Digital empathy secures Frankenstein's monster

Raymond Bond[1], Felix Engel[2], Michael Fuchs[3], Matthias Hemmje[4], Paul Mc Kevitt[5], Mike McTear[1], Maurice Mulvenna[1], Paul Walsh[6], and Huiru (Jane) Zheng[1]

[1] Faculty of Computing, Engineering & Built Environment, Ulster University (Jordanstown), Northern Ireland
{rb.bond,mf.mctear,md.mulvenna,h.zheng}@ulster.ac.uk
[2] Research Institute for Telecommunication & Cooperation (FTK), Dortmund, Germany
fengel@ftk.de
[3] Software Engineering & Computer Science, Wilhelm Büchner University of Applied Sciences, Darmstadt, Germany
michael.fuchs@wb-fernstudium.de
[4] Computer Science, Fern University, Hagen, Germany
matthias.hemmje@ferruni-hagen.de
[5] Faculty of Arts, Humanities & Social Sciences, Ulster University (Magee), Northern Ireland
p.mckevitt@ulster.ac.uk
[6] SIGMA Research Group, Cork Institute of Technology, Cork, Ireland
Paul.Walsh@cit.ie

CEUR-WS.org/Vol-2348/paper24.pdf

Abstract. People's worries about robot and AI software and how it can go wrong have led them to think of it and its associated algorithms and programs as being like Mary Shelley's *Frankenstein monster*. The term *Franken-algorithms* has been used. Furthermore, there are concerns about driverless cars, automated General Practitioner Doctors (GPs) and robotic surgeons, legal expert systems, and particularly autonomous military drones. Digital Empathy grows when people and computers place themselves in each other's shoes. Some would argue that for too long people have discriminated against computers and robots by saying that they are only as good as what we put into them. However, in recent times computers have outperformed people, beating world champions at the Asian game of Go (2017), Jeopardy (2011) and chess (1997), mastering precision in medical surgical operations (STAR) and diagnosis (Watson), and in specific speech and image recognition tasks. Computers have also composed music (AIVA), generated art (Aaron), stories (Quill) and poetry (Google AI). In terms of calling for more Digital Empathy between machines and people, we refer here to theories, computational models, algorithms and systems for detecting, representing and responding to people's emotions and sentiment in speech and images but also for people's goals, plans, beliefs and intentions. In reciprocation, people should have more empathy with machines allowing for their mistakes and also accepting that they will be better than people at performing particular tasks involving large data sets where fast decisions may need to be made, keeping in mind that they are not as prone as people to becoming tired.

We conclude that if digital souls are programmed with Digital Empathy, and people have more empathy with them, by doing unto them as we would have them do unto us, this will help to secure Shelley's monster.

1 Introduction

In recent times the fears and anxieties about robots and AI have come to the fore again with academics, but particularly industry and the military-industrial complex, working on conversational chatbots, healthcare assistants, driverless cars, humanoid robots including sexbots and military robots including autonomous drones. Many would argue that the fears are unfounded as these systems are currently no where near the level of human intelligence envisaged in *strong* AI and are more at the level of *weak* or now *narrow* AI, solving fixed problems in limited domains [1]. However, in response to people's fears and a need for ethics in design and production of AI we have seen a rise in the formation of institutes addressing ethical matters in respect of AI such as the *Future of Life Institute* [2], championed and funded by Elon Musk.

What appears to make people anxious about robots and AI is the possibility of robots displacing employment and putting people out of jobs, the fact that robots may get *too big for their boots* and control people who become their slaves, and that building robots that are like people should only be the work of *God*. In effect, the fears that people have is that in the strive to create lifelike machines, monsters like Dr. Frankenstein's [3, 4] (Figure 1) will inadvertently be created.

In this paper we explore people's fears on the rise of AI and how more *digital empathy*, where people and robots can put themselves in each other's shoes and work in harmony, will help to secure AI from becoming Frankenstein's monster. In section 2 we discuss some historical background to the field of robots and AI exploring people's relationship with it in philosophy and literature. Section 3 discusses people's fears leading to possible silicon discrimination. Successes and failures in AI which have added fuel to people's fears are discussed in Section 4. Section 5 discusses key efforts to bring ethics to bear with AI and robots. Section 6 discusses how digital empathy can help to secure people's fears and finally section 7 concludes with avenues for future work.

