COMPUTER LAW & SECURITY REVIEW ■■ (2016) ■■–■■

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Stifling artificial intelligence: Human perils

Gonenc Gurkaynak [a,*], Ilay Yilmaz [a], Gunes Haksever [b]

[a] ELIG Attorneys-at-Law, Istanbul, Turkey
[b] IBM Turkey, İstanbul, Turkey

Keywords:
Artificial intelligence (AI)
Regulations
Economic efficiency
Singularity
Sci-fi
Creative destruction
Existential threat

A B S T R A C T

Although scientists have calculated the significant positive welfare effects of Artificial Intelligence (AI), fear mongering continues to hinder AI development. If regulations in this sector stifle our active imagination, we risk wasting the true potential of AIs dynamic efficiencies. Not only would Schumpeter dislike us for spoiling creative destruction, but the AI thinkers of the future would also rightfully see our efforts as the 'dark age' of human advancement. This article provides a brief philosophical introduction to artificial intelligence; categorizes artificial intelligence to shed light on what we have and know now and what we might expect from the prospective developments; reflects thoughts of worldwide famous thinkers to broaden our horizons; provides information on the attempts to regulate artificial intelligence from a legal perspective; and discusses how the legal approach needs to be to ensure the balance between artificial intelligence development and human control over them, and to ensure friendly artificial intelligence.

Our technology, our machines, is part of our humanity. We created them to extend ourselves, and that is what is unique about human beings. – Ray Kurzweil[1]

1. Introduction

The Chinese cardboard game "Go" is one of the most complex strategy games humankind invented. Go was considered so important, there are myths indicating that ancient kings played Go between their armies in the battlefield to resolve the conflict in peace. Computers prevailed against humanities best in many zero-sum, perfect-information, partisan, deterministic strategy games[2] before, with the exception of Go, which was something to be proud of.

The strategy aspect of Go is very complex and emphasizes the importance of balance on multiple levels and has internal tensions. A game of Go cannot be won by using brute force: calculating every possible move, similar to what IBM®'s then state of the art AI, Deep Blue® used to win over Gary Kasparov. To manoeuvre through the countless possible moves on the Go board and chose the most efficient path, one requires capabilities beyond the conventional computing powers; capabilities only our minds have (or so we thought), such as extremely accurate image and pattern recognition and insight, all of which we thought granted us superiority over the artificial minds we created.

In October 2015, a software called "AlphaGo®" became the first computer to beat a professional human Go player in an un-handicapped game of Go (Silver and Hassabis, 2016). AlphaGo's victory is probably one of the most significant demonstrations of the capabilities of an AI. Firstly, it shows that AIs are beginning to surpass us at things where success is dependent on strategy as well as calculation. Things we classify as a "game", from stock exchange to conflicts, from

* *Corresponding author.* ELIG Attorneys-at-Law, Çitlenbik Sokak, No: 12, Beşiktaş, İstanbul, Turkey.
E-mail address: gonenc.gurkaynak@elig.com (G. Gurkaynak).

[1] Author, computer scientist, inventor, futurist, and director of engineering at Google.
[2] These are chess, checkers and reversi.

http://dx.doi.org/10.1016/j.clsr.2016.05.003
0267-3649/

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

contract negotiations to hostage situations. Second, AlphaGo developed strategies on its own, through playing millions of games against itself. These feats sent the chills down the spines of those who fear that AIs will overpower us in the future.

We humans accelerate the future with our minds. This is a strength and a weakness. Often, our predictions of the future are highly inaccurate. Based on predictions from a book called 'The World in 2010', published in 1976, we should have been living above and below the surfaces of three planets as of five years ago. Predictions regarding the future of AI are equally likely to be off base.

To avoid premature regulation over AI, we should be studying and searching for the meaningful point in time when a broader anxiety about AI becomes a genuine concern. The study of a point of ripeness, a 'threshold ability test,' asks when AI could really bring about concrete disadvantages that might counter-balance the demonstrated contribution to economic efficiency and welfare.

In the absence of such an objective benchmark marking the point in time when AI becomes a competitor with the human mind, regulators could easily jump the gun in regulating AI, which would lead to irreparable harm in total welfare of human societies.

Most of what we consider AI today is really our own intelligence re-formatted and re-cycled, with the help of computers lacking any skill of learning or consciousness of being. Regulation at this stage would be perverse. The economic efficiency potentials of AI should be set entirely free at this point in time, allowing us to actively and aggressively research appropriate goals for them which would not result in the extinction of humankind.

If you think our future robot overlords will one day thank us for ignoring the risks and under regulating, think again. On the one hand, any issues we may face from AIs will likely result from humanity failure to effectively direct AIs to our needs, not because we switched to a defensive AI regulation regime too early. On the other hand, at some point of time in the not too distant future, natural, human-related or external factors may threaten the fate of the Earth, and we may need AI to save the planet and us. One hopes that society has not pulled the hand brakes on the wheels of AI too early, fearing our own active imagination.

2. Artificial intelligence

Human kind is constantly seeking to identify qualities distinguishing itself from other animal species, trying to prove that certain qualities makes us 'human'. Religion took the way claiming that 'god' created us to be the dominant 'unique' species on Earth, while science is still debating on what separates us from the rest of the 'life as far as we know it' in our universe. René Descartes stated the fundamental difference between humans and animals in his famous words: 'Je pense donc je suis. ('*I think therefore I am.*'). According to Descartes, being aware of thought process is not only the sole thing a human can rely on while evaluating its existence, but also a stepping stone in human kind's evolution process; the stage where 'Homo Sapiens' ('The Man Who Knows') evolved into a higher form of existence; 'Homo Sapiens Sapiens', ('The Man Who Thinks on Thought').

