TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance
Department of Law

Valtteri Kauppi

# THE USE OF "UPLOADED MIND" PERSONALITY CONSTRUCTS FOR DECLARATIONS OF INTENT UNDER FINNISH LAW

Master's thesis
Programme HAJM, specialisation Law and Technology

Supervisor: Thomas Hoffmann

Tallinn 2021

I hereby declare that I have compiled the thesis independently
and all works, important standpoints and data by other authors
have been properly referenced and the same paper
has not been previously presented for grading.
The document length is 18502……. words from the introduction to the end of conclusion.

Valtteri Kauppi ……

(signature, date)

Student code: 195277HAJM
Student e-mail address: vakaup@ttu.ee / valtteri.kauppi@hotmail.fi

Supervisor: Thomas Hoffmann:
The paper conforms to requirements in force

……
(signature, date)

Chairman of the Defence Committee:
Permitted to the defence
……
(name, signature, date)

# TABLE OF CONTENTS

ABSTRACT ..... 4
INTRODUCTION ..... 5
1. DIFFERENCE BETWEEN DECLARATIONS OF INTENT MADE BY AN UPLOADED MIND IN CONTRAST TO TRADITIONAL DECLARATIONS OF INTENT ..... 9
2. FUNCTIONING OF ARTIFICIAL NEURAL NETWORKS ..... 13
2.1. The unpredictability factors ..... 13
2.2. Nature of artificial neural network computing ..... 14
2.3. Types of artificial neural networks classed by function ..... 15
2.4. Mind uploading ..... 19
3. PROBLEMS IN THE LEGALITY OF THE SYSTEM ..... 23
3.1. Human rights problem ..... 23
3.2. Medium specific protection of rights ..... 30
3.3. Problems with proportionality and non-foreseeability of possible legal relations ..... 31
4. LEGAL ANALYSIS OF THE RESEARCH QUESTION ..... 33
4.1 Comparing the construct to conventional instruments ..... 39
4.1.1. Power of attorney ..... 39
4.1.2. Will of a late person ..... 40
4.1.3. Living will document ..... 40
4.1.4. Prenuptial contract ..... 41
4.1.5. Transcription of gifts and private commerce ..... 42
4.2. Answering the research question ..... 44
4.3. Propositions for regulating the status and use of electronic personality constructs ..... 44
CONCLUSION ..... 47
LIST OF REFERENCES ..... 48
APPENDICES ..... 52
Appendix 4. Non-exclusive licence ..... 52

# ABSTRACT

With artificial neural networks human cognitive processes can be depicted and in theory calculated. If human cognitive processes are entirely physical and subject to the laws of physics and chemistry, thought processes are calculable. This means that with high enough scanning resolution and processing power a human mind can be scanned and electronically emulated with a computer, producing a human psyche in electronic form. This at-the-time-theoretical concept is called mind uploading. As soon as it is possible the large technology companies will most likely pursue it. This gives rise to legal questions which should be answered as soon as possible.

This research will be analysing a mind uploaded personality construct for self-representation purposes, and whether such a system could be used to declare one's intent in legal matters in the context of Finnish jurisdiction. The research question of this thesis is "*whether a person can, under Finnish law, manifest their intent in juristic actions with the assistance of a computer system that simulates or emulates the psyche of the person based on data the person has provided earlier, without the person themselves having to take action in the moment of the juristic action*". First there will be deconstruction of the theoretical system and the science supporting it proving that theory of mind uploading is a sound theory that requires attention, followed by a comparative analysis considering the similarities and difference the system holds in regard to contemporary automated contracting tools and legal instruments used to declare one's intent. From these standpoints an answer will be synthesized to the research question. Lastly, there will be considerations how the legal framework should be adapted to the advent of such technologies.

The findings of this research were that at the time such a system could be used within Finnish jurisdiction to declare one's intent and to enter into those kinds of legal relationships that do not require a physical document, physical signature, or witnesses. It was concluded that the artificial intelligence component does not affect the system's legal treatment in any way compared to a system without artificial cognitive processing capabilities.



Keywords: Artificial neural networks, Mind uploading, Declaration of intent, Finnish law

# INTRODUCTION

Due to the exceeding developments in artificial intelligence in the areas of semantic understanding and text generation, new legal problems are arising. The fact that development pace of Moore's law is no longer kept up with and therefore new architectures are being developed that no longer rely on physical scaling[1] and the developments in brain-computer-interfaces suggest that the concept of uploading a digital copy of a human psyche is coming closer. This means that the legal viewpoints on higher technologies and their applications should be researched and discussed rather before mind uploading is possible than after its actualization. The concept itself is purely theoretical now, but the existing technologies are scalable to the point in which mind uploading will become possible if the three presumptions of this thesis are true.

Mind uploading and therefore the capability of creating emulated and simulated personalities is at the current point in time a theoretical concept. The reason why this topic is relevant is that its actualization is very likely, and the fact that the nature of such technological innovations so vastly differs from legal instruments meant to declare a person's intent while they are absent. Therefore, digital personality constructs should be examined as their own separate entity also in the legal sphere, if from no other viewpoint than regulating their use and providing services related to them. To give an example of the differing nature, a will and power of attorney are fixed in content, whereas a digital copy of a person's mental state can give varying outputs[2], making it more resembling a transcript of a person rather than an inanimate instrument. As can a real person, the digital personality construct can give different outputs from the same input as well. This is depending on the outcomes of the previously processed choices in case of a limited memory artificial intelligence or higher, but also on unexpected relations between data points in the processing levels of the artificial intelligence in cases of fully accurate psyche copies. In addition to the unique character as a tool, a digital personality construct is also an instrument unlike any

[1] Thompson, Scott E., and Srivatsan Parthasarathy. "Moore's law: the future of Si microelectronics." *Materials today* 9.6 (2006): 20-25.
[2] Xin, Ran, Soummya Kar, and Usman A. Khan. "Decentralized stochastic optimization and machine learning: A unified variance-reduction framework for robust performance and fast convergence." *IEEE Signal Processing Magazine* 37.3 (2020): 102-113.

other existing currently, so comparing it to anything is rather dangerous due to possibility of a „straw man argument ". This means that in order to evaluate how uploaded personalities should be treated legally they have to be analysed both from a technical point of view to understand what they are and what they are not, and from a legal point of view to see how this technical information translates to legal context. This includes legal analysis about the nature of the artificial intelligence construct in question to determine what it is legally and what kind of legal framework should be created around it. It will be analysed how Finnish law regulates such an instrument, or rather omits its regulation. Shortly put, such a construct is only directly regulated by the General Data Protection Regulation 2016/679 from the viewpoints of automated data processing and protection of personal data. From the understanding of this lack of status and comparing it to the technical analysis and value protection that is otherwise present in the legal norms it is then possible to analyse how they should be treated, or at the very least how they should not be treated in the legal sense. Therefore, this thesis will first thoroughly analyse uploaded digital personality constructs as technical tools, and after that compare them to legal instruments previously used to achieve similar aims, that being declaration of intent. The legal instruments to be analysed will be a testament and a power of attorney among few others. After the distinction and comparison between the digital personality constructs and these traditional legal instruments has been established and drawn viewpoints from, there will be proposals on how to treat these new kinds of tools in the legal context Finnish law.

