is something what has no ‘real’ agency, it can be used in ethically-relevant way. Knife has no ethical impact until it is used to kill. Implicit ethical agent has some possibility to influence reality in such way. Airbag in car is designed to serve ethical purpose and it can be activated in right moment. Technically it is designed to be ‘moral’ agent without possibility to choose immoral option. Explicit ethical agent has options to do morally relevant actions in both directions. Autonomous military drone can choose about target or not to shoot at all [Sparrow, 2007, Hallevy, 2013]. Full ethical agent has furthermore all other options which explicit ethical agent lacks. Simply, nowadays only full ethical agent is human.

Now back to previous claim — we ‘naturally’ feel that the latter two are ‘more human’. They describe some ‘creature’ which has options to choose relevant action. Both of them have ‘free will’ in some way. This feeling is supported by human-like word ‘agent’. This discourse brings us more to anthropomorphic representation than before. “Importantly, new artificial actors do not enter the stage from nowhere. We design them. We can give them the mask, the appearance we want. Since we love ourselves as humans, it is very likely that we will give them our own mask. Or the mask of the animals that remind us of our own, human infants.” [Coeckelbergh, 2009] says Coeckelbergh and it also becomes one of problems in the field.

Generally we can see here (and whole ethics of AI is) tendency to interpret such complex machines in terms related to social world. Sharkey quotes psychological study [Sharkey and Sharkey, 2006] which studied people evaluating other peo-

[10] There is also criticism to this [de Beer, 2012] and other approaches [Schiaffonati, 2003].

[11] As e. g. [Yampolskiy and Fox, 2013] notes, there most possible AIs according tu current state-of-the-art will be of non-anthropomorphic design.

ple (and in which case they were more positive if evaluated person was present) compared to people evaluating computers (in which case they were more positive if computer was present). Simply people treated computers like persons in this case. This has to fall in section about materiality as these 'things' directly change way how we think and behave only by their mere presence. "Indeed we are often quite willing to ascribe intentionality to much less human-like things e. g. 'My car doesn't want to start this morning. My computer is thinking about it.'" [Adam, 2008, p. 152] Such approach leads to problems that human traits are ascribed to machines and in the field of ethics it means (as easiest way) that human-targeted ethic (as Kant's is) is used on them. Some authors of course noted this problem and Adam, which is quoted here, proposes: "The combination of the arguments of Latour on the attribution of agency, even moral agency to things, Dennett's 'as if' intentionality, Magnani's moral mediators and Collins' and Kusch's delegation of action all lend support to IE as an ethics where things have as much place as humans." [Adam, 2008, p. 154] It is reflected by Latour's program: "So long as humanism is constructed through contrast with the object that has been abandoned to epistemology, neither human nor the nonhuman can be understood." [Latour and Porter, 1991, p. 136]. Adam goes this way, while most of the authors simply ignores it and continues with anthropocentric approaches.

If ethicists have intention to create AMA (autonomous moral agents) as they call their holy grail, they are coming to trap what 'ethical' means. If it has to be implemented in machine they have these options: a) design ethical model and implement it b) let ethical behaviour emerge from AI itself. Mostly from securitization perspective option *b* is out of the game as there is no imagined way how to prevent potential damage[12] [Davis, 2015]. So, only option *a* is relevant. This is the time when ethics hits the wall. "Free will has to be 'technologically modelled' first, in robot/AI design or in imagination, before we can fully work out the philosophical difficulties." [Coeckelbergh, 2009, p. 184] It is one of the problems. AI needs free will 'defined' and 'implemented'. Only after that there is a possibility to implement ethical behaviour. There are tons of proposals how to do that and every each of them is limited by technical possibilities. It is not by chance, that AI geeks are into Kant. It is simply too tempting to have global categorical imperative in AI. It seems to be simplest to program.

Now taking bridge to Kant. Andersons states, that "Deontic logic's formaliza-

[12] Of course except for AI zealots. In [Verdoux, 2011] is paraphrased Bostrom in way, that it makes no sense to design ethical behaviour for AI as it is better equipped than human to do so. Which is in this case interesting. Because I can hypothetize that not even materiality itself but also only potential vision of materiality (some future AI) has direct impact on society. Here we can see, that it removes shackles of responsibility from AI designers. According to Bostrom they shouldn't even think about what is and what is not ethically right, because there will be something better (I feel urge here to say 'God-like') what will outperform them.

tion of the notions of obligation, permission, and related concepts, make it a prime candidate as a basis for machine ethics." [Anderson and Anderson, 2007, p. 8] and further says that "one of the central issues in machine ethics is trust and 'mechanized formal proofs are perhaps the single most effective tool at our disposal for establishing trust.'" We can easily see here, that parameter for choosing ethical theory or principles is not based on anything else than on some plausibility (use something which has legitimate status) and it has to be algorithmically easily described. When we set such rules, there is not many systems which satisfy them. We already limit ethical systems almost to only deontic ones and it seems that only Bentham's utilitarianism [Anderson and Anderson, 2007] and Kant's categorical imperative [Tonkens, 2009, Allen et al., 2000] are favourites. These systems are not so well thought out for AI and have internal contradictions, but they are still top candidates because of their compatibility with programming.[13] Other types like *virtue ethics* have much more technical problems, that they are not considered at all or only as a supplement to one of the previous [Wallach et al., 2007].

There is of course a lot of another material limitations and questions of related ethics, such as memory issues [Vargas et al., 2011], biologically-derived ethics [Barandiaran and Egbert, 2014], human comfort [Cole et al., 2012] or actual real-world problems as reappearance of 'trolley problems'[14] with autonomous cars [Bonnefon et al., 2015]. Whole field of materiality-impacted theory is also linked to embodiment and problems of AI with or without body [MacCormack, 2012, Kaur, 2013, Seaman and Rossler, 2008].

2.1.3 Rules of a Field

From Bourdieu's view [Bourdieu, 1996, Bourdieu, 1984], we've to count on field of science. How science and especially positivist science is done make very large impact to how AI is constructed. Let's make take an example from Velik:

> In 1969, for instance, M. Minsky and S. Papert published their famous work on Perceptrons [84] criticizing computational models of the nervous system, which showed "with unanswerable mathematical arguments that such models were incapable of doing certain important computations" [70]. This killed most research in neural computing for the following 15 years.