2 Historical background and related work

People's fears about scientists inadvertently creating monsters in trying to create life run right back to Mary Shelley's, *Dr. Victor Frankenstein* [3], from 1818. In her novel the monster is referred to as *creature*, *monster*, *demon*, *wretch*, *abortion*, *fiend*, and *it*. Speaking to Dr. Frankenstein the monster states: "I ought to be thy Adam, but I am rather the fallen angel." Dr. Frankenstein, whilst based at the *University of Ingolstadt*, excels in chemistry and other sciences and develops a secret technique to impart life to non-living matter and then creates the humanoid which is 8 feet in height and large. The creature escapes and lives alone in the wilderness finding that people were afraid of and hated him due to

Fig. 1. Frankenstein's monster, Boris Karloff, 1931 [3]

his appearance which led him to fear and hide from them. The creature learns to speak and read and when seeing his reflection in a pool realised his physical appearance was hideous and it terrified him as it terrifies normal humans. The creature demands that Dr. Frankenstein creates a female companion like himself and whilst doing so he suffers fears that the two creatures may lead to the breeding of a race that could plague mankind and so destroys the unfinished female creature. In the end Victor dies and the creature drifts away on an ice raft never to be seen again. Shelley travelled through Europe in 1814 and along the Rhine in Germany with a stop at Gernsheim which is 17 km away from the Frankenstein Castle at Pfungstadt, where an alchemist and theologian, Johann Conrad Dippel, was born on 10 August 1673 and was engaged there in experiments on alchemy and anatomy performing gruesome experiments with cadavers in which he attempted to transfer the soul of one cadaver into another. He created an animal oil known as *Dippel's Oil* which was supposed to be an *Elixir of Life*. Soul-transfer with cadavers was a common experiment among alchemists at the time and Dippel supported this theory in his writings [5]. It is rumoured that he dug up bodies and performed medical experiments on them at Frankenstein Castle and that a local cleric warned the parish that Dippel had created a monster that was brought to life by a bolt of lightning.

The title of Shelley's book also makes a reference to *Prometheus*, in Greek Mythology a Titan culture hero and trickster from 800 BC who is credited with creation of man from clay and who defies the gods by stealing fire and giving it to humans. He is then punished by eternal torment and bound to a rock where each day an eagle feeds on his liver which grows back but is then eaten again

the next day. In Greek Mythology the liver contains emotions. Hence, it is clear that Shelley is bringing into our consciousness, that in man's overreaching quest for scientific knowledge, there is the inadvertent or unintended consequences of tinkering with the work of the Gods. In Shelley's novel there is also the gender theme of man attempting to create life without involvement of woman and Shaw's play *Pygmalion* [6] also investigates gender in how man (*Professor Henry Higgins*) attempts to transform a poor flower girl (*Eliza Doolittle*) so she can pass as a duchess by teaching her to speak properly the Queen's English, which also has an unhappy ending. Pygmalion is a reference to the Greek story of *Pygmalion*, catalogued in *Ovid's Metamorphoses* from 8 A.D., who fell in love with one of his sculptures which then *Aphrodite* brought to life. The story of breathing life into a statue has parallels with Greek myths of *Daedalus* using quicksilver to put voices in statues, *Hephaestus* creating automata and *Zeus* making *Pandora* from clay. Gender power matters are also explored again in Greek Mythology in 400 B.C. the *Gorgon* monster *Medusa*, a winged human female with venomous snakes in place of hair, has a power where gazers upon her face would turn to stone.

Der Sandmann (*The Sandman*) [7] is a short story in a book titled *Die Nachtstücke* (*The Night Pieces*) which appeared in 1816, around the time of Shelley's Frankenstein. In *The Sandman*, *Nathanael* tells of his child terror of the sandman who stole the eyes of children who would not go to bed and fed them to his own children who lived in the moon. Nathanael calls his fiancée *Clara* an "inanimate accursed automaton" and Nathanael's Professor, *Spallanzani*, creates a daughter automaton called *Olimpia*, who Nathanal becomes infatuated with and is determined to propose to, where there is also the gender matter of no mother or woman involved. In 1870 elements of *The Sandman* were adapted as the comic ballet *Coppélia*, originally choreographed by *Arthur Saint-Léon* to the music of *Léo Delibes* with libretto by *Charles-Louis-Étienne Nuitter*. *Die Puppe* (*The Doll*) [8] is a 1919 fantasy comedy film directed by Ernst Lubitsch also inspired by *The Sandman*. Objections to machines displacing employment go back to Leviathan [9] in 1651 which discusses humans as being sophisticated machines. Then we have the *Luddites* of 1811, English workers who destroyed machinery mainly in cotton and woollen mills which they believed was threatening their jobs. *Marx* and *Engels* also objected to the weaving automota upset by sights of children's fingers being chopped in machines and visions of them being eaten by these machines. The related theme of slaves and masters between machines and humans comes leads back to slaves in ancient Egypt where we find some of the first automata in the form of moving statues. Egyptian Kings, and kings throughout the ages, displayed their power by demonstrating moving statues, clocks and fountains at public entertainment events [10–14] such as *Jacques de Vaucanson's* mechanical duck (1739) which could eat, drink and go to the toilet. This theme of slavery comes up again *in Fritz Lang's Metropolis* [15]. Alternatively, others have argued that machines will do all the hard work liberating people to pursue creative pursuits and leisure, a kind of Utopia [16].