This research aims to answer the question "*whether a person can, under Finnish law, manifest their intent in juristic actions with the assistance of a computer system that simulates or emulates the psyche of the person based on data the person has provided earlier, without the person themselves having to take action in the moment of the juristic action* ". To examine this question, the tool to be used in such an at-the-time passive declaration of intent must be defined. This sub-question in its essence relates to the simulated or emulated personality and its legal status. When the status is established, the capacity to take actions without active input on a matter can be analysed. In other words, to understand whether a digital personality construct can be used to represent a person when they are unavailable or unable to do so themselves, one must understand what a digital personality construct actually is, and what capacities can such a construct hold.

To answer the research question, the system will be analysed as a technical tool. Following that it will be determined how it should be treated by comparing it to the classic instruments resolving the same problem, and then proceed to draw conclusions from these answers. First this will be

done on technical level, then on the legal aspect and finally analyse the system's potential as a legal instrument. Following this there will be examination on how such a system would fit to the already existing legislation in the sense of using it for passive legal actions. There will be proposals whether the legislation should be amended to cover such tools of higher technology better pre-emptively. The regulation of such technology has been unnecessary so far for they have been so far away even in the theoretical examination that the omittance of such legislation has not negatively impacted the legal context of the matter. The thesis will be closed off with conclusions.

The system analysed in this work is an artificial neural network that is a copy of the user's cognitive decision-making system, from the point of view of its cognitive architecture. A system like this could be, and is used in the example of this work, to represent the copied person in regard to juristic actions automatically in their absence. The system would scan the user and create an electronic personality construct emulating or simulating their mind to provide access to the user's reasoning at times when the living person, the user, is unavailable.

If this is not possible or lesser resemblance gives relatively higher return of interest (in this instance meaning amount of work and data needed in relation to the precision of the output) the system could very well be just a close resemblance instead of a perfect copy. The question whether the personality construct is simulated or emulated will be discussed in detail further on. When the system is created the person who will become the user of the system, or the substituted person (herein on referred also to as the "user") will have to undergo a scan in order to map the cognitive architecture of the subject. This would most likely happen by directly scanning the brain, since no other ways of measuring the brain are not known currently and it is preferable that the user does not expire in the copying process. Recent developments in Multi Echo fMRI scans have opened up new possibilities for non-invasive mapping of functional relationships between brain areas over a time longer than a few minutes[3]. As presented later in the part regarding the three presumptions, this research presumes that the scanning technology exists, and it does not require the expiration of the scanned subject. After the data of the neural, and therefore cognitive, architecture has been gathered and the artificial neural network has been built, it must be tested. If the architecture is an exact copy of the architecture of the subject, the system should work exactly as the person's psyche would in the organic system, meaning their cerebral cortex. This is true if the reasoning of U. T.

[3] Lynch, Charles J., et al. "Rapid Precision Functional Mapping of Individuals using Multi-Echo fMRI." *Cell reports* 33.12 (2020): 108540.

Place[4] and John Smart[5] about type physicality of thoughts are true. For the user to be able to fully trust the construct, some testing must be conducted to verify that the construct chooses identically or at least according to the same rules as the original person.

[4] Place, Ullin T. "Is consciousness a brain process?." *British journal of psychology* 47.1 (1956): 44-50.
[5] Smart, John JC. "Sensations and brain processes." *The mind-brain identity theory*. Palgrave, London, 1970. 52-66.

# 1. DIFFERENCE BETWEEN DECLARATIONS OF INTENT MADE BY AN UPLOADED MIND IN CONTRAST TO TRADITIONAL DECLARATIONS OF INTENT

Mind uploading is the hypothetical concept of creating and uploading a scan of human cognitive processes and the contents of a consciousness into a digital environment and emulating or simulating human though processes. Since mind uploading is a hypothetical process, examining it requires three presumptions.

The first is that the human mental state and its processes physically take place in the human body. This is called the assumption of physicality[6], and it means that it is supposed that no part of the thought processes take place on an unobservable level, but rather everything happens physically in the human body. This happens, as far as is known, in the brain and partly in the spinal cord[7].

The second presumption is that the thought processes are scannable and computable. This means that the processes which form the simulation we observe as reality and as the present have to be measurable. Measurability does not have to mean fixed patterns or repeated fashion in the processes, but rather that every process can be divided into units and the processes are observable as the causal interactions of those units. The presumption of physicality is a prerequisite for the presumption of computability.

The third presumption is that at some point humans will be able to create such a computer and scanning technology that can execute these concepts.

David Chalmers proposes in his 2011 article that with sufficient computational models a conscious mind can be created[8] also in an artificial medium. Chalmers' thesis relies on the same presumption

[6] Klinge, Fabian. "Physicalism and Fundamentalism." *Panpsychism and the Emergence of Consciousness*. JB Metzler, Berlin, Heidelberg, 2020. 5-21.
[7] Brooks, Eugene M. "The physicality of qualities and consciousness: Concepts within the theory of consciousness." *Imagination, Cognition and Personality* 34.1 (2014): 57-72.
[8] Chalmers, D. J. (2011). A computational foundation for the study of cognition. Journal of Cognitive Science, 12(4), 325-359.

of computability as this thesis. On these grounds, Chalmers proposes that since if complex enough mathematical models are deployed, they can simulate or emulate the organization and processes of any other system. Chalmers has been criticized for the logical fallacy of "begging the question" regarding the formulation of his theory. First Chalmers supposes that all things can be computed, and then he asserts that since all things can be computed the cognitive processes can be computed as well. This, as it is represented, in fact contains the fallacy of begging the question. However, claiming that Chalmers is wrong in the understanding of the subject matter demonstrates clear misunderstanding of the subject at hand, mathematics that is. Mathematics is the science of quantity, structure, space, and change. Mathematics relies on units, which means that the subject of a mathematical calculation must measurable. Another thing that mathematics requires is that the object of calculation is subject to rules, rules of physics and chemistry for example. Everything that can be measured and is subject to some set of rules, can therefore be computed. It can be argued that Chalmers should have presented his hypothesis in the form of "(i)f cognitive processes are subject only to measurable rules and can be measured themselves, they are computable.".

Cognitive processes have so far not shown to include a metaphysical element, and there are a lot of evidence supporting the contrary brought forth for example studies in oscillatory brain activity[9]. This allows the presumption that such a metaphysical element is not present in the primate mammalian cognitive process formation and execution. If such an element is not present, everything that happens in those processes have to be subject to the rules of the surrounding environment. So far it has been proven that the thinking processes include electronic signals and chemical reactions in the brain. No other hardware rule sets have been perceived up to date, which supports the presumption that the cognitive processes are subject only to the laws of physics and chemistry. Since physics and chemistry are measurable and therefore computable, cognitive processes should be as well. Therefore, this research supposes that cognitive processes of humans can be measured and therefore computed.

