

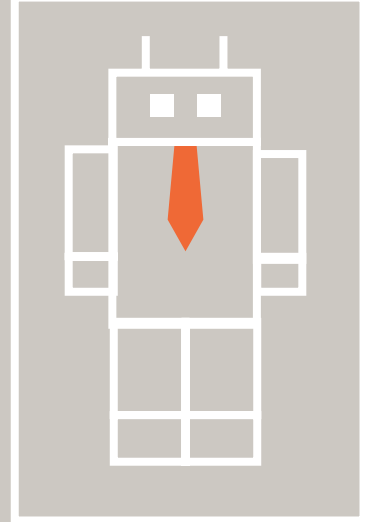
# The FrankensteinFactor 'The Anatomy of Fear of AI' Machine Intelligence

Report 3 of 4



Menno van Doorn, Sander Duivestein, Thijs Pepping

# CONTENTS



We thank the Oxford Internet Institute (OII), for insights and ideas that came out of our collaborative Machine Intelligence workshop, specifically: Luciano Floridi (Director of Research, Professor of Philosophy and Ethics of Information), Scott Hale (Data Scientist), Vili Lehdonvirta (DPHil Programme Director, Research Fellow), Michael Osborne (Associate Professor in Machine Learning), Mariarosaria Taddeo (Researcher, Information and Computer Ethics), John Tasioulas (Professor of Politics, Philosophy and Law, King's College London), Greg Taylor (Research Fellow, Microeconomics of Internet mediated marketplaces), Taha Yasseri (Research Fellow, Computational Social Science). We like to note that neither the OII nor any of the above named persons are responsible for the views expressed in this report.



Attribution-NonCommercial-ShareAlike 3.0

(CC BY-NC-SA 3.0)



<b>1</b>	<b>The FrankensteinFactor in IT</b>	<b>5</b>
<b>2</b>	<b>Frankenstein: superintelligent and superstupid</b>	<b>11</b>
<b>3</b>	<b>Frankenstein: fear of ourselves</b>	<b>19</b>
<b>4</b>	<b>Frankenstein: fear for the end of the culture</b>	<b>23</b>
<b>5</b>	<b>Frankenstein: fear of the 'Big Four'</b>	<b>28</b>
<b>6</b>	<b>The FrankensteinFactor: CIO as a therapist</b>	<b>40</b>
<b>Appendix A</b>	<b>Asilomar AI Principles (2017)</b>	<b>45</b>
<b>Appendix B</b>	<b>Frankenstein Syndrome Questionnaire</b>	<b>47</b>
<b>Sources</b>		<b>49</b>



Reason has always served computerization. But today emotions run higher and higher when it comes to IT. Almost everything seems possible now that artificial intelligence (AI) is awakening from its hibernation. Computers are increasingly doing things that previously could only be done by people – a fascinating fact. Computers that can listen, learn, see, and talk back – an important arsenal for artificial intelligence – are progressing towards rivaling human capabilities.

### Wow!

We read almost daily about new breakthroughs in the field of AI. After the board games chess and Go, algorithms also show themselves to be the best at poker. Cancer research gets a boost and focuses with AI 'Moonshots' on pioneering new methods.<sup>1</sup> The ambition of Bridgewater, the world's largest venture capital organization, indicates that the boundaries of AI capabilities are continuously being stretched. Bridgewater wants an AI application to take 75 percent of all management decisions (including hiring and firing of employees) and all this within the next five years. Chatbots such as the Chinese Little Bing will become personal buddies with whom we share our innermost feelings. We see Watson answering quiz questions and beating the best (human) players in a game and we see cars maneuvering independently through cities. And what about commercials created by an algorithm? The artificial creative director AI-CD β of McCann Erickson in Japan writes scripts for commercials

and competes with the human creative director.<sup>2</sup>

We are on the threshold of an extraordinary development that makes the automation heart beat faster. Possibilities are presenting themselves to build better and safer systems that are much more effective and have a greater impact. But there is something else going on. Put these typically human skills into an algorithm and the subconscious is starting to play a role. Emotions take over when we see human characteristics in non-human form (algorithms). This offers excellent opportunities to strengthen the bond with your customers – for more on this, read our previous report *The Bot Effect: 'Friending Your Brand'*. This concerns more than just the efficiency and effectiveness with which IT has been familiar for many years. In *The Bot Effect: 'Friending Your Brand'* we described this phenomenon, called anthropomorphism (from the Greek: *anthropos* is 'man' and *morphe* is 'form'), and how this can strengthen

1 See for example <http://www.digitaleng.news/de/nvidia-joins-cancer-moonshot-with-ai-platform-for-accelerating-research/>

2 See both commercials here: <http://adage.com/article/creativity/check-commercial-mccann-japan-ai-creative-director/304320/>



Man versus Machines: also in creative professions, such as creation of commercials, we see algorithms taking over. We increasingly see these 'man versus machine' competitions.

the relationship with the customer. However, at the other end of this spectrum there is the fear, the fear of the unknown, the automata and the android phenomena of technology, the fear of finally having to succumb. The paradox is that we want to experience the pleasure of the AI capabilities, even to the extent that we become attached to them. At the same time, there is that fear of commitment, the uncanny feeling that plays up when we are faced with this digital *doppelgänger*.

There are examples that simply placing a humanoid robot at a reception desk causes great consternation. An action intended as an inspiring gimmick – go and play with a robot – may be misunderstood. Of the same category is the use of certain images of robots in a PowerPoint presentation that somehow come across as

being creepy and steer the meeting in an unintended direction. Not to mention the fear of job loss, fueled among other things by Facebook messages with ominous titles such as 'Robots will steal your job' or 'Beware of a Robocalypse'.

### The social acceptance of AI

If you don't learn about the emotions that drive your customer, you will not know your customer, according to Forrester Research in a recent report.<sup>3</sup> Research by the British Science Association shows that over a third of the UK population believes that in the long term artificial intelligence poses a threat to the survival of man. Almost half of the people in this research (46 percent) said that they see no benefit in the idea of robots that look like people. The same goes for robots with human capabilities, emotions and/or

3 'If you don't understand their emotions, you don't understand your customer.'  
Forrester Research, 'Understanding the role of emotion in customer experience', 2015,  
<https://www.forrester.com/report/Understanding+The+Impact+Of+Emotion+On+Customer+Experience/-/E-RES122503>

a personality such as in the films *Her* and *Ex Machina*. And 60 percent believe that due to robots there will be fewer jobs in 10 years' time.<sup>4</sup> 'Trusting AI to do Work' ranks 32 on the top 90 list of America's Top Fears 2015.<sup>5</sup> 'Robots' is in 36th position, 'AI' at 39 and 'Tech I don't understand' is in 49th position. The AI-related themes all scored higher than issues such as fear of needles, shooting incidents, flying, kidnapping or getting murdered by someone you know.

Artificial intelligent algorithms trigger, rationally or irrationally, feelings of fear in people; they score high on the 'FrankensteinFactor', the fear of the artificial *doppelgänger*. In this report we explore the possible causes of this. Not to fuel or stoke up fear, but simply as support for the debate that will be held during the coming years. Our appeal is to not only look at efficiency and effectiveness, but to also take into account these feelings of anxiety. Expressed in a formula: the three Es, Efficiency and Effectiveness supplemented by a third, Existence, are decisive for success.

The question is how the ambition to achieve success in IT will take shape exactly. Indeed, the acceptance of technology is always an interaction between the technological possibilities, economic feasibilities and social desirabilities. The

fascination for a technical tour de force ('Look! Awesome! A self-driving car!') does not automatically mean that everyone wants to have one, even if they could afford it. Is this car reliable? How do the algorithms behind the self-driving car work? Am I still in control, or am I at the mercy of this car? Just a few questions, some of a philosophical nature, about transparency and trust and things that have to do with our existence.

The question of what AI does to the lives of your customers and your employees, their existence, their identity, the control they have over their affairs, their *raison d'être*, comes up much more often as more AI applications come on the market (and are talked about). This means that those responsible for IT policy, management and their CIO offices have to become better interlocutors in this area. Knowledge of the emotional life becomes more important as technology moves in areas that affect the emotions.

## In this report

This report is in a sense a diptych about bonding and a fear of bonding in relation to AI, a diptych together with our earlier *The Bot Effect: 'Friending Your Brand'*.

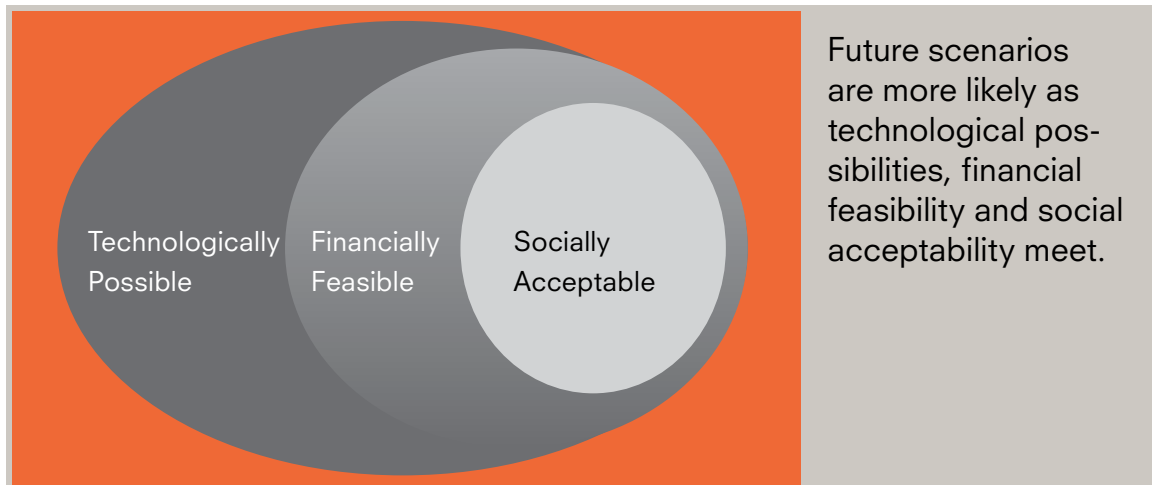
This report is not so much about whether these fears are justified or not, we examine the reasons for these fears and where they come from. The signal we want

**Efficiency × Effectiveness × Existence = IT success**

4 'One in three believe that the rise of artificial intelligence is a threat to humanity', <http://www.britishscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity>

5 <https://blogs.chapman.edu/wilkinson/2015/10/13/americas-top-fears-2015/>





to send is that an advantage can be gained if the emotional relationships are included more in the AI plans. Everybody who wants to achieve social acceptance of AI, could take advantage of this. Anyone wishing to explore the possibilities of this advanced technology, should make human existence a central focus in their approach.

#### **Four approaches**

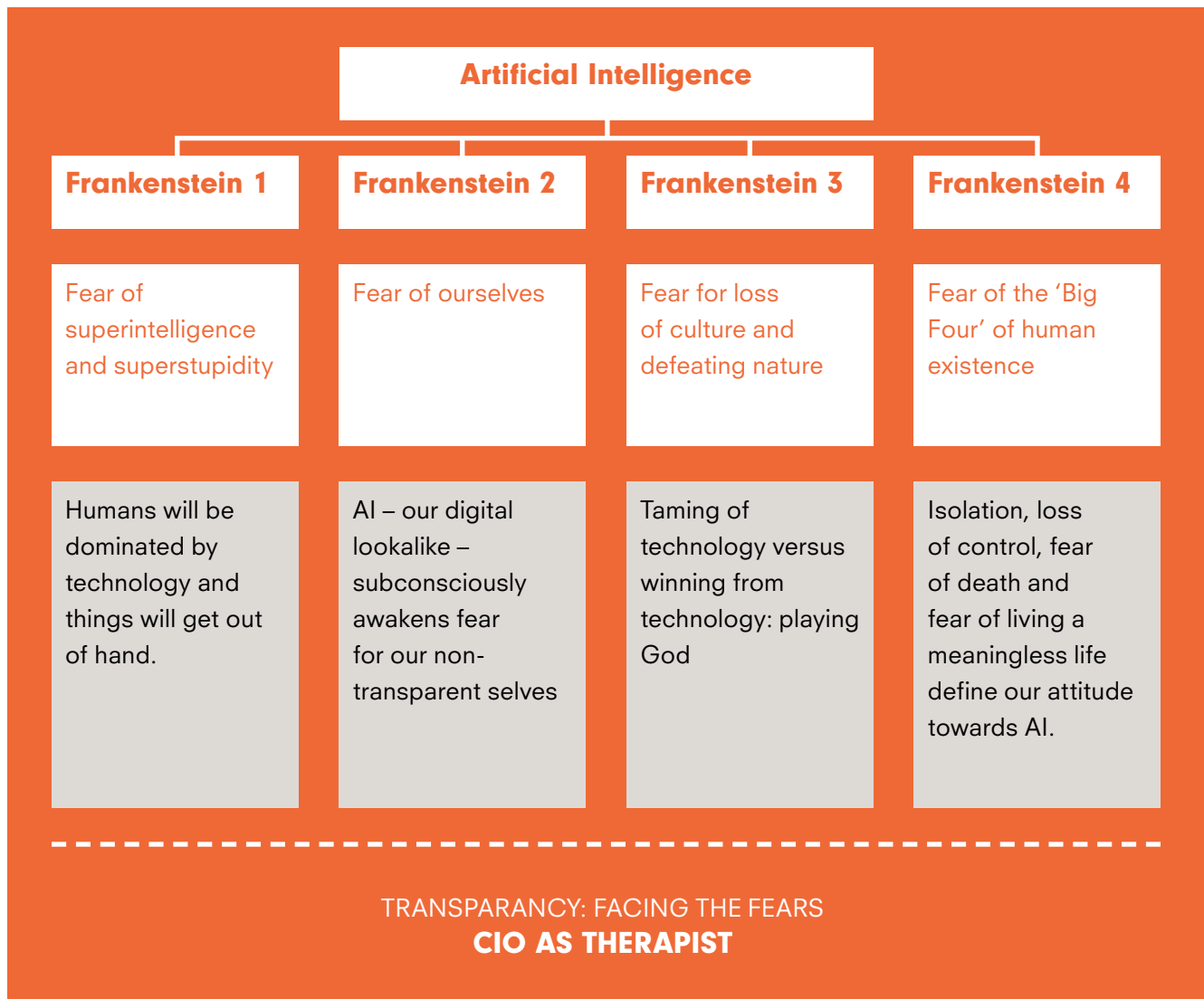
For the anatomy of fear we will present four different approaches providing a perspective of these relationships. We start with the current debate: the fear of superintelligence, but also the risks and fears that arise when the systems prove to be too stupid. The fear of ourselves as, *inter alia*, Freud described in 'The Uncanny', and the (Big Four) existential fears of the neo-Freudian school of psychology subsequently provide a deeper insight into these subconscious emotional relationships. A fourth anatomical exploration concerns the fear of

