
TANEL KERIKMÄE,

*Tallinn Law School, Tallinn University of Technology,
Akadeemia tee 3, 12618 Tallinn, Estonia
e-mail: tanel.kerikmae@ttu.ee*



PEETER MÜÜRSEPP,

*Tallinn Law School, Tallinn University of Technology,
Akadeemia tee 3, 12618 Tallinn, Estonia
e-mail: peeter.muursepp@ttu.ee*



SANDRA SÄRAV,

*Tallinn Law School, Tallinn University of Technology,
Akadeemia tee 3, 12618 Tallinn, Estonia
e-mail: sandra.sarav@ttu.ee*



ARCHIL CHOCHIA,

*Tallinn Law School, Tallinn University of Technology,
Akadeemia tee 3, 12618 Tallinn, Estonia
e-mail: archil.chochia@ttu.ee*



UDK 304.08:174

Ethical Lawyer or Moral Computer – Historical and Contemporary Discourse on Incredulity between the Human and a Machine

***Abstract:** Contemporary society is largely influenced by the digital revolution. Modern computer technology contributes to our everyday existence, yet in a way simultaneously directs our lives. The penetration of digital solutions has gone far beyond merely providing assistance;*

somewhat unexpectedly, its impingement has reached even to most humane spheres, such as justice. Could, for instance, the entire procedure of legal decision-making be automated? A few decades ago, this question would have seemed completely ill-suited. Nevertheless, we acknowledge that the situation today has transformed drastically. Bearing these aspects in mind, given article observes how humankind has come to this point of digitalisation, while further attempting to pinpoint the potential threats related to automatisisation of justice. However, the article does not refer back to the years to the Antikythera mechanism¹, as little is known about its construction; thus it will be confined by the developments of the last centuries, leading up to the present moment.

Key words: Artificial intelligence, intelligence augmentation, automatisisation of law, mechanised thinking.

The origins of mechanised thinking

Considering the focus point of the article, it is reasonable to start with referring to Leibniz. This law educated German mastermind from the XVII-XVIII century was at least sometimes thinking like a modern person, for his tools for precisions and accuracy included mechanical equipment. Leibniz's dream was to create a universal symbolic language (*lingua characteristica universalis*) and the symbolic logic using this tool (*calculus ratiocinator*) (Tamme et al 1997). With the help of the latter, it ought to have been possible to derive new veritable propositions and check the accuracy of any given speculation. It goes without saying that a tool like that would be very helpful especially for a lawyer. Leibniz believed that these derivations and inspections could be done with a specific machine.

The logical system that Leibniz created in 1680's is similar to the system created by George Boole back in 1847. Yet, it is Boole who is considered to be

the founder of mathematical and symbolic logic. The reason is probably due to Leibniz having been ahead of his time. His ideas did not have tremendous impact on the following hundreds of years. The academic minds started comprehending Leibniz's innovative thinking only with the birth of the Boolean logic. It is well known that Leibniz was not only a thinker but also an engineer and a constructor. Amongst others, he constructed one of the first calculating machines to commit arithmetic operations.

The world at the beginning of 18th century, not even the academic circles, was certainly not ready for the automatisisation of human thinking. The views started gradually changing by the mid-19th century. The innovative algebraic logic of George Boole, as it emerged, rather quickly caught the attention of his colleagues, resulting in referring to it as Boolean algebra. Boole's approach was indeed similar to Leibniz's because similarly to the latter, the intention of the former was to reach the arithmetic of thinking. However, the novelty of the Boole's approach derived from his concentration on propositional calculus, i.e. sentential logic (Tamme et al 1997). It is true that Boole did not see beyond sentential lo-

¹ Most possibly it was a device calculating astronomical phenomena that in today's meaning did not have a direct practical value, but in the ancient world could have been much help for everyday planning.

gic¹ and neither did he or Frege build computers. Charles Babbage began to take this into account in the beginning of 19th century. In 1822, Babbage completed the first prototype for programmable computer. When Blaise Pascal's computer in the 17th century could only sum up and subtract and Leibniz's one could additionally multiply, divide and take a square root, then Babbage's computer was able to take tasks that the machine could mechanically follow and reach the wanted results².

Thus, the first substantial steps towards artificial intelligence were made in the middle of the-19th century. Naturally, the mechanical machine was not flexible enough to model human thinking. Furthermore, nor Pascal's, Leibniz's or Babbage's computers were very reliable, that is if we can talk about reliability at all with them. Electronic equipment that became available to the computer enthusiasts in the middle of the 20th century was needed. Here, we are going to exclude the possible reasoning analysis for now and concentrate on the interest of mathematicians and logicians for the theory of building up algorithms and programming. One does not need a functioning computer for working with the theory. An abstract imaginary computer is enough. This kind of abstract computers and their programming theories were created independently from each other, by an American Alonzo Church (1903–1995) and an Englishman Alan Turing (1912–1954).

¹ The predicate calculus was invented by Gottlob Frege, who in 1879 published a revolutionary work «Begriffsschrift».

² As already said, Babbage's machine as an innovation was programmable. But not Babbage conducted the programming work but his colleague Ada Lovelace, an especially talented mathematician.