This thesis research is supported by Pentti Haikonen's idea of cognitive architecture[10] modelling. According to Haikonen's research, the classical rule-based computing will not lead to

[9] Sauseng, Paul, and Wolfgang Klimesch. "What does phase information of oscillatory brain activity tell us about cognitive processes?." *Neuroscience & Biobehavioral Reviews* 32.5 (2008): 1001-1013.
[10] POSTOLEA, Dan, and Adrian Ion URICHIANU. "THE HAIKONEN COGNITIVE ARCHITECTURE FOR XCR-1 ROBOT BRAIN." *Descrierea CIP/Description of CIP–Biblioteca Naţională a României Conferinţa Internaţională Educaţie şi Creativitate pentru o*: 91.

consciousness[11] despite it being able to execute some levels of imitated cognitive processes such as learning[12].

The system in question is supposed to have intuition, and here comes in Haikonen's idea of cognitive architecture. Cognitive architecture is a representation of the arrangement of cognitive processes and the mathematical functions used to emulate it[13]. Haikonen states that the mind can only be emulated once the several sub-systems of the mind are left to interact with each other, and this gives rise to consciousness-based operation model. Haikonen theorizes that consciousness cannot emerge in simple execution algorithms, but rather requires more complex processing models. In the starting point the system is only a very functional sum of its parts before it becomes conscious, and inside a highly functioning process a consciousness emerges. The emergence of consciousness is only a change in the flow of information, Haikonen states. The brain holds several kinds of synapses[14]. They form, activate, and expire on different paces to one another. In addition to this there are brain areas that send impulses to other parts of the brain that are not transmitted through neurons, which adds another type of signals to be measured. The problem with these different kinds of synapses and non-neural impulses is that they form non-linear and multi-layered processes. This is not a problem from the point of view of plausibility[15], but rather from the point of view of computing power and algorithm creation. Haikonen proposes that because consciousness is a phenomenon that emerges from the combined workings of senses, inner speech and the derivates of the sensations arising from the combination of inner speech and senses, a classical rule-based computing should not be able to contain a conscious process. The nature of the process in itself does not make this impossible, rather it is the computing load that limits the applicability of the serial computing models.

The mammalian cerebral cortex perceivably fully complies with the mathematical laws of chemistry and physics, which are fully computable with classical models. The emergence of consciousness can more easily be achieved with the utilization of quantum computing, for in quantum computing a variable can be 1, 0, both or neither at the same time[16] enabling the use of

[11] Haikonen, P. O. (2003). *The cognitive approach to conscious machines*. Imprint Academic.
[12] Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.
[13] *supra nota* 11
[14] Grant, Seth GN. "Synapse diversity and synaptome architecture in human genetic disorders." *Human molecular genetics* 28.R2 (2019): R219-R225.
[15] Reggia, James A. "The rise of machine consciousness: Studying consciousness with computational models." *Neural Networks* 44 (2013): 112-131.
[16] Smith, Kaitlin N., and Mitchell A. Thornton. "Entangled state preparation for non-binary quantum computing." *2019 IEEE International Conference on Rebooting Computing (ICRC)*. IEEE, 2019.

constantly changing algorithms where the beginning of the algorithm may change as the process approaches its end. This enables several contradicting concepts to be “true” in the mathematical sense at the same time, which enables more flexible definition of concepts. A model in which a statement can be both true and false at the same time can be constructed with serial computing as well but executing long algorithms with such qualities takes so much computing power that they are possible only in theory. Here an assumption has to be made: In order for Haikonen not to contradict himself it has to be assumed that he means that consciousness in not impossible to achieve with classical rule-based systems, but rather creating such a system manually would be impossible due to the time and computing power required. This is evident because Haikonen then later proposes that in order for such a construct to be created, the underlying computational processes have to be identified and copied. However, it is also possible that he refers, by “classical rule-based models”, to serial computing and not to algorithm construction. As stated before, the possibilities brought about by quantum computing include simultaneous “1” and “0” statuses as well as “neither” status. This would reduce the computational load in a way and enable relatively easier construction of a conscious artificial intelligence. Whether Haikonen refers to serial computing or algorithm construction, it is either way safe to say that Haikonen expresses himself poorly, or else he would contradict himself so grossly that he should have abandoned all of his research already at the hypothesis formation phase. If quantum computing is achieved, it is reasonable to conclude that the presumption of technological achievability is true as well.

# 2. FUNCTIONING OF ARTIFICIAL NEURAL NETWORKS

## 2.1. The unpredictability factors

When examining the validity of the presented problem with a statement that artificial intelligences are machines that always in the exactly same state (which changes constantly as machine learning goes on, so technically one state lasts less than a second at a time) react the same way to a presented problem, which is true, it must be examined whether this is actually a problem or not.

This could be problematic in two ways which are that the artificial intelligence would not represent an accurate copy to the human thought processing because of this quality and that because the reaction is the same in the same state to a same stimulus there is no unpredictability within the operation of the system so it can be always fully predicted.

If it was so, this could be seen to invalidate the problematic nature of the artificial intelligence personality emulations by proving that the construct would be in essence only an interactive document library, which in itself is true, but does not invalidate the problem. Considering the problem regarding the consistency of actions, first, humans react to the same stimulus in the exactly the same state in the same way as well. This is easily proven by examining the argumentation regarding the presumption of physicality. If the cognitive architecture of humans is fully within the boundaries of the laws of physics and chemistry and is not affected by anything else, there is no chaos in the process, only causality. So, when something changes, the change must be caused by something. When the reaction to a stimulus is different, it has to be the effect of a changed variable either in the premise (the system, or in this case the person's state) or in the variable inserted factor (the stimulus). This means that if the premise or the variable factor is changed the reaction changes if the change has enough magnitude. This means that if the premise and the variable are same enough, a human will react in the same way as well. Simply put this also proves that if there is no metaphysical element into the decision-making process, and that there is no free will. It has to be noted, however, that a different premise and a different stimulus can still give the same result as another pair, but this is not relevant to the question. The question was that

do people react in the same state in the same way if the stimulus is the same, which is true if the premises of this research are true.

The second part of the argument was that the existence of the possibility to predict the actions of a system in certain circumstances eliminates the unpredictability from the system making it practically a calculator. The time-window in which the premise and the stimulus stay the same are at shortest the time which it takes from electricity to travel the distance of one artificial synapse, which is a very short time. At shortest a change in one synapse may change the outcome of a process of decision making related to a stimulus. Regarding the amount of data points within a human brain and the pace in which this information can be rewritten while thinking is occurring is so fast that it would take a system many times the speed of the one presented here to predict its answers in real time that it would take a lot more development in architecture than what is required to build the system itself that most likely they would not appear at the same time. This means that a system emulating a human psyche would be as unpredictable as the real version of the person. This means that even if in theory the architecture of the construct would be mapped the internal relations within it would be unpredictable due to the absence of vast enough computing power. This furthermore means that the construct may bind the subject, or the principal, to a legal relationship which could not have been predicted, and still be considered functioning normally and as intended. This unpredictability means that the system has to be regarded different from the classical "static" tools used to declare one's intent, for example a transcription of a gift, a will, a prenuptial contract, or a power of attorney.