[13] They still have problems when it comes to algorithmization. "Given the tremendous implications of failure, the system must avoid not only bugs in its construction, but also bugs introduced even after the design is complete, whether via a random mutation caused by deficiencies in hardware, or via a natural event such as a short circuit modifying some component of the system." [Yampolskiy and Fox, 2013, p. 222] This makes basic requirement of deontic system — that agent will behave according to rules without exceptions technically impossible.

[14] Family of ethical problems, when actor is faced with problem to choose who has to be killed/harmed in situation when there is no win-win solution.

> Much later, S. Paper admitted in an interview [70]: “Yes, there was some hostility behind the research reported in Perceptrons ... part of the drive came from the fact that funding and research energy was dissipated ... money was at stake.” In recent times, these conditions have turned out to be particularly hard for scientists working in Brain-Inspired AI. [Velik, 2012, p. 44]

Marvin Minsky — one of the loudest proponents of AI caused so-called ‘AI winter’ for almost twenty years not only from scientific right to critique, but also from political reasons like funding. We can’t close eyes in front of this. It is likely that if in this case Minsky reflected this push, there still have to be many more examples when scientists are not able to recognize these non-voluntary pressures. Two biggest can be still easily imagined — military and business. Illusion that AI is independent of anything and it is purely scientific project is something what discourse builds on, but in the same time it is mere illusion.

The ‘Pure Science’ trope is building on its independence, but as already Merton noted,

> Science includes disinterestedness as a basic institutional element. Disinterestedness is not to be equated with altruism nor interested action with egoism. Such equivalencies confuse institutional and motivational levels of analysis... It is rather a distinctive pattern of control of a wide range of motives which characterizes the behavior of scientists. [Merton, 1973]

Ethos of scientific truth as goal for itself successfully hides most external influences and reifies itself building false legitimization base for AI scientists. Typical example where this issue is noted, but not reflected properly is in Campa’s text [Campa, 2008] where he states “quest for pure knowledge *is* and *can* be part of the transhumanism agenda”.[15] There is even not much reflection why transhumanism should, or could, be part of science at all. Important is that it is somehow compatible with *pure science* and that means, that it is good.

2.2 Discursive Construction of AI

Term itself — Artificial Intelligence was coined in 1956. Of course, there is a long history of how term was chosen and another long history how its meaning has been changing until today.

[15] italics from original

2.2.1 Case: Turing

Imitation Game [Turing, 1950] made a significant impact on AI term. I am not going into deep why Turing did this and did this in this exact way. I still would anticipate that he never expected what such relatively simple article will do in next hundred years. What was so extraordinary with this 'AI test', that it is still seriously considered even if it has quite a lot of methodological flaws (e. g. [Berrar et al., 2013]) that renders it completely unusable from scientific point of view?

Let's start with simple transcription. The crucial question which should be answered by proposed test is "Can machines think?" Human interrogator is set to position where he can ask two other 'creatures' any question. By investigating answers, reaction times, mood, generally whatever he has to decide which of creatures is computer and which is human. Turing further sets estimation that in fifty years there will be smart-enough machine that human interrogator will not be able to guess correctly in more than 70% of cases after five-minute interrogation. From present it seems that many computers already passed this test as 30% threshold is quite low. We know from practical life, that there are many people who believe to spambots and even can be forced to do some 'self-destructive action' like sending money to these virtual identities.

What probably makes paper special in that time is next section with responses to *anticipated* objections. It is noteworthy to emphasize word anticipated as selection of these question shows us state of the time. And we will see, that most of them defined debate for almost a hundred years now. There is nine of them and first is 'Theological Objection'. We're set in Great Britain, 1950. What engineer selects as an objection to AI? Religion and Immortality. It is not question about possibility of thinking but about Right of Creation and makes it part of Story of Religion and Alchemy.

Next argument is about dangerousness of such machines and it prefigures *Story of Securitization (3)*. Mathematical objection is mentioning the superiority of human over machine by *defining* it as such. Interrogator is asking machine logical paradox. Machine can't answer such question correctly. Such situation gives human feel of an advantage — not until time when he realizes that human is in the same situation.

Argument from consciousness is not only about consciousness but also about emotions. Turing quotes Jefferson "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.” Turing responds to this argument that machine can counterfeit such feelings and thus it is not important for the test. What is interesting here is opening another Pandora’s box. Question of emotions and feelings quickly became one of the crucial ones. In fact in second half of twentieth century it is becoming the primal one in secular society. The machine now differs only in its abilities to understand and ‘feel’ emotions. While in the time of Turing crucial trait was ‘creativity’ now it is ‘emotionality’ as creativity was deconstructed and killed by postmoderns. Emotionality now strives to be definition of human identity in western world. Where we see this topic most embraced is art (especially sci-fi). I would stress only one example from 2015 — movie *Ex Machina* [Garland, 2015] as it deals with Turing test directly. Humanoid AI is presented to interrogator and whole test is reduced to investigating if AI’s emotions are ‘real’ or ‘simulated’. It is quite illustrating how AI is being treated if it is stripped of its securitization aura. Of course there are still everlasting topics of responsibility of god-creator, primal sin and other Christian overload. But still — difference between human and machine is ‘real’ emotionality here.

It would not be an AI text if it would not mention Lady Lovelace. Objection attributed to her is one about creativity. “The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform” (quotation by Turing, italics by Lovelace 1842). This is typical issue of that time. Humanity as resulted from Enlightenment project was defined by creativity. To be human means to produce new meanings. Turing response is critical here. He contradicts term of ‘originality’. There is nothing new without building on some foundation. If there is no independent ‘idea from nothing’ then learning machine could produce same results. Turing is trying to beat the enlightenment beasts here. Creativity was definitely a topic for him as also other people noted ([Berrar et al., 2013, p. 249]).