Descartes believed that humans could verify their existence via their thought processes moulded by experience, while animals merely follow prefixed programs (Boesch, 2007). This concept is often labelled as 'tabula rasa' ('blank slate') and can be traced back to Aristotle (Polansky, 2007),[3] the Stoic school in Ancient Greece and Avicenna ('Ibn Sina')[4] of Persia (Rizvi, n.d.). The renowned philosopher of the modern age John Locke then modernized the idea (Locke, 1690).[5]

Computer scientists have adapted this concept 'tabula rasa' to computer science as development of autonomous agents that have the capability to reason and plan towards their goal without any built-in knowledge base of their environment. This resembles early AI initiatives.

As one commentator put it, AI researchers are 'the children of Rene Descartes, trusting in absolute power of logic and mathematics as they push their religion of Cartesian Dualism on the rest of us.' (Ormandy, 2015).

John McCarthy[6] introduced the term 'Artificial Intelligence' in 1955, defining it as 'the science and engineering of making intelligent machines, especially intelligent computer programs'. However, AI is a broad concept with myriad applications from the 'intelligent' assistants in our cell phones (i.e. Siri® of Apple®) to 'intelligent' home appliances to any future technologies that may cause a paradigm shift in our understanding of life (Urban, 2015). Therefore, due to its broad definition and applications, as well as the phenomenon McCarthy states as, 'as soon as it works, no one calls it AI anymore'

[3] Aristotle uses the term 'unscribed tablet.' In one of the more well-known passages of this treatise he writes that: '*Have not we already disposed of the difficulty about interaction involving a common element, when we said that mind is in a sense potentially whatever is thinkable, though actually it is nothing until it has thought? What it thinks must be in it just as characters may be said to be on a writing-tablet on which as yet nothing stands written: this is exactly what happens with mind.*'

[4] Avicenna argued that the '. . .human intellect at birth resembled a *tabula rasa*, a pure potentiality that is actualized through education and comes to know,' and that knowledge is attained through '. . .empirical familiarity with objects in this world from which one abstracts universal concepts,' which develops through a '. . .syllogistic method of reasoning; observations lead to propositional statements, which when compounded lead to further abstract concepts.' He further argued that the intellect itself '. . . posesses levels of development from the static/material intellect (*al-'aql al-hayulani*), that potentially can acquire knowledge to the active intellect (*al-'aql al-fa'il*), the state of the human intellect at conjunction with the perfect source of knowledge'.

[5] In Locke's philosophy, tabula rasa was the theory that, at birth, the (human) mind is a 'blank slate' without rules for processing data, and that data is added and rules for processing are formed solely by one's sensory experiences.

[6] American computer scientist and cognitive scientist, professor at Stanford University, Massachusetts Institute of Technology, Dartmouth College and Princeton University. Died on October 24, 2011.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

(Vardi, 2012), AI is often considered as a topic merely geeks[7] and science fiction enthusiasts are interested in.

In actuality, it is one of the hottest topics in the world today, attracting tech magnates such as Stephan Hawking, Ray Kurzweil, Bill Gates, Elon Musk and Bill Joy.[8]

The first thing that comes to the mind when we hear the term AI is usually the world commonly depicted by Hollywood, where humanity is either enslaved by intelligent machine overlords,[9] faces the threat of imminent extinction at their hands[10] or human creators face their death at the hands of their creations.[11] This simplistic view ignores the various types of artificial intelligence, much of which is benign. In order to provide a better understanding on what AI means when it comes to reality, we find it worthwhile to classify the 'intelligence level' of AI under three main groups (Urban, 2015): Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence.

2.1. *Artificial Narrow Intelligence ('ANI')*

ANIs are AIs specialized in a specific area, such as IBM's Deep Blue®, the supercomputer that beat Gary Kasparov, the reigning World Chess Champion in May 1997. As with most ANIs, Deep Blue's abilities were limited. Playing chess was the only thing Deep Blue® could do. McCarthy was actually disappointed despite Deep Blue's achievement. Criticizing the fact that Deep Blue's success was entirely due to its sheer computational power rather than a deep understanding of chess itself. He stated, 'The fixation of most computer chess work on success in tournament play has come at scientific cost' (Vardi, 2012). In line with McCarthy's critique, humanity has created supercomputers exceeding the calculation capacity of the average human brain, which is around 10 quadrillion[12] calculations per second. China's Tianhe-2 can do 34 quadrillion calculations per second while consuming 24 megawatts of power and taking 720 square meters of space. The human brain on the other hand, runs on a puny 20 watts, approximately a millionth of the power Tianhe-2® requires. These computers can solve complex problems in the blink of an eye but they do not have any preconception of things other than the information provided to them by their creators. In a sense, ANI's reality is limited to their pre-determined capabilities of observation. Just like us.[13]

ANIs surround us today and they work in a similar manner. The intelligent thermometers of 'Nest®', Apple's 'Siri®', video games, search engines, social networks, web cookies, online advertising services, data miners and data scrapers, autopilots, traffic control software, automated phone answering services and so on; neither of which can initiate thought processes in order to provide queries falling outside the scope of their predetermined operations. This is the level of AI humanity has achieved thus far. In a sense, ANIs represent 'tabula rasa' at their current state.

However, ANI's talents are constantly getting better and more impressive. Speech recognition and processing allows computers to convert sounds to text with greater accuracy. Google® is using AI to caption millions of videos on YouTube®. Likewise, computer vision is improving so that programs like Vitamin D Video® can recognize objects, classify them, and understand how they move. Narrow AI is not just getting better at processing its environment, it also understands the difference between what a human says and what a human wants. IBM®'s Watson® was so good at understanding questions and matching them with the facts that it beat humans in Jeopardy (Markoff, 2011).

World's leading tech companies, such as IBM®, Google® and Microsoft®, are effectively working on creating an AI that processes information with software based on our physical, biological and chemical thought process, and spending substantial resources on research for technologies like electronic neural networks, cognitive computing algorithms and artificial neo-cortex through software. Through these technologies and research, tech companies hope to create an AI that may become on par with our ability to encode information.