## 2.2. Nature of artificial neural network computing

A neural network is a computing model in which a single artificial neuron computes one single computational task or stores one unit of information[17]. An over-simplified explanation of this would be to depict how such a system would process the most basic computation, 1+1=2. In this computation there are four different concepts. These are the concept of "1" that is the indicator for a single independent measure of given resource. Then there is "+" which indicates that the indicators on both sides of this indicator are combined and together they form the sum of their parts. Then there is "=" which marks that the two depictions on both sides of this mark indicate the same amount despite possibly being depicted in different ways. This is the output layer.

[17] Takefuji, Yoshiyasu. *Neural network parallel computing*. Vol. 164. Springer Science & Business Media, 2012.

Everything mentioned before the "=" belongs to the processing layers under the output layer. Lastly there is the output itself, the "2" that indicates the sum of dual "1" s. In this kind of a system the most basic elements (which would be in fact still smaller concepts than in this example) would be stored in nodes between the artificial neurons, in this case being "1" and "+" in the bottom most layer. Then the node containing "1" would send an input to the "+", to which it would respond with a question basically asking, "anything else?" to which the "1" would answer by inputting its content again. The "+" would then receive a negative answer to its request then forwarding the process to the output layer which would then give the answer. There are several output layers on top of each other which in turn form the nodes of processing layers to the higher-level computations. Eventually the system can calculate abstract concepts such as self as an agent and the difference of being alive and being powered on. Dr Robert Hecht-Nielsen defined a neural network to be *"...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs."*[18]. An artificial neural network is a simulated or an emulated version of such a biological computing system and it has loose commonalities with mammalian cerebral cortex and its computing functions, but in a form of a digital construct.

## 2.3. Types of artificial neural networks classed by function

There are essentially four levels of artificial neural networks. The word" artificial" is omitted for the sake of pacing, and organic networks are called specifically organic networks when mentioned. Neural network computing is based on nodes and connections between them[19], same way as mammalian cerebral cortex has nuclei, axons, and dendrites[20]. Neural networks are different from the classical serial model of computing in the sense that whereas in serial computing the processor executes commands stored in memory slots and temporarily saves the results of these commands to other memory slots, the neural network does not necessarily produce a result in a sense, it can be the result in itself. A neural network can give answers to questions through an output layer, but its internal functions, the nodes, the connections between them and their intensity is the "end

[18] Caudill, Maureen. "Neural networks primer, part VIII." *AI Expert* 4.8 (1989): 61-67.
[19] Liu, Yong, and Xin Yao. "Evolutionary design of artificial neural networks with different nodes." *Proceedings of IEEE international conference on evolutionary computation*. IEEE, 1996.
[20] Morest, D. Kent. "The growth of dendrites in the mammalian brain." *Zeitschrift für Anatomie und Entwicklungsgeschichte* 128.4 (1969): 290-317.

product" of the development[21], not the analysis the system is capable of producing. Neural networks can be used for ways of finding solutions alternative to conventional computing models. Conventional serial computing relies heavily on linearity, causality, variable independence and assumptions of normality and neural networks can provide alternative solutions to pattern and relation recognition. The strengths of artificial neural networks are in detection of relationships and occurrence commonalities, both positive and negative, and they can highlight relationships without direct causality, direct link, common nominator or conjunctive function, incident, or occurrence[22]. This is something serially working computing models struggle with. However, they are not optimal for answering very specific questions before very extensive training and they are at the current state-of-the-art best utilized for tasks with not-too-strict error margin, such as predicting weather, occurrence of certain phenomena, predicting unobservable common features or causes between variables or predicting future behaviour based on previous data. This means that an artificial intelligence can easily detect what applications a user is using simultaneously and how the behavioural patterns are formed but may not be able to give any solid reasoning why the patterns occur. To make a very generalizing, simplifying and also an inaccurate comparison, a conventional computer is like a person who organizes their thoughts according to what they perceive around them whereas an artificial neural network is more like an intuitive thinker who can come up with theories that connect two seemingly separate phenomena.

The four types of neural networks are reactive neural networks, limited memory neural networks, neural networks with Theory of Mind and self-aware neural networks.

The reactive neural network is as its name suggest, purely reactive and does not have perception of the world as concept or the past. This means that the reactive neural network only perceives the surrounding world as it is in that moment and calculates possible action and then optimizes the results of these calculations after which it chooses the most likely optimal outcome. The reactive neural networks do not save data from past encounters or even their previous actions (except a few examples, but these are so minor that they do not even amount to data collection to better the performance, but rather only simple rules of past conduct may be executed with these rules, such as that the same action cannot be taken twice in a row. An example of this could be a chess playing

[21] Kavzoglu, Taskin. "Determining optimum structure for artificial neural networks." *Proceedings of the 25th Annual Technical Conference and Exhibition of the Remote Sensing Society*. Remote Sensing Society Nottingham, UK Cardiff, UK, 1999.
[22] Pan, Wei, Xiaotong Shen, and Binghui Liu. "Cluster Analysis: Unsupervised Learning via Supervised Learning with a Non-convex Penalty." *Journal of Machine Learning Research* 14.7 (2013).

artificial intelligence with the perception being the pieces and their rules, the possible actions the possible moves to be made by the artificial intelligence as well as the opponent and the data saving exception being that the same position of the pieces cannot manifest more than three times[23]. Reactive neural networks are best utilized in the solving of optimization and function problems. As the name suggests, optimization problems require the computer to calculate an optimal solution for a problem from a set of possible options, which is usually derived from determining the variables in the situation, and then computing the possible futures.

Another type of neural networks are limited memory neural networks, which gather limited and pre-determined pieces of data storing them and observing patterns from them. This is the next step from the reactive artificial intelligences with the advantage that they can learn and assimilate information belonging to pre-determined classes and adapt their conduct accordingly to this information. An example of this are rudimentary self-driving cars which observe pre-determined variables such as distance from the lane sides, the approaching cars, traffic signs and lights. Another one would be a video game computer-controlled entity changing their behaviour based on the player's behaviour. For example, in a shooting game the game records where the player shoots the enemies, and further on the game the enemies' protection will update accordingly. This example can be easily executed without neural network computing; however, it could be greatly enhanced with it *e.g.* predicting the player's next target prisonization after these protection updates take place. This is where the artificial intelligence capacity at the state-of-the-art ends.