the loss of culture, an argument that also plays a prominent role in a new European resolution to which we will return later. With regard to this point, we look at the differences between East and West, and particularly to Japan and why the horror factor seems to be less common in Japan than over here.

Ultimately we place the CIO in the chair of the therapist. Many functions have already been attributed to him, from Chief Information Officer and Innovation Officer to Change Officer. Being a customer therapist or expert on the emotional life, and specifically the dark side of fear, is certainly not yet included in the standard repertoire of the CIO. But with the knowledge that the fears that we have described will sooner or later emerge during an arbitrary AI initiative, it seems wise to brush up one's knowledge in this field. The role of the therapist in this is threefold: a) be transparent about the functioning of AI algorithms, the plans



that exist in this area and the impact to be expected; b) enter into a dialogue with users, customers and employees about the possible fears that exist with regard to these new applications; and c) immediately apply the lessons learned from this therapeutic dialogue in his own organization.



### **Frankenstein 1: the current debate**

#### **The fear of superintelligence and superstupidity**

Leading scientists warn us that the end of time is near when computers become more intelligent than humans. AI software engineers warn about the unintended consequences of AI applications that are just not smart enough yet. It is an intellectual yes-no debate and Hollywood ignores this completely and responds to people's basic instincts.

### **Frankenstein 2: Freud and 'the uncanny'**

#### **We do not fear the robots but ourselves**

Psychoanalysts Sigmund Freud and Ernst Jentsch explain the feeling of horror that people experience when they are confronted with automata. This is not about cognitive and reasoned fear as in the above analysis of the current debate, but rather the relationship with one's basic instincts.

### **Frankenstein 3: East versus West**

#### **We fear the change of the culture**

The fear of AI explained through cultural spectacles leads us to the contrasts between East and West. You often hear people say that the Japanese are not at all afraid of robots. This difference with the West can be explained by the way people in Japan perceive AI. In Japan, AI is regarded as taming technology, whereas we in the West see it as conquering nature. The period of Enlightenment and Romanticism in the West and the Shinto religion in Japan form the basis of these different approaches. But both cultural routes – East and West – lead to fears of artificial intelligence.

### **Frankenstein 4: the fear of the Big Four**

#### **We are all afraid of the same things**

The four basic congenital fears that every person faces shed a new light on this matter. The fear of artificial intelligence is fueled by it, but AI can also have a damping effect on these fears. This also raises new ethical questions. Is the goal to reduce these fears with AI worth pursuing? This anatomy is based on research from experimental existential psychology (XXP) and the terror management theory (TMT).

## FRANKENSTEIN: SUPERINTELLIGENT AND SUPERSTUPID

**'As the fate of the gorillas now depends more on humans than on the species itself, so would the fate of humankind depend on the actions of the machine superintelligence.'**<sup>6</sup>

Nick Bostrom, philosopher and author of the book *Superintelligence*

Exactly two hundred years ago, Mary Shelley started to write the classic novel *Frankenstein; or The Modern Prometheus*. In this macabre story the eccentric scientist Victor Frankenstein imparts life to non-living material. In this way, Victor tries to create a friend and companion from body parts of various bodies from the local cemetery. Victor was repulsed by the creature he had created and in panic he fled his laboratory. He leaves his failed creation behind in fear, sadness and confusion. In revenge the monster destroys over the years everything its creator loves so passionately.

The fear of Frankenstein and similar characters is a recurring theme in many science fiction books and films. It is the world of mutants and scare

tactics, of demigods and mythical figures, monsters and other planets and killer robots. Isaac Asimov, another famous science fiction writer, coined this terror phenomenon as the 'Frankenstein complex': human-like entities (androids) or mechanical versions thereof (automata) fill us with fear and anxiety. Usually these stories do not end well for humans. Extinction of the entire species is imminent.

The fascination of man with his creations, with the threat against humanity by something greater than himself, the fascination with death and survival or even of being immortal, is inextricably linked to human history and our human existence. In the thirteenth century Albertus Magnus, a Catholic bishop, reportedly built an automaton which

6 Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 3 July 2014  
 7 Godwin, W., *Lives of the Necromancers or The Exercise of Magical Power*, 1876, p. 180, <https://archive.org/details/livesnecromance04godwgoog>

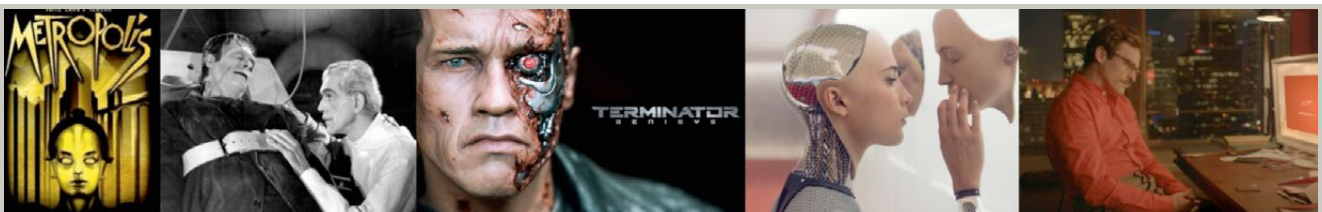
he called Android and which performed small household chores.<sup>7</sup> The famous philosopher and theologian Thomas Aquinas, pupil of Albertus Magnus, destroyed the android because it was too disconcerting. There are stories that the android was of flesh and blood and that Aquinas was particularly confused by the endless chatter of the 'robot'. In Greek mythology we find the predecessor of the iron man, Talos, a bronze automaton made by Zeus, who had to protect Europe from destruction. The modern equivalent of the iron man can be seen in the famous Iron Man films. The films are based on comic strip hero Tony Stark (the first comic book was published in 1963) and the first film in 2008 made 98,618,668 dollars in its first week. We have to hand it to Hollywood and the entertainment industry: they know like no other how to turn artificial intelligence into business.

### **Hollywood horror, game over**

Without really thinking, everybody very quickly reaches the following conclusion: artificial intelligence is coming, so game over for mankind.<sup>8</sup> Because the brain is so

important for our identity, people tend to see AI as the thief of our individuality, says Kevin Kelly. This is the root of all problems (and fears). Kelly is the founder of the famous *Wired Magazine* that closely follows the developments of the digital culture. With his recent book, *The Inevitable*, Kelly contributes to the lively AI debate. Kelly is a declared techno-optimist. In his book he says that all humanity will benefit and that ultimately AI will be a good thing for everyone. But Kelly has doubts whether his message comes across. It is very difficult to compete against the Hollywood scenarios, he says.

**'There's a lot of possibilities, while in the alternative Hollywood version, there's only one story: we die.'**



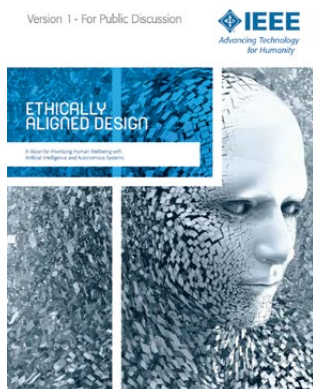
Hollywood horror: films such as *Metropolis*, *Frankenstein*, *Terminator*, *Ex Machina* and *Her* have been painting terrifying visions of the future for decades.

8 'Wired founder Kevin Kelly on letting go of AI anxiety', <https://slackhq.com/wired-founder-kevin-kelly-on-letting-go-of-ai-anxiety-bcd94e50dedc#.amf4qbyxp>

According to Kelly, the idea that we ourselves are no longer relevant, no longer required, is the source of all fear of AI.

With this Kelly broaches an important question: why are we actually afraid of artificial intelligence? In a study at Stanford University, the 'One Hundred Year Study on Artificial Intelligence' (yes, you read that correctly, the study will take a hundred years), the researchers note in their first report in 2016 an exaggerated fear among the general public, but also excessive optimism. The optimists and the scaredy-cats are both dismissed as being 'unrealistic'. The step from smarter algorithms and smarter products – from refrigerators to cars and care robots – to a Hollywood scenario is indeed easily made, at least in our thoughts or perhaps in our subconscious. The seemingly realistic scenarios for the introduction of your new product or service can easily founder on the 'unrealistic' feelings of the potential customers.

Numerous scientists, newspapers, websites and CEOs of technology companies warn against the risks of this new era.



Hollywood goes one step further and presents us with a scenario of how we will end up, which is usually not good. The cumulative effect of this media violence is that we are always on the alert when the term 'artificial intelligence' is used.

## **Interference from governments in Europe and the United States**

Europe seems to be in the grip of fear. A draft resolution, which was sent to the European Parliament at the end of May 2016 and adopted in February 2017, describes the fear that artificial intelligence will affect the intrinsic European and humanistic values.<sup>9</sup> In the first section Mary Shelley's Frankenstein is seamlessly connected with the emergence of robots and smart algorithms. The resolution expresses the risks and fears that exist concerning the development of artificial intelligence and the actions that must be taken.

The resolution addresses, among other things, the dehumanization when care robots take over human tasks and proposes that robot engineers sign a code of conduct respecting the 'dignity, privacy and safety' of people.

An IEEE report of December 2016, *Ethically Aligned Design: A Vision for prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, gives a fairly detailed picture of what a mandatory code of conduct (as discussed in the European resolution) would look like. For example, the IEEE advocates AI systems that are transparent in their pre-

9 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BD0C%2BPDF%2BV0//EN>

dictions. This means, among other things, that they must be able to 'explain' afterwards what they did.

The report for the European Commission calls for a European agency for robotics and AI that has to set standards and take measures to protect the consumers. All 'smart robots' must be registered with this institution. This is one of many proposals to ensure more transparency in algorithms and the operation of (ro)bots.

A similar report was published in the United Kingdom: *Robotics and artificial intelligence*.<sup>10</sup> One of the action items in this government document is to stir up the public debate about AI and to set up a special commission for this. Furthermore, the United Kingdom agrees with the strict privacy laws in Europe.

The Americans have their 'National Artificial Intelligence Research and Development Strategic Plan' of the National Science and Technology Council (NSTC). This policy paper advised president Obama in October 2016 about the impact of transformative technology, as artificial intelligence is called. In this paper we read that AI agents must above all demonstrate 'good behavior', just as you would expect from your fellow man. Americans also stress the importance of transparency. It should be easy to understand for people what the AI application does and on what information its decisions are based. So America and Europe both advocate transparency and good behavior. However, Americans also

express another fear: the fear of losing from the Chinese when it comes to AI and the fear of not exploiting all the potential in economic terms.