There have always been religious objections to AI and robotics, man of them based on the fact that creating life is the work of God and only people can have souls. Descartes [17, 18] who focused on rationalism and logic whilst watching through his Amsterdam window Dutch people walking in street saw no difference between them and automata and produced his well known statement: "Je pense, donc je suis" ("Cognito ergo sum"; I think, therefore I am). He emphasized that people are rational and only they can have souls, not animals or automata which are mechanical. Leibniz, another rationalist, and Hobbes an empiricist, had similar views. However, religion has also used statues to demonstrate God's power where at the *Cistercian Boxley Abbey* in Boxley, Kent, England there were moving and weeping statues with nuns inside manufacturing the tears. There have also been arguments that the design of people's hands is proof of the existence of God.

3 Silicon discrimination?

People's fear of AI and robots has led to what could be called *discrimination* against them with common colloquial sayings such as *computers are only as good as what we put into them.* In 2017 Jack Ma, the founder of Alibaba said that there is IQ (Intelligence Quotient), EQ (Emotional Intelligence) and LQ (Love Intelligence) and that people have all three of these but robots cannot have LQ. However, Ma does not take into account a particular type of humanoid robot that we will visit below in Section 4. People also say that robots have no creativity and no soul as they are not created by God.

Searle [1, 19, 20] makes the distinction between human level intelligence envisaged in *strong* AI and *weak* or now *narrow* AI where programs are solving fixed problems in limited domains. In arguing against strong AI, Searle proposed the *Chinese Room Argument* [1, 19] GedankenExperiment arguing that AI programs are like a person inside a room who uses a large rule book to handle messages with written Chinese instructions and responds to them, but has no understanding of Chinese at all. There is also Harnad's *Symbol Grounding Problem* [21], asking the question how symbols in AI programs are grounded in the real world. Dennett in the *Intentional Stance* [22] discusses a *Ladder of Personhood* where computers can have the ability to perform language processing, reasoning, but cannot have stance (ascribing intentions to others), reciprocity and consciousness, which only people can have. However, Ballim & Wilks [23] discuss in detail how computers can have beliefs and nested beliefs about other people's beliefs. Dreyfus [24, 25] points out that computers cannot have common sense like humans. Weizenbaum, an AI researcher, explains the limits of computers and that anthropomorphic views of them are a reduction of human beings and any life form [26]. Penrose [27] argues from the viewpoint of Physics, that AI cannot exist.

In terms of employment displacement fears, in many states there is fair employment discrimination legislation enabling tribunals based on religious, sexual orientation, nationality, race, and gender discrimination. In a future where in-

telligent robots do exist will they request legislation on silicon discrimination? and could there be robots sitting on Tribunal panels?

4 AI successes and failures

There have been successes with AI and robots over the years. There was the checkers (draughts) playing program developed by Arthur Samuel at IBM in 1961. IBM also built Deep Blue which beat the world chess champion Gary Kasparov in 1997 and in 2011 IBM's Watson beat the world champions at the quiz show game of *Jeopardy!*. Google's *AlphaGo* beat the world champion in Korea at the Asian game of *Go*. However, there have also been failures with Driverless cars and particular image recognition tasks.