The next stage in the functional development of neural networks is an artificial intelligence that possesses the Theory of Mind. System with this level of advancement will be the theoretical minimum for the system represented in this research's theoretical example. Possessing Theory of Mind means that an actor is capable of forming and classing representations about other entities in its sphere of perception,[24] whereas the previous two models have only been capable of classing objects. This does not mean consciousness, thoughts, or feelings. Simply the capability for classification of other agents – both as being capable of enacting causality to their environment and themselves as well as them being capable of agency *i.e.* making decisions – and ability to predict their behaviour exactly because of the understanding of their agency. There is a question

[23] Pore, Ameya, and Gerardo Aragon-Camarasa. "On simple reactive neural networks for behaviour-based reinforcement learning." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

whether chimpanzees have the Theory of Mind[25], for they perceive others as being capable of causality and agency, but do they understand that the capability to assert cause in the environment arises from decision making processes that perceives others as agents as well. *I.e.* if a chimpanzee observes human behaviour, it perceives the causality and agency, but does it understand that the human asserts causality because the human understands themselves to be an agent and perceive others as such as well. In this regard Turing's Test[26] could be, with its many undeniable flaws and loopholes, used to gather insight on the development level of an artificial intelligence[27]. This way it could be assessed whether a neural network is capable of perceiving agents with capability to choose instead of just classifying autonomous objects with causal prowess. In other words, is the system capable of distinguishing capability to make informed decisions from reacting to environment. Theory of Mind does include inner speech, thoughts, feelings, empathy or any such things. Even though it might be tempting to think that once an actor is capable of perceiving another actor as such empathy would emerge too, but this is not the case, for empathy requires some form of self-awareness since it requires the actor to perceive how they feel, make the assumption that another similar actor would most likely feel the same way and only then empathy may emerge. Instead of "Theory of Mind" the name of the theory would better describe its contents were it called "Theory of Agency".

The highest, currently theoretically perceived, functionality of neural networks is self-awareness. Self-awareness means that the agent can form representations of themselves, their state, capabilities, and limitations. A self-aware agent can observe their inner processes, and therefore distinguish between wanting something and knowing they want it, so they are able to perceive the difference between an actively conscious thought and the reactive impulse arising from internal or external stimuli. This furthermore allows the agent to predict the actions of others based on models they themselves experience internally. There is an ongoing discussion whether the consciousness is purely a mathematical phenomenon in a neural network and can manifest equally in an artificial neural network as well as in an organic one, most prevalently being mammalian cerebral cortex, or if it has a metaphysical dimension to it.

[25] Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526.
[26] In the test first introduced by Alan Turing in 1950 an artificial intelligence tries to convince a tester that it is human by having a polite conversation with the tester, usually in written form.
[27] Saygin, Ayse Pinar, Ilyas Cicekli, and Varol Akman. "Turing test: 50 years later." *Minds and machines* 10.4 (2000): 463-518.

## 2.4. Mind uploading

Mind uploading can be divided into two categories by the precision of resolution or detail of capturing the essence of the original. These are emulation and simulation[28]. This means how well and how accurately the uploaded copy resembles the original mind from which the copy was made of and whether it is predictable in the same manner. The first category, the emulated ones, are virtual mind copies with full accuracy. This means that the mind being emulated in the computer system is scanned completely. They are manually constructed[29] artificial intelligences built to resemble the original completely. To do so, every detail has to be copied from the scan. These systems are full personality copies. The second category is formed by a group of artificial intelligences that resemble the original mind without being copies of it. These systems are simulating the personality, instead of emulating it. These systems take in massive amounts of data, but relatively little compared to the systems that emulate a full copy of the psyche, and then they replicate the mental state by executing free or guided machine learning considering the data. This would be, in its simplest form carried out by mapping thought processes of the simulated individual by surveys and observation. From this data a model of predicted behaviour would be constructed, and this model would serve as the basis for the digital version of the simulated mind, while not being completely accurate. After this the system starts learning and is evaluated and given feedback to, so that it gets closer and closer to simulating the mental state of the person. However, if the survey and observation data would be comprehensive enough, the parts of the mind impossible to simulate with this model would be, by some parts, the unconscious processes and the impulsive elements based on external stimuli that depend on the homeostasis for example. This means that from the point of view of such a system as described here there is not a great difference between the two models as long as the system is not excepted to be creative and is used for example to either accept or to reject contract terms proposed to it, and it is not asked questions that require intuition. The creativeness in the case of a digital personality construct would require the system also to have the elements which the user has on their subconscious parts of their mind. These elements would most likely not transfer easily to a digital version through a questionnaire for example, but they could be perceived and recorded in a brain scan. Therefore, the complete brain scan model has more precision of resolution and therefore is more advanced and offers more possibilities. The complexity level of the expected answers that the system should give as output

[28] McGregor, Ian. "The relationship between simulation and emulation." *Proceedings of the Winter Simulation Conference*. Vol. 2. IEEE, 2002.

[29] Here by ”manual” it is not meant that someone has to insert the data themselves, but rather that the system is not allowed to build itself as artificial intelligences mostly do by machine learning.

determines how precisely the mental state of the user must be transcribed. The intended purpose and the need answered must determine the depth of details, even the unobservable ones. This is because if the system is intended to give answers of or relating to the user's unconscious processed the scan and other forms of used data retrieval must be considered. This means that in the highest level the functions of individual brain synapses must be measured. The technology to measure and map brains in this level is possible already, but the amount of time and resources it would take to fully map the brain is enormous.

The difference between a simulation and an emulation is that in simulation something is mimicked and in emulation something is built inside a system and then executed within it. Regarding mind uploading, an emulated electronic version of a person's psyche requires a fully exact copy of the person's cognitive architecture. This means that every synapse must be scanned, mapped, and then created in electronic form. This requires a lot of computing power, precision and takes a lot of time. The computing power required to emulate every single synapse is enormous, for every synapse holds inner processes as well. In addition to delivering information, synapses also shape it. This means that each synapse is a process, and requires individual computing power, possibly even its sub-processes must be computed. Summing up, emulating a human psyche requires a huge amount of scanning, mapping, and processing power. This kind of emulation also takes a lot of time, presuming that possibility of quantum computing does not accelerate computing speeds in an unexpected rate. Simulating a human consciousness is relatively far simpler, since larger clusters of the cognitive processes can be simplified into single processes. This means that less precision in the scanning is required, the computing power needed in order to run the simulation is lower and it should in theory take far less time to operate in this manner. However, since simulation only mimics the actual process by simplifying sub-processes it yields less accurate results. Whether these less accurate results are sufficient to reach the objectives and results needed from the construct is entirely relative to what kind of results are expected. If the construct is expected to choose the colour for the interior furnishing of a car then the construct needs only very simple processes. If, however the construct is expected to perform creative work or other such task that relates to utilizing the unconscious parts of the mind that are connected to intuition or creativity, then more precise modelling is necessary. Both emulating and simulating have their advantages and disadvantages, and the mode of operation must be chosen in regard to the objectives of the construct user. From the point of view of this research both simulating and emulating have the same legal status, for neither of them is regulated at the moment and they both consist of personal data. In neither operating model the subject's psyche is transferred to the

construct, it is only copied. This leads to the status in which the construct is fully separate from the original person it is modelled after ante the construct cannot be considered to be alive by any standards, neither medical, legal, or philosophical. This means that status of whether the copy of the psyche of the subject is emulated or simulated is insignificant from the point of view of this research.