The reader has probably heard about the Turing machine. It is a simple hypothetical computer that consists of an endless tape that performs the tasks of a memory, of a reader-writer's input and a table that contains the controlling program. Turing's position was that with his machine one can simulate however complex computer. Therefore, Turing machine presents a universal computer and its abilities can be studied without the machine itself existing as hardware. Particularly was Turing interested in the question of what can be calculated in principle, meaning what kind of solution held problems have an algorithm for the result? (Turing 1950) For example, Turing proved that predicate calculus is not solvable, meaning it is not always possible to decide if whatever statement written in the language of predicate calculus is right. Propositional calculus, however, is solvable. Indeed, that was known already before Turing.

By keeping in mind the aims of this paper, we are more interested in the so-called Turing test and everything involved with it, rather than the Turing machine. According to the standard approach of the test (Turing 1950) it is an experiment where a person in the role of a judge communicated with two distinct partners out of whom one is also a human but the other one is a computer. All participants are in separate rooms to avoid any visual contact. The judge asks partners questions with the intention to find out which of the respondents is human and which one is a computer. If the judge is not able to guess correctly in a certain period of time who is who, then we can declare that the computer has

passed the test. Should we take a position that a machine like that can think? Turing did not think so. It is important to notice that the goal is not to necessarily give the right answers but only those that a human would normally give. The computer taking the Turing test doesn't have to correct a human but it only needs to look as close to human as possible¹. Until this day prevails an opinion that no computer has yet passed the Turing test. It is known that the best results showed Eugene Goostman by tricking one third of the judges, but that program didn't simulate a grownup person but a thirteen-year-old Ukrainian boy (Warvick et al 2016). There are cases where the number of «scammed» judges reaches to nearly half of the participants, but these results have been gained in a situation where the judges didn't know that one of the communication partners might be a computer. Furthermore, they had even no reason to suspect that.

While we are discussing the decisions made during a real judgement making and its automation, then what is really important? Probably their accuracy with existing jurisdictions, validity and justice, as much as the latter can be considered. If a computer can fill these requirements without being recognized, then is there even a problem?

If and how can (analytic) philosophy assist? Naturally, there are points of contact between logic that is necessary for the creation of the computer world and analytic philosophy. For example, Got-

¹ With the standard approach to Turing test it is important that the judge is aware of the situation that one of the communication partners is a computer and another one a human and that the whole event is directed to recognize the computer.

lobFrege these days is considered to be the founder of modern logic and also analytic philosophy. In his only known public lecture given in 1929 or 1930, Ludwig Wittgenstein (1965), perhaps the most influential analytic thinker of all times masterfully marked the difference of the factual world and the normative world of ethical judgments. Wittgenstein explains addressing his audience: «Suppose one of you were an omniscient person and therefore knew all the movements of all the bodies in the world dead or alive and that he also knew all the states of mind of all human beings that ever lived, and suppose this man wrote all he knew in a big book, then this book would contain the whole description of the world; and what I want to say is, that this book would contain nothing that we would call an ethical judgment or anything that would logically imply such a judgment» (Wittgenstein 1965). It seems that to this day machines cannot reach the normative world, where the ethical judgements dwell. The computers function in the factual world better and better. Perhaps one day we can even build a computer that is omniscient like this hypothetical man in Wittgenstein's metaphoric example. However, even an omniscient computer may not be able to orient itself among ethical judgements so important to humans. The computer might even not recognize the demarcation line between the two worlds. To borrow some more from Wittgenstein: we can tolerate a bad pianist or a bad tennis player but we cannot tolerate a bad person who is not interested to stop lying (Wittgenstein 1965). The computer can either tolerate all of those or not tolerate any. Can any machine ever reach the level of

recognizing the principal difference between the two worlds, the descriptive and normative ones? Shortly, we shall see that John Searle is probably ready to answer this question positively.

Then again, we are facing a question whether juridical decisions and explanations should at all exit the borders of the factual world. If not, then with legal discussions we can count on the strictly regulated world description provided by the logical positivists. However, Stephen Toulmin (1969) claims that the logical positivists only modified the metaphysics of Hume and Mach into the symbolism of Whitehead and Russell. If so, then it is not so easy to get rid of the normative world.

Alan Turing and his test are often connected to Alfred Ayer's (2001) approach that he presented in his most known work. According to Ayer, the only way to tell a conscious minded human from a machine or from a dull person is by committing empirical tests. That sounds a lot like Turing. Yet, it is not known if Turing was familiar with Ayer's philosophy at all. The main practical problem seems to be compiling selective tests successfully. Modern computers have been smart for already decades. For instance, they are capable of giving not only direct concrete answers to presented questions, but also they can in a way evaluate situations and hence generate results independently, for which no direct information has been implemented into the computer. Computers like that existed already in second half of the 1970's. It seems that when a computer can evaluate situations and generate answers regarding corresponding indirect questions that mostly even humans would

answer the same, then what else there is to wish. From here, we can deduct that a computer like this understands things, meaning it is thinking. But is it necessarily so?