4.1 Conversational chatbots

One of the popular areas of AI is the development of natural language processing conversational AI programs which can interact in, usually typed and not spoken, dialogue with people. These chatbots have been applied in many areas including pychotherapy, companions, call-centre assistance, and healthcare. One of the first programs developed at MIT by Joseph Weizenbaum in 1964 [26, 28] was *ELIZA*, named after *Eliza Doolittle* discussed in Section 2 above. ELIZA simulated conversations by using pattern-matching that gave users an illusion of understanding by the program. Scripts were used to process user inputs and engage in discourse following the rules and directions of the script. One of these scripts, *DOCTOR*, simulated a Rogerian psychotherapist. Many individuals attributed human-like feelings to ELIZA including Weizenbaum's secretary, with users becoming emotionally attached to the program forgetting that they were conversing with a computer. In 1972 ELIZA met another conversational AI program named PARRY developed by Ken Colby [29] where they had a computer-to-computer conversation. PARRY was intended to simulate a patient with schizophrenia. Since ELIZA there have been many chatbots developed and *Hugh Loebner* launched the *International Loebner Prize* contest in 1990 for chatbots which could pass the *Turing Test* [30], a test proposed by Alan Turing where programs which could fool human judges that they were humans would be deemed to be intelligent. The most recent winner in 2018 of the annual bronze medal for the best performing chatbot is *Mitsuku*, which has now won the contest four times (2013, 2016, 2017, 2018). No chatbot has won the Loebner Gold Medal Prize, where it fools all of the four judges that it is human.

4.2 Medical assistants

Chatbots are being used in healthcare and last year *Babylon*, a healthcare chatbot with the goal of making healthcare universally accessible and affordable, giving online consultation and advice effectively acting as a General Practitioner

Doctor (GP), has received a £100 million investment in September, 2018 creating 500 jobs in London. Babylon caused controversy by claiming that in tests it performs medical diagnosis on a par with GPs achieving medical exam scores in the MRCGP (Royal College of General Practitioners) exam.

IBM's *Watson* mentioned above is also being applied to a number of application domains including healthcare. Watson uses natural language processing, hypothesis generation and evidence-based learning to support medical professionals as they make decisions. A physician can use Watson to assist in diagnosing and treating patients. Physicians can pose a query to Watson describing symptoms and other related factors and then Watson identifies key pieces of information in the input. Watson mines patient data and finds relevant facts about family history, current medications and other conditions. It combines this information with findings from tests and instruments and examines existing data sources to form hypotheses and test them. Watson incorporates treatment guidelines, electronic medical record data, doctor's and nurse's notes, research clinical studies, journal articles and patient information into the data available for analysis. Watson provides a list of potential diagnoses along with a score that indicates the level of confidence in each hypothesis.

DeepMind was founded in London in 2010 and acquired by Google in 2014, now part of the Alphabet group. DeepMind is a world leader in AI research and its application. DeepMind Health is focused on helping clinicians get patients from test to treatment faster. DeepMind Health works with hospitals on mobile tools and AI research to help get patients from test to treatment as quickly and accurately as possible. *Streams* is an app developed and in use at the Royal Free London National Health Service (NHS) Foundation Trust using mobile technology to send immediate alerts to clinicians when a patient deteriorates. Nurses have said that it is saving them over two hours each day meaning they can spend more time with those in need.

The STAR (Smart Tissue Autonomous Robot) in 2017 [31] beat human surgeons at a flesh-cutting task in a series of experiments. It made more precise cuts than expert surgeons and damaged less of the surrounding flesh. Previously, in 2016 STAR had sewed together two segments of pig intestine with stitches that were more regular and leak-resistant than those of experienced surgeons.

The SenseCare (Sensor Enabled Affective Computing for Enhancing Medical Care) [32, 33] project is developing a new affective computing platform providing software services applied to the dementia care and connected health domain providing intelligence and assistance to medical professionals, care givers and patients on cognitive states and overall holistic well-being. Data streams are integrated from multiple sensors fusing these streams together to provide global assessment that includes objective levels of emotional insight, well-being and cognitive state where medical professionals and care can be alerted when patients are in distress. A key focus of SenseCare is to detect the emotional state of patients from their facial gestures and a web service has been developed which outputs emotional classifications for videos [32, 33]. A sample screenshot of the application is shown in Figure 2. Results from analysed emotions giving senti-

ment (positive, negative) and timestamps are shown in Table 1. The MIDAS (Meaningful Integration of Data, Analytics & Services) [34] project is also addressing connected health from the point of view of data analytics and services.