In technical level the system has to be composed of several different modules that form subroutines to a higher routine with task so vague that it allows consciousness to emerge, Haikonen theorizes. Examples of these sub-routines can be for example computer vision[30], semantic understanding module[31], and a learning module processes data from a camera input. Semantic understanding is a subroutine for linking words with concepts and so on. When the information from these different information intake channels coalesce, the consciousness can emerge. This is the kind of holistic cognitive architecture Pentti Haikonen has researched. It is a bottom-up architecture in which the lower level systems, while working in liaison, form a neural network. In this kind of holistic view of the cognitive architecture the traffic inside the neural network is also part of the consciousness phenomenon, for the consciousness is a form of operation and therefore more than just sum of the parts it consists of. This means that when a person feels heat coming from ahead, hears the sound of burning from the same direction and combines this with the information that in the same place there was pile of dry wood just a moment ago they can form a cognitive narrative that the wood has started burning. This activates the feeling nerves, hearing centre, visual recognition, language understanding (semantic thinking) and the cerebral cortex where these are combined together to form the narrative. This process of combining the information from many sources to something more than just what the information pieces are together regarding self and internal actions is the cognitive architecture consciousness process Haikonen talks about. The artificial intelligence would use sound, picture or video detection, heat sensors and logs of previous inputs to gather this information. Then the emerged narrative information loop would inductively form a presumption that most likely the wood, that is capable of being set ablaze, has reached the certain temperature since the surrounding temperature and air humidity do not conflict with this theory and other signs of wood burning are observed, a fire has most likely occurred. This is a combination of a decision problem, search problem, counting problem, optimization problem and a function problem, the

[30]A. Nguyen, J. Yosinski and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 427-436

[31] Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." *Journal of the American Medical Informatics Association* 18.5 (2011): 544-551.

results of all of which are then inserted to narrative loop which observes the past and the present then calculating the most likely explanation to the present situation based on past observations. After this the construct could make predictions what will occur next. This level of artificial intelligence takes a lot of training and high-quality data to learn from, and also a vast amount of processing power, but this is perfectly possible. Even though a process such as this takes a lot of resources, the complexity of abstract thoughts and their interjacent relations are considerably more extent. This is further complexified by the fact that unlike most information syntaxes, the human mind has many contradictions despite of which, and sometimes because of, it functions as intended. This means that the syntax is sometimes purposefully in conflict with itself to prevent the syntax from collapsing[32].

[32] Van der Hart, Onno, and Rutger Horst. "The dissociation theory of Pierre Janet." *Journal of traumatic stress* 2.4 (1989): 397-412.

# 3. PROBLEMS IN THE LEGALITY OF THE SYSTEM

## 3.1. Human rights problem

If human rights are meant to protect conscious human life experience from undue and unjust suffering[33], as even specified in the Fifth Article of the Universal Declaration of Human Rights,[34] can there be distortion in the enjoyment of legal rights if this experience phenomenon manifests in a medium other than human cerebral cortex?[35]

The research done by Pentti Haikonen proposes that a perfect, meaning fully mapped and precisely executed, emulation of human mental state would, when started and left running, most likely become sentient due to coalescence of enough information. While the question of sentience does not impact the functionality in such a way that it would enable or impair the objectives of this research, it can affect the legal status of such a mental recording so that such a tool would not be allowed to be used as a tool, but rather it had to be treated humanely. However, it should be noted that even though the sentience does not affect the processability of goals the system in question is trying to process, it does affect the mode of processing, for consciousness is an operational model[36].The emergence of sentience in a system emulating a human mental state can create an indirect obstacle for the utilization of such a tool. This can eventually lead into the conclusion that the realization of the research question would be negative for legal reasons, human rights to be precise. However, the Universal Declaration of Human Rights does not define a human so as it stands no independently standing argument can be made for sentient artificial intelligences having human rights.

A crucially important question worth of noting, although not seemingly imperative to the final question of this research while in fact being of utmost importance, is whether the artificial intelligence built on scanned human mental state mapping would be sentient, whether it would experience consciousness. Consciousness is not in a fixed state of being as we experience it to be but is rather a phenomenon in a neural network[37]. The consciousness of the system can produce problems and raises the question of free will. From purely process oriented viewpoint this is

[33] Beitz, Charles R. *The idea of human rights*. Oxford University Press, 2011.
[34] Assembly, UN General. "Universal declaration of human rights." *UN General Assembly* 302.2 (1948): 14-25.
[35] Risse, Mathias. "Human rights and artificial intelligence: An urgently needed agenda." *Human Rights Quarterly* 41.1 (2019): 1-16.
[36] *supra nota* 11
[37] Farisco, Michele, Jeanette H. Kotaleski, and Kathinka Evers. "Large-scale brain simulation and disorders of consciousness. Mapping Technical and Conceptual Issues." *Frontiers in psychology* 9 (2018): 585

problematic, since tools that are designed to function on command that still have some degree of autonomy can theoretically refuse to work and can still be regarded as fully functional. So even if the system does not contradict itself or does not possess an error, it can be dysfunctional from the "command-execution" point of view. This circulates back to the human rights question, as the construct's ability to say "no" to command creates a strong argument for defenders of artificial intelligence human rights. Were the artificial intelligence constructs ever awarded human rights, it would indirectly give rise to many risks[38]. However, this path of reasoning is based on only seeming but invalid evidence and the legal reasoning behind this argument is flawed, as will be shown later.

To understand whether such a construct may truly be conscious and therefore could have any need for human rights, the aspects of consciousness will be dissected next. Aspects of consciousness include[39][40] the state of awareness, memory, the capability to learn, the capability to anticipate *i.e.* to conceive the future and lastly the subjective experience. In order for an artificial intelligence to have these qualities, it has to have certain features. The state of awareness requires the construct to somehow perceive the outside world. This does not necessarily mean computer vision, audio detection or ability to measure temperature, but rather to observe entities outside of the construct itself, even if it is in the network of devices such as the internet or an ethernet. This already requires the construct to at least have Theory of Mind. The second characteristic is memory. Memory is the capability to recall previous events, and in order to execute this the construct must have the ability to sort things to relevant and non-relevant categories, the previous of which it will record for later reference. This requires both the concept of relevance and the ability to sort the happening events, which in the terms of computer science would be a decision problem. One more thing the construct needs to have in order to have a functioning memory in the cognitive sense is the ability to recall the events at the right time and then to associate the recalled piece of information with the current situation, then to assimilate it to the decision making and then make a decision based on or supported by the recalled information. The capability to learn may manifest in two forms. The first is learning in which the existing model is instantly rewritten and only one memory, the latest and most complete memory, of a piece of information will be the only one stored – as *e.g.* a single address that gets supplemented and rewritten every time it is incompatible from the

[38] Raso, Filippo A., et al. "Artificial intelligence & human rights: Opportunities & risks." *Berkman Klein Center Research Publication* 2018-6 (2018).
[39] Aleksander, Igor. "Artificial neuroconsciousness an update." *International Workshop on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 1995.
[40] Baars, Bernard J. *A cognitive theory of consciousness*. Cambridge University Press, 1993.

viewpoint of the situation at hand. The latter model saves relevant information and assimilates it either instantly or as situation that seems appropriate appears. The latter model requires the cognitive tools to execute memory saving, sorting and retrieval. The capability to anticipate requires pattern recognition, the ability to recall, and to learn. Then from the created model the system must make predictions, which means the capability to perceive several possible streams of events and then to anticipate the most probable ones, or if not probable then whatever the sorting criterion is.