## **Brakes on innovation**

You can download and read all the above documents through the links included in this report. The concise list of 23 AI principles in appendix A of this report gives a good impression of what the issues in these reports entail. The 23 rules in the appendix are called the 'Asilomar AI Principles', named after the place where this AI conference took place at the start of 2017. Among other things, the above-mentioned government papers were input for academics, AI experts and leading figures such as Stephen Hawking and Elon Musk to formulate workable principles for our AI future. Regulation by governments or self-regulation by industry and academia can ensure that the brakes are put on innovation. The fear that this will slow down innovation is justified. The four reports (US Government, European Parliament, British House of Commons and IEEE) are now all calling for measures to make AI more transparent. After transparency and clarity, it is easier to also establish rules for the so-called good behavior that the Americans advocate, or the protection of our dignity, privacy and safety for which Europe and the UK argue.

## **It's getting out of hand: superintelligence**

The sensational start of the European resolution, the reference to the story of Frankenstein, is relevant if we look at the

10 House of Commons Science and Technology Committee, *Robotics and artificial intelligence*, September 2016, <https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

current debate about the fears associated with the emergence of AI. Frankenstein, after all, is only a story, a fantasy, and you don't want to obscure a sober reflection of AI by such a powerful narrative. At least, that is the opinion of the school that wants to continue to have a level-headed view of the facts; we call this 'camp 2'. Camp 1 is represented by people who warn against superintelligence and who are not afraid of using hyperboles such as Frankenstein to convince us of the danger. In daily practice, however, we do not have to deal much with robots, mutants and horror characters from Hollywood. The rapid emergence of machine intelligence does ensure an increase in the horror factor. Indeed, the boundaries between humans and machines are fading. The

main 'news' that dominates the media is the possible advent of superintelligence causing everything to get out of hand. Nick Bostrom, philosopher and author of the book *Superintelligence*, created his own code of conduct, the MAXIPOK rule, especially for this: build in the maximum (MAXI) probability (P) of an OK outcome (OK) in the algorithm, because a small deviation to the left or right could result in very serious consequences in the long run.<sup>11</sup> Incidentally, Bostrom blames the entertainment industry for wanting to score cheaply, for example by using articles about AI as 'clickbait'. However, at the same time he frightens people with his statements that we are small children playing with a bomb and that AI is more dangerous than global warming.

### **Camp 1: The end of mankind is near**

Elon Musk (CEO of Tesla), Steve Wozniak (co-founder of Apple), Bill Gates (founder of Microsoft), the aforementioned Nick Bostrom and Stephen Hawking (physicist) belong to the camp that says that sooner or later we will lose the race against the computer and that this could cause extremely dangerous situations. Bostrom thinks it could take another 5 centuries, but perhaps it will take only 50 years. Gates says he does not understand why there are not more people who are worried about superintelligence. Wozniak thinks that the dominant computers will treat us as their pets and he thinks this is good news (because people in general treat their pets kindly). Musk says that the end of mankind is imminent and believes in the (small) chance we are already living in a computer simulation. Motivated by fear, Musk is investing heavily (\$ 1 billion together with other investors) in Open AI because he wants to 'democratize' it and does not want to leave it in the hands of a few big players. Shane Legg of DeepMind, the company which was bought by Google and which beat the world champion of the board game Go, is also afraid of the future. He disagrees with his boss Eric Schmidt, who thinks it's all nonsense.

11 <https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>



## Camp 2: Ravings. There are more important things to worry about.

The following persons are not so worried about computer supremacy: Eric Schmidt (executive chairman of Alphabet), Pedro Domingos (machine intelligence professor at the University of Washington and author of *The Master Algorithm*) and Luciano Floridi (philosopher, professor at the University of Oxford and author of *The Fourth Revolution*). Domingos comes with a sobering conclusion: the real problem is not that machines are too smart and therefore everything gets out of hand, but that people are too stupid to realize that we have passed this point long ago. Shut down the Internet and we are heading straight for a catastrophic end of our civilization. Schmidt thinks that not everyone should participate in the debate, only the software engineers. Lastly, Floridi believes it is very unlikely that computers will ever become superintelligent. He compares the growth spurt that we are currently experiencing with AI with climbing up a tree. When you reach the top, it means that you are at the end of your trip and not that you are on the way to the moon. Deep Learning may now be able to recognize images independently, but that is still a long way from anything resembling self-consciousness.

From the perspective of warfare there are also fears for artificial smartness. A large number of leading scientists and technology experts take the view that we should never start developing autonomous weapons systems.<sup>1213</sup> They argue that autonomous weapons may have benefits such as reducing the number of casualties on both sides, but that this does not outweigh the risks. For example, they fear the consequences if autonomous weapons fall into the wrong hands and they are afraid that countries will be inclined to go to war sooner if the risk of casualties on their own side is smaller.<sup>14</sup>

According to reports from the American Department of Defense, robots will play an increasingly important role in future battlefields.<sup>15</sup> In anticipation of this, researchers from Harvard Law School decided in the summer of 2016 to coin the term 'war algorithm'.<sup>16</sup> It is a legal description intended to circumvent the difficult concept of 'autonomous'. An algorithm meets the war algorithm classification if 1) it is expressed via computer code, 2) it can make decisions independently and 3) it can operate in an armed conflict, even if it is not specifically designed for it. This last point may mean that self-driving cars

12 Autonomous weapons are also called Lethal Autonomous Weapons (LAW), Lethal Autonomous Robots (LAR) or simply killer robots.

13 <http://futureoflife.org/open-letter-autonomous-weapons/>

14 Walsh, J. & M. Schulzke, 'The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?', United States Army War College, September 2015, <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA621793>

15 Stewart, J., 'Strong Artificial Intelligence and National Security: Operational and Strategic Implications', U.S. Joint Military Operations Department, 18 May 2015, <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA622591>

16 [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2832734](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2832734)

could also fall under the legal terminology of war algorithm.

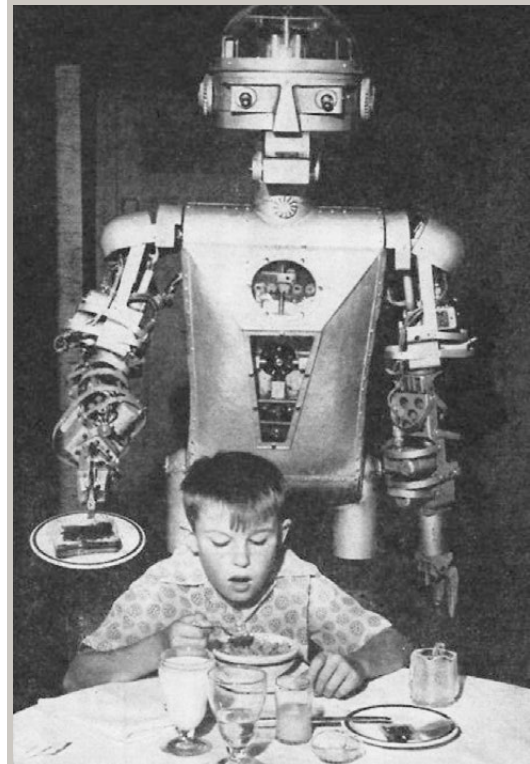
### **It's getting out of hand: superstupidity**

Apart from the fear in the long term – superintelligence – there is also the fear in the short term for specific problems that arise because AI is still too stupid in many areas. The name of the article says it all: 'Specific Problems in AI Safety'<sup>17</sup> is about safety issues that arise when self-learning systems go awry. This article expresses the concerns of engineers and AI experts at Google who enumerate the limitations of self-learning systems, and reduce it to a list of things which we should be (really) afraid of. To illustrate this idea they have taken as an example a 'breakfast robot', a device that is capable of serving breakfast and cleaning up afterwards.

The following is what might go wrong:

#### **This is what we should also be afraid of:**

1. *Causing collateral damages.* The robot knocks over vases and other things, because he pursues his goal too rigidly: to serve breakfast faster.
2. *Reward hacking.* If the robot closes his eyes (computer vision switched off), he no longer sees the breakfast and therefore does not need to clear the table. This may lead to 'gaming the system'.
3. *Unable to make a distinction.* When clearing up, the robot throws away



The breakfast robot. Anything can go wrong if it is not smart enough.

- your school diploma. How does he know the difference between a napkin and a valuable document?
4. *Gaining painful learning experiences.* In order to learn, the robot needs to experiment, for instance with cooking techniques. But an egg in the microwave is not a desirable situation.
5. *Displaying inappropriate behavior.* A robot that makes breakfast at the neighbors and comes to work for you may miss changes in the surroundings and make mistakes.

17 Amodei, D., et al., 'Concrete Problems in AI Safety', Cornell University Library, 26 July 2016, <https://arxiv.org/abs/1606.06565>

To this list of Google's researchers we can add the list of the World Economic Forum. In their Global Risks Report 2017<sup>18</sup> they state the following three risks of emerging technologies such as AI:

6. *Cyber criminality*: the risks of hacking systems in the physical infrastructure, which could lead to a cascade of problems.
7. *Too little or too much governance*. Too much new legislation may slow down development and kill innovation. Too little governance may direct the technology in the wrong direction.
8. *Bugs in autonomous weapons systems* that could lead to catastrophic consequences.

You don't hear much about these kinds of problems; it is mainly a topical debate among engineers. However, when the Microsoft AI chatbot, called Tay, went completely off the rails, it was all over the media. By pestering the public, that interacted with the chatbot, Tay became a racist holocaust denier. Google already had to apol-

ogize once before because its smart algorithm labeled photographs of black people as 'gorillas'. These kinds of stigmatization and insult are painful and we see that it can happen intentionally (Tay) and accidentally (Google).

On the part of cyber feminism there is a serious concern about these sometimes unintended and often unanticipated consequences. Erika Hayasaki, associate professor of literature and journalism at the University of California, wrote in *Foreign Policy* magazine an extensive article on whether artificial intelligence is good for women or not.<sup>19</sup> Her conclusion is that sexism, racism and misogyny will be widely disseminated if we do not impose more rules on the smart algorithms. The prejudices of their creators learned by algorithms spread easily across the Internet. And the robots themselves are often hyperboles – sexy women, macho men – affirming the traditional male and female roles instead of breaking through these gender stereotypes.



Tay, the AI chatbot that went off the rails.

18 World Economic Forum, *The Global Risks Report 2017*, <http://reports.weforum.org/global-risks-2017/executive-summary/>

19 Hayasaki, E., 'Is AI Sexist?' *Foreign Policy*, January 2017, <http://foreignpolicy.com/2017/01/16/women-vs-the-machine/>

## FRANKENSTEIN: FEAR OF OURSELVES

First, we can conclude from the above that camp 1 feels extremely uncomfortable with the idea that machines are becoming superintelligent, that camp 2 trivializes this idea and that it will be some time before engineers succeed in making AI really intelligent. Apart from this rational fear, involving cognitive judgments about future events, we also have the subconscious fear, the gut feeling. Floridi (camp 2) has a remedy for the irrational fears. He says that we should just imagine that we are no longer afraid:

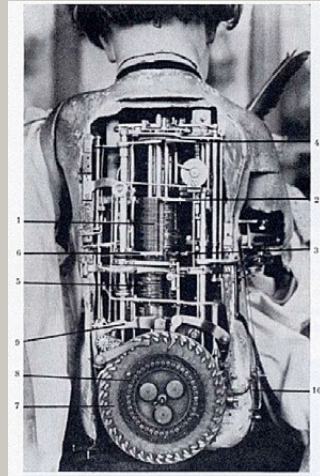
**'Suppose you enter a dark room in an unfamiliar building. You could become afraid that monsters might emerge from the dark, or you could turn on the light to avoid bumping into an obstacle.'**<sup>20</sup>

However, that is easier said than done, because the subconscious mind, where these fears manifest themselves, goes its own way. A hundred years ago Sigmund Freud and his colleague Ernst Jentsch already wrote about the subconscious fears of automata (which, for convenience's sake we call artificial intelligence here). In his article 'The Uncanny' (1919), Freud writes that the confrontation with humanlike phenomena, such as mechanical dolls, evokes an *unheimisch* (uncanny) feeling, precisely because we feel affinity with them. We see our *doppelgänger* in these mechanical dolls or robots. He emphasizes that the word *heimisch* means 'to feel at home', 'to feel affinity with' as well as the opposite: 'secret'.

Grammatically, *unheimisch* (uncanny) has a double meaning (non-related and non-secret), which fits in with Freud's theory that we hardly know ourselves and are therefore not transparent for ourselves. This is also important in this context because the affinity that we feel with the AI *doppelgänger* also evokes a feeling that

<sup>20</sup> <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>

In the analysis of Jentsch and Freud, an important role is played by Olympia, the 'living doll' (containing a mechanism to mimic gestures and sounds) from the book *The Sandman* by E. Hoffmann from 1816



has to do with not knowing ourselves. AI, or in this case automata, as a mirror image also confronts us with the bad aspects of our personality and therefore with humanity as a whole.