The other arguments deal more with technicalities and are not so interesting for us. Argument from disabilities simply claims, that machine will have some flaws. Quick answer is, that every man can’t do something. It is not about thinking, but about magnitude of thinking (or doing). Argument from continuity in nervous system says that human body is analogous compared to digital computer. As computer/human is allowed only to communicate in written way, this argument is for Turing irrelevant. Development of next decades obsoleted it completely. On one side analogous computers has risen and on the other side human body is no longer treated like analogous machine more like quantum digital space. Argument from informality of behaviour is reduction of learning problem — if computer can learn to read, then it can also learn to behave according to situation. Last argument is from extrasensory perception. Even such thing is Turing evaluating — telepathy, precognition and psychokinesis. He proposes to bracket telepathy out, as it infringes the test definition (communication only via written text). The rest of text is dedicated to description of learning machines.

I have already pointed few topics which have tremendous impact on whole AI discourse. Of course that Turing has not invented them, only distilled them from contemporary scientific discourse. But terminology and approach to AI (word was not used yet in 1950) determined AI as anthropomorphic problematic. It is not important that machine can think but that 'thinking' is defined as an exclusive human activity. Turing is routinely using other anthropomorhic words as 'learning', 'playing' or 'mental act' when speaking about machine. I would dare to say that Turing test is so successful concept only because it treats AI as human and setting human as a measure to it. This definition by relation without investigating what runs under the hood of AI seems to be compatible with Habermas' communicative action [Habermas and Outhwaite, 1996] which suggests possibility of its success. Every other definition of AI is complex and even incomprehensible, so this one (which is not in fact definition at all) is still successful. Here starts the mainstream discourse of human-like AI. Ironically, AI scientists sometimes understands text as in this example: "We believe that Turing wanted to encourage us to abandon our anthropocentric view and to adopt a broader view on these phenomena." [Berrar et al., 2013, p. 249].

On one side Turing was thinking about machines as something what will merge with people or maybe even something what is identical to people, but on the other side he was not able to escape its humanistic origins. One of such quotes is here: "It is natural that we should wish to permit every kind of engineering technique to be used in our machines [...] Finally, we wish to exclude from the machines men born in the usual manner." [Berrar et al., 2013, p. 254]. This dichotomy between machine and man is hunting AI science until transhumanism is born in late sixties which finally transcends shackles of humanism.

Thin thread connecting text from 1950 with our present could be seen in AI discourse. IBM created Deep Blue in 1997 to play chess and beat mankind represented by Garry Kasparov. As Berrar quotes him and Martin "he sometimes saw deep intelligence and creativity in the machine's moves [...] The decisive game of the match was Game 2 [...] we saw something that went beyond our wildest expectations [...] The machine refused to move to a position that had a decisive short-term advantage — showing a very human sense of danger." [Berrar et al., 2013, p. 245]. Anthropomorphic approach is still present and legitimized by authority of science, especially by Turing which was put on pedestal decades after his death.

2.2.2 Evolution

Image would not be complete if we will not examine how AI term's meaning is changing through the time. We've to found the actors (as there are probably some new according to Alexander's note: "The exigencies of time and space create specific aesthetic demands; at some historical juncture, new social roles like director and producer emerge that specialize in this task of putting text 'into the scene'" [Alexander, 2004]) and see in which way they construct this term.

'Intelligence' is base term which are we working with. In public it is still something taken for granted, but we can note that even the word is starting to have contemporary meaning only in last few hundred years: "Though it is most often assumed to be an ahistorical concept, the construction of one's intellectual capacity as being 'normal' and another being 'abnormal' only began to appear in the English language during the era of industrialization." [Nina Lester and Gabriel, 2014, p. 778]. Special term which allowed psychology to take 'expert' position on this [Goodey and Dawson Books, 2011].

I will not follow this path further and turn to 'artificial' where robot is starting anthropomorhization point. It is interesting, that in Czech (from where 'robot' word comes [Čapek, 2004]) is difference between general robot intended for any work and robot which is human-like. This is given by grammatical gender in which the world is used (there are only small differences e. g. in plural form 'roboty' vs. 'roboti'[16]) but it includes awareness that there are more types of robots. This is fading to grey, when it comes to intelligence.

One issue is with false implications which connection of 'artificial' and 'intelligence' allows. As Gozzi notes, construction of public AI legitimacy involved metaphors which were used to describe it. Most of them lead to anthropomorphization trap as they were used e. g. to compare computer to brain. Simple example [Gozzi Jr., 1994, p. 234] show this false implication. 'Computer is brain' with minor premise 'Thinking is computing' leads to conclusion, that "Sophisticated (AI) programs will give the computer a mind, with consciousness, intelligence and other human attributes". As Gozzi puts in his research, he sees diminishing usage of strong metaphors as AI term gets more and more established. It was no more needed as field was set up as legitimate and metaphors can move further, typically to Kurzweil's Singularity and transhumanism program. Another reason is that AI term can be broadened and re-thinked as something what is not limited by its human image. Greer for example thinks about intelligence whose subset is human intelligence. So now, AI is freed from human limits and is part of 'something bigger' and should not be treated as such.

Limits of AI can be seen here. As at first there was no need to compare human and AI status, there was simply AI aspiring to human level. From this discourse could be selected this (under)definition "AI can be defined on the basis of the factor of a thinking human being and in terms of a rational behavior: (i) systems that think and act like a human being; (ii) systems that think and act rationally." [Čerka et al., 2015]. AI is here simply human-equal entity.

[16] `http://prirucka.ujc.cas.cz/?slovo=robot`

As transhumanist ideas appeared in the field, problem of comparison appeared as well. Transhumanist agenda tried to define AI as something better than human, which could be seen in terms of *strong AI*. "The 'strong AI' theory asserts that an AI is capable of having a mind, as well as mental states; versus the 'weak AI' position, which suggests that a machine is only capable of behaving intelligently." [Laufer, 2013, p. 1]. Ironically this distinction come from the opponent's side [Searle, 1980] as part of argument why it can't work. It provided name for new project where AI is condemned to transcend humanity.