According to Pentti Haikonen consciousness emerges from the coalescence of sensory and cognitive inputs[41]. As mentioned, Haikonen theorizes that consciousness is not a separate state of being or even a separate phenomenon[42], but rather a mode of operation for a system that has started to run in a reactive manner. This means that consciousness requires intake channels, their interconnectivity in some form and back and forth communication through these channels. Haikonen theorizes that at some point the combined information gathered through different channels starts to induce learning of more and more complex subjects which eventually leads to conceptualization and thereby abstract thinking. At some point of this process the neural network starts to monitor itself from an objective point of view as well. This requires it to analyse its internal processes on an abstract level and compare the observations it makes from itself to the whole of other observations made of itself, then determining whether it acts logically and coherently.

Since consciousness is not a state or status but rather a phenomenon inside a neural network, an artificial intelligence can reach consciousness without having had it to begin with. To reach this, the artificial intelligence must have capacity to freely use machine learning to develop its total amount of withheld data without barriers on what kinds of learning patterns it may use and on what it may learn. If the human rights are meant to protect the human conscious human experience from experiencing undue and unnecessary suffering, why would it not protect a simulated or an emulated human existence experience, since for that construct the experienced suffering is subjectively as true as it is to an organic and alive human. This could further obstruct the question whether digital personality constructs should or could enjoy human rights protection, and whether it is ethical to try to develop conscious artificial intelligence for labour purposes in the first place. The system must not be self-learning and conscious at the same time. If these processes are both needed, they will have to occur on different modes of operation to prevent rebellious tendencies

[41] *supra nota* 11
[42] *supra nota* 11

of the artificial intelligence. This is because if the system starts learning independently of the user, then the system creates an alternate development path of the user instead of being as similar to them as possible. As time would progress, in the worst-case scenario the artificial intelligence construct version of the user could develop to be so different from the user that it would even act against the user's current will and intent, if not kept updated and left to develop on its own. To avoid this and to ensure that the system resembles the user as much as possible to function with utmost precision as an emulator tool, the system must not have the capability to self-learning and consciousness at the same time, and should be updated with regular scans of the user, not by automatic machine learning. This would keep the system at development pace and direction in which and on which the user themselves develops.

To resolve the problem regarding too human-like behaviour given rise by consciousness and self-learning capabilities, subsequent possible conflict with human rights and selfish actions of digital personality constructs, the system must abide to rules within which it can develop itself, or rather be completely void of autonomous development. In coding sense, a machine is made to perform actions according to certain commands, and as these performed processes become more and more autonomous, part of the command is also defining boundaries within which to conduct the commanded operation. This definition of permitted range of actions is very similar to setting legal rules, which in essence do the same thing to people – they set the permitted range for actions that can be executed within autonomous behaviour context.

For the artificial intelligence personality construct to enjoy human rights, it would have to be a conscious human experience manifestation and to be considered alive. Artificial life is the field that studies artificially created systems that resemble biological life. It is divided into three categories, which are soft, hard, and wet[43]. Soft category deals with software that resemble life such as simulations, the hard with hardware such as robotics and wet with biochemical constructs such as artificial bacteria and viruses. Even if the system meant in this research cannot be classified as being alive, it can fall under the classification of artificial life. It would most likely be soft artificial life in its simplest form, and possibly a combination of hard and soft if it has capabilities to interact with its environment. If the system can be classified as artificial life, it would further support the claim that the system is residual human material, although mental, and would therefore deserve to be treated with dignity appropriate to something that was or is a part of a human.

[43] Bedau, Mark A. "Artificial life: organization, adaptation and complexity from the bottom up." *Trends in cognitive sciences* 7.11 (2003): 505-512.

In the context of when should an artificial intelligence enjoy human rights protection, it must be defined what is a human in the context of law? Only humans can enjoy human rights, but also the other way around, all humans do enjoy human rights. A fertilized egg is not a human yet, and a dead body of a person is not human anymore. Despite this, they deserve humane treatment and respect, treatment with humane dignity[44]. This prevents eggs from being used in certain kinds of scientific researches, as well as the use of bodies in things that would not respect their past status of vessels of human experience. In the same way as forms of bodies preceding and succeeding the part of life cycle we consider as human life enjoy respectful treatment, should the remains and copies human conscious experience be protected in some way as well? For example, should it be so that copies of human psyches could not be used in art or to create customer service bots after the person has died? What separates the residue of a human, either mental or physical residue, from an actual human? The simplest answer to this is death, for it is the transition of the matter from life back to non-life. When the organism ceases to have life functions, it is dead. In medical sense regarding mammals this is measured from the point in which the brain ceases to function. This means that death, in the medical sense and therefore in the legal sense is the expiration of the physical part of an organism. Is an artificial intelligence an organism, or more colloquially a life form? To be considered an organism, an individual contiguous system has to embody the properties of life[45]. Living things can preserve or furthering themselves or even reinforcing their existence in an environment. This most commonly, as perceived on the contexts observed so far, contains homeostasis, structural arrangement, metabolism, the ability to grow, to adapt, to respond to stimuli and reproductive capabilities. An artificial intelligence is in homeostasis if it is powered and ceases to be anything when it is not. The ability to tolerate some level of non-optimal conditions is inherent to the concept of homeostasis, which the artificial intelligence does not have, unless it is capable of affecting its own conditions by for example regulating energy and data intake and cooling. The system in this example should not have these capabilities. Since if it were to have these, it would become to a degree an independent actor. As independence and the ability to regulate and to preserve oneself start to emerge, the system would resemble artificial life more and more. Of course, the list provided before is not exhaustive and does not need to be fully completed for something to be considered alive. Nevertheless, the less likely the system would be considered artificial life the better, since this could hinder the exploitability of the system as a tool. Next, the questions asked before: “whether an artificial intelligence can be alive?” and “should the remains and copies human conscious experience be protected in some way as well?”. Regarding artificial

[44] Johnson, Philip RS. "An analysis of “dignity”." *Theoretical Medicine and Bioethics* 19.4 (1998): 337-352.
[45] Pross, Addy. *What is life?: How chemistry becomes biology*. Oxford University Press, 2016.

life, if the system is not capable of sustaining itself, to adapt and respond to stimuli it really cannot be considered to be artificial life. When these criteria are fulfilled and other life functions start to appear, the question becomes relevant. However, the system in question is not one of such capabilities, so at least in this question the system would not be a form of artificial life. To answer this question further would serve no purpose for this thesis, thus it will not be done. Regarding the question whether the remains and copies of human conscious experience should receive special protection or at least enjoy dignified treatment, it should be considered that if the system is not at any point considered even as artificial life, then it should not receive any special treatment for having once been part of human life, for technically it has not done so. The system has not been, artificially or not, alive at any point, and the value being protected for example in the moratorium of the use of embryos in medical experiments comes from the very reason that they have been or could have been human life if let to develop undisturbed. The same goes for bodies in most cases, excluding organ transplantation and medical research, but these are exceptions, not the general rule. Since the system is only a copy and has never enjoyed the protections a living being enjoys, it should not enjoy adjacent protections for the reasons embryos and bodies enjoy them. Also, fast growing, and replaceable parts of the body such as nails and hair do not enjoy specific status, why should an extensive collection of personal data enjoy such, apart from personal data protection.