2.2. *Artificial General Intelligence ('AGI')*

AGIs represent 'Human-Level AIs', computers as smart as humans in every aspect and capable of performing all intellectual tasks humans can. AGIs are expected to be capable of solving various complex problems in various different domains with the ability of autonomous control with their own thoughts, worries, feelings, strengths, weaknesses and predispositions (Goertzel and Pennachin, 2007). Performing tasks that involve complex calculations requiring substantial effort, time and dedication for humans are very simple for AIs. However, tasks that seem so simple for us, such as voice and image recognition, movement, anticipation and perception are extremely hard for AIs, or the coders of AIs, mainly due to the difficulty of presenting certain predetermined conditions for these AIs to identify when external conditions appear random. Donald Knuth elaborated this phenomenon in these delicate words

[7] The word geek is a slang term originally used to describe eccentric or non-mainstream people; in current use, the word typically connotes an expert or enthusiast or a person obsessed with a hobby or intellectual pursuit, with a general pejorative meaning of a 'peculiar or otherwise dislikeable person, especially one who is perceived to be overly intellectual.

[8] Former CTO of now-defunct Sun Microsystems and author of the article claimed that AI was the greatest threat to humanity's existence: 'Why the Future doesn't Need Us' published on Wired Magazine in April 2000 (http://archive.wired.com/wired/archive/8.04/joy.html).

[9] In the movie series 'The Matrix®' humans waged a war against intelligent machines they had created in the 21st century and blocked the machines' access to solar energy, the machines began to harvest the humans' bioelectricity as a substitute power source.

[10] In the movie series 'The Terminator®', an artificial intelligence defence network known as Skynet becomes self-aware and initiates a nuclear holocaust to end human life on earth.

[11] In the 2015 movie "Ex Machina®", Ava, the AI that is being put through the Turing Test to determine whether she is self-conscious, manipulates her tester into escaping the captivity of her creator, murdering him in the process.

[12] 1,000,000,000,000,000, one thousand million million; 10^{15}.

[13] In a recent study at The Australian National University, the physicists concluded that measurement is everything and reality does not exist if you are not looking at it, at the quantum level. For more information and the press release on the results of the study, see http://www.sciencedaily.com/releases/2015/05/150527103110.htm.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

while answering Nils J. Nilsson's question on how the leading scientist in the AI field think about AI's achievements: 'AI has by now succeeded in doing essentially everything that requires 'thinking' but has failed to do most of what people and animals do 'without thinking', that, somehow, is much harder!' (Nilsson, 2010) Ask Siri® to review Stanley Kubrik's 2001: A Space Odyssey® for example, which requires immense amount of cognitive effort, an amazingly simple task for humans, and at best it will tell you "This is what I found" and provide you with URL addresses to human reviews on the web.

There is a heated debate among world's leading AI scientists as to when humanity can achieve an AGI; the majority claiming it is as near as 2030, some claiming that it will not happen in this century and some arguing that such a day will never come. However, the same scientists predict that once we achieve ANIs, it will not be long before we reach technological singularity (Kurzweil, 2005). When the current state of AI development is considered, it is clear that we still have a long way to go.

2.3. *Artificial Superintelligence ('ASI')*

ASIs represent AIs 'much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.' (Bostrom, 2006). ASIs are the AIs that many fear, will optimize Earth, aiming to fulfil their goals, by eventually removing human kind from the face of it (Yudkowsky et al., 2010). The majority of AI scientists (including Bostrom (Urban, 2015)) foresee that, after the development of an AGI, it will evolve itself into an ASI very quickly (Yudkowsky et al., 2010), as a result of an exponential growth loop. This phenomenon is also known as an 'inteligence explosion' or 'singularity'.[14]

Bostrom dissects superintelligence into three major forms: Speed Super Intelligence, Collective Super Intelligence[15] and Quality Super Intelligence,[16] claiming that all three can coexist within the same ASI entity. Bostrom goes on to argue that any one of the three super intelligences is capable of creating the other two (Bostrom, 2014).

Gordon Moore, co-founder of Intel came up with 'Moore's Law' in 1975, calculating that the world's maximum computational power doubles every two years. This means that it grows exponentially. In Moore's Law exponential growth scenario, proven accurate for forty years and which Gordon Moore himself updated only once, the doubling time is two years, meaning that the growth in the world's maximum computational power every two years is greater than the total of all preceding growth, according to the rule of exponential functions.

Kurzweil predicts, based on Moore's Law and his own evolutionary theory on technological change, Law of Accelerating Returns (Kurzweil, 2001), that singularity will arrive around 2045 (Kurzweil, 2005). Kurzweil stresses that: "Moore's Law is one paradigm among many in computation and computation is one example among the many of the law of accelerating returns" (Kurzweil, 2014) and the history of technology shows that technological progress grows exponentially.[17]

Dr. Albert A. Barlett, Professor Emeritus at Department of Physics, University of Colorado, has presented the following scenario (Arithmetic, Population, and Energy, 2007) to explain the power of exponential growth:

> *Imagine bacteria growing steadily in an empty bottle. They double in number every minute. Now imagine we put one bacterium in the empty bottle at 11:00 am and the bottle becomes full at 12:00 am. At what time was the bottle half full? The answer is 11:59 am, simply because the number of bacteria is doubling every minute. If you were an average bacterium in this bottle, at what time would you first realize that you were running out of space? Let us look at the numbers indicating how full the bottle was at the last 5 minutes. At 11:55 am, the bottle was merely 3% full. Now how many of you will think that there is a problem 5 minutes before 12:00 am? . . . You simply don't need any more arithmetic than this.*

This example also hints the cognitive biases we humans possess, where we create our own subjective social reality from our perception of the information we obtain. In other words, we almost never think that something will substantially change our lives until it actually changes our lives. The evolution of the internet and social media affected us in a very similar fashion. While scientists such as Vint Cerf[18] may have anticipated what Internet might evolve into while they were working on ARPANet®,[19] most people saw it as merely one of the myriad defence projects that the United States' government ran. We can observe cognitive bias when the evolution of Facebook® is considered. Early on, Facebook® had a hard time finding adequate investment for its business, as perhaps only its founder Mark Zuckerberg and first president Sean Parker,[20] was aware of its potential. Similarly, our cognitive bias almost pre-

[14] Joseph Schumpeter was an Austrian-American economist and political scientist. He briefly served as Finance Minister of Austria in 1919. In 1932 he became a professor at Harvard University where he remained until the end of his career. One of the most influential economists of the 20th century, Schumpeter popularized the term 'creative destruction' in economics. He derived the term Karl Marx's works and popularized it as a theory of economic innovation and the business cycle, also known as 'Schumpeter's Gale'.