Let's give the proposed question a closer look. In 1980 John Searle published a paper with breaking significance with its core idea known as the Chinese room argument. Searle claims directly that his arguments of every person's mental phenomenon simulations can be applied on a Turing machine (Searle 1980). Let us have a look at the content of the Chinese room. The use of a computer is carried out on a traditional input-output method. We input data into the computer, program it to complete a certain task and then receive the output. If the delivered result is close to expectation then we probably don't suspect anything and we peacefully take the delivered results into use. Meanwhile there is no control over how the results occurred in the computer. The fact that we are able to for example receive correct answers in Chinese when we have written and entered the questions into the computer with Chinese characters does not mean that the computer necessarily understands Chinese language. Naturally, one can claim that the automated approach without understanding the content is not an issue if the results are correct. If it is so, then it has to be especially certain that the automation works reliably and universally, especially in regards of influencing people's destiny when making juridical decisions.

Can a machine like that even exist? Subsequently, Searle discusses in his article's concluding part that human thinking is intentional; if we are thinking then

we think of something; therefore human brains and probably also animal brains are built up so that intentionality is the product of a brain's causal characteristics. Certain processes of a brain are possibly enough for intentionality. But functioning of a computer program can't be enough for intentionality. Naturally, a human could also realize a program without possessing the needed intentionality. From that appears that in order to explain intentionality caused by a brain, something more is needed than just realization of the program. Searle calls this indispensable addition a causal force. Can a machine have a causal force like that? Probably it can if the machine has a close build-up to a human brain. Searle's answer to a question of whether a machine could think seems even unexpected at first place. Searle confirms: «Here an argument is developed that only a machine could think and only a very special kind of machines, particularly brains; hence the machines that have an inner causal force equivalent to brain's» (Searle 1980). Therefore, a machine that imitates human brain's causal force is needed to automate human thinking as an activity, to ensure the intentionality of thinking. This is what it means to understand a situation. Such machine would perhaps also be able to distinguish between the descriptive and the normative.

**Simulation of justice:
Artificial Intelligent (AI)
or Augmented Intelligence (IA)?**

Concerning the environment surrounding us, we cannot only refer to the digital influence. Perhaps most importantly, we must acknowledge the society, where

the rule of law is superseding all societal aspects. Whereas in the latter order, the bearers of crucial roles are the legislators, interpreters of law, as well as the judiciary. Could we welcome the use of technology in legal decision-making? Amongst conservative lawyers, one could presume that the use of automated logic in law and in practice could be subjected to heavy criticism. We should probably agree with Menne (1964) on that «the logical development of law is more possible than the automation of legal decisions». Next we will discuss the questionable triumphs and relative failures in legal technologization practice through philosophical and historical prism.

First of all, one ought to refer to legal positivism, i.e., the domination of written norm above general values and the ever-changing interpretations. Moritz Schlick, Ernst Mach, Ludwig Wittgenstein, Bertrand Russell and August Comte – the fathers of (logical) positivism – were not lawyers but physicists, mathematicians, sociologists, yet their platform was the same as for legal philosophers: John Austin, Jeremy Bentham, Hans Kelsen, H.L.A Hart and Joseph Raz. It is often assumed that positivism in general, as a line of methodology, can be attuned with common understanding of law. According to Keat (1971) for example: «[f]or the positivist, it is the aim of science to provide us with predictive/explanatory knowledge concerning these privileged entities». and «[s]cientific theories are to be seen, primarily, as sets of highly general, law-like statements, preferably taking the form of mathematically expressed functional relationships between measurable variables». Meanwhile, the majority

of the use of the historical method of idealism/intuition in the social sciences (and here it ought to be mentioned that law is certainly a part of social sciences) and the use of comparative and teleological methods in law, give us a reason to assume that in legal research, the creation of legal norms, adaptation and interpretation are involved with numerous social variables. Although law may indirectly mean a positivist model, then the automation of decision-making is highly complex and insecure.

In the globalising world, the potential of uniform automatization of law further depends on the geographical location (a specific legal system). To concede with King (2011), «[l]aws vary substantially from one jurisdiction to the next, such that content or services may be legal in one jurisdiction and unlawful in another. This variation creates a tremendous demand for geolocation technologies that can accurately screen users by jurisdiction, so as to allow online vendors to do as much business as possible without breaking the law». Reidenberg (2015) is rather blunt by adding roughly that: «technologically-created ambiguity challenges sovereign jurisdiction», and believes that the use of ICT-tools per se creates excessive tension between conservative justice and a lot more liberal digital world. To paraphrase him, the rule of law should stand higher than technological determinism (Reidenberg 2015). Geographical location is problematic in these regional jurisdictions such as the European Union and the Digital Single Market that the Union is struggling to effectuate in cooperation with the Member States. Does the regulation of already existing or potential

e-technologies and e-services in reasonable abstractedness require more effort in regards that 28(27) national peculiarities of legislation have to be considered?