Twenty years later this project splits to *AGI* (Artificial General Intelligence) which is more technically-oriented and seriously taken in AI science, while *Superintelligence*[17] concept of N. Bostrom is some more general and tries to open hands for other implementation and criticism. Superintelligence tries to break stigma of 'artificiality' and tries get beck to Lamarckian discourse of evolution. Bostrom is thus one of biggest proponents of 'naturality' of such process. This discursively dissolves intelligence perception more generally as it is detached from natural persons [Greer, 2014]. Superintelligence (with its paternalistic tone) also opens way to study things like "Charismatic Robot Leader: The Superintelligence" [Gladden, 2014]. It somehow removes stigma of *apocalyptic AI*.

Term 'intelligence' make software more tolerable than others. As some applications of software systems which nowadays could be hardly called as AI. Schwartz [Schwartz, 1989] shows that using the world allows easier deployment of surveillance systems which otherwise will cause bigger social resistance. Their intelligence make them more human-compatible and less state power tool. Similar examples are taken from places where these 'expert systems' would replace some human workers.

Completely different chapter is link with science fiction. As Geraci [Geraci, 2011] points even in title 'Transhumanist Evangelism in Science Fiction and Popular Science', sci-fi is entangled with AI discourse. It on one side serves as generator of ideas and on the other it consumes scientific discourse and disseminates it in public. From Marxist, more especially Althusser's, position we can see it as its tool of power. AIs are present in sci-fi almost from the beginning and shaped

[17] "By a 'superintelligence' we mean an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills. This definition leaves open how the superintelligence is implemented: it could be a digital computer, an ensemble of networked computers, cultured cortical tissue or what have you. It also leaves open whether the superintelligence is conscious and has subjective experiences.
Entities such as companies or the scientific community are not superintelligences according to this definition. Although they can perform a number of tasks of which no individual human is capable, they are not intellects and there are many fields in which they perform much worse than a human brain — for example, you can't have real-time conversation with 'the scientific community'." [Bostrom, 1998]

the discourse. Asimov’s laws of robotics are influential until today and on the other side, concepts of friendly, apocalyptic, strong AI or superintelligence is there. Discourse is widely shaped by literature and movies. They’re not constituent part of scientific story, but still quite interconnected with it.

3 Story of Securitization

> Suppose someone were to say, "Imagine this butterfly exactly as it is, but ugly instead of beautiful."
>
> *(Ludwig Wittgenstein[1])*

Securitization is term borrowed from International Relations created by so-called Copenhagen school [Buzan et al., 1998]. It means politicization of some topic via its relation to security issues. Important is that it is talking about security as about speech acts and in our case it is a view which allows us to see how (potential, perceived or imagined) AI security issues influence real-world and science progress. Anthropomorphization in this chapter goes very closely with discourse. If AIs as human beings combined with threats they represent is combination which leads to unexpected ideas as thinking about AI punishment. Topic of security is used by some groups (apocalyptic) to fortify their position and to denounce AI development even if their 'real' objection is e. g. religious.

Concerns from AI and science in general is leitmotif here. AI as a grave danger is framing of almost any public discussion of the topic. It does not depend if it is a religious counterargument or simple positivist calculation how much time is needed for uncontrolled self-replication nanobot to destroy whole planet [Kurzweil, 2005].

Fear of being destroyed by own progeny once again follows old myths represented e. g. Zeus' revolt against Titans. In relation to AI will be more interesting modern incarnation in Čapek's novel *R.U.R.* [Čapek, 2004] or in the later world-known novel *Do Androids Dream of Electric Sheep?* from P. K. Dick [Dick, 1996][2]. AI supersedes humanity and turns against mankind often resulting in human apocalypse. These myths are forged to concrete fears which are based on specie or at least group rivalry. If AI will be created what is the probability that it will be human-compatible? What will stop AI from enslaving mankind? Will AI have any cognitive capacity to even perceive humans? Whole this category of questions is leading us to two sides of quarrel — first is abolitionist movement which wants to stop everything before self-improving AI will be unstoppable. On the other side is an effort to make AI compatible from design and in second row to build ethical-legal frame for coexistence.

[1] [Wittgenstein, 1967]

[2] Better known via movie adaptation *Blade Runner* [Scott, 1982]

Especially question of legal framing is quite interesting. Technicist branch is still rooted in Asimov's three laws of robotics [Asimov, 1977, Khalil, 1993] and is set in defensive mode. Compared to it humanist line is based on human/AI equality thesis and in its weaker mode on animal rights argumentation. This line could be also split to part which opposes Anglo-Saxon legal order together with Christian tradition where man is qualitatively different and such order is formed to his defence. Main objection is that if animal rights were already admitted as substantially independent (so not as defence of their owner's rights) the taboo is already broken and there is no reason why to not extend these rights to AIs. Second line is rooted in discourse promoted by P. Singer [Singer, 1975]. In conneciton with human rights the new interesting mix is born. Some rights are confirmed, while not every one of them [Ashrafian, 2015]. Field is quite large and has origins in anthropocentrism, biocentrism or ecocentrism with all their historical baggage [Torrance, 2013]. Whole discourse almost logically expands to bigger issues like 'when man is still man' and 'when AI is ready for its rights'. This is the point where sociologically relevant authors enters the field: Barthes, Baudrillard or Husserl with human identity as a social construct, etc. [Hrişcă, 2012]

This partly scientific and partly public discourse is further propagated to legal norm proposals. Their genesis and construction are also interesting. One of the origins is already mentioned Asimov (note, that this is the 'art' input) and second quite influential is historical experience with legal norms for genetic manipulations [Habermas, 2003] and virus treatment. Proposals which are dedicated to transformation to laws are brought to live [Veruggio, 2007, Ashrafian, 2014, Field, 2012]. They are still not real laws, but media works with them as witch such. It is more connected to discursive construction than to securitization. Practical norm which at least someone adheres to are *Principles of Robotics* [EPSRC, 2010] or EU proposals [RoboLaw, 2014].