[15] A system composed of a large number of smaller intellects such that the system's overall performance across many very general domains vastly outstrips that of any current cognitive system.

[16] A system that is at least as fast as a human mind and vastly qualitatively smarter.

[17] It is important to note that Kurzweil was accurate with an impressive ratio of 86% (129 out of 147) in all predictions for 2009 under the book "The Age of Spiritual Machines" he wrote in the 1990s, based on the law of accelerating returns (Kurzweil, 2010).

[18] Vinton 'Vint' Cerf is one of the founding fathers of the internet, the person who developed the TCP/IP protocol with Bob Kahn. TCP/IP protocol enabled computers to exchange information although they are not within the same mainframe network.

[19] The Advanced Research Projects Agency Network (ARPANET) was an early packet switching network and the first network to implement the protocol suite TCP/IP. Both technologies became the technical foundation of the Internet. ARPANET was initially funded by the Advanced Research Projects Agency (ARPA, later Defence Advanced Research Projects Agency, DARPA) of the United States Department of Defence.

[20] An American entrepreneur who cofounded the file-sharing computer service Napster.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

vented us from mapping the human genome, when the creditors complained that the scientists could only unearth 1% of the human genome in the first 7 years of the 15-year project, thinking that the progress is linear and it will take 700 years to complete. However, scientists completed the project on time, 7 years later, proving that progress was exponential.

Therefore, we are likely to dwell in our intuitive linear view on progress, and fail to notice the arrival of ASIs, until they suddenly become a part of our lives, unless we grasp the concept of the law of accelerating returns.

3. Protecting human dominance through regulation or setting tailored goals to maintain human existence

Having a timeless and robust definition of AI is of paramount importance when thinking of regulating AI. One cannot regulate a certain subject without establishing a robust definition of what it regulates. The ambiguity of the definition of AI is mainly due to the "I", "intelligence" of the AI. Concepts like "intelligence", "consciousness", "free will" and "soul" accompanying it are yet to have deterministic definitions although the greatest minds of our planet have tackled them for thousands of years (Burkeman, 2015).[21]

Neither any of the foregoing definitions of AI, nor many other definitions in the academia presents adequate definitions that can be satisfactory when regulation techniques are considered. In addition, the lack of definition is only one of the problems regulators will face; they will need to tackle liability gaps, control and transparency problems (Danaher, 2015).

In light of the foregoing, our primary statement stands firm: it is very early to begin thinking about regulating AIs or AI studies, particularly if such regulations may hinder developments that could prove essential for human existence. The turning point in AI development will probably be the development of ANIs, which should be encouraged through regulation, not restricted. However, if humanity fails in establishing adequate safe guards for ANIs, science fiction may turn into reality. Goertzel and Pitt (2012) call this the 'AGI Sputnik moment'.

3.1. *The great AI hype of 2015*

Elon Musk's and Stephen Hawking's fears, Bill Gates' cautious approach, Kurzweil's optimistic take and Bostrom's realistic analysis on the future that will probably be painted by AIs point to a single fundamental and existential dilemma: Are we going to be extinct because of AIs or will we maintain our existence with the help of AIs?

The cycle of extinction and rise of species may be the greatest success of evolution: ensuring the continuity of life. Over 90% of all species that ever existed on Earth went extinct and humanity's fate will be no different, unless we come up with methods to achieve transcendence over evolution.[22]

Urban (2015) also treats this concept with a less theatrical manner and stresses two major outcomes for a possible 'ASI Sputnik moment'. He states that either the introduction of ASIs will make immortality possible for our species or it will drive the human race into extinction.

Evolution has granted us our strongest instinct: survival. Instinctively we are in a never-ending war with nature, aiming to prolong our existence. In the abstract, the field of medicine solely exists for this purpose. Therefore, instinctively we will either try to eliminate the existential threat that ASIs might pose against us when we face the threat itself or try to eliminate a potential threat prematurely and in so doing cause our own extinction.

3.2. *Reshaping perception on law*

We may be living in the dawn of the age of artificial intelligence today. Consequently, the legal landscape surrounding our lives will require rethinking, as the case was with every big leap in technology. The industrial revolution brought conveyor belts and mechanical manufacturing processes operated by workers for longer and longer hours, which ended in myriad clashes between proletariat and employers. Hence, we developed labour laws, bringing a humanitarian minimum standard for the workers that were suffering from extreme working conditions. Similar legislative efforts followed each time when technologies required us to adapt new paradigms they introduced, technologies such as electricity, telegraph, telephone, railroad, automotive, television, and computers and so on. . .

Below we will seek answers to some exemplary questions as to how AI might reshape our thinking, in terms of certain matters of current and prospective law.

3.2.1. *Liability on damages*

There are very few laws or regulations that address the challenges raised by AIs, and no courts appear to have developed standards so far, addressing who is legally responsible if an AI causes harm.

[21] In 2014, seeking to move matters forward, Dmitry Volkov, a Russian technology billionaire, convened a summit on board a yacht of leading philosophers, including Daniel Dennett, Paul Churchland, and David Chalmers. Perhaps unsurprisingly, they did not reach a consensus, and Chalmers suggested that it was unlikely to emerge within the next century.