Huhn (2002) has concluded following: «All rules of law may be stated in the following hypothetical form: *If* certain facts are true, *then* a certain legal conclusion follows». Artificial intelligence (AI) seems to be the right tool to draw conclusions on concrete facts. Prior to making premature assumptions on legal reasoning regarding artificial intelligence, one needs to be familiar with the prevailing critique of the field. One good example is an essay written by pioneers Buchanan and Headrick in 1970, during the time when experts were attempting to study «thinking machines» and consequently apply them to decision making (see for instance the «Taxman» project (McCarty 1977)). The authors present a primary problem that proceeds from the different mentalities of professional tribes in their field of interest. They claim that «lawyers have viewed the computer as, at most, a storehouse from which cases and statutes might be retrieved by skilfully designed indexing systems». While the computer scientists consider the law to be a collection of facts and «correct» legal principles and «they have assumed that the computer can be most helpful to the lawyer if it can retrieve the right answers quickly» (Buchanan et al 1970). This conception leads to an assumption that legal professionals rather see themselves rather as authors of a unique argument, whereas the argument would base on the substance that only partially originates from a (computerised) database or an archive. More than 30 years later, Sunstein (2001) presented an-

other dichotomy, offering that a computer and AI cannot reason based on analogy, insofar as «they are unable to engage in the crucial task of identifying the normative principle that links or separates».

Today we are facing an abundance of «Law Practice Software Products» that are mutually usable for their software¹, including streamline, clouding, built-in reminder and invoicing systems and calendars, but also those that administer certain legal areas. Yet the main discourse regarding artificial intelligence in terms of legal justification promptness has remained similar to what Turing had in mind (thinking machine vs. imitation of human mind). Possibly most known software amongst others is the well-known IBM Watson², that by using natural language treatment and machine learning analyses unstructured data and selects the most important information from the documents. Nevertheless, IBM Watson has several implementations outside of law area and it is only one form of augmented intelligence (IA)³, therefore it is not exactly an artificial intelligence, meaning an intellect system like that is meant to complement human activity through computer-human interaction whereas decision

making remains to the human. One could draw a parallel here with how calculators initially worked in the hands of engineers and architecture professionals (Xia et al 2013).

Ergo, in law practice legal science, IBM Watson contributes to faster and more effective (time-wise) research by grasping, collecting and analysing the data based on entered inquiries, although its intellect is limited to operating only upon commands. Strictly more law-oriented is a so-called artificially intelligent lawyer (we would call it an intelligence augmenting lawyer – in the IA form), ROSS⁴, that has been built on IBM Watson's platform and works on a research basis, relying on direct inquiries in a question form and giving immediate answers while at the same time being an independent learning system. Similarly to Watson, it is a supporting system, leaving the legal justifications to the command giver by default. An example from Estonia – in order to improve juridical (e)-services and abandon archaic *modus vivendi*, the former Estonian chancellor of justice introduced his start-up company Avokaado, which enables the clients to create drafts of standard legal documents, such as contracts, online. He claims that the goal is to make regular forms more easily accessible and affordable (The Baltic Course 2016). Teder, who currently works as an attorney at law, believes that the traditional legal services field is a late bloomer in terms of implementing innovation and stresses that «it is no longer acceptable to ask for a tailor made price for standard solutions» and suggests that

¹ See examples from: <http://www.captterra.com/law-practice-management-software/>.

² See the description and official website at: <http://www.ibm.com/watson/>.

³ The conceptual framework for augmenting human intellect was initially introduced in the early 1960s by Engelbart (1962), where it was referred to as increasing the human intellectual capacity to approach and solve particular problems. See his original work: Augmenting Human Intellect: A Conceptual Framework at: <http://www.dougenelbart.org/pubs/augment-3906.html>. Engelbart's work has been customised to contemporary human-computer interface by Cambridge and MIT scholars' joint work: Xia and Maes (2013).

⁴ See more about the ROSS at its official website at: <http://www.rossintelligence.com/>.

the area of legal services will change dramatically within the following years (The Baltic Course 2016).

Lippe and Katz argue that though, «Many imagine Watson might displace lawyers for legal reasoning. We believe that systems like Watson are very unlikely to displace the reasoning processes of lawyers» (Lippe et al 2014). Nevertheless, software systems, like the aforementioned ones, are potential reduction or exclusion means in the process of systematisation of legal order. Even though computers can be ridiculed in the sense of them ever-replacing legal decision-makers, they can be rather successful in structuring the legal knowledge. They could:

a) Explicate the sources of legal norms and their hierarchic order and find contradictions and overlaps;

b) Analyse lawyers' arguments from the viewpoint of presented values and principles by using the «big data» method and by that move closer to solid and valid system of values;

c) Analyse the methods of textual interpretation and their applicability in practice;

d) Categorise cases, difficult cases and pick out elements from the arguments that were influenced by legally external facts;

e) Be «the identifier of facts» for processing the digital legal documents¹.

We tend to think that a computerised brain is not (yet) ready to compile arguments, and, by nature, remain rigid. Why has creating the ideal digital decision-maker so far failed (or not received enough support)? We consider the pri-

¹ A term, used by Dewitz more than two decades ago. See: *Dewitz* (1995).

mary reason can be found in the internal systematic imperfection of justice and legal system and their reluctance to obey to mechanical and transparent estimators that with great probability may not sense justice as «living phenomenon» (similarly to the Chinese room argument). Simultaneously with digitally separate fields, representatives of different narrow specialities such as prosecutor, attorney or an in-house lawyer might express diverse approaches or attitudes towards the use of artificial intelligence.