Beside these proposals which are more utopia than reality what we hit is problematic of software (and robot) responsibility. First trials are being run, where judges tries to answer question about who — if its operator, manufacturer or drone itself — is responsible for drone's actions [Hallevy, 2013]. Part of the field is completely serious discussion about how AI should be persecuted if found guilty. Proposals for limiting some of their rights, jury's right to turn it off (as a euphemism to capital punishment) puts a mirror to how these rights, ethical and democratic foundations are perceived in concrete communities. Habermas' theory of communicative ethics [Habermas and Outhwaite, 1996] can enlighten some aspects. For example even if AIs are in this discourse treated as completely human-equal they are not permitted to anyhow participate on legal system. Only some extremist positions hold on that AI will be part of a democratic process [Kurzweil, 2005]. Question of political rights of AIs is for most of the actors completely unacceptable and this non-acceptance is in direct contradiction with their egalitarian position.

AI as a danger (and here I omit other variants as an ecological threat, etc.) is main argumentation line of AI antagonists. Ethical or religious issues takes place behind this flagship.

"Nor is there, to be sure, any lack of wild speculation. A handful of freaked-out intellectuals is busy reading the tea leaves of a naturalistic version of posthumanism, only to give, at what they suppose to be a time-wall, one more spin - 'hypermodernity' against 'hypermorality' — to the all-too-familiar motives of a very German ideology." [Habermas, 2003, p. 22]

The big problem is once again anthropomorphization. Let's say with Sharkey, that "Like other cultural myths, it can be harmless in casual conversations in the lab. But it is a perilous road to follow in legal and political discussions about enabling machines to apply lethal force." [Sharkey, 2012, p. 791] See an example where military equipment was treated more like comrades. "The *Washington Post* reported that soldiers on the battlefield using bomb disposal robots often treat them as fellow warriors and are sometimes prepared to risk their own lives to save them. They even take them fishing during leisure time and get them to hold a fishing rod in their gripper." [Sharkey, 2012, p. 792] or human-like discourse soaking to the highest ranks of U. S. military "Gordon Johnson, former head of the Joint Forces Command at the Pentagon, told the *New York Times* that robots 'don't get hungry. They're not afraid. They don't forget their orders. They don't care if the guy next to them has just been shot.'" (Ibid). Robot is taken for soldier here and treated as such. Even if we admit that there will be human-like AI, we're not there yet and these 'soldiers' are simple machines with no free will. I would quote Sharkey's once again as he brightly summarizes the problem which emerges here:

> This is not just being picky about semantics. Anthropomorphic terms like 'ethical' and 'humane', when applied to machines, lead us to making more and more false attribution about robots further down the line. They act as linguistic Trojan horses that smuggle in a rich interconnected web of human concepts that are not part of a computer system or how it operates. Once the reader has accepted a seemingly innocent Trojan term, such as using 'humane' to describe a robot, it opens the gates to other meanings associated with the natural language use of the term that may have little or no intrinsic validity to what the computer program actually does. [Sharkey, 2012, p. 793]

Direct result of such subversion, is responsibility slowly virtually shifting to machines. Man who is driving drone feels no more full responsibility for its actions, as drone is semi-autonomous. This quality of anthropomorphization effect is on the second side exploited in business and in therapy. Human-like robots are better sold [Robertson, 2007, Larson, 2010] or they can be used as companions for e. g. elderly people [Hakli et al., 2014, Kerstin, 2014, Kernaghan, 2014, Pfeifer et al., 2012]. The question of trust to AIs is repainted as 'commitment'

[Michael and Salice, 2014]. This targeted use makes border between what we pretend and what we believe even more fluid.

Generally there is a technicist discussion if and how is AI able to destroy mankind [Evans, 2007]. Second part of it is AI will have intention to do it. “The concern is that the machines will view us as an unpredictable and dangerous species.” [Panda et al., 2015, p. 111] shows typical argumentation. On the other side transhumanists and scientists are trying to oppose such fears [Kurzweil, 2001]. Compared to direct destruction there is an eternal idea of society decay in the realm where everything is provided [Rossano, 2001].

Part of securitization discussion is also topic of AI as a citizen. We’ve already covered it in other chapters, so just adjusting some topics how they are reflected from this point. At first capability of AI to be a citizen is not coupled with its human status. There are tendencies to define new criteria and redefine traditional Turing test to ‘citizenship test’ [Erden and Rainey, 2012, p. 143]. Such activities can cast some light on how this duality humanity-citizenship is treated in today’s society. Here is sometimes seen tendency to evade such problem by “. . . best not to make AIs extremely human-like in appearance, to avoid erroneous attributions that may blur the bright lines we set around moral categories (Arneson 1999). If such confusion were to develop, given the strong human tendency to anthropomorphize, we might encounter rising social pressure to give robots civil and political rights, as an extrapolation of the universal consistency that has proven so central to ameliorating the human condition.” [Yampolskiy and Fox, 2013, p. 224] If scientist are not willing to give AIs civil rights, they are not against lending them at least some ‘non-harming’ rights like intellectual property rights [Davies, 2011]. Of course, there is still discussion running if it is not possible to grant them more rights (with hidden goal to provide them also political rights, but ‘somehow’ limited) [Kunneman and Derkx, 2013].

4 Conclusion

AI anthropomorphization theme has shown as quite wide. It is permeating AI science itself but other — on first sight unrelated — fields like ethics, religion or law. All these ‘independent’ and ‘rational’ areas seems to be more or less affected by unconscious choices given the agency of things. It is nothing new, but worth to study on concrete examples. Difficulty of holistic dealing with the topic is on the other side seen also in the structure of work. It is not easy to talk about AI religion without touching its alchemistic side or looking to limits of programming. What can be taken as a lesson here is that potential empiric research should take in game all these effects. Otherwise, it may will describe some facets of the topic, but probably it will create distorted image. AI is not clear hard science. AI is human project as most others and is influenced by and influences living world also in purely social ways.

Originally I have selected five narratives from whose only two are now present in text. I have dropped religious, alchemist and art stories due to lack of space. I have also investigated them and these chapters are available from me on demand. The selected ones are still broad enough to show different aspects which can be studied from sociological point of view.