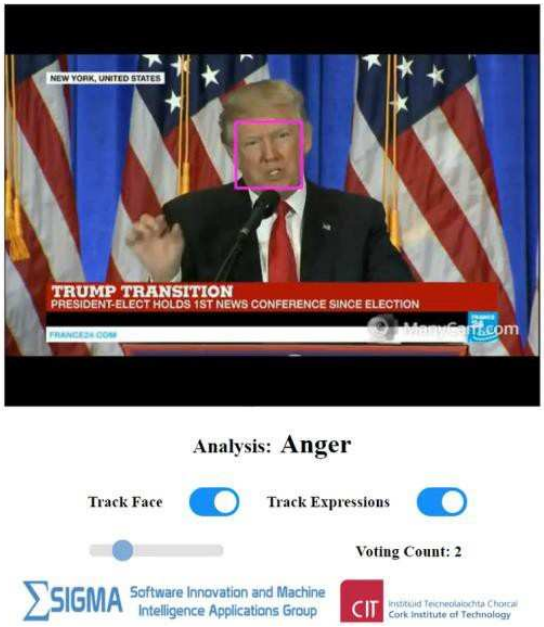

Fig. 2. Detecting emotional states in facial expressions [32, 33]

Table 1. SenseCare timestamped emotion analysis.

Positive			*Negative*		
Happiness	*Interested*	*Enthusiasm*	*Bored*	*Angry*	*Frustrated*
00:50:30	00:21:00	02:00:40	00:35:12	01:10:15	01:48:20
01:32:00	00:40:30		00:41:10		
01:35:55	01:00:35		01:20:41		
	01:05:20		01:25:15		
	01:07:15		01:45:30		
			01:51:30		

4.3 Driverless cars

In the car industry the race is on to develop the first driverless car with many computer companies such as Google, car companies such as Volkswagen, Daimler AG, and even Taxi companies such as Uber working on the problem. Google's *Waymo (New Way Forward in Mobility)* is a self-driving technology company with a mission to make it safe and easy for people and things to move around. Waymo has more than 25,000 autonomous miles driven each day, mainly on

complex city streets on top of 2.7 billion simulated miles driven in 2017. Vehicles have sensors and software that are designed to detect pedestrians, cyclists, vehicles, road work and more from up to three football fields away in all 360 degrees. However, there have been some failures with a *Tesla Model S* car driver being killed in May, 2016 when the car (and driver) did not detect the white side of a truck against the brightly lit sky, with no application of the brake, on *US Highway 27a* and in March, 2018 a *Tesla Model X* electric SUV crashed into a *US Highway 101* barrier killing the driver whilst on autopilot. Also, in March, 2018 at Tempe, Arizona, USA an Uber driverless Volvo SUV killed a female pedestrian walking with her bicycle across the road when the driver was not paying attention after the car attempted to hand over control to her.

4.4 Humanoid robots

Industrial robots have already been used in the car and other manufacturing industries for decades. More recently there has been a focus on more humanoid robots detecting and exhibiting emotions with numerous applications such as companions and healthcare assistants. Examples of such robots are *Erica* developed at Osaka University which is a humanoid female robot and *Pepper* developed by *Softbank*, which can detect and react to people's emotional states. Companies such as *Abyss creations* are developing Sexbots such as *RealDoll* sex doll and sex robot *Harmony*, with customisable AI and swappable faces, which are life-like and can move and speak to their users. Customers can have conversations with Harmony.

4.5 Military robots

Military robots for warfare in the field are also being developed by companies such as Boston Dynamics. These robots which can be animal-like have the ability to move fast over difficult terrain and pick themselves up again after slipping on ice. Robotic drones such as the *Reaper*, have also been developed for warfare, which are currently controlled by people, but there are military moves towards having them operate autonomously. Alternatively, there are drones which are used in the service of society such as environmentalism and search and rescue. Such drones have been used to deliver vital medical supplies to the field in remote regions. Drone waiters have even been used to deliver drinks and food at parties!

4.6 Creativity

The mathematician *Lady Ada Lovelace*, daughter of *Lord Byron*, mentioned above in Section 2, is acknowledged as the first computer programmer. She recognised the creative and artistic potential of computers in the 1840s suggesting that computers "might compose elaborate and scientific pieces of music of any degree of complexity" [35]. There has been work in AI on modelling creativity in art, music, poetry, storytelling [36, 37]. In respect of art, AARON has

been developed since 1973 by Harold Cohen [38, 39] and creates original artistic images. Initial versions of AARON created abstract drawings that grew more complex through the 1970s. In the 1980s more representational figures set in interior scenes were added, along with colour. In the early 2000s AARON returned to abstract imagery, this time in colour. Cohen has used machines to enable AARON to produce physical artwork.

AIVA (Artificial Intelligence Virtual Artist) [40], developed by Pierre & Vincent Barreau in 2016 composes music and is the world's first computer program to be recognised by a music society, *SACEM (Société des Auters, Compositeurs et Éditeurs de Musique)*. By reading a large collection of existing works of classical music written by human composers such as *Bach, Beethoven*, and *Mozart*, AIVA can understand concepts of music theory and compose its own. AIVA is based on deep learning and reinforcement learning AI techniques. AIVA is a published composer, with its first studio album, *Genesis*, released in November, 2016 and its second album, *Among the Stars* in 2018.