The «Rigidity-argument» is probably not only subjective opposition of the legal practitioners, but also a grounded apprehension of the legislators. Prakken and Sartor (2015) convince their readers in the so-called backwards effect: (even) legislators cannot always fully predict on which specific matters a certain law should be applied. They also refer to the abstractedness, a general category of exceptions that can be interpreted in specific cases that create uncertainty and leave room for disagreement. Therefore, two important elements of jurisdiction can be derived from the abovementioned authors' work: a) law's uncertain orientation towards the future; b) its institutional nature, which makes the legal justification be connected to context (the addressee, drafter, implementer, enforcer) that are conditioned by the premise that nobody is able to foresee the tomorrow's world. Here, the connection with digitalisation and the age of information technology is even more evident, as the technical innovations derive from inventions that cannot be planned ahead. Poscher (2011), who is looking for reasons why legal interpretation is rarely predictable, refers to the fact

that law cannot be more concrete than the undefined life itself and finds that with «difficult cases» that are «phenomena on the borderlines of law» cannot have predictable mechanical assessments or legal interpretation, but rather strictly depend on the decision-maker.

Why do we even think and discuss about the likelihood of using digital means in the sphere of law? Naturally, cost-effectiveness¹ (time, human resource), avoiding the extra-legal elements in the process of ensuring the rule of law (politics, ideologies); yet also better predictability of legal decision (legal certainty).

Computers were first created to calculate, not for dealing with social processes. Yet sometimes it seems that the leading law firms have a sort of an «arms race» for creating the most efficient software, whereas the goal is to sell their product to competitors and not reduce the burden of their colleagues. Meanwhile, as mentioned before, the lawyers themselves are not always excited about being dependent of the new dimensions of the digital world and at times even discuss about «taking a break» or «disconnecting» from technology, but usually find that instead of ignoring technology they could reduce the overload or minimise the effect of technology².

Perhaps the confrontation of lawyers versus computers is based on conservative legal education. Sandgren (2000) argues: «lawyers' one sided training in the legal method makes them poorly equipped to

¹ «Doing more with less» as stated by *Puron-Cid* (2013).

² Listen, for example, podcast of The Kennedy-Mighell Report at Legal Talk Network, available at: <http://legaltalknetwork.com/podcasts/kennedy-mighell-report/2016/08/taking-break-technology/>.

use other methods and creates a mental block against the use of such methods». His multi-disciplinary approach can be explained with empiricism. The above-mentioned author sees empirical approach through social roles, such as an employer, an employee, a consumer, a woman and a refugee. In the context of digitalisation and vast development of technology, we should not only include different disciplines (IT engineeris and architects, policy makers, consumers, lawyers), but also the stakeholders directly related to application of digital technologies. To go even further, empiricism can also be a useful factor, as seen from the «dark side» – therefore, in order to strengthen the efficiency of legal regulation, we should consider the experience of hackers, criminals and terrorists. As Sandgren suggests: «Empirical material can also contribute to “a shift of perspective within the law”» (Sandgren 2000).

The computerisation or automatisa-tion of any system of symbols is directed at finding better solutions and most adequate means of processing certain data. The discussions set forth above indicate that artificial intelligence is rather focused on the question «what», not so much on the question «how» (Poole 2010). Let us presume that consciousness is one precondition for legal justification – does it not then mean reductionism that could work with well-computerized brained *homo faber*? In an ideal world, law should be a living instrument; however, often it is treated as an hermeneutic system, that legalises something only through licensed professionals, who take certain measures or apply specific norms on in their seemingly tautological world. Therefore, the

method of restriction holds its place if the aim is to avoid the exterior, «outer space» influence of this tautological world, such as reinterpretation of politicians, entrepreneurs and any other stakeholders. To paraphrase Cohen (1932), if we cannot be like Gods, knowing the absolute difference between the good and the bad, then at least we can have knowledge of our limits.

Morality and ethics of the computers

Knepper (2007), who leans on Nobel Prize winner Becker (1976), brings us the dilemma that would be presumably inherent only to human nature: «people decide whether or not to engage in criminal activity by comparing the benefits and costs of criminal and legitimate activities». This paradigm fits into the attempt of Karl. R. Popper's situational analysis (Knepper 2007). In legal science, it is called utilitarianism or rational behaviour. According to Popper «elements of a social situation define the appropriate line of action, and given the rationality postulate, the theory predicts that actors will adopt this line of action» (Hedström et al. 1998)

Although Popper relies on rationality or the rational choice, the question remains – can it be determined or foreseen? Can ethical standards be a tool for the rational social model that Popper was looking for? As «justice» and «law» are not synonyms, positive norm as such is not ethical *per se*. For example, Tamanaha (2007), using social approach claims that «law is a form of concentrated social power that claims to be moral». Mikhail (2008) provokes further: «could a computer be programmed to make moral judgments about the cases of intentional

harm and unreasonable risk that match those judgments people already make intuitively». Beside of his versions of the periodic tables of moral elements, Mikhail asserts that there are several variables such as neurological activity, reaction time etc. (elements that a computer don't possess and cannot forecast). The question of collision and confrontation of ethical behaviour *vs.* intuitivity is, of course a separate research area. We may still assume that at least certain part of the behaviour is dependent on ethical standards due to the cognitive prototype model, explained by Larson (2017). Lets take the most dramatic decision making area – international humanitarian law: researchers have analysed application of the independently operating technologies during an armed conflict but the questions remain. The main dilemma is «drafting a law through morality versus having morals interpret law?» or even «technology rules law versus law rules technology?» (Burgess 2015).