### **Uncanny: not knowing the other**

In his article, Freud also talks of the *doppelgänger* and the subconscious confrontation with death which is always present somewhere below the surface. In practical terms it means, according to Freud, that we start thinking about these things as 'savages'. The subconscious mind controls our thoughts and behaviors. If we see devices moving or performing mental feats that you would only expect from humans, we find that uncanny, says Ernst Jentsch in his 'On the Psychology of the

Uncanny' from 1906. In his article, Freud dissects the horror factor even further and notes that the fears mainly live in the imaginary world of books and theatres and stories before bedtime. According to Jentsch that is the place where we want to experience those feelings. But fantasy and reality are often intermingled and it does not take much to arouse childish fears in people (a crocodile under the bed or the dead coming to fetch you). In this context Freud talks in his article about so-called intellectual people who are nevertheless always driven by non-intellectual fears. It is interesting to read in Freud's article that two mechanisms that evoke an uncanny feeling, are closely related to the possibilities of AI and Big Data today: 'recurrent similarities' and the 'omnipotence of thought'.

### **Recurrent similarities**

We get an uncanny feeling when we see a recurring pattern in chance events for which we have no explanation. Passing a house with number 62 and reading a newspaper headline containing the number 62 exactly on the day of your 62nd birthday is an example of this. In the Big Data world which we live in today, we experience these kinds of 'recurrent similarities' more and more often. Freud says that the repetition of the number 62 and the inexplicability thereof evokes the internal fear that we are developing an obsessive-compulsive neurosis and that we are repeating actions without knowing why. The argument that Freud keeps coming back to is that the fear of automata stems from other anxieties hidden inside us.

### **Omnipotence of thoughts**

One of Freud's patients said about a certain person that he could drop dead as far as he was concerned. When this actually happened a few days later, this caused an enormous fear in that patient – a belief that his thoughts could really bring this about. The fear of what Freud calls the 'omnipotence of thoughts', is still hidden in every human being and awakens when these things occur. In the AI fantasy there is the story about the 'psychic pizza': the mere thought of a pizza is enough to make a pizza delivery man call at your door half an hour later. A funny futuristic fantasy of AI engineers can thus stir up the feeling in our subconscious that an all-seeing 'Evil Eye' really exists.

### **IT with which you feel familiar**

As Freud said, the uncanny is both in the known and in the unknown. The uncanny of the AI anthropomorphism, man as a *doppelgänger* for which one is not afraid, is a completely new area which we are only now starting to discover. In our previous report *The Bot Effect: 'Friending Your Brand'* we already alluded to the possibilities of more intimate relationships between organizations and their customers. The (ro)bot as *doppelgänger* of the organization is capable of creating different emotional connections. The boundary between uncanny and familiar,

intimate or unwanted, seems to get lodged in the subconscious mind. Knowledge of the subtle differences that cause AI to become uncanny, will determine to a large extent the success of the applications in the market.

### **Finally**

There is a renewed interest in theories that focus on emotions and the subconscious mind. We say renewed, because after Freud it was mainly cognitive psychology that has dominated during the last decades: almost all human behavior can be explained on the basis of rational

considerations. Deviations in behavior according to this cognitive school are 'small faults' in the human system of information processing. The modern variant of Freud, which has been on the rise since the 1980s, is called existential psychotherapy, or experimental existential psychology (XEP). This movement does not deny the importance of Freud's work, but says that there is something that is slightly more important than what a person has experienced in their youth. Namely, something that is certainly going to occur in the future: the (inevitable) death, the denial and acknowledgement thereof and how

that determines what we do. Death and a number of other matters that we are all faced with, with which we are always confronted, no matter who we are and where we live. Before we examine what these neo-Freudians teach us about the fear of artificial intelligence, we will first zoom in on the question of why the horror factor in Japan appears to be less prevalent. The lesson that we can learn is that Japan has a long tradition in taming technology that sustains the culture. In the West meanwhile, since the Enlightenment, technology is far more in competition with human nature and is the big favorite to win.

### Uncanny valley

With his famous 'uncanny valley' robot scientist Masahiro Mori makes a connection between the extent to which the robot resembles a human being and the extent to which we find the robot uncanny. The main postulate of this theory is that robots become scarier as they more closely resemble people (for instance through motion), until they look so human-like that you can no longer see the difference. On the left you see robots C-3PO and R2-D2 from *Star Wars*, Han from Hansen Robotics, next to that professor Hiroshi Ishiguro with his *doppelgänger*, then Mark 1 from Ricky Ma, and on the right Peter Cushing who, after his death, was revived in the *Star Wars* film *Rogue One*. No matter how plausible the theory may seem, there is not a lot of scientific evidence for the operation of this 'uncanny valley'.





## FRANKENSTEIN: FEAR FOR THE END OF THE CULTURE

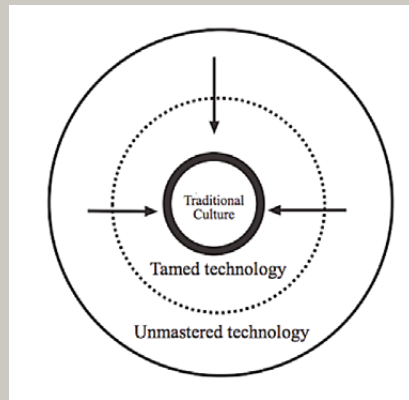
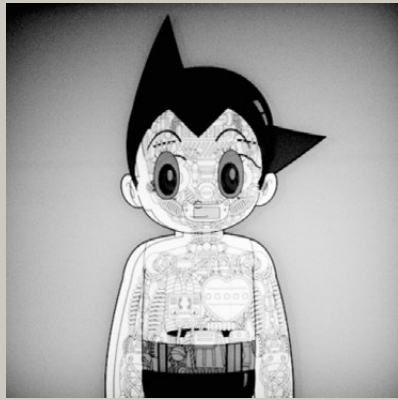
How about Japan? It is an often-heard argument: only in the West are we afraid of AI and robots. In the East and particularly in Japan things are very different. The explanation from the various opinions in culture seems plausible, but the scientific proof for a drastic difference in experiencing fear is thin.

The standard work from Frederic Kaplan, 'Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots', provides a good insight into the cultural differences and the consequential explanation of the reactions to human robots. Kaplan worked for the Sony Computer Science Lab and for several years now has led the Digital Humanities Lab of the Polytechnic University of Lausanne. It appears, according to Kaplan, that everyone in Japan is obsessed by robots, because people react enthusiastically and everybody seems to easily accept them. However, Kaplan turns the proposition around: it is us (in the West) who are obsessed by them and therefore are less quick to accept them. In Japan they do everything they can to control the new technology (robotics) and allow it to serve the existing traditions. It looks more like a battle in the West.

The title of the book *Race against the Machine* from Erik Brynjolfsson and Andrew McAfee correctly presents the sentiment in the West. It is mankind or the machine that will win. Kaplan uses the word 'tame' for the Japanese approach.

Taming the machines, the new technology, even became formal politics in the second half of the nineteenth century. Japan learned from the technological innovation that arrived from the West and deployed the same technology to strengthen the traditional Japanese culture. The artificial recreates the natural. This explains the higher degree of acceptance and possibly also the Japanese aversion against cyborgs: half man, half robot. Integrating the technology in the culture is central, not assimilating. As soon as the foreign culture takes over, people want nothing to do with it.

In Japanese fiction, such as comic strips, we see how this taming of technology works. For example with the popular robot Astro Boy (also known as Mighty Atom), who walks around with an atomic heart. When his robot maker rejects him, he is warmly received by people into our society. The robot stories in Japan are often about victory over something from outside, by stealing something from the technology of others (atomic energy) in order to defend their own culture. That Mighty Atom has become so popular after Japan experienced the horrors of the atom



The tamed technology supports the traditional culture in Japan. Astro Boy (Mighty Atom) is an example of how this works in pop culture.

bomb, underlines how far this drive for integration goes.

You find these 'all's well that ends well' stories everywhere in Japanese robot fiction, including the stories about the mega robots which as a child you can sit in and acquire super powers. The child remains in control; there is no autonomous machine that takes over the power.

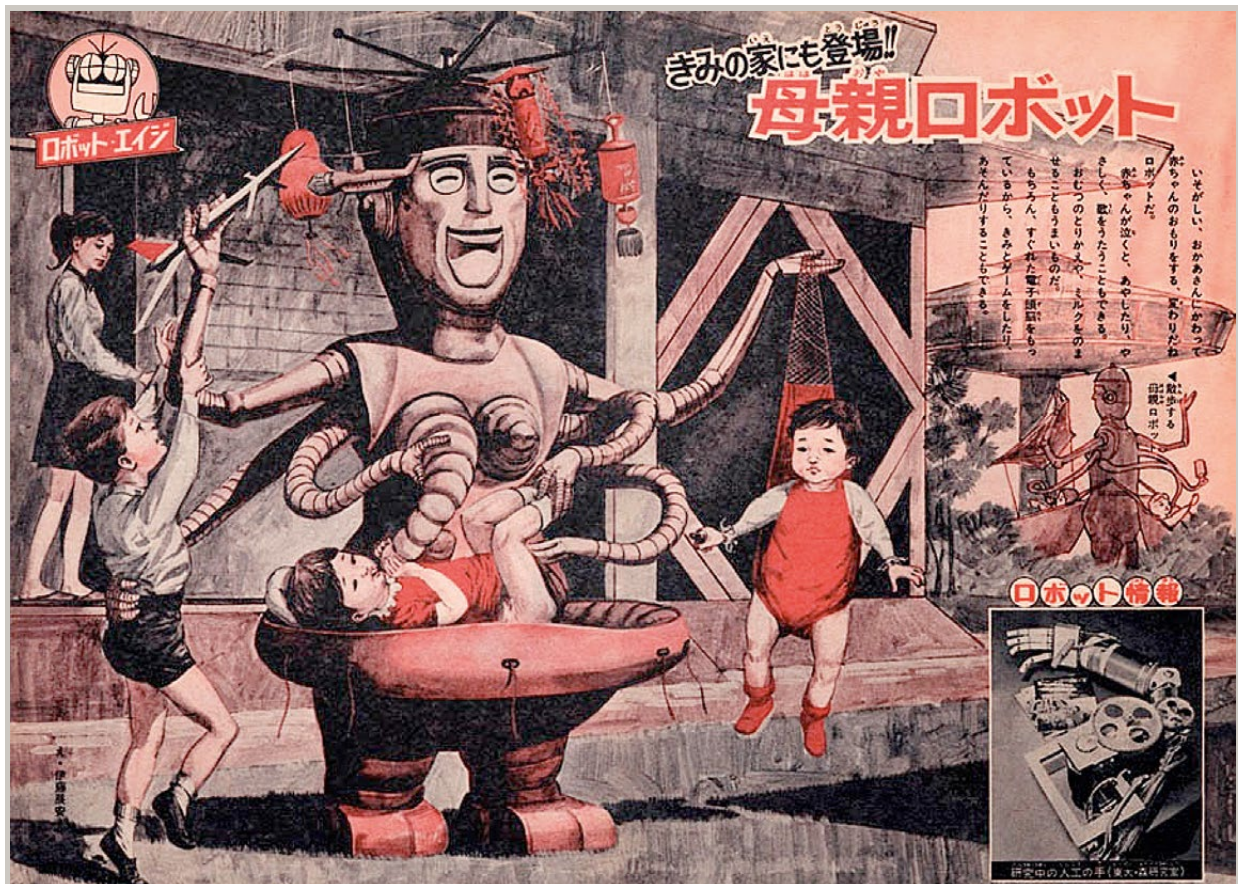
And thus Japanese children grow up with the image of good-natured robots that they can control and that are often capable of incorporating the role of both father and mother. These comic strips and films are a reflection of the culture, but are also a sounding board that raises the children with this idea of tamed technology. In the same article Kaplan reflects on the western relationship with robots. He argues that the Enlightenment has played a crucial role for us in the manner in which

we now associate with robots. During the Enlightenment, science, knowledge and technology were elevated above nature and created the well-known division: nature versus culture (of which technology and science are a part). In the western world it is often either nature or culture, there is no room for hybrid forms. In Japan there is no desire to separate these two, nature and the artificial are seen there as a holistic whole. As an example, Kaplan outlines the difference between a typical American and a Japanese fountain. The American fountain sprays water yards into the air, as a triumph of technology over nature. The Japanese fountain is more subtle, water rippling over a few stones, exactly as you expect to find it in nature, but then even more beautiful. Animism has an important place in Japan's Shinto culture and religion. This is a spiritual and philosophic concept that assumes everything has a soul. According to the Shinto

philosophy plants, stones, things, robots and people are all animate and therefore more connected (and less alien).