[22] The movie 'Transcendence®' treated this concept in 2014 (spoilers ahead). In the movie, a team of scientists led by Dr. Will Caster (played by Johnny Depp) works on developing a sentient quantum computer,[1] Dr. Caster gets murdered by an anti-AI terrorist group and his wife, who is also a scientist, somehow uploads Dr. Caster into the quantum computer and the computer adopts Dr. Caster's personality. However, Dr. Caster undergoes an intelligence explosion the moment it connects to the Internet and evolves into an ASI, becoming another entity, transcending humanity and achieving immortality through nanotechnology. Dr. Caster ends up with obtaining enormous power and enslaving humans he cures, which leads to his own destruction through a computer virus injected by no one other than his own wife. However, at the end of the movie, we see that Dr. Caster and his wife have both achieved immortality through the nano-particles developed by Dr. Caster, despite the total destruction of everything Dr. Caster created in his time as an ASI.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

The diversity and richness of individuals and firms that participate in the creation of an AI will make it difficult to identify the persons under liability. Certain technologies used in the development of an AI may date back to years before such AI is developed. Further, the developers of such technology may never have thought that one day, someone might incorporate their creation into any AI system. In such circumstances, it would be unfair to hold the developer of such technology responsible for a possible tort.

National and international laws do not recognize AI as a legal person. Therefore, current legal systems cannot hold them liable for the damages they might cause. However, what if an AI was fully autonomous and aware of its actions, causing harm knowingly and willingly?

This brings us back to the debate on consciousness. A conscious AI should naturally be liable for its actions. However, how can that be possible if we keep refraining from coming up with an adequate definition of what an AI is as far as legal 'beings' are considered? Should we ascribe legal personhood to them? (Paulius et al., 2015).

3.2.2. *Intellectual property*

IP law and its application places human initiative at its core. Berne Convention of 1886[23] requires an 'author' and an 'artistic work' to begin talking about intellectual property. While there is no limitation as to what form a 'work' can assume as long as humans can perceive it, an author must be a 'human'. A San Francisco court applied and materialized this concept in 2015 by deciding in a lawsuit by PETA, the renowned organization defending animal rights, against David John Slater, a professional photographer, that a macaque money cannot own copyright to a selfie it took using the photographer's camera (Kravets, 2016). What about AIs though? Can they own copyrights to the artistic works they create? Should law consider them as 'individuals'?

3.2.3. *Copyright and AI*

Currently, a handful of AI applications are capable of producing works that resemble 'art', such as Deep Dream and the Cybernetic Poet.

Google's® researchers developed DeepDream® to create a human-like image recognition software to identify certain things through mimicking human cognitive abilities. DeepDream uses Google's artificial neural networks protocol to discern and process images of things to learn what they look like, such as a cat.

Google's developers taught DeepDream what a cat looks like by showing millions of images of cats. Then they put DeepDream's learning and identifying abilities to test by asking it to identify cats in pictures with cats and if found amplify them, introducing a feedback loop to work on. Then the developers introduced a random image to DeepDream and asked it to enhance the image in such a way as to elicit a particular interpretation. This method enabled the developers to understand whether DeepDream understood the essence of the things it learns. As a result, DeepDream searched in the images provided for all the things the developers trained it to recognize and when it found the tiniest bit of reference, it enhanced the relevant reference to make it look like the thing it found similar.

The resulting images were surprisingly close to works of art. Few predicted this phenomenon, including DeepDream's developers.[24]

Ray Kurzweil developed a poem software in mid-80s, a computer-implemented method of generating a poet personality that reads poems and generates analysis models to build its personality, and ultimately writes poems; the 'Cybernetic Poet'. Cybernetic Poet is "provided with an input file of poems written by a human author or authors. It analyses these poems and creates a word-sequence model based on the poems it has just read. It then writes original stanzas of poetry using the model it has created." (Bridy, 2012)

Now, who owns the copyrights of the artistic works created by these AIs?

As explained, current law cannot vest ownership of the copyrights to an AI, as it is not 'human'. However, the laws of the United Kingdom make express provision for copyright in computer-generated works and introduce the following definition: 'works generated by a computer in circumstances such that there is no human author'.[25] The copyright in such works under UK law vests in 'the person by whom the arrangements necessary for the creation of the work are undertaken'. Concordantly, Irish Law adopts the same principles.[26]

However, the UK and Irish approaches to the issue surrounding copyright ownership of computer-generated works and not the works of an AI. Therefore, they overlook the possibility of 'non-human' copyright ownership, ruling out the possibility of an AI that develops its own creative abilities. Who will have the ownership then?

3.3. *Regulate and dominate?*

A regulatory oversight and governmental intervention is a need when the development of AI is considered.[27] It is not common to hear a Silicon Valley entrepreneur who operates on the frontiers technological advancement, urge governments to directly intervene with a developing technology in the hope of preventing humanity to do 'something stupid'. When such thing happened in October 2014, it created a ripple effect and caused 'The Great AI Panic of 2015' (Sofke, 2015), which eventually led an institution called 'Future of Life Institute® (FLI)' to issue an

[23] The Berne Convention, adopted in 1886, deals with the protection of works and the rights of their authors. It provides creators such as authors, musicians, poets, painters etc. with the means to control how their works are used, by whom, and on what terms. The convention bases on three basic principles and contains a series of provisions determining the minimum protection to be granted, as well as special provisions available to developing countries that want to make use of them.

[24] Please see #deepdream at Twitter, Google + and Facebook to see such images.

[25] Copyright, Designs and Patents Act, 1988, c. 48, § 178 (U.K.).

[26] Copyright and Related Rights Act 2000, Part I, § 2 (Act. No. 28/2000) (Ir).

[27] Stated by Elon Musk, one of the bravest entrepreneurs in the world (brave enough to take on the entire automotive and oil industry by mass-producing electric cars substantially better than its oily counterparts).