There have been attempts to examine ethics and moral reasoning of the robots (Rennselaer AI and Reasoning Lab), but the complex study is far from being able to give clear conclusions. One could recall a Somerset Maugham prize awarded Michael Frayn's (1965) novel «The Tin Men» written more than 40 years ago, where the expert of the Royal Institute of the Automation Research, Mr. Macintosh attempts to construct moral robots by positioning them on a drowning raft and confronting with a dilemma of rescuing *vs.* sacrificing. The experiments clearly failed: the robot sacrificed itself, regardless of what was on the raft. Samaritan II was programmed to sacrifice itself

only when the other creature was on the same level of intelligence - in that case both equally intelligent robots drowned. A new version, Samaritan III sacrificed itself only for a more intelligent individual (based on the brain's size i.e. the size of one's head). But the human expert directly made the commands and the programming of the robots and the size of someone's head cannot most likely be seen as adequate criteria to be more valuable for society.

Ideally lawyers and interest groups would agree to cooperate only with hypothetical Samaritan IV as independent agent who would agree with the interpretation methods of humans on «rule-based or precedent-based reasoning» (Branting 2000), and would leave the values, morality and ethics (explanations and argumentation) to flesh and blood commanders.

The symbolic battle of morality and ethics in the scientific technology is represented also in Mary Shelly's work «Frankenstein» that questions the «intellectual property rights of the God» – are we entitled to create something that has its own moral code?

Those who use artificial intelligence software as digital assistants (Apple's Siri, Microsoft's Cortana) or legal unstructured data analysis (ROSS) or something similar, don't usually imagine themselves living now in an Orwell-type dystopian society but rather being the masters of the army of high-tech agents (without always comprehending the big dilemma on ethics when provided with the information that leads them to decision-making). When you ask your iPhone Siri today, «is it ethical to kill people», you will not get straight answers but be rather be led to the

websites that are discuss the matter of the issue philosophically. This is far from virtual couching but can we still be sure of the virtual ethics code of the core information (based supposedly on Big Data or commerce), deriving from the vulnerable free internet as Siri's source of the delivered data.

Several authors are seeing the process of creating artificial intelligence as «our biggest existential threat» (Musk 2014). Noted experts of artificial intelligence and technology signed a public letter in 2015, at an international artificial intelligence themed conference in Buenos Aires, Argentina (Future of Life Institute). Though it was concentrated on the morality of using artificial intelligence in military weapons that could pose a fatal outcome to the whole society, three signatories proposed to carry out extensive research regarding the effect of AI to the society, including the field of law and ethics, reflecting that «the development of systems that embody significant amounts of intelligence and autonomy leads to important legal and ethical questions whose answers affect both producers and consumers of AI technology. These questions span law, public policy, professional ethics and philosophical ethics, and will require expertise from computer scientists, legal experts, political scientists and ethicists» (Russell et al 2015). There are still many questions to be analysed. As Finlay (2015) emphasises the main questions: who is the beneficiary of the automatized decision-making and how immutable is the data in use. In legal decision-making, the primary beneficiary should be the individual, seeking for a justice. The current software solutions are rather tools

for lawyers assisting them to systematise data and compose better arguments. But they are not benefitted to become more ethical when advised by robots. By Wihlborg et al (2016): «Professionals can either make an alliance with the automated system or the client. This choice of strategy is related to the issues of legitimacy and professional competences». Taking account the previous discussion, the lawyers and other decision-makers whose judgments may affect individual's rights, should know their limits when applying the software and follow their professional ethics code that could be complemented with the articles related to AI based operational systems.

Summary

The on-going discussion about the use of automated logic in law and practice and its practical opportunities, limitations and threats have by today started to drift away from the science fiction and are set into the frames of science. The current debate is based on strong philosophical studies of Leibniz (*calculus ratiocinator*), Turing machine, metaphors of Wittgenstein, and Searle's assumption that computers lack intentionality that would lead to what Reidenberg calls «technologically-created ambiguity». There is a remarkable tension between conservative legal society and liberal digital space. Although, lawyers are officially welcoming the developments in scientific studies, it has become quite visible that they are also concerned about whether computers could, at least partially, replace them. AI seems to be a good tool to draw conclusions on concrete facts or just become a storehouse for legally relevant data. However, law-

yers are carefully accusing the engineers in their attempts to create positivist doers who are incapable of judging human values, ethics and take into account law's «living» nature. It is still not expected that the computers will displace lawyers in the reasoning process because of «rigidity-argument» and the fact that the AI is focused rather on «what to do» instead on «how to proceed».