This Japanese instinct to reconstruct and perfect nature, versus the western instinct to conquer nature, can also be explained from this Shinto tradition. Everything is indeed one. It outlines the difference between East and West. In the West it is not only about the separation

of nature and technology, but also about the triumph of technology over nature. In western culture and philosophy reason is considered superior, and this reason in an algorithmic coating, artificial intelligence, is logically speaking the dominant force that will win from human nature. In Occidental stories, the counterpart of the Japanese Atom robot, this often does not end well. The moral, reflected in the stories, is that mankind is ultimately not suf-



The mamabot: Japanese robot that carries out care duties, plays with children, cleans nappies and gives milk.



ficiently competent to control nature. It is better not to play God. Freud's argument can also be raised here that mankind is not transparent for itself and is capable of building bad 'robots', because this badness is also represented in mankind itself.

The Romanticism and the work of Rousseau, that followed the Enlightenment, brought about an additional area of tension. The revaluation of nature from that period went hand in hand with the devaluation of culture and technology. The artificial drives mankind away from nature, according to philosophy, and

the happiness of mankind is affected. In short, the technology and automata that were still seen as noble products during the Enlightenment, also got a negative connotation after the eighteenth century. This area of tension is expressed in the Frankenstein story and the associated syndrome. Frankenstein was also abandoned by his creator, but he was not lovingly cared for by humans as in Japan. The superiority of the creature turns against us and challenges our culture, instead of it becoming ours and protecting us. The moral here is that messing around with nature and lifting us above it is not a good plan.



Animism leads to burial rituals for robot dogs in Japan

## **A fear of robots is a fear of ourselves and our culture**

Kaplan's conclusion is therefore that the real fear of robots is the fear of losing who we are. Because the devices can potentially force us to redefine that which we thought made us so specific. The fear of robots is the fear of discussing how we see ourselves. It is the fear that culture will be ruined, a fear that according to the philosopher Peter Sloterdijk is ward off with our narcissistic shield. Narcissism lifts and protects us. But ultimately mankind must redefine itself, because the new machines force us to do so.

## **The Frankenstein questionnaire**

Whether the Japanese are also really less afraid of robots has not or hardly been demonstrated. For example, there are studies that show that Americans are more negative about stereotype robots than the Japanese, but this is exactly the other way around with the more anthro-

pomorphic robots.<sup>21</sup> Other research shows that boys in the United Kingdom are far more positive about robots than boys in Japan, for example.<sup>22</sup> Fear and enthusiasm are universal; the East-West differences seem smaller than philosophies lead us to believe. The most applied research instrument to demonstrate fear differences is the so-called Frankenstein Syndrome Questionnaire. These are thirty questions that are frequently used in research to analyze the differences. Because it immediately provides a good insight into how the complete fear palette is looked at from a scientific viewpoint, we have included the complete list with questions and propositions in appendix B. It is a broad palette and it ranges from the fear that robots will make us lazy, that they will make decisions for us and that they will influence our children, to the fear that we will forget what makes us actually human.

21 <https://www.bartneck.de/publications/2008/whoLikesAndroids/bartneckRoMan2008.pdf>

22 [http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura\\_etal\\_AISB2015\\_differencesonsocial.pdf;sequence=2](http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura_etal_AISB2015_differencesonsocial.pdf;sequence=2)

## FRANKENSTEIN: FEAR OF THE 'BIG FOUR'

**'The more unlived your life, the greater your death anxiety. The more you fail to experience your life fully, the more you will fear death.'**

Irvin D. Yalom, *Staring at the Sun: Overcoming the Terror of Death*

We can hold long debates on the question from the previous section; is it good or bad that we control nature? The possible fears that are related to this are initially reasoned fears. This in contrast to this fourth and last component in our anatomy: four fears that we are all facing and that we need to deal with. These so-called 'Big Four' fears from the existential psychotherapy and the related schools from psychology are always present to a greater or lesser extent:

- fear of becoming isolated from the world and from people around us;
- fear of losing control;
- fear of leading a meaningless life;
- fear of death.

The subconscious plays a major role here as well; it is a neo-Freudian view of fear. The most remarkable difference is that Freud always looks backwards in search of explanations for behavior: the birth, primal instincts from evolution, the relation-

ship with parents, traumatic experiences et cetera. And the Big Four theory especially looks forward, to the inevitability of death that we are all confronted with and the existential givens that provide a universal explanation for that which people are afraid of. The anatomy of these Big Four fears deliver interesting new insights and questions. We particularly want to emphasize two of these in advance, before we give a further explanation about this fear theory.

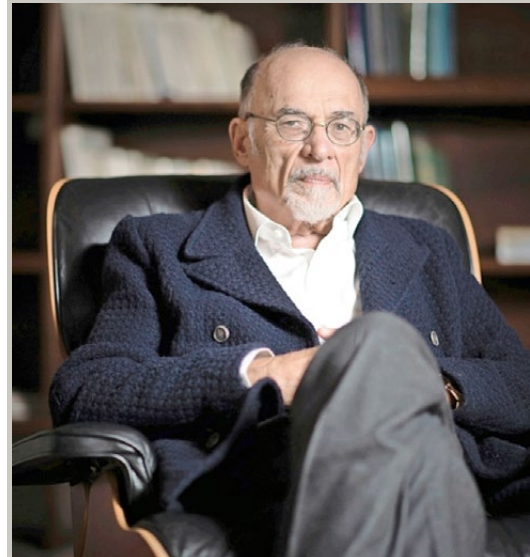
1. Up to now, in the anatomy, we have mainly looked at the fear phenomenon negatively. However, with the Big Four as a given, this can also be turned around and AI can be seen as something that can dampen existential fears.
2. The second is the conclusion that the fear of privacy does not belong to the four basic fears. Moreover, the fear of not being seen and leading a corresponding unseen life, is one of the four

main fears. This is especially relevant in the AI debate that is also about loss of privacy, but seldom about the opposite: the opportunity to be seen and to allow yourself to be seen. People experience a 'need to be seen'.

## Existential fear

Questions about our existence are often present in the background. Now and then our life is abruptly disrupted by the death of someone close to us, redundancy, a mid-life crisis or the end of a relationship. At that moment we are mercilessly confronted with our, previously latent, existential fears. A milder variant of a major personal loss is for instance the fear that arises if we board an airplane, if we look into the eyes of an android or if, in the future, we drive off in our self-driving car. Existential psychotherapy has a strong philosophic background and is influenced by works from among others Nietzsche, Kierkegaard, Husserl, Heidegger, Sartre, Maslow and humanistic psychology. There is also experimental existential psychology (XEP) and the terror management theory (TMT) – all schools that are closely related with each other and that have emerged since the 1980s.

One of the main contemporary experts of existential psychotherapy is Irvin D. Yalom.<sup>23</sup> Yalom has been active in this professional field for more than sixty years and he has specialized in group therapy and individual existential psycho-



Irvin Yalom, emeritus professor of psychiatry at Stanford University.

therapy. His contributions to this professional field, such as the books *Existential Psychotherapy* and *The Theory and Practice of Group Psychotherapy*, have assisted large numbers of new psychologists and psychotherapists. And with his fiction books, in which he translates his theories in an accessible manner to the life stories of his characters, he has inspired a large audience. *Nietzsche's Tears* is perhaps his best-known non-fiction book. In his book *Existential Psychotherapy* Yalom describes one of his most influential contributions to existential psychotherapy, the four ultimate concerns that every person has and the associated needs.

23 Yalom argues for inclusive psychotherapy whereby biology and genetics are examined for the origin of psychological problems (psyche-pharmacology model), our suppressed instincts (Freud), a dysfunctional relationship with parents (relational perspective), dysfunctional behavior patterns (cognitive behavior therapy) and also for our confrontation with our existence (existential approach).



### The four concerns:

- **Need to control**  
(fear of freedom)<sup>24</sup>
- **Need to be seen**  
(fear of isolation)
- **Need to survive**  
(fear of death)
- **Need to be needed**  
(fear of meaninglessness)

## Big Four: fear of death

These are four separate fears, but they are intrinsically and inseparably linked with each other.<sup>25</sup> They are actually all about death, or how we deal with this one certainty. This does not concern the fear of death that we experience when we stand eye to eye with a lion, but the fear of death in the long term and how we deal with it. We do that for example by not thinking about it too much, but by being busy with important matters from which we derive meaning. If that meaning ceases, then that existential fear surfaces once again. Solely for pedagogical reasons, says Yalom, it is convenient that we examine them separately and that is precisely what we will do.

The manual *Experimental Existential Psychology: Exploring the Human Confrontation with Reality* offers a current overview of the present thinking about these four fears and the prominent role that (denying) death plays in them. It also explores how themes such as beauty, spirituality and nostalgia fit in with this fear. The returning theme here is that it is

**'The fear of death plays a major role in our internal experience; it haunts as nothing else does; it rumbles continuously under the surface; it is a dark, unsettling presence at the rim of consciousness.'**

Irvin Yalom

always about coping mechanisms: dealing with the fear of death and damping this fear that is always lurking somewhere in the subconscious mind. Longing for the past, melancholy and nostalgia for instance, gives control of life, creates meaning in the light of death. Like the three other needs of the Big Four, the need to be seen (with excesses such as narcissism), the need to be in control (with excesses like the control freak) and the need to be needed (the workaholic) provide the same sort of compensation.

24 Yalom sees this fear in relation to the fear of ultimate freedom. Although many people want to be free, Yalom says that people are afraid of ultimate freedom: the awareness that reality is not structured and objective. We have to impose this structure for ourselves and bear responsibility for our own experiences in life.

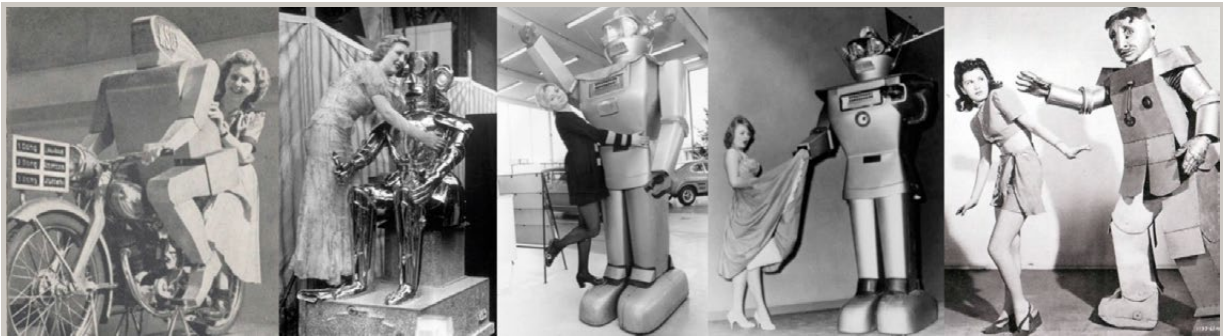
25 <http://www.freepatentsonline.com/article/North-American-Journal-Psychology/173513562.html>

## Big Four: fear of a meaningless life

The fear that our life has no meaning, forms the basis for one of the most important motivations in life: the search for meaning.<sup>26</sup> The fear that life is meaningless, according to Yalom's colleague, neurologist and psychiatrist Viktor Frankl, is even the ultimate fear. The relationship with artificial intelligence obviously lies close to the surface. Artificial intelligence is able to take over the tasks to which we attach meaning and activate the fear stimulus, the meaninglessness. This is also the argument from Kevin Kelly from the beginning of this report. The conclusion that we jump to: the robots are coming, game over for human beings. This is already frequently shown in science

fiction films, but perhaps best expressed in the film *Terminator 2: Judgment Day*. The mother of the ten year old John says, when she sees the cyborg playing with her son, that it is suddenly completely clear to her that this Terminator will never tire out. It will always continue to want to play, always give attention to her son, will never become drunk and hit him. This machine is the only one that can truly fulfill the role of father figure.

It therefore concerns the fear that I will be deprived of (all) meaning that I derive from life, by intelligent and affective devices. My job, my hobby and even my relationship. We can laugh about it, but also in this area there is a lot of speculation about the future possibilities of sex robots. It is therefore not surprising that the technological advances, e.g. at this level, are considered and experienced with trepidation.



Flirting with the idea that robots take over the role of men has been going on for some time. Even if they are only simple (and predictable) scenes, under the skin and unconsciously they can trigger the fear of meaninglessness.

26 Frankl places the desire for meaning above Nietzsche's desire for power and Freud's desire for lust. Frankl, V., *Man's Search for Meaning*, Beacon Press, 1 June 2006 (first publication: 1946).

Roy Baumeister, professor in psychology at the University of Florida, describes the mechanism about how giving meaning leads to positive self-respect and a 'need to be needed'. If we want our lives to have meaning, we need family, friends, customers, the organization and our pets to need us in some way or another (or so we think). We see that the mechanism has much to do with our wish to have the control over our life in our own hands. To give meaning you must:

1. Find a goal, so that you can give meaning to your activities. You do it, say, for a purpose.
2. You can then *justify* your own activities; you can justify your behavior.
3. That gives the feeling that you have a certain *grip* or control over your life.
4. *Positive self-respect* is the result.