Story of Science has shown, that AI is not deliberate concept which is valid in any rational space. It was constructed especially in western civilization during last few hundred years and it has shown that in other intellectual and socio-political situation such concept need not exist at all.

Story of Securitization on the other side shows that when such concept exists it can be on one side used as a political weapon nevertheless what real impact is. Or on the other side in combination with anthropomorphization it leads to weird human behaviour and laws which merely reflects ‘the objective reality’. Agency of things itself can modify how judge will decide its case as we have seen.

When we would dive to religious stories, we would have seen that AI science field is driven by western Christianity and even non-consciously embracing eschatological discourse. This is something which on one side has big impact to securitization story and on the other side warns us a lot. As AI scientists would never say they are religious believers in AI, they behave like their followers. We can think about us as sociologists if we don’t behave in same way with some other deity.

Also many questions were opened to be tested by rigorous hypothesis, some of them can be verified by experiments, some by deeper field research. Anthropomorphic AI can seduce us to thinking about it as only a small fragment of sociology of science. But from other stories we see, that it is not only sociology of science, but also sociology of religion, sociology of things, . . . Sociology is also not the only tool which can be used to understand it — anthropology can give us a helpful hand with its tradition of observation. What we also need is independent point of view which can reflect on role of possible sociology of sociology of AI (in same meaning as meta-sociology of science or so-called second-order science [Müller and Riegler, 2014]). So let's explore the borders of constructed science.

Bibliography

Adam, A. (2008). Ethics for things. *Ethics & Information Technology*, 10(2/3):149.

Alexander, J. C. (1989). *Structure and meaning : relinking classical sociology.* New York : Columbia University Press, c1989.

Alexander, J. C. (2004). Cultural Pragmatics: Social Performance between Ritual and Strategy. *Sociological Theory*, 22(4):527.

Allen, C., Varner, G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3):251–261.

Anderson, M. and Anderson, S. L. (2007). The status of machine ethics: a report from the AAAI Symposium. *Minds & Machines*, 17(1):1–10.

Ashrafian, H. (2014). Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights. *Science and Engineering Ethics.*

Ashrafian, H. (2015). AIonAI: A Humanitarian Law of Artificial Intelligence and Robotics. *Science & Engineering Ethics*, 21(1):29–40.

Asimov, I. (1977). *I, Robot.* Fawcett Crest, New York.

Barandiaran, X. E. and Egbert, M. D. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, 20(1):5–28.

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. (2000). Open problems in artificial life. *Artificial Life*, 6(4):363–376.

Berrar, D., Konagaya, A., and Schuster, A. (2013). Turing Test Considered Mostly Harmless. *New Generation Computing*, 31(4):241–263.

Bickhard, R. (2014). Robot Sociality: Genuine or Simulation. In *Sociable Robots and the Future of Social Relations : Proceedings of Robo-Philosophy 2014*, Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam.

Blank, R. H. (2013). *Intervention in the Brain : Politics, Policy, and Ethics.* Basic Bioethics. MIT Press, Cambridge, Mass.

Bobrow, D. G. (1964). *Natural Language Input for a Computer Problem Solving System.* Doctoral Thesis, Massachusetts Institute of Technology.

Bonnefon, J.-F., Shariff, A., and Rahwan, I. (2015). Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars? *arXiv:1510.03346 [cs].* arXiv: 1510.03346.

Bostrom, N. (1998). How long before superintelligence? *International Journal of Futures Studies*, 1998(2).

Bourdieu, P. (1984). *Distinction: A Social Critiques of the Judgment of Taste.* Harvard University Press, Cambridge.

Bourdieu, P. (1996). *The Rules of Art: Genesis and Structure of the Literary Field.* Stanford University Press.

Brouwer, L. E. J. and Heyting, A. (1980). *Collected works. 1, 1,.* North-Holland Publishing Company, Amsterdam [etc.].

Buzan, B., Wæver, O., and Wilde, J. d. (1998). *Security: a new framework for analysis.* Lynne Rienner Pub, Boulder, Colo.

Campa, R. (2008). Pure Science and the Posthuman Future. *Journal of Evolution & Technology*, 19(1):1–7.

Čapek, K. (2004). *R.U.R.*

Čerka, P., Grigiene, J., and Sirbikyte, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review: The International Journal of Technology Law and Practice.*

Chomsky, N. (2002). *Syntactic Structures.* Walter de Gruyter.

Churchland, P. M. (2001). *Matter and consciousness : a contemporary introduction to the philosophy of mind.* Bradford book. Cambridge, Mass. : MIT Press, 2001.

Coeckelbergh, M. (2009). Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents. *AI & Society*, 24(2):181–189.

Cole, R. J., Bild, A., and Matheus, E. (2012). Automated and human intelligence: direct and indirect consequences. *Intelligent Buildings International*, 4(1):4–14.

Cristianini, N. (2014). On the current paradigm in artificial intelligence. *AI Communications*, 27(1):37–43.

Davies, C. R. (2011). An evolutionary step in intellectual property rights – Artificial intelligence and intellectual property. *Computer Law and Security Review: The International Journal of Technology and Practice*, 27:601–619.

Davis, E. (2015). Ethical guidelines for a superintelligence. *Artificial Intelligence*, 220:121–124.

de Beer, F. (2012). Responsibility in the age of the new media: Are cyborgs responsible beings? *Communicatio: South African Journal for Communication Theory & Research*, 38(1):15.

de Chardin, T. (1966). *Man's Place in Nature: The Human Zoological Group.* Collins, London, fontana books edition.

DeBaets, A. M. (2014). Can a Robot Pursue the Good? Exploring Artificial Moral Agency. *Journal of Evolution & Technology*, 24(3):76–86.

Dennett, D. C. (1992). *Consciousness explained.* Allen Lane The Penguin Press, London.

Dick, P. K. (1996). *Do androids dream of electric sheep?* Ballantine Books, New York.

Eco, U. (1995). *The Search for the Perfect Language.* Wiley.

EPSRC (2010). Principles of robotics - EPSRC website.