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

open letter signed by Elon Musk, Stephen Hawking, hundreds of AI researchers in addition to many individuals representing U.S. government (Russel et al., 2015). FLI urged expanded research on how to contain AI systems within the walls of human benefit, including premature regulation. However, FLI used statements such as 'AI systems must do what we want them to do', 'We should identify research directions that can maximize societal benefits' and 'AI super-intelligence will not act with human wishes and will threaten humanity' while providing a research roadmap for AI researchers.

While the 'we' hints at a desired ownership over a technology under development (i.e. AI) and the 'we' implies superiority over 'others' in determining how a technology will be socially beneficial for humanity. It also begs the questions, 'Who are you to claim that you have the capacity to force your desires over the entire human race, and who are you to claim that you can decide what is socially beneficial for us?' Stating that an ASI will definitely be against the humanity's welfare is an unexpectedly ignorant claim, allegedly coming from some of the greatest minds on Earth.

We experienced this line of thought when the Internet reached the masses, disrupting the status quo by lifting the boundaries of communication and information exchange and blurring the sense of control over disseminated information and access to such. The idea of an open interconnected network of networks that is not in anyone's control or under any jurisdiction challenged lawmakers, policy makers and judiciary bodies and it still does. We have still been unable to set out universal rules on Internet (except DNS policies, where all stakeholders over Internet govern these policies through ICANN, a non-governmental organization) for almost 60 years. It would be very naïve to think that we can regulate AI policies, while AI is still in its infancy.

There is almost a consensus within the scientific AI community that definitive predictions on the future of ASI are impossible at this stage, simply because we are so far from creating an ASI, let alone understanding its implications.

3.3.1. *Current and prospective regulatory efforts*

Trying to anticipate ASI's desires from where we stand now in terms of AI development is very similar to a chimpanzee trying to anticipate our motives when we crush an onion to remove its skin. Therefore, aiming to establish regulations to prevent ASIs from obliterating us is a hopeless endeavour.

However, this line of thought may eventually lead regulators to prevent AI research from developing an AGI, fearing that it will break free from the chains of our capacity and become an ASI by itself. For example, John Frank Weaver, an attorney working in the field of AI law, praised the regulators at California when they intervened with Google's self-driving cars and required test drivers to be present in these cars. He even claimed that this as a 'wonderfully swift governmental response to autonomous technology and artificial intelligence' while further supporting four states (Mississippi, Florida, Nevada and California) for passing restrictive regulation on autonomous cars that are not even on the market yet (Weaver, 2014).

3.3.1.1. Legislative efforts for autonomous vehicles. Nevada is the first U.S. state to enact a legislation authorizing the operation of autonomous vehicles in 2011 and was then followed by six other states, with many other states in still pending status with reference to their respective autonomous vehicle legislations. Tennessee among those who did enact such legislations stands out with its enabling and refreshing legislation wherein it prohibits local governments from banning the use of motor vehicles equipped with autonomous technology (Legislatures, 2016).

Throughout the world, legislators are working to incorporate autonomous (driverless) vehicles into their legislations to allow this thriving technology bloom and develop further, which brings hope.

The Convention on Road Traffic,[28] of the United Nations, ratified by 73 countries, is in the process of amendment to allow automated vehicles on roads in many countries. European Road Transport Research Advisory Council published the roadmap for automated driving for Europe.[29] German Federal Highway Research Institute published a report on the status of German legal landscape pertaining to vehicle automation technologies, indicating the areas of improvement on research, legislation and involvement of government agencies.[30] Netherlands, Sweden, Japan and many other developed countries are actively working on improving the conditions of economic and legislative environment to enable swift development and consequently to reap the benefits of being involved in the forefront of innovative technologies.

While governments are honing in on preparing the legislative grounds for the operation of autonomous vehicles, academia adopts a wider approach and handles the concept in a wider manner, and works on determining the adequate policies for robotics and AI.

3.3.1.2. The RoboLaw project. The main objective of the RoboLaw project ("Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics") is to understand the legal and ethical implications of emerging robotic technologies and to uncover whether existing legal frameworks are sufficient in light of the rapid expansion of robotics technologies.[31]

The project was launched in March 2012 and funded by the European Commission (Paulius et al., 2015). The project produced the "Guidelines on Regulating Robotics", which was then presented to the European Commission, to create the legal framework surrounding the development of robotic technologies in Europe.

The RoboLaw Project considered industrial robots, domestic robots, care robots, medical and surgery robots, autonomous vehicles, and humanoids/animaloids. The report discussed five

[28] Namely the Vienna Convention on Road Traffic is an international treaty designed to facilitate international road traffic and to increase road safety by establishing standard traffic rules among the contracting parties.

[29] Please see the full text of the "Automated Driving Roadmap" at the url address http://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf.

[30] Please see the full report, "Legal Consequences of an Increase in Vehicle Automation, Consolidated Final Report" at the url address http://bast.opus.hbz-nrw.de/volltexte/2013/723/pdf/Legal_consequences_of_an_increase_in_vehicle_automation.pdf.

[31] Please see the url address http://www.robolaw.eu/projectdetails.htm for more details on the project.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

essential legal areas for robotics: (i) health, safety, consumer, and environmental regulations; (ii) liabilities; (iii) intellectual property rights; (iv) privacy; and data protection and (v) capacity for legal transactions (Anon, 2015).

3.3.1.2.1. *Health, Safety, Consumer and Environmental Regulation*. The report identifies that common usage of robotics in hospitals, homes, commercial areas and our daily lives will require a new wave of legislations to cope with the prospective health and safety matters.

3.3.1.2.2. *Liability*. The report argues that imposing substantial liability on manufacturers, owners or users of robots for damages caused to third parties may increase safety while inducing wider social acceptance of robots. However, the report also argues that such approach on a liability regime may result in the displeasure of tech industry, consumers and, in the end, the general public, and may slow down the development of AI and robotics technologies. Therefore suggests a balanced approach between the interests of manufacturers, users, and third parties, and between risk regulation and stimulation of innovation, to encourage research, innovation and experimentation on these technologies, for increasing welfare in health, transport, commerce and other areas of business.