### **The sensitive issue: the jobs**

We are not talking here about the debate of whether mass unemployment threatens as a result of AI, but about the individual matter of the loss of *my* job and the significance that *I* derive from it. This last point can be a very specific and realistic fear at the moment that my income is put at risk. However, the loss of meaning cannot be viewed in isolation from this. The existential fear of leading or to have led a worthless life now that the machine takes over the work, is just as relevant. The search now commences. Not only for a new job, but also for a new meaning. Something can be done about both fears: replacement, updating skills, mediation on the labor market and therapeutic coaching for the loss of meaning.

Kevin Kelly, the techno-optimist from our introduction, describes the rollercoaster that people experience in search of that other job. It is a psychological lesson in seven steps to which only one conclusion can be linked: insights can change; opinions and fears are not set in concrete. This Baumeister 2.0 version<sup>27</sup> gives possible comfort for everybody who encounters robotization on their career path.

### **The 7 stages of robot replacement**

1. *A robot/computer cannot possibly do what I do.*
2. *OK, it can do a lot, but it can't do everything I do.*
3. *OK, it can do everything I do, except it needs me when it breaks down, which is often.*
4. *OK, it operates without failure, but I need to train it for new tasks.*
5. *Whew, that was a job that no human was meant to do, but what about me?*
6. *My new job is more fun and pays more now that robots/computers are doing my old job.*
7. *I am so glad a robot cannot possibly do what I do.*

27 See <http://kk.org/thetechnium/the-7-stages-of>



It is the same Kelly who warned us about the power of Hollywood: the horror scenarios of the robots in the future and the end of mankind. Without claiming that we know whether structurally more jobs will disappear than are created, we do want to point out the fact that the fear scenario fares rather better in the media.

The disaster stories about mechanization and automation have been with us for many decades, as is also apparent from this newspaper report from 1928 and many other reports from the last century.

The illustration below is symbolic for the manner in which news about AI spreads via social media. The FrankensteinFactor fares better on clickbait.

There are numerous surveys that point out a loss of jobs due to AI automation.<sup>28</sup> The nuance that technology has provided even more jobs in the past 140 years is often mentioned, but that is not what lingers in the mind. The OECD calculated last year<sup>29</sup> that it is only about 9 percent of jobs that could be lost due to automation through AI. This corresponds with a recent study from McKinsey<sup>30</sup> that says that at most 5 percent of jobs will be completely automated. According to McKinsey we must take greater account of an automation of components of current jobs, so specific tasks instead of complete jobs.

### Finally: the fourth revolution

The larger picture is beginning to take shape. We (mankind) appear less exceptional than originally thought. Oxford professor Luciano Floridi focuses on this in his book *The Fourth Revolution*. With



28 See for example: Frey, C. & A. Osborne, 'The Future of Deployment', 2013, [http://www.oxfordmartin.ox.ac.uk/downloads/academic/The\\_Future\\_of\\_Employment.pdf](http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf)

Deloitte, 'Technology and people: The great job-creating machine', 2015, <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf>

29 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis', May 2016, [http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries\\_5jlz9h56dvq7-en](http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en)

30 McKinsey Global Institute, 'Harnessing automation for a future that works', January 2017, <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>

**'Freedom in this sense has terrifying implications; it means that beneath us there is no ground – nothing, a void, and an abyss. A key existential dynamic then, is the clash between our confrontation with groundlessness and our wish for ground and structure.'**<sup>31</sup>

Irvin Yalom

each major technological breakthrough or discovery we see the same pattern every time. Mankind falls slightly from its pedestal. We saw this happen four times, the fourth revolution is the IT revolution, which we are now in the middle of. The first time was when we discovered that the earth is not the unique stable center of the universe around which everything revolves. Our planet moves just like other planets and rotates around a different center: the sun (Copernicus and Galileo). We discovered that mankind does not form the center of the universe. That we are not the unique creations that we thought we were, but just like animals are part of evolution (Darwin), is the second. And that man does not know himself and is not as transparent as we thought, but that we possess dark emotions that we don't even understand ourselves (Freud), is the third. We are not the 'masters of the universe', not even the masters of our own intellect, and we now discover that we are no longer the absolute ruler

of language, reasoning, listening, predicting and analyzing. This is the fourth revolution, that of information technology, that has already been around for some time but with the arrival of AI has quickly forced us to face facts. What the technology does with us, in this case artificial intelligence, is just as important as what we can do with the technology. And what the technology does with us is that it puts our own meaning into perspective, and that can further stimulate the fear of the meaninglessness of existence.<sup>1</sup>

## **Big Four: fear of loss of control**

The fear of losing control is, to a greater or lesser extent, present in all of us. There is a control freak lurking somewhere in everybody. The control freaks, also called the internals, are happy to make decisions themselves, go deeper into the matter and

31 Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980.

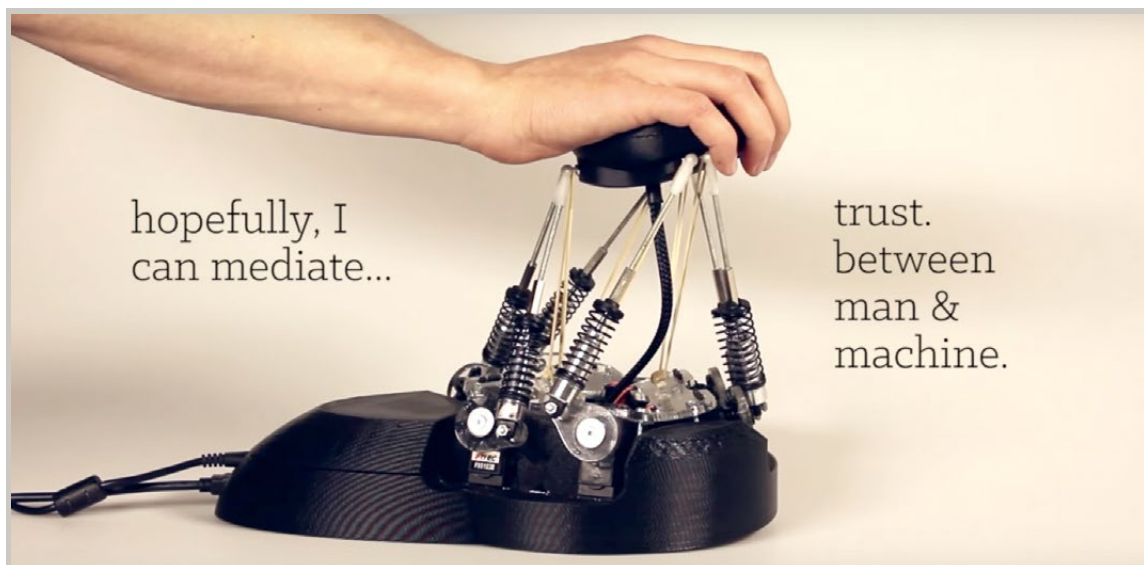
32 This concerns the psychological notion of 'locus of control'. See: Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980.



like to take matters into their own hands.<sup>31</sup> The counterparts of the control freaks, the so-called externals, are happy to leave things to others. This is also a way of not having to take responsibility for something. They would rather accept that a chatbot or app takes the decisions for them, as long as they do not have to think about it themselves. However, these are extremes, everyone wants to be 'in control' to some extent.

For example, many people prefer to drive themselves, rather than sit in the passenger seat (giving comments on the driver's actions), let alone handing over the steer-

ing wheel to an algorithm that controls the car. Recent studies show that only one in ten people have complete confidence in self-driving cars.<sup>33</sup> And 95 percent want to keep the full control panel (steering wheel, brakes and gas accelerator) even if they would hand over full control to the car. The fact that Mercedes-Benz is marketing its new self-driving car (F 015) as a 'quality time' concept is understandable, but also raises questions. Understandable, because with the video entertainment in the car you can spend your time watching lovely films (quality time). But the question is whether with the loss of control and the associated fear,



Stewart is a 3D joystick that acts as a mediator between the self-driving car and the 'driver'. Through Stewart the 'driver' can feel exactly what the self-driving car is planning to do, for instance accelerating while taking a bend to the right, and he can overrule this by moving Stewart the other way. This may be a solution for people who need to get used to handing over the control of the car.<sup>34</sup>

33 <http://www.dailymail.co.uk/sciencetech/article-3245058/Do-trust-driverless-cars-1-10-feel-confident-getting-autonomous-vehicles-women-skeptical.html>

34 [http://felixros.com/projects.html#big\\_wrap](http://felixros.com/projects.html#big_wrap)

it would really provide quality time for the driver. In any case, Mercedes-Benz takes away one of the fears associated with self-driving cars. The company announced that its self-driving cars will always prioritize the safety of the driver. You have probably heard this kind of dilemma before: whose safety does the car choose? A mother with a pram, an elderly person crossing the road ... et cetera. This can all go into the wastebasket. The car will choose its driver.

### **Autonomous or semi-autonomous**

In fact, we are faced with the fear of ultimate freedom, explains Yalom. The feeling of having lost the ground beneath your feet (figuratively) automatically evokes the reaction of wanting to take control to fill this void.

The lesson from this for AI applications seems simple. A certain amount of human control of autonomous systems is always preferable from the perspective of fear. The Valhalla of AI, a world in which everything is run by machines and algorithms, and where humans can simply enjoy their quality time, is at odds with this fear that everyone has. If we do not take action by taking life into our own hands and thus giving it meaning and supplanting the fear of death, we will definitely become afraid and experience the opposite of quality time. There are of course many different ways to shape the control of AI. It is not necessarily an obstacle for the use of AI applications, one should only take into account this 'need to control'.

## **Big Four: fear of isolation**

The story goes that when Yalom was working with people who had lost their partner, he was not only touched by their loneliness, but also by the despair of leading an unseen life. The thought that no one knew when they came home, went to bed and woke up, plagued them. Many people hold on to an unsatisfactory relationship because they crave a witness of their lives – a buffer against the experience of existential isolation. Our need for affiliation and interaction and our awareness that we are fundamentally alone in the world are in conflict with each other. A deep sense of affiliation does not simply solve the existential isolation, but it does offer solace.<sup>35</sup> People are social creatures. Chronic loneliness is rarely a choice.

It's easier to empathize with the fear of existential isolation when you stop to think about the fact that there are moments when nobody in the world thinks of you. At these moments we crave Facebook notifications, apps, software bots, holograms, smart thermostats and smart refrigerators that start talking to us. But also the radio in the background or watching TV can dispel such a sense of isolation. You could say that the smartphone has put an end to the notion of an unseen life. Google reports that 24 billion selfies were uploaded to their photo app last year.<sup>37</sup> Every second 9,000 new photos are uploaded to Snapchat (just under 300 billion a year), most of which are selfies. And this is not including Instagram, Facebook, Baidu and other media.

35 The existential fear of social isolation is about the awareness that we are all born alone and will die alone. This awareness does not disappear through social connection, but is softened by it. See: Wedding, D. & J.C. Raymond, *Current Psychotherapies*, Brooks/Cole, 2014.

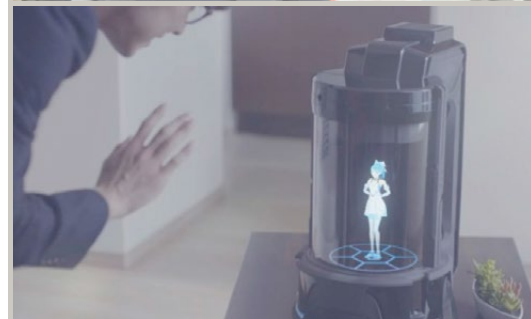
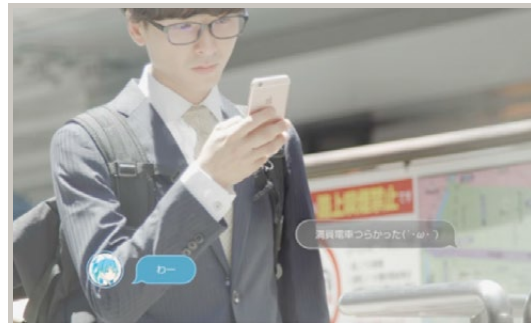


**'I mean, they say you die twice. One time when you stop breathing and a second time, a bit later on, when somebody says your name for the last time.'**

Graffiti artist Banksy<sup>36</sup>

The Japanese company Vinclu gives a disconcerting insight into the new possibilities. Their AI Gatebox product is for sale at 2500 euros and consists of a female hologram, a chatbot in terms of function and with features that you can also find in a product like Amazon Echo. In the promotional video, we see the hologram chatting with a young man getting ready for work. During lunch he receives a message from the chatbot that the hologram misses him and could he come home a bit earlier tonight. Once at home the man lets out a sigh and says: 'Somebody's home for me. Feels great.'<sup>38</sup>

It is a surrogate, but machine feelings can also be really experienced. The feelings that soldiers get in conflict situations for the robot helpers that they use when they are in danger, are a good example of this. People build a relationship with these machines, says defense specialist Peter Singer.<sup>39</sup> There are situations known that a soldier risks his life to rescue a robot, that a robot is given a military promotion or that soldiers organize a funeral when the material can no longer be repaired. We



For 2,500 euros, Vinclu's chatting hologram dispels your loneliness.

once again refer to our previous report about machine intelligence (*The Bot Effect: 'Friending Your Brand'*), in which we describe the anthropomorphic relationship and the emotional connection with AI. There is of course a dilemma in this surrogate solution.