Erden, Y. J. and Rainey, S. (2012). Turing and the Real Girl. *New Bioethics*, 18(2):133–144.

Evans, J. H. (2014). Faith in science in global perspective: Implications for transhumanism. *Public Understanding of Science*, 23(7):814.

Evans, W. (2007). Singularity Warfare: A Bibliometric Survey of Militarized Transhumanism. *Journal of Evolution & Technology*, 16(1):161–165.

Field, C. (2012). South Korean Robot Ethics Charter 2012.

Floridi, L. (2004). Open Problems in the Philosophy of Information. *Metaphilosophy*, 35(4):554–582.

Garland, A. (2015). Ex Machina.

Geraci, R. (2014). A Novel Society: Science Fiction Novels as Religious Actors. *Implicit Religion*, 17(4):417–431.

Geraci, R. M. (2007). Robots and the Sacred in Science and Science Fiction: Theological Implications of Artificial Intelligence. *Zygon: Journal of Religion & Science*, 42(4):961–980.

Geraci, R. M. (2008). Apocalyptic AI: religion and the promise of artificial intelligence. *Journal of the American Academy of Religion*, 76(1):138–166.

Geraci, R. M. (2011). There and back again: transhumanist evangelism in science fiction and popular science. *Implicit Religion*, 14(2):141–172.

Gladden, 2), M. . . (2014). *The social robot as 'charismatic leader': A phenomenology of human submission to nonhuman power*, volume 273 of *Frontiers in Artificial Intelligence and Applications*. IOS Press. 329.

Goodey, C. F. and Dawson Books (2011). *A History of Intelligence and 'intellectual Disability' : The Shaping of Psychology in Early Modern Europe*. Ashgate Publishing Ltd, Burlington, VT.

Gozzi Jr., R. (1994). A Note on the Metaphorically Charged Discourse of Early Artificial Intelligence. *Metaphor & Symbolic Activity*, 9(3):233.

Greer, K. (2014). Is Intelligence Artificial? *arXiv:1403.1076 [cs]*. arXiv: 1403.1076.

Habermas, J. (2003). *The future of human nature*. Cambridge : Polity Press, 2003.

Habermas, J. and Cronin, C. (2008). *Between naturalism and religion : philosophical essays*. Cambridge : Polity, c2008.

Habermas, J. and Outhwaite, W. (1996). *The Habermas Reader*. Polity Press.

Hakli, R., Nørskov, M., and Seibt, J. (2014). *Sociable Robots and the Future of Social Relations : Proceedings of Robo-Philosophy 2014*. Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam.

Hallevy, G. (2013). *When Robots Kill : Artificial Intelligence Under Criminal Law*. Northeastern University Press, Boston.

Hodder, I. (2012). *Entangled: An Archaeology of the Relationships Between Humans and Things*. John Wiley & Sons.

Houtman, D. and Aupers, S. (2010). *Religions of Modernity : Relocating the Sacred to the Self and the Digital.* International Studies in Religion and Society. Brill, Leiden.

Hrişcă, A. M. (2012). Artificial Body: Between “to Be” and “to Have”. *Studia Universitatis Babes-Bolyai, Philosophia*, 57(2):121–135.

Jones, S. (2003). Resolving classical experience and the quantum world. *Technoetic Arts: A Journal of Speculative Research*, 1(2):143–164.

Kaur, G. D. (2013). Kant and the simulation hypothesis. *AI & SOCIETY*, 30(2):183–192.

Kernaghan, K. (2014). The rights and wrongs of robotics: Ethics and robots in public organizations. *Canadian Public Administration*, 57(4):485–506.

Kerstin, D. (2014). Social Robots As Companions: Challenges and Opportunities. *Frontiers in Artificial Intelligence and Applications*, pages 9–10.

Khalil, O. E. M. (1993). Artificial Decision-Making and Artificial Ethics: A Management Concern. *Journal of Business Ethics*, 12(4):313–321.

Kleene, S. C. (1952). *Introduction to metamathematics.* Van Nostrand, New York.

Knight, C. C. and Murphy, N. C. (2010). *Human Identity at the Intersection of Science, Technology and Religion.* Ashgate Science and Religion Series. Ashgate Publishing Ltd, Farnham, Surrey, England.

Kunneman, H. and Derkx, P. (2013). *Genomics and Democracy : Towards a 'lingua Democratica' for the Public Debate on Genomics.* Life Sciences, Ethics and Democracy. Brill Academic Publishers, Amsterdam.

Kurzweil, R. (2001). Promise and Peril – The Deeply Intertwined Poles of 21st Century Technology. *Communications of the ACM*, 44(3):88–91.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology.* Penguin.

Kurzweil, R. (2014). Don’t Fear Artificial Intelligence. *Time*, 184(26/27):28–28.

Larson, D. A. (2010). Artificial Intelligence: Robots, Avatars, and the Demise of the Human Mediator. *Ohio State Journal on Dispute Resolution*, 25(1):105–163.

Latour, B. (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory.* OUP Oxford.

Latour, B. and Porter, C. (1991). *We have never been modern.* Cambridge, Mass. : Harvard University Press, c1991.

Laufer, M. S. (2013). Artificial Intelligence in Humans.

Laukyte, M. (2014). *Artificial agents: Some consequences of a few capacities*, volume 273 of *Frontiers in Artificial Intelligence and Applications.* IOS Press. 115.

Leuenberger, G. (2014). Universal Algorithmic Ethics. *arXiv:1404.1718 [cs].* arXiv: 1404.1718.

Levine, A. (2014). Sociality Without Prior Individuality. In *Sociable Robots and the Future of Social Relations : Proceedings of Robo-Philosophy 2014*, Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam.

MacCormack, P. (2012). *Posthuman Ethics : Embodiment and Cultural Theory.* Ashgate Publishing Ltd, Burlington, VT.

Megill, J. (2014). Emotion, Cognition and Artificial Intelligence. *Minds & Machines*, 24(2):189–199.

Menabrea, L. F. (1842). Sketch of The Analytical Engine.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations.* University of Chicago Press.