Google AI, working with Stanford University and University of Massachusetts, developed a program in 2016 that accidentally produces poetry [41], after attempts to digest romance novels, using an AI technique called RNNLM (Recurrent Neural Network Language Model).

Quill developed by *Narrative Science* in 2015 [42] is a natural language generation storytelling program that analyses structured data and automatically generates intelligent narratives. Narrative Science received some early criticism from journalists speculating that it was attempting to eliminate the jobs of writers, particularly in sports and finance.

5 Ethical AI

Isaac Azimov in *I, Robot* in 1942 [43], provided three laws of robotics:

- A robot may not injure a human being or, through inaction, allow a human being to come to harm
- A robot must obey orders given it by human beings except where such orders would conflict with the *First Law*
- A robot must protect its own existence as long as such protection does not conflict with the *First* or *Second Law*.

Later in 2013, Alan Winfield suggested a revised 5 *Principles of Robotics* published by a joint Engineering & Physical Sciences Research Council (EPSRC)/Arts & Humanities Research Council (AHRC) working group in 2010 [44]:

- Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.
- Humans, not Robots, are responsible agents. Robots should be designed and operated as far as practicable to comply with existing laws, fundamental rights and freedoms, including privacy.

- Robots are products. They should be designed using processes which assure their safety and security.
- Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.
- The person with legal responsibility for a robot should be attributed.

In response to people's fears and a need for ethics in design and production of AI we have seen a rise in the formation of institutes addressing ethical matters in AI such as the *Future of Life Institute* [2], founded in 2014 by Max Tegmark, Elon Musk, Stuart Russell, and Stephen Hawking, the OpenAI institute [45] founded in 2015 by Elon Musk, Microsoft, Amazon, Infosys, the *Future of Humanity Institute* [46] at Oxford University founded in 2014 by Nick Bostrum and the *Centre for the study of existential risk* [47] at Cambridge University founded in 2012, by Jaan Tallinn (founder of Skype) and Seán Ó hÉigeartaigh and the *Foundation for Responsible Robotics (FRR)* [48] at The Hauge in The Netherlands founded in 2015 by Aimee van Wynsberghe (President) and Noel Sharkey (Treasurer) with Shannon Vallor as Secretary. The FRR has as its mission:

> to shape a future of responsible robotics and artificial intelligence (AI) design, development, use, regulation and implementation. We see both the definition of *responsible robotics* and the means of achieving it as ongoing tasks that will evolve alongside the technology.

where *responsible robotics* means that it is up humans to be accountable for the ethical developments that necessarily come with the technological innovation. Recently the FRR and professional services network Deloitte have announced they will be launching a quality mark for robotics and AI to promote transparency and trust in AI products which will match a set of standards to receive the quality mark. Criteria will include environmental protection, sustainability, worker treatment, safety and security. The FRR in partnership with Deloitte will give products a rating out of three. The FRR has supported the *Curbing Realistic Exploitative Electronic Pedophilic Robots (CREEPER) Act*, introduced in the USA on December 14th, 2017 by Congressman Dan Donovan with a bipartisan coalition of 12 original cosponsors, to ban importation and distribution of child sex dolls. Similar bans exist in Australia and the UK.

The *International Committee for Robot Arms Control (ICRAC)* is a Non Governmental Organisation (NGO) of experts in robotics technology, AI, robot ethics, international relations, international security, arms control concerned with the dangers that military robots pose to peace and international security and to civilians in war. A key component of ICRAC's mission statement is:

> the prohibition of the development, deployment and use of armed autonomous unmanned systems; machines should not be allowed to make the decision to kill people

where it is on the Steering Committee for the *Campaign to Stop Killer Robots*, launched in London in April, 2013, an international coalition working to preemptively ban fully autonomous weapons. Recently the European Union (EU)

has passed a resolution supporting a ban on the use of weapons that kill autonomously.

Wilks [49] and Lehman-Wilzig [50] discuss responsible computers and how blame and punishment might be applied to computers and how they might be said to take on social obligations. Wilks notes that humans behind machines and programs can be identified to carry the blame, or the companies who have produced them. However, this can be tricky due to the fact that large teams can be involved in developing software which has been edited and updated over many years and some of these people may also have passed away. Wilks points out that a machine can be turned off and smashed, and the software with it, or burned separately making sure that one has all the copies. He notes that *Joan of Arc's* body was punished but not her soul which reminds us of the discussion on Descartes in Section 2 above.