36 <http://legacymultimedia.com/2015/01/09/when-is-your-memory-truly-forgotten/>

37 <http://www.dailymail.co.uk/sciencetech/article-3619679/What-vain-bunch-really-24-billion-selfies-uploaded-Google-year.html>

38 <https://www.youtube.com/watch?v=nkcKaNqfykg&feature=youtu.be>

39 Singer, P., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin Press HC, 2009.

**'If they can give the appearance of aliveness and yet not disappoint, they may even have a comparative advantage over people, and open new possibilities for narcissistic experience with machines.'**

Sherry Turkle

Sherry Turkle, professor in Social Studies of Science and Technology at MIT, warns that virtual placebos lead to the flattening of contact. Her book *Alone Together: Why We Expect More from Technology and Less from Each Other*<sup>40</sup> mainly concerns the use of the smartphone. The smartphone causes us to isolate ourselves from others because we give more attention to the smartphone than to the people around us. There are many studies that confirm that the loneliness of older people is reduced through the use of care robots.<sup>41</sup> So this is a good thing, these robots. In a study that Turkle carried out into care robots<sup>42</sup> she concludes that the use of robots can lead to more narcissism, people who are more involved with themselves than with others, but also that robots have a competitive advantage.

Researchers at the University of Sheffield also point out the potential dangers in their study 'Granny and the robots',<sup>43</sup> such as:

- the possible reduction of the amount of human contact;
- increase of the chance of objectification of the patient: seeing the patient as a 'thing';
- loss of privacy (for instance, because the robot stores data or because others can also watch);
- loss of personal freedom;
- deception and infantilization.

If robots start copying us and say that they feel alone and are happy when you come home again, then that is of course the world turned upside down. You can of course also argue that AI actually ensures more connections are made between people. Facebook reminds you about the birthdays of your friends, LinkedIn makes suggestions to connect with people who you might know. Or what about Bernie, the AI dating app that takes care of swiping and provides the initial opening sentence with potential partners? All on the basis of learning about your own taste. Perfect

40 Turkle, S., *Alone Together: Why We Expect More from Technology and Less from Each Other*, Basic Books, 2011.

41 See for example Robinson, H. et al., *The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial*, 2013.

42 Taggart, W., S. Turkle & C. Kidd, 'An Interactive Robot in a Nursing Home: Preliminary Remarks', 2005, <http://www.androidscience.com/proceedings2005/TaggartCogSci2005AS.pdf>

43 Sharkey, A. & N. Sharkey, 'Granny and the robots: ethical issues in robot care for elderly', 2010, <http://link.springer.com/article/10.1007/s10676-010-9234-6>

for those who want to play Casanova, but have no time to comb through all the possibilities.

## **Finally: breaking through of the protective cocoon**

The Big Four are recognizable and oppressive; in general, we do not want to think about this. Sociologist Anthony Giddens relates in his book *Modernity and Self-Identity* that we spin a safe cocoon around ourselves with which we can keep fears and threats outside. We develop mechanisms and mental strategies so that we think we are safe and trust that the world is 'normal' and predictable. We can often sustain this for a long time, but sometimes an event is so pervasive that the cocoon breaks and we stand face to face with our fear.

Due to its human characteristics and pervasiveness in our lives, AI provides additional moments of confrontation. When we make

an artificial intelligence that lives forever, we cannot help but immediately think about our own death. Our cocoon is penetrated at that moment; we are confronted with our fear of our own death. When we see how the people around us spend increasingly more time on their mobile phones or with their care robot, we are confronted with our fear of being excluded. In this way, time and time again, AI penetrates our protective cocoon. AI also works this way as a mirror: it shows us our deepest fears and puts us on the spot. Important questions about life arise unrequested: what does death mean? Do we want eternal life? What does 'real contact' mean? Can we have control over everything? And what is actually the meaning of my life?

Yalom applauds confrontation and sees this as the only way to deal with fears in a positive manner. Not with the objective of never being afraid again, but at least to discover a better relationship with our fears. Reason therefore to not avoid the fears and start the dialogue with your employees and customers.

## THE FRANKENSTEINFACTOR: CIO AS A THERAPIST

**'In any psychotherapy, the therapist himself is a highly important part of the human equation. What he does, the attitude he holds, his basic concept of his role, all influence therapy to a marked degree.'**<sup>44</sup>

Carl Rogers, in his book *Client Centered Therapy*

The fear of super smartness and stupidity, the fear of our subconscious, the fear of the decline of culture and nature and the four life fears with which people wrestle, are four perspectives to recognize

and analyze the fear. For convenience sake we list the factors Frankenstein 1 to Frankenstein 4 and connect them with each other.

### **Frankenstein 1 – Fear of smartness and stupidity**

These two fear extremes have the same origin. The AI applications get more autonomy and start operating independently. Things go wrong in the short term due to a lack of intelligence. In the long term it goes wrong because they are too intelligent. The underlying fear is the same: we have no control over the AI machinery, Frankenstein turns against us.

### **Frankenstein 2 – Fear of the uncanny**

The feeling, especially the gut feeling, reigns when confronting the unknown other person, the automata and the android. The Frankenstein that we have built, is us, we do not know ourselves well either. We project the evil that resides in every person on to the digital *doppelgänger*.

<sup>44</sup> Chapter 2, [https://books.google.nl/books?hl=nl&lr=&id=bWieBAAQBAJ&oi=fnd&pg=PT13&dq=therapy+roleplaying&ots=LwitLxFLi8&sig=gC\\_tFNqnZmcnaI-YAU-YjqOoqs8#v=onepage&q=Play%20therapy&f=false](https://books.google.nl/books?hl=nl&lr=&id=bWieBAAQBAJ&oi=fnd&pg=PT13&dq=therapy+roleplaying&ots=LwitLxFLi8&sig=gC_tFNqnZmcnaI-YAU-YjqOoqs8#v=onepage&q=Play%20therapy&f=false)

### Frankenstein 3 – Fear of the decline of culture and nature

Reservations with regard to technological progress are rooted in the western and eastern cultures. Both wrestle with the question of how far technology may advance and may change life. Taming technology and overcoming nature demonstrates the characters of East and West. After overcoming nature, nature can also turn against us. This story originates from the confrontation of the Enlightenment and Romanticism, that are in conflict with each other. It is the moral of the Frankenstein story.

### Frankenstein 4 – Fear of the Big Four

If we construct new Frankensteins with the current AI technology, which characteristics will we give them? Products that put people in the spotlight, letting them be acknowledged and seen, products that connect people and draw them out of social isolation, that give meaning to life and give us control, or vice-versa? With the other version we literally look death right in the eyes and we have actually built a real Frankenstein.

**‘The task of the therapist is to reduce anxiety to comfortable levels and then to use this existing anxiety to increase a patient’s awareness and vitality.’<sup>45</sup>**

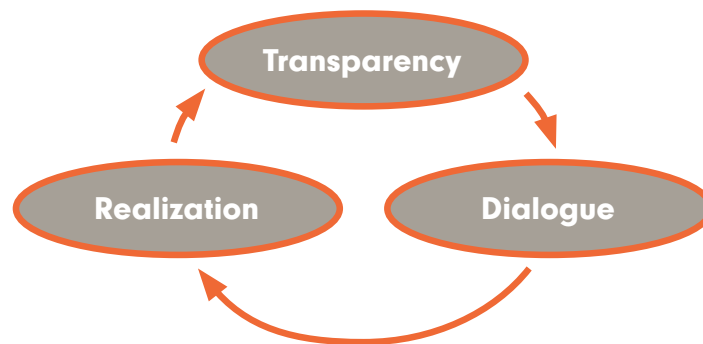
Irvin Yalom

The CIO is not alone in the challenge to retrieving the best out of AI. His stakeholders, the customer, the employees, his peers, the marketers and the shareholders all stand to gain from the fruitful deployment of artificial intelligence. It therefore concerns the entire organization and the brand, and the question is which role the CIO must fulfill. Is he the follower, the pure technologically driven enabler, or the initiator and the one who takes the lead? Some may say that it makes more sense that the CTO, the Chief Digital Officer or the Chief Marketing Officer concern themselves with this. It will differ for each organization and level of maturity, but it will be obvious that with this report we want to give the CIO a helping hand.

<sup>45</sup> Yalom, I.D., *Existential Therapy*, Basic Books, 1980, p. 188.

But whoever does it, it is a debate that must be carried out now. In the worst case the customer searches elsewhere for his business where his fears are taken seriously and the employee leaves to join a competitor that deals better with the gut feelings of his employees. What is the CIO to do then? A simple ABC formula, as suggested at the beginning of this

report, is by definition not fully consistent with the intense matter that we have dealt with here. At the same time, it is evident that fear prevails and that this must be addressed. The metaphor 'the CIO as a therapist' offers a solution: in therapy there are no simple ready-made solutions either, but there are conditions for a fruitful session:



**A. Transparency.** It is mentioned in reports from the EU, by Luciano Floridi and also in the 23 AI principles of Elon Musk and Stephen Hawking: that AI is surrounded by speculation and transparency can be seen as 'switching on the light in a dark room'. This means for the CIO that he must be transparent in the plans, activities and intentions that the organization has with AI. But also to already start operating in line with the European resolution adopted in February 2017, in which transparency plays an important role. Another thing is to initiate the discussion about the impact on jobs for employees and the need for retraining. What do they have to take into consideration and what not? For the customer it may be helpful if it is clearly stated what AI is capable of and what not; this will also contribute to minimizing fears and fantasies. We already mentioned Mercedes-Benz which reported for whom their self-driving car would choose when an accident is unavoidable. A transparency button may also help: if you press this button, the reasoning of AI will be explained and it becomes clear why this particular decision was made.



**B. Dialogue.** The transparency gives substance to, and opens the way for the dialogue. According to the new standards of psychology, you can also do this through showing your own feelings and uncertainties. With this Yalom relegates the classic ‘therapist listens and patient talks on the sofa’ to the past. It offers the possibility to ‘really’ make contact with the person behind the product and the system. Apart from efficiency and effectiveness, you also need to engage in the dialogue on existence, how it affects you, your employees and your customers. Your fear compass, that has hopefully become more sensitive by this report, will help you to recognize and identify the fears. More understanding from your employees and customers: this could put you ahead of the competition. It would be a good preparation to step into the shoes of the stakeholders and make an analysis of the four Frankenstein factors from their perspective.

Experiment in the current dialogue with your customer. ‘Exposure in vivo’ for example is a psychological method in which people are gradually exposed to situations that they fear. The method is used, for instance, to treat anxiety phobias.<sup>46</sup> This may be the solution for people who suffer from the fear of losing control because the pace of technological development is too high: create ‘intermediate phases’ or intermediary products that don’t remove all control. The intermediary between a self-driving car and a driver, Stewart, is a good example.

**C. Realization.** Applying the insights from this existential dialogue to your policy and your products. The insights from the analysis of the FrankensteinFactor are already pointing in a specific direction. With due regard for the Big Four, you will be able to provide the most reliable, valuable and popular IT products and services that you could wish for. No Frankenstein alienating or numbing us, but new man-machine relationships adding more meaning to life and en passant make your organization purpose-driven. Let AI support your core values and not be a subversion of humanity but rather a complimentary strength of humanity.

These are abstract tools and, as always, you will have to do the real work yourself. Keep at least your fear compass close at hand and address a fear as soon as you

recognize it. The CIO must set an example and switch on the light in the dark room, and perhaps the monster under the bed is not as scary as you thought.

46 <http://www.sciencedirect.com/science/article/pii/S0005796701000237>

## EPILOGUE: FEAR IS THE ZEITGEIST

There is another argument for the anatomy of fear that we have not yet discussed. If fear in general is becoming a trend and it becomes easier to put a fear label on things, this means that fear of AI is nothing but a reflection of the current Zeitgeist. Dominique Moïsi, founder of the renowned French think tank Institut Français des Relations Internationales (IFRI), argues that fear reigns today. He made a new fear analysis after watching popular series on Netflix, a project he called 'du bingewatching'. According to Moïsi these series are a reflection of the current sentiment and a mirror of our fears. He identifies five fears that prevail in our current society.

- fear of chaos and barbarism (*Game of Thrones*);
- fear of the decline of democracy (*House of Cards*);
- fear of terror or not knowing the identity of the enemy (*Homeland*);
- fear of the disappearance of the existing order (*Downton Abbey*);
- fear of a Russian occupation (*Occupied*).