Michael, J. . . . and Salice, A. . . . (2014). *(How) Can robots make commitments? A pragmatic approach*, volume 273 of *Frontiers in Artificial Intelligence and Applications.* IOS Press. 125.

Minsky, M. L. and Papert, S. A. (1972). *Perceptrons: an introduction to computational geometry.* The MIT Press, Cambridge/Mass., 2. print. with corr edition.

Moravec, H. (1988). *Mind Children: The Future of Robot and Human Intelligence.* Harvard University Press.

Moravec, H. (1991). The Universal Robot.

Mossman, K. L. and Tirosh-Samuelson, H. (2012). *Building Better Humans? : Refocusing the Debate on Transhumanism.* Beyond Humanism: Trans- and Posthumanism. Peter Lang, Frankfort au Main.

Muggleton, S. (2014). Alan Turing and the development of Artificial Intelligence. *AI Communications*, 27(1):3–10.

Müller, K. H. and Riegler, A. (2014). Second-Order Science: A Vast and Largely Unexplored Science Frontier. *Constructivist Foundations*, 10(1):7–15.

Newton, Sir, I. (1723). *Philosophiae naturalis principia mathematica.* Amsterodam : Sumptibus societatis, 1723.

Nina Lester, J. and Gabriel, R. (2014). The discursive construction of intelligence in introductory educational psychology textbooks. *Discourse Studies*, 16(6):776.

Panda, S. S., Panda, M., Das, R. R., and Mohanty, P. K. (2015). The top species will no longer be humans: Robotic surgery could be a problem. *Journal of Minimal Access Surgery*, 11(1):111–111.

Pfeifer, R., Lungarella, M., and Iida, F. (2012). The Challenges Ahead for Bio-Inspired 'Soft' Robotics. *Communications of the ACM*, 55(11):76–87.

Pratt, V. (1987). *Thinking machines: the evolution of artificial intelligence.* B. Blackwell.

Raatikainen, P. (2015). Gödel's Incompleteness Theorems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy.* Stanford University, spring 2015 edition.

Rescher, N. (2011). *Philosophical Episodes.* Walter de Gruyter.

Robertson, J. (2007). Robo sapiens Japanicus: Humanoid robots and the posthuman family. *Critical Asian Studies*, 39(3):369–398. 369.

RoboLaw (2014). Guidelines on Regulating Robotics.

Rosental, C. (2003). Certifying Knowledge: The Sociology of a Logical Theorem in Artificial Intelligence. *American Sociological Review*, 68(4):623–644.

Rossano, M. J. (2001). Artificial intelligence, religion, and community concern. *Zygon*, 36(1):57–75.

Schiaffonati, V. (2003). A Framework for the Foundation of the Philosophy of Artificial Intelligence. *Minds & Machines*, 13(4):537–552.

Schopenhauer, A. (2008). *Studies in Pessimism, On Human Nature, and Religion: a Dialogue, etc.* Digireads.com.

Schwartz, R. D. (1989). Artificial intelligence as a sociological phenomenon. *Canadian Journal of Sociology*, 14(2):179.

Scott, R. (1982). Blade Runner. IMDB ID: tt0083658 IMDB Rating: 8.2 (427,643 votes).

Seaman, B. and Rossler, O. (2008). Neosentience a new branch of scientific and poetic inquiry related to artificial intelligence. *Technoetic Arts: A Journal of Speculative Research*, 6(1):31–40.

Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 1980(3).

Sharkey, N. and Sharkey, A. (2006). Artificial intelligence and natural magic. *ARTIFICIAL INTELLIGENCE REVIEW*, 25(1-2):9–19.

Sharkey, N. E. (2012). The evitability of autonomous robot warfare. *International Review of the Red Cross*, 94(886):787–799.

Shelley, M. W. (2012). *Frankenstein; Or, The Modern Prometheus.* Lackington, Hughes, Harding, Mavor & Jones, London.

Singer, P. (1975). *Animal liberation: a new ethics for our treatment of animals.* New York review : distributed by Random House.

Skinner, B. F. (1976). *About behaviorism.* Vintage Books, New York.

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1):62–77.

Stephenson, N. (2005). *The system of the world.* Harper Perennial, New York.

Thagard, P. (2010). *The brain and the meaning of life.* Princeton Univ. Press, Princeton, N.J.

Tonkens, R. (2009). A Challenge for Machine Ethics. *Minds & Machines*, 19(3):421–438.

Torrance, S. (2008). Ethics and consciousness in artificial agents. *AI & Society*, 22(4):495–521.

Torrance, S. (2013). Artificial agents and the expanding ethical circle. *AI and Society*, 28(4):399–414. 399.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59(236):433–460.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, 58(345-363):5.

Vargas, P., Fernaeus, Y., Lim, M., Enz, S., Ho, W., Jacobsson, M., and Ayllet, R. (2011). Advocating an ethical memory model for artificial companions from a human-centred perspective. *AI & Society*, 26(4):329–337.

Velik, R. (2010). Quo vadis, Intelligent Machine? *Quo vadis, Intelligente Machine?*, 1(4):13–22.

Velik, R. (2012). AI Reloaded: Objectives, Potentials, and Challenges of the Novel Field of Brain-Like Artificial Intelligence. *BRAIN: Broad Research in Artificial Intelligence & Neuroscience*, 3(3):25–54.

Verdoux, P. (2011). Emerging Technologies and the Future of Philosophy. *Metaphilosophy*, 42(5):682–707.

Veruggio, G. (2007). EURON Roboethics Roadmap.

Wallach, W., Allen, C., and Smit, I. (2007). Machine morality: bottom-up and top-down approaches for modelling human moral faculties. *AI & SOCIETY*, 22(4):565–582.

Whitehead, A. N. and Russell, B. (1963). *Principia Mathematica Volume I*. Cambridge University Press, Cambridge, 2 edition.

Wittgenstein, L. (1967). *Zettel, 40th Anniversary Edition*. University of California Press.

Yampolskiy, R. and Fox, J. (2013). Safety Engineering for Artificial General Intelligence. *Topoi*, 32(2):217–226. 217.