Plug & Pray [51] is a 2010 documentary film about the promise, problems, and ethics of AI & robotics with the main protagonists being MIT professor Joseph Weizenbaum mentioned in relation to ELIZA in Section 4 above, who died during the making of the film, and the futurist Raymond Kurzweil. Kurzweil dreams of strong AI where machines will equal their human creators where man and machine merge as a single unity. However, Weizenbaum questions society's faith in the redemptive powers of new technologies and their ethical relationships to humans.

6 Digital empathy

It is clear that with the recent rise of developments in AI, and particularly by industry, there is a need more for *digital empathy* between machines and people and people and machines. First, if we take machines and people then failsafe mechanisms such as the *Laws of Robotics* discussed in Section 5 will need to be included as a *backstop* (safety net) in robots and AI, in cases where they have not preemptively had their wings already clipped. Such laws will need to be programmed into the robots and AI so that they make rational decisions in respect of for example making split second decisions whilst avoiding an accident, deciding whether to turn off a life-support system for a patient, or leniency in legal decision making. As emotions and beliefs about others are closely related to empathy, robots and AI will need to have better mechanisms for detecting, representing and responding to people's emotions in speech, gestures, and facial expressions and people's goals, plans, beliefs and intentions. Second, people will have to have more empathy with robots and AI, accepting that they will make mistakes from time to time, but also accepting that they will be better than people at performing particular tasks which involve very large amounts of data where fast decisions may need to be made, also keeping in mind that they are not as prone as people to becoming tired.

7 Conclusion

Here, we have discussed people's fears on the rise of robots and AI in relation to employment displacement, loss of control to robots where people become their slaves, and that this really should only be the work of *God*. Otherwise scientists and industry could inadvertently create Dr. Frankenstein's monster. We have covered what may be deemed silicon discrimination where people have been critical of developments, the successes of AI which have given fuel to people's fears, and efforts to define ethical laws for robots and AI so that they do not get out of control. Future work includes developing further more accurate methods enabling robots to better detect, represent and respond to people's emotional states through improved image and speech processing and people's goals, plans, beliefs and intentions whilst also imbuing them further with ethical principles and laws. Frankenstein's monster can be secured with more *digital empathy*, where people and robots place themselves in each other's shoes.

Acknowledgement. This research was partly supported by an EU Horizon 2020 Marie Skłodowska-Curie RISE Project, *SenseCare*, and an Invest NI Proof of Concept (PoC-607) R&D Project, *Broadcast Language Identification & Subtitling System (BLISS)*.

References

1. Searle, J.R.: Minds, brains and programs. In Behaviour and Brain Sciences, 3: 417-424 (1980)
2. FLI: http://futureoflife.org/
3. Shelley, M.: Frankenstein; or, The Modern Promethius. Lackington, Hughes, Harding, Mavor & Jones, January 1st (1818)
4. Smith: A.: Franken-algorithms: the deadly consequences of unpredictable code. Guardian Newspaper, August 30th (2018)
5. Dippel, J.C.: Maladies and remedies of the life of the flesh (1736)
6. Shaw, G.B.: Pygmalion. Play, Vienna, Austria (1913)
7. Hoffmann, E.T.A.: Der Sandmann (The Sandman). In *Die Nachtstücke* (*The Night Pieces*) (1816)
8. Lubitsch, E.: Die Puppe (The Doll). Film (1919)
9. Hobbes, T.: Leviathan (1651)
10. Sharkey, N.: I ropebot. New Scientist, 195 (2611), 32-35 (2007)
11. Truitt, E.R.: Medieval Robots: Mechanism, Magic, Nature, and Art. University of Pennsylvania Press (2015)
12. Truitt, E.R.: Preternatural Machines. Aeon (2015)
13. Kohlt, Franziska: In the (automated) eye of the beholder: Automata in culture and the enduring myth of the modern Prometheus. Marvellous Mechanical Museum (Compton Verney Press) (2018)
14. Cave, Stephen, Dihal, Kanta: Ancient dreams of intelligent machines: 3,000 years of robots. Nature, July (2018)
15. Lang, Fritz (Director): Metropolis. Film (1927)
16. More, T.: Utopia. (1516)
17. Descartes, R.: Discourse on the method (1637)