Thus, one could claim that lawyers prefer to see themselves as the ultimate commanders of systems like IBM Watson and avoid giving too much autonomy to the automated systems.

Unsolved dilemmas on the different characters of human nature and artificial intelligence, proposed theories and discussions over the emerging software take us to a conclusion that although the computers are being ridiculed for as if they would be able to replace the juridical decision makers, they could still structure legal knowledge and regulate technologies themselves. They would specify the sources of legal norms and their hierarchic order, analyse lawyers' arguments from the standpoint of inserted values and principles and use the big data method in categorising the textual part of law, interpretation methods and applicability in practice. They may categorize cases by picking out the external elements that influence justice making, being the «identifiers of facts» when processing digital legal documents.

John Searle's experiment that is called the Chinese room is about a closed room where a person who, according to a corresponding manual, finds the right answers to the questions presented in Chinese. Until a mistake is made, everything is

fine with the assumption that the manual is flawless. There needs to be just a single human mistake by the person taking part of the experiment and it becomes clear that he/she is someone who actually doesn't understand Chinese. The situa-

tion will probably change if in the future a computer is built to be able to exclude the Chinese room argument that means a machine that really understands life circumstances, «a Chinese language», and is not just an augment of an intellect.

References

Academic sources:

1. Ayer A. J. (2001) *Language, Truth and Logic*. Penguin.
2. Becker, G. S. (1978) *The economic approach to human behaviour*. Chicago: University of Chicago Press.
3. Branting, L. (2000) *Reasoning with Rules and Precedents: A Computational Model of Legal Analysis*. Springer.
4. Buchanan, B. G., Headrick, T. E. (1970) Some Speculation about Artificial Intelligence and Legal Reasoning. In *Stanford Law Review*. Vol.23, No 1, pp 40–62.
5. Cohen, N. (1932) *Philosophy and Legal Science*. In *Columbian Law Review*, No. 7, Vol. XXXII, pp 1103–1127.
6. Dewitz, S. (1995) *Using Information Technology as a derterminer of Legal Acts «in Informatics and the Foundations of Legal Reasoning»*, ed by ZenonBukowski, Ian White and Ulrika Hahn. Springer.
7. Engelbart, D. C. (1962) *Augmenting Human Intellect: A Conceptual Framework*. Stanford Research Institute.
8. Frayn, M. (1965) *The Tin Man*, William Collins Sons and Co.
9. Hedström, P., Swedberg, R., Udéhn, L. (1998) Popper's Situational Analysis and Contemporary Sociology. In *Philosophy of the Social Science*. Sage Publications, Vol. 28, No 3, pp 339–364.
10. Huhn, W. (2002) The Use and Limits of Deductive Logic in Legal Reasoning. In *Santa Clara Law Review*, Vol. 42, pp 813–862.
11. Keat, R. (1971) Positivism, Naturalism, and Anti-Naturalism in the Social Sciences. *Journal for the Theory of Social behaviour*, Vol 1 (1), pp 3–17.
12. King, K. F. (2011) Personal Jurisdiction, Internet Commerce, and Privacy: The Pervasive Legal Consequences of Modern Geolocation Technologies. In *Albany Law Journal of Science and Technology*, Vol. 21, pp 61–124.
13. Knepper, P. (2007) Situational logic in social science inquiry: From economics to criminology. *Rev Austrian Econ*. Springer Science.
14. Larson, C. A. (2017) A Cognitive Prototype Model of Moral Judgment and Disagreement. In *Ethics & Behavior* Vol. 27, Issue 1. <http://www.tandfonline.com/doi/abs/10.1080/10508422.2015.1116076> Accessed January 11 2017.
15. McCarty, T. L. (1977) Reflections on «Taxman»: An Experiment in Artificial Intelligence and Legal Reasoning. In *Harvard Law Review* Vol. 90, No. 5, pp 837–893.
16. Menne, A. (1964) Possibilities for the Application of Logic in Legal Science. *MULL: Modern Uses of Logic in Law*, pp 135–138.
17. Mikhail, J. (2008) *Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge*. In *The Psychology Of Learning And Motivation: Moral Cognition and Decision Making*, D. Medin, L. Skitka, C. W. Bauman, D. Bartels, eds., Vol. 50, Academic Press; Georgetown Public Law Research Paper No. 1163422.
18. Poole, D., Mackworth, A. (2010) *Artificial Intelligence: Foundations of Computational Agents*, Cambridge University Press.
19. Poscher, R. (2011) Ambiguity and Vagueness in Legal Interpretation. In *Oxford Handbook on Language and Law*, Lawrence Solan & Peter Tiersma, eds., Oxford University Press.