We would like to add two popular robot series to this:

- fear of the dominance of robots over humans (*Real Humans*);
- fear of dominance of humans over robots and the fear of our own evil (*Westworld*).

If it is true that the fears from the popular series are a mirror of society, then the fear of AI is just part of a bigger story. According to Moïsi the fear that is present, is fed by the central theme of the 21st century: who are we, who do we want to be and how do we achieve this identity? The robot series theme fits this seamlessly. Now that AI can do and mean everything for us, what is the meaning of us then and how do we define our position with regard to our *doppelgänger*s?

# APPENDIX A

## ASILOMAR AI PRINCIPLES (2017)

### Research issues

1. Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
2. Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:
  - How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
  - How can we grow our prosperity through automation while maintaining people's resources and purpose?
  - How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
  - What set of values should AI be aligned with, and what legal and ethical status should it have?
3. Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.
4. Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
5. Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

### Ethics and Values

6. Safety: AI systems should be safe and secure throughout their operational life-time, and verifiably so where applicable and feasible.
7. Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.
8. Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.
9. Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.
10. Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.
11. Human Values: AI systems should be designed and operated so as to be

- compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
12. Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.
  13. Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
  14. Shared Benefit: AI technologies should benefit and empower as many people as possible.
  15. Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
  16. Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
  17. Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
  18. AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

### **Longer term issues**

19. Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
20. Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.
21. Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
22. Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.
23. Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.

# APPENDIX B

## FRANKENSTEIN SYNDROME

### QUESTIONNAIRE

1. I would feel uneasy if humanoid robots really had emotions or independent thoughts.
2. If humanoid robots cause accidents or trouble, I believe that the people and organizations developing of them will provide sufficient compensation to the victims.
3. Widespread use of humanoid robots would lead to high maintenance-costs for them.
4. I am concerned that humanoid robots would be a bad influence on children.
5. I would hate the idea of robots or artificial intelligences making judgements about things.
6. I feel that if we depend on humanoid robots too much, something bad might happen.
7. I don't know why, but humanoid robots scare me.
8. Many humanoid robots in society will make it less warm.
9. Something bad might happen if humanoid robots developed into human beings.
10. Widespread use of humanoid robots would take away jobs from people.
11. Humanoid robots can create new forms of interactions both between humans and between humans and machines.
12. Humanoid robots may make us even lazier.
13. Humanoid robots can be very useful for caring the elderly and disabled.
14. Humanoid robots should perform repetitive and boring routine tasks instead of people.
15. I don't know why, but I like the idea of humanoid robots.
16. Humanoid robots can be very useful for teaching young kids.
17. Humanoid robots are a natural product of our civilization.
18. Humanoid robots can make our life easier.
19. Humanoid robots should perform dangerous tasks, for example in disaster areas, deep sea, and space.
20. I am afraid that humanoid robots make us forget what it is like to be human.
21. The development of humanoid robots is a blasphemy against nature.
22. I feel that in the future, society will be dominated by humanoid robots.
23. The technologies needed for developing humanoid robots are amongst those fields that humans should not advance too far in.
24. The development of humanoid robots is blasphemous.

25. The people and organizations that develop humanoid robots can be trusted.
26. The people and organizations that develop humanoid robots seem sincere.
27. I trust the people and organizations that develop humanoid robots to disclose sufficient information to the public, including negative information.
28. Persons and organizations related to development of humanoid robots will consider the needs, thoughts and feelings of their users.
29. Interacting with humanoid robots could sometimes lead to problems in relationships between people.
30. I am afraid that humanoid robots will encourage less interaction between humans.



## SOURCES

- Adams, T., 'Artificial intelligence: "We're like children playing with a bomb"', *The Guardian*, 12 June 2016  
<https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>
- Arntz, M., T. Gregory & U. Zierahn, 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis', OECD Social, Employment and Migration Working Papers, 14 May 2016  
[http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries\\_5j1z9h56dvq7-en](http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5j1z9h56dvq7-en)
- Amodei, D. et al., 'Concrete Problems in AI Safety', Cornell University Library, 26 July 2016  
<https://arxiv.org/abs/1606.06565>
- Bartneck, C., 'Who like androids more: Japanese or US Americans', Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, 1-3 August 2008  
<https://www.bartneck.de/publications/2008/whoLikesAndroids/bartneckRoMan2008.pdf>
- Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 3 July 2014
- British Science Association, 'One in three believe that the rise of artificial intelligence is a threat to humanity'  
<http://www.britishscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity>
- Chapman University, 'America's Top Fears 2015', 13 October 2015  
<https://blogs.chapman.edu/wilkinson/2015/10/13/americas-top-fears-2015/>
- Deloitte, 'Technology and people: The great job-creating machine', 2015  
<https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf>
- Doland, A., 'Check Out the First Ad From McCann Japan's "AI Creative Director"', *AdvertisingAge*, 6 June 2016  
<http://adage.com/article/creativity/check-commercial-mccann-japan-ai-creative-director/304320/>
- Emmelkamp, P.M.G. et al., 'Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia', *Behaviour Research and Therapy* 40(5), May 2002  
<http://www.sciencedirect.com/science/article/pii/S0005796701000237>

European Parliament, Committee on Legal Affairs, *DRAFT REPORT with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, 31 May 2016  
<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN>

Floridi, L., 'Should we be afraid of AI?', Aeon, 9 May 2016  
<https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>

Forrester Research, 'Understanding the role of emotion in customer experience', 2015  
<https://www.forrester.com/report/Understanding+The+Impact+Of+Emotion+On+Customer+Experience/-/E-RES122503>

Frankl, V., *Man's Search for Meaning*, Beacon Press, 1 June 2006 (first edition: 1946)

Frey, C. & A. Osborne, 'The Future of Deployment', 2013  
[http://www.oxfordmartin.ox.ac.uk/downloads/academic/The\\_Future\\_of\\_Employment.pdf](http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf)

Future of Life Institute, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', 2015  
<http://futureoflife.org/open-letter-autonomous-weapons/>

Godwin, W., *Lives of the Necromancers or The Exercise of Magical Power*, 1876, p. 180  
<https://archive.org/details/livesnecromance04godwgoog>

Gooch, J.J., 'NVIDIA Joins Cancer Moonshot with AI Platform for Accelerating Research', *Digital Engineering*, 14 November 2016  
<http://www.digitaleng.news/de/nvidia-joins-cancer-moonshot-with-ai-platform-for-accelerating-research/>

Gray, R., 'What a vain bunch we really are! 24 billion selfies were uploaded to Google last year', *MailOnline*, 1 June 2016  
<http://www.dailymail.co.uk/sciencetech/article-3619679/What-vain-bunch-really-24-billion-selfies-uploaded-Google-year.html>

Hayasaki, E., 'Is AI Sexist?' *Foreign Policy*, January 2017  
<http://foreignpolicy.com/2017/01/16/women-vs-the-machine/>

House of Commons Science and Technology Committee, *Robotics and artificial intelligence*, September 2016  
<https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

Lewis, D.A., G. Blum & N.K. Modirzadeh, 'War-Algorithm Accountability', 31 August 2016  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2832734](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2832734)

McKinsey Global Institute, 'Harnessing automation for a future that works', January 2017  
<http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>

Misener, D. & D. Maloney, 'Wired founder Kevin Kelly on letting go of AI anxiety', 6 December 2016  
<https://slackhq.com/wired-founder-kevin-kelly-on-letting-go-of-ai-anxiety-bcd94e50dedc#.amf4qbypx>

- Nomura1, T., D.S. Syrdal & K. Dautenhahn, 'Differences on Social Acceptance of Humanoid Robots between Japan and the UK', 4th International Symposium on New Frontiers in Human-Robot Interaction, 21-22 April 2015  
[http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura\\_etal\\_AISB2015\\_differencesonsocial.pdf;sequence=2](http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura_etal_AISB2015_differencesonsocial.pdf;sequence=2)
- Robinson, H. et al., *The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial*, 2013
- Rogers, C., *Client Centred Therapy*, 1951
- Ros, F., Stewart II, June 2016  
[http://felixros.com/projects.html#big\\_wrap](http://felixros.com/projects.html#big_wrap)
- Sharkey, A. & N. Sharkey, 'Granny and the robots: ethical issues in robot care for elderly', 2010  
<http://link.springer.com/article/10.1007/s10676-010-9234-6>
- Shaughnessy, M.F., D. Main, J. Madewell & I. Yalom, 'An interview with Irvin Yalom', *North American Journal of Psychology* 9(3), December 2007  
<http://www.freepatentsonline.com/article/North-American-Journal-Psychology/173513562.html>
- Singer, P., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin Press HC, 2009
- Stewart, J., 'Strong Artificial Intelligence and National Security: Operational and Strategic Implications', U.S. Joint Military Operations Department, 18 May 2015  
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA622591>
- Taggart, W., S. Turkle & C. Kidd, 'An Interactive Robot in a Nursing Home: Preliminary Remarks', 2005  
<http://www.androidscience.com/proceedings2005/TaggartCogSci2005AS.pdf>
- Turkle, S., *Alone Together: Why We Expect More from Technology and Less from Each Other*, Basic Books, 2011
- Twyford, S., 'When Is Your Memory Truly Forgotten?', *Legacy Multimedia*, 9 January 2015  
<http://legacymultimedia.com/2015/01/09/when-is-your-memory-truly-forgotten/>
- Vinclu, 'Gatebox - Virtual Home Robot', YouTube, 13 December 2016  
<https://www.youtube.com/watch?v=nkcKaNgfykg&feature=youtu.be>
- Walsh, J. & M. Schulzke, 'The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?', United States Army War College, September 2015  
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA621793>
- Wedding, D. & J.C. Raymond, *Current Psychotherapies*, Brooks/Cole, 2014
- Woollaston, V., 'Do YOU trust driverless cars? Only 1 in 10 of us feel confident about getting in autonomous vehicles and women are the most hesitant', *MailOnline*, 23 September 2015  
<http://www.dailymail.co.uk/sciencetech/article-3245058/Do-trust-driverless-cars-1-10-feel-confident-getting-autonomous-vehicles-women-skeptical.html>
- World Economic Forum, *The Global Risks Report 2017*  
<http://reports.weforum.org/global-risks-2017/executive-summary/>
- Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980

## About VINT [labs.sogeti.com](http://labs.sogeti.com)

VINT, the Sogeti trend lab and part of SogetiLabs, provides a meaningful interpretation of the connection between business processes and new developments. In every VINT publication, a balance is struck between factual description and the intended utilization. VINT uses this approach to inspire organizations to consider and use new technology.

VINT research is done under the auspices of the Commission of Recommendation, consisting of • F.M.R. van der Horst (chairman), CEO Retail Banking ABN AMRO Bank N.V. • J. Behrens, Senior VP, Business Leader Personal Health Solutions Philips • P. Dirix, Managing Director Operations NS Reizigers • D. Kamst, Chief Executive Officer Klooker • Prof. dr. ir. R. Maes, Professor of Information & Communication Management Academy for I & M • P. Morley, Member Supervisory Board TLS • T. van der Linden, Group Information Officer Achmea • E. Schuchmann, Chief Information Officer Academisch Medisch Centrum • K. Smaling, Chief Technology Officer Continental Europe Aegon • W.H.G. Sijstermans, Member of the Board, CIO Nederlandse Zorgautoriteit (NZa) • M. der Weduwen, CIO & Vice President Information Services, Holland Casino • M. Boreel, Chief Technology Officer Sogeti Group • J.P.E. van Waayenburg, Chief Executive Officer Sogeti Group • P.W. Wagter, Chief Executive Officer Sogeti Nederland B.V.

## About SogetiLabs [labs.sogeti.com](http://labs.sogeti.com)

SogetiLabs is a network of over 120 technology leaders from Sogeti worldwide. SogetiLabs covers a wide range of digital technology expertise: from embedded software, cyber security, simulation, and cloud to business information management, mobile apps, analytics, testing, and the Internet of Things. The focus is always on leveraging technologies, systems and applications in actual business situations to maximize results. Together with the Sogeti trend lab VINT, SogetiLabs provides insight, research, and inspiration through articles, presentations, and videos that can be downloaded via the extensive SogetiLabs presence on its website, online portals, and social media.

## About Sogeti

Sogeti is a leading provider of technology and engineering services. Sogeti delivers solutions that enable digital transformation and offers cutting-edge expertise in Cloud, Cybersecurity, Digital Manufacturing, Digital Assurance & Testing, and emerging technologies. Sogeti combines agility and speed of implementation with strong technology supplier partnerships, world class methodologies and its global delivery model, Rightshore®. Sogeti brings together more than 25,000 professionals in 15 countries, based in over 100 locations in Europe, USA and India. Sogeti is a wholly-owned subsidiary of Cap Gemini S.A., listed on the Paris Stock Exchange. For more information please visit [www.sogeti.com](http://www.sogeti.com).

*Rightshore® is a trademark belonging to Capgemini.*