18. Descartes, R.: Passions of the soul (1649)
19. Searle, J.R.: Minds, brains and science. London: Penguin Books (1984)
20. Searle, J.R.: Is the brain's mind a computer program? In Scientific American, 262: 26-31 (1990)
21. Harnad, S.: The symbol grounding problem. In Physica D., 335-46 (1990)
22. Dennett, D.C.: The Intentional Stance. The MIT Press, Cambridge, USA (1987)
23. Ballim, A., Wilks, Y.: Artificial Believers: the ascription of belief. Psychology Press, London, England (1991)
24. Dreyfus, H.L.: What computers can't do: the limits of artificial intelligence. Harper Collins, London, England (1972)
25. Dreyfus, H.L.: What computers still can't do: a critique of artificial reason. MIT Press, Cambridge, USA (1992)
26. Weizenbaum, J.: Computer Power and Human Reason: From Judgment to Calculation. W.H. Freeman and Company, New York (1976)
27. Penrose, R.: The Emperor's New Mind: Concerning computers, minds and the laws of Physics. Oxford University Press, Oxford, England (1989)
28. Weizenbaum, J.: ELIZA a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9: 3645, (1966)
29. Colby, K.M., Watt, J.B., Gilbert, J.P.: A Computer Method of Psychotherapy. The Journal of Nervous and Mental Disease, 142 (2), 14852 (1966)
30. Turing, A.: Computing machinery and intelligence. Mind LIX (236): 433-460 (1950)
31. Strickland, E.: In flesh-cutting task, autonomous robot surgeon beats human surgeons. IEEE Spectrum, October (2017)
32. Donovan, R., Healy, M., Zheng, H., Engel, F., Vu, B., Fuchs, M., Walsh, P., Hemmje, M., Mc Kevitt, P.: SenseCare: using automatic emotional analysis to provide effective tools for supporting wellbeing. In Proc. of 3rd International Workshop on Affective Computing in Biomedicine & Healthcare (ACBH-2018), IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2018), Madrid, Spain, 3-6th December, 2682-2687. IEEE Computer Society, New Jersey, USA(2018)
33. Healy, M., Donovan, R., Walsh, P., Zheng, H.: A machine learning emotion detection platform to support affective well being. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2018), Madrid, Spain, 3-6th December, 2694-2700. IEEE Computer Society, New Jersey, USA(2018)
34. Cleland, B., Wallace, J., Bond, R.R., Black, M., Mulvenna, M., Rankin, D., Tanney, A.: Insights into Antidepressant Prescribing Using Open Health Data. Big Data Research, 12, 41-48 (2018)
35. Lovelace, A: Translation of article, 'Sketch of the Analytical Engine Invented by Charles Babbage by Italian military engineer Luigi Menabrea on Babbages Analytical Engine. Richard & John E. Taylor, Red Lion Court, London (1843)
36. Hofstadter, D.: Gödel, Escher, Back: an eternal golden braid. Basic Books, New York, USA (1979)
37. Mc Kevitt, P., O Nuallain, S., Mulvihill, C. (Eds.): Language, vision and music, Advances in consciousness series. John Benjamins Publishing Company, Amsterdam, The Netherlands (2002)
38. Cohen, H.: The art of self-assembly of art. Dagstuhl Seminar on Computational Creativity, Schloß Dagstuhl, Germany (2009)
39. Cohen, H.: Driving the creative machine. Orcas Centre, Crossroads Lecture Series, September (2010)
40. Barreau, P.: How AI could compose a personalised soundtrack to your life, TED 2018: the Age of Amazement (2018)

41. Gibbs, S.: Google AI project writes poetry which could make a Vogon proud. Guardian Newspaper, May 17th (2016)
42. Levy, S.: Can an algorithm write a better news story than a human reporter?. Wired, June 6th (2014)
43. Asimov, I.: Runaround. In I, Robot. Doubleday & Company, New York (1950)
44. Winfield, A.: Five roboethical principles – for humans. New Scientist, No. 2811 (2011)
45. OpenAI: https://openai.com/
46. FHU: https://www.fhi.ox.ac.uk/
47. CSER: https://www.cser.ac.uk/
48. FRR: https://responsiblerobotics.org/
49. Wilks, Y.: Responsible computers? International Joint Conference on Artificial Intelligence (IJCAI), 1279-1280. Los Angeles, CA, USA (1985)
50. Lehman-Wilzig, S.N.: Frankenstein unbound: towards a legal definition of Artificial Intelligence. Futures, 107-119 (1981)
51. Schanze, J. (Director): Plug & Pray. Documentary film, Masha Film (2010)