20. Prakken, H., Sartor, G. (2015) Law and logic: a review from an argumentation perspective. In *Artificial intelligence*, Vol. 227, pp 214–245.
21. Puron-Cid, G. (2013) Interdisciplinary Application of Structuration Theory for E-Government: A Case Study of an IT-Enabled Budget Reform. In *Government Information Quarterly* 30, pp S46–S58.
22. Reidenberg, J. R. (2015) Technology and Internet Jurisdiction. In *University of Pennsylvania Law Review*, Vol 153, pp 1951–1974.
23. Russell, S., Dewey, D., Tegmark, M. (2015) Research Priorities for Robust and Beneficial Artificial Intelligence. In *AI Magazine of Association for the Advancement of Artificial Intelligence*, pp 105–114.
24. Sandgren, C. (2000) On Empirical Legal Science. In *Scandinavian studies in law*, No. 40, pp 445–482.
25. Searle, J. R. (1980) Minds, Brains, and Programmes. In *The Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417–424.
26. Sunstein, Cass. R. (2001) Of Artificial Intelligence and Legal Reasoning. In *University of Chicago Public Law & Legal Theory Working Paper No 18*, pp 1–10.
27. Tamanaha. B. (2007) The Contemporary Relevance of Legal Positivism. In *Legal Studies Research Paper Series. Paper #07-0065*.
28. Tamme, T., Tammet, T. Prank, R. (1997) *Loogika: mõtlemisest tõestamiseni*. Tartu: Tartu Ülikoolikirjastus.
29. Toulmin, S. (1969) From logical analysis to conceptual history. In *The Legacy of Logical Positivism, Studies in the Philosophy of Science* (Peter Achinstein & Stephen F. Barker, eds.). Baltimore: The Johns Hopkins University Press, pp. 25–52.
30. Turing, A. (1950) Computing Machinery and Intelligence. In *Mind*, vol. 59, no. 236, pp 433–460.
31. Warvick, K., Shah, H. (2016) Can machines think? A report on Turing test experiments at the Royal Society. In *Journal of Experimental & Theoretical Artificial Intelligence*. Vol. 28 Issue 6, pp 989–1007, p. 19.
32. Wihlborg, W. E., Larsson, H., Hedström, K. (2016) The Computer Says No! – A Case Study on Automated Decision-Making in Public Authorities. In *49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, pp. 2903-2912 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=7427547&isnumber=7427173> Accessed December 28, 2017.
33. Wittgenstein's Lecture on Ethics (1965) *Philosophical Review*, LXXIV, No. 1, pp. 3–16.
34. Xia, C., Maes, P. (2013) The Design of Artifacts for Augmenting Intellect. In *Proceedings of the 4th Augmented Human International Conference (AH'13)*, Association for Computer Machinery, pp 154–161.

Other:

1. Burgess, L. (2015) Autonomous Legal Reasoning? Legal and Ethical Issues in the Technologies of Conflict. *Intercross Blog*. <http://intercrossblog.icrc.org/blog/048x5za4aqeztdiu3r8f96s8m7lzom/> Accessed December 20, 2016.
2. Finlay, S. (2015) Ethical Risk Assessment of Automated Decision Making Systems. <http://www.odbms.org/2015/02/ethical-risk-assessment-automated-decision-making-systems/> Accessed January 16, 2017.
3. Future of Life Institute. An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. <http://futureoflife.org/ai-open-letter/> Accessed January 10, 2017.
4. Lippe, P., Katz, D. M. (2014) 10 predictions about how IBM's Watson will impact the legal profession. http://www.abajournal.com/legalrebels/article/10_predictions_about_how_ibms_watson_will_impact/ Accessed January 14, 2017.
5. Musk, E. (2014) Interview for MIT students at the AeroAstro Centennial Symposium. <http://aeroastro.mit.edu/aeroastro100/centennial-symposium> Accessed January 12, 2017.
6. Rensselaer AI and Reasoning Lab within a project «Moral Reasoning & Decision-Making: ONR: MURI/Moral Dilemmas». <http://rair.cogsci.rpi.edu/> Accessed January 8, 2017.
7. The Baltic Course (2016) Former Estonian Chancellor of Justice established Law Firm Teder. In *The Baltic Course* http://www.baltic-course.com/eng/markets_and_companies/?doc=115102 Accessed January 18, 2017.

**Етичний юрист, або Моральний комп'ютер:
історичний та сучасний дискурс про недовіру між людиною та машиною**

Сучасне суспільство багато в чому перебуває під впливом цифрової революції. Сучасні комп'ютерні технології впливають на наше повсякденне існування, але в той же час задають напрям нашого життя. Проникнення цифрових рішень виходить далеко за рамки простого надання допомоги; децю несподівано, що це посягання (вплив) досягло навіть найгуманніших сфер, таких як справедливість. Чи можна, наприклад, автоматизувати всю процедуру прийняття законних рішень? Кілька десятиліть тому це питання було б цілком доречним. Проте сьогодні ми визнаємо, що ситуація кардинально змінилася. Беручи до уваги всі аспекти, в даній статті простежується, як людство досягло цієї точки оцифровки, в той же час намагаючись визначити потенційні загрози, пов'язані з автоматизацією правосуддя (прийняття законних рішень). Однак стаття не стосується років механізму Антікітри, оскільки мало відомо про його конструкції (будови); тому вона буде обмежена подіями останніх століть, аж до теперішнього часу.

Ключові слова: штучний інтелект, збільшення інтелекту, автоматизація права, механічне мислення.