3.3.1.2.3. *Intellectual Property Rights*. RoboLaw Project indicates the lack of legal provisions that specifically apply to robotics. RoboLaw Project states that further research would be beneficial to determine whether the current application of intellectual property rights sufficiently meets the needs of the robotic industry and society.

3.3.1.2.4. *Privacy and Data Protection*. The RoboLaw Project suggests implementation of legal requirements into the robot's software and interface through the 'privacy by design' approach, such as data security through data encryption and data access control in order to comply with the data protection requirements.

3.3.1.2.5. *Capacity for Legal Transactions*. The report stresses the lack of legal personality of robots and indicates that robots are seen as 'mere tools' to carry out commands that can, directly or indirectly, be attributed to human beings. Consequently, this approach requires the legal responsibility for robot actions to rest with their human 'masters'.

It is possible to attribute legal personality to robots through legislative effort. Non-humans such as corporations, associations, and foundations gain their legal personalities through registration. The registration principle could be extended to robots and AIs (including requirements how robots can prove their registered identity); the capability of owning property is less easy to create, although legal constructions could be devised to accommodate this.

The report concludes with indicating that if these issues concerning legal personality are resolved at a certain point in time, more practical requirements and rules pertaining to legal acts will come into play, such as implementing legal conditions into the machines to make it possible for them to enter into a contract.

Lawmakers need to familiarize themselves with the potential benefits of AIs. Strict rules may prevent humans from the possible damages of AIs. However, these rules will also dampen possible improvements. Therefore, lawmakers should consider the balance between protection of humanity and development in technology.

4. Conclusion

When aiming to regulate currently non-existent technologies, we must avoid this approach at all costs. Putting restrictions on developing technologies based on our personal presumptions might indeed help us to avoid extinction at the hands of 'evil robots', but it might also cause our extinction due to natural reasons, such as evolution by making it harder for the human race to use technology to adapt.

Based on the statements of Elon Musk, Steve Wozniak, Bill Gates, Bill Joy, Stephen Hawking and FLI's open letter, it is clear that what they all fear is an 'unfriendly AI' and what they all want is a 'friendly AI' in the abstract.

The terms 'friendly' and 'unfriendly' do not refer to a personal trait of an AI system. These terms refer to whether the actions of an AI will have a positive or a negative impact on humanity (Urban, 2015). This is because AIs are computers and they do not have human values. We tend to anthropomorphize[32] AI and attribute them with our moral values such as 'good and evil', 'moral and immoral' that are formed by our consciousness. These attributes developed only after thousands of years of social interaction. AIs will not share these human traits unless we specifically create them to do so. They operate on a task and goal oriented manner. To illustrate this point, for instance, there is an AGI, whose main task is to ensure that trees in a certain pine tree plantation are under protection from alien spores to keep the tree DNA as pure as possible. We should not be surprised when such an AGI takes drastic measures as far as obliterating the entire flying bug population in the area. One who is unaware of the goals of this AGI might easily label it as 'evil' and a 'danger to humanity' as he/she has no preconception on what the AGI's motives or goals were. Similarly, a chimpanzee fearing that the crushing of an onion is a sign of aggression might attack us. Ironically, this view is very similar to the perspective of those who propose premature regulation of AIs.

This goal-oriented approach applies to law and legal institutions as well. Ryan Calo[33] argues that AI (and robotics) will introduce unprecedented legal issues on unparallelled concepts and scenarios, making these issues exceptional. Which means law should handle them in an exceptional manner, as their introduction into the mainstream will require a systematic change to the law or legal institutions to reproduce the existing balance of legal values (Calo, 2015). Here we can draw from Lawrence Lessig's[34] three lessons on regulation over cyberspace where he successfully foresaw the issues Internet faced with respect to application of law: Limits on law's power over cyberspace, transparency and narrow tailoring

[32] Projecting human values on a non-human entity.

[33] Assistant Professor, University of Washington School of Law. Faculty Director, University of Washington Tech Policy Lab. Affiliate Scholar, Stanford Law School Center for Internet and Society.

[34] Lester Lawrence "Larry" Lessig III (born June 3, 1961) is an American academic, attorney, and political activist. He is the Roy L. Furman Professor of Law at Harvard Law School and the former director of the Edmond J. Safra Center for Ethics at Harvard University.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

(introducing very narrow scoped regulation over specific properties of a subject) (Lessig, 1999). Through these lessons, we might anticipate AI's future struggles with the law. These three lessons point us that during the development of this new technology, the regulators will try to govern the core of this technology, which is software code for AIs, without resorting to restrictive regulations. This will then introduce transparency issues, since governing the software code of AI will mean regulating it subtly, keeping it out of judicial review. Finally, regulators will aim to put AIs under strict, very specific and narrowly tailored secondary regulations. These honed legal texts might hinder the development process of AIs. Unless they solely represent technical, safety and health standards that are essential and not precautionary measures, just like we currently experience in the automotive sector.

In this regard, we need to investigate the values that illuminate the entire law and draw the cores of these values there, then seek ways to apply these values to our newly created Law of Artificial Intelligence, where we will discuss the applicability of these core values to the AI scene.

AIs will likely act to fulfil the goals provided by its programmers (Urban, 2015). They will not have the capacity to develop new goals by themselves unless required by their code. However, the method an AI may adopt to fulfil these goals might cause it to be 'unfriendly', based on the consequences introduced through the process.

Our best bet in ensuring human existence with the help of AIs would be by introducing regulations that support AGI research, focusing on creating models of AI risks and AI growth trajectories (Yudkowsky et al., 2010). Research should also look at how to set safe goals for AIs, and cover all philosophical, ethical and scientific issues surrounding AI development and behaviour. Above all, we should refrain from anthropomorphizing AIs. AIs are not human, but they may be humankind's greatest hope.[35]

Until we come up with solid proposals, we should let the scientists to their science and nurture AI into the technology that might one day save us from extinction.

REFERENCES

Anon. Final Report Summary – ROBOLAW (Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics), s.l.: European Commission, 2015.
Aristotle. Politics. s.l.: Dover Thrift Editions, 2000.
Arithmetic, Population, and Energy. [Film] s.l.: s.n., 2007.
Boesch C. What makes us human (Homo sapiens)? The challenge of cognitive cross-species comparison. J Comp Psychol 2007;121:227–40.
Bostrom N. How long before superintelligence? Linguistic and Philosophical Investigations 2006;5(1):11–30.
Bostrom N. Superintelligence: paths, dangers, strategies. Oxford: Oxford University Press; 2014.
Bridy A. Coding creativity: copyright and the artificially intelligent author. Stanf Technol Law Rev 2012;5(2012):1–28.
Burkeman O. Why can't the world's greatest minds solve the mystery of consciousness? [Online], <http://www.theguardian.com/science/2015/jan/21/-sp-why-cant-worlds-greatest-minds-solve-mystery-consciousness>; 2015 [accessed 19.05.16].
Calo R. Robotics and the lessons of cyberlaw. Calif Law Rev 2015;103(3):513–63.
Danaher J. Is regulation of artificial intelligence possible? [Online], <http://hplusmagazine.com/2015/07/15/is-regulation-of-artificial-intelligence-possible/>; 2015 [accessed 19.05.16].
Goertzel B, Pennachin C. Contemporary approaches to artificial general intelligence. In: Cognitive technologies. Rockville: Springer; 2007.
Goertzel B, Pitt J. Nine ways to bias open-source AGI toward friendliness. J Evol Technol 2012;1:116–31.
Kravets D. Law & disorder / civilization & discontents. [Online], <http://arstechnica.com/tech-policy/2016/01/judge-says-monkey-cannot-own-copyright-to-famous-selfies/>; 2016 [accessed 19.05.16].
Kurzweil R. The law of accelerating returns. [Online], <http://www.kurzweilai.net/the-law-of-accelerating-returns>; 2001 [accessed 19.05.16].
Kurzweil R. The singularity is near. London: Penguin Group; 2005.
Kurzweil R. Kurzweil accelerating intelligence. [Online], <http://www.kurzweilai.net/images/How-My-Predictions-Are-Faring.pdf>; 2010 [accessed 19.05.16].
Kurzweil R. Google I/O Biologically Inspired Models of Intelligence [Interview], 25 June 2014.
Legislatures NCS. Autonomous | Self-Driving Vehicles Legislation. [Online], <http://www.ncsl.org/research/transportation/autonomous-vehicles-legislation.aspx>; 2016 [accessed 19.05.16].
Lessig L. The law of the horse: what cyberlaw might teach. Harv Law Rev 1999;113(2):501–49.
Locke J. An essay concerning humane understanding. In: Four books. London: Eliz. Holt for Thomas Basset; 1690.
Markoff J. Computer wins on 'Jeopardy!': Trivial, It's Not. [Online], <http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all.&_r=0>; 2011 [accessed 19.05.16].
Nilsson NJ. The quest for artificial intelligence. New York: Cambridge University Press; 2010.
Ormandy R. Google and Elon Musk to decide what is good for humanity. [Online] <http://www.wired.com/insights/2015/01/google-and-elon-musk-good-for-humanity/>; 2015 [accessed 19.05.16].
Paulius Č, Jurgita G, Gintarė S. Liability for damages caused by artificial intelligence. Comput Law Secur Rev 2015;31(3):376–89.
Polansky R. Aristotle's De Anima. s.l. Cambridge University Press; 2007.
Rizvi SH. Avicenna (Ibn Sina) (c. 980–1037). [Online], <http://www.iep.utm.edu/avicenna/>; n.d. [accessed 19.05.16].
Russel S, Dewey D, Tegmark M. Research priorities for robust and beneficial artificial intelligence. [Online], <http://futureoflife.org/static/data/documents/research_priorities.pdf>; 2015 [accessed 19.05.16].
Scherer MU. Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harv J Law Technol 2016;29(2).
Silver D, Hassabis D. Google research blog. [Online], <http://googleresearch.blogspot.co.uk/2016/01/alphago-mastering-ancient-game-of-go.html>; 2016 [accessed 19.05.16].
Sofke E. An open letter to everyone tricked into fearing artificial intelligence. [Online], <http://www.popsci.com/open-letter-everyone-tricked-fearing-ai>; 2015 [accessed 19.05.16].

[35] We would like to thank Edward Kenny for his very productive cold-eye review and comments on our article, as well as Noyan Utkan, for his diligent research and input.

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003

The Internet Encyclopedia of Philosophy (IEP). Avicenna (Ibn Sina) (c. 980–1037). [Online], <http://www.iep.utm.edu/avicenna/>; n.d. [accessed 19.05.16].

Urban T. The AI revolution: the road to superintelligence. [Online], <http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>; 2015 [accessed 19.05.16].

Vardi MY. Artificial intelligence: past and future. s.l.: association for computer machinery, 2012.

Weaver JF. We need to pass legislation on artificial intelligence early and often. [Online], <http://www.slate.com/blogs/future_tense/2014/09/12/we_need_to_pass_artificial_intelligence_laws_early_and_often.html>; 2014 [accessed 19.05.16].

Yudkowsky E, Salamon A, Shulman C, Kaas S, McCabe T. Reducing long-term catastrophic risks from artificial intelligence. [Online], <https://intelligence.org/files/ReducingRisks.pdf>; 2010 [accessed 19.05.16].

Please cite this article in press as: Gonenc Gurkaynak, Ilay Yilmaz, Gunes Haksever, Stifling artificial intelligence: Human perils, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.05.003