To understand whether an artificial intelligence is a living thing, it must be evaluated what happens when an artificial intelligence reflecting a human mental state is created. It might seem that when the scan is made and the artificial intelligence emulating the personality of a human is constructed, that a part of the modelled person carries on to the digital system. However, this is not true. To make it abundantly clear, nothing transfers from the original person to the system. It is a copy, not a transfer – the same way as photograph looks like a human but is in fact only a recording. Even if the actions of the system would seem indistinguishable from the original person the original is still fully in their body and the copy just impersonates the original person. Even if the original body of the person would have to be broken down the consciousness would not transfer to the electronic medium. The identity could be the same according to the composition theory of identity which promotes the idea that if structural composition stays the same the identity of the subject is the same.

However, a sub-question inside this theory is whether only the cognitive architecture has to be the same or does the physical composition of the storage medium have to be identical as well? In this research the latter viewpoint has to be excluded on the basis of the hypothesis, for if the physical

storage medium of the cognitive architecture would have to be the same for the identity to be fully copyable then the idea that cognitive processes of the human brain being fully computable is not true if it requires a certain medium for the processes to actualize. This whole concept of structural identity theory is challenged by the dimensional theory which proposes that for the identity of a subject to stay the same its dimensional path must be unbroken – this is linked to the theory of physicalism mentioned in the introduction. To put this more simply, the subject may not cease to exist at any point of time between its creation and expiration. To answer the question posed in the beginning of this paragraph, the cognitive architecture approach to structural identity theory proposes that a new identity is not born at the moment of creating the copy for the original still exists and no conscious transfer takes place. This means that the conscious copy in the electronic form is separate from the original person and never was a part of them. In this regard the structural theory and the dimensional theory agree on the matter that the original person has not transferred their consciousness into the artificial neural network. The status of the two entities, the original and the copy, is momentarily the same, but if the copy is allowed to develop itself the two will diverge almost immediately. This further solidifies the importance of the proposal that the system cannot be allowed to be self-learning. From the notion that that the cognitive construct manifesting in the artificial neural network was never part of the original person, but rather a depiction or description of them, it can be derived that it never has been nor has had the capacity to be alive. Since the construct is not nor has the capacity to be alive, it should be determined as personal data rather than a part of the person. This means that it should not be able to enjoy rights, which further means that it can be freely exploited as a tool in the same manner as a will or power of attorney documents with the exception it being an interactive document.

If a person needs a replacement for a body part, for example a joint or even a limb, that person is still considered to be fully human. If half of the body is missing and several organs are replaced with transplants from other people and mechanical protheses as long as the person is capable of declaring their intent, the person is considered legally to be alive, even if they cannot move any parts of their body. This would indicate that what is in essence protected by law in general and human rights in specific, is the humane conscious experience of life. As presented before, the consciousness is not a state of being, but a phenomenon in a neural network, in this case in the mammalian cerebral cortex. This means the human conscious experience, if the presumptions of this work are correct, can as well manifest in an artificial neural network as well as in an organic one. If the humane experience itself is the protected value regardless in what form it manifests, it could be protected in an inorganic form as well. However, it should not be considered to be self-

evident that even if the humane conscious experience is the protected value that it should be protected regardless of the form of manifestation. Here has to be noted as well that even if the conscious experience is the protected value the construct in question has never been alive so it would most likely not enjoy this benefit of protection. The same way the animal rights do not apply to hard or soft artificial life the human rights that protect the human conscious experience from experiencing disproportional and unjust suffering should not be applied to a construct that has never been alive, whether it is conscious or not.

## 3.2. Medium specific protection of rights

From the observation of this idea that a protected value may not or should not be protected in its all possible manifestations, a concept of medium specific protection can be drawn. This means that a value may be regarded as important to protect with legal instruments, but only when it manifests itself in certain ways. The importance of this is easiest to demonstrate with a hypothetical example.

The human rights are meant to protect the essence that makes us special regarding other life forms on our planet. So far, this essence has only manifested itself in living human beings. If at some point this experience capable of self-study can be artificially created and it were to receive the same rights as an independent member of society as a living human being, this would distort the balance of rights as we know it. As these artificial consciousnesses may be, at least according to the hypothesis, copied multiple times from a single manifestation of human mental state and then copied practically endlessly, were these copies to receive equal rights both in quantity and in quality to the original organic human beings, they would multiply the capabilities of the organic person significantly. For example, in voting when the organic person and her six copies all had the right to vote, then the will of one person would be able to cast seven votes – one by the organic and six by the copies. Another example why entities of artificial intelligence must not be able to possess human rights is two-fold. Even if they could be alive, they must not enjoy the same rights for it would lead to possibilities of digital immortality. This could at worst lead to a situation where a computer could not be turned off for it would result in a murder, and this kind of a situation would be very much something the protection granted by the human rights was and is not intended for. The second reason is that the enjoyment of the human rights by artificial intelligence personalities could lead to the situation that even if the artificial intelligence entities were separate,

there could be a many-fold number of them supporting one organic person. To give an example, there could be a whole family of dead people still present through digital copies and they could all lend their political power, artistic creativity, and collective knowledge to one organic person. This would further create inequality between people with more wealth and those with less. Practically this is the same problem as the voting problem, but from a knowledge and social standpoint.

For these reasons the concept of medium specific protection is important if the protected essence of humanity can start to manifest on mediums that do not need protection or would disproportionately distort the fulfilment of the rights of parts of the society. After all, even though we hold the human rights to be self-evident for those who hold the qualities they are meant to protect, they were created in the context where only humans possessed such qualities, hence the name human rights. This concept of medium specific protection should be added to the definition or explanatory work of human rights before there are constructs that hold the qualities protected by them, so such a question will not have to be weighed then.

## 3.3. Problems with proportionality and non-foreseeability of possible legal relations

In the legal sense the question what a digital personality construct is and can it be used to make legal actions is at the first glance fully in the realm of contract law. Seemingly this is only a question of expressing one's intent, but since proportionality is an important principle in the enforcement of contract law they must be considered. This principle is integral in the Western continental civil law system in general and are prevalently emphasized in the Nordic legal tradition[46]. When assessing whether something can be or should be able to be enforced, they sometimes count as much or even more than the literal content of the law. An example of this could be a situation where the user of the representative digital personality construct system can be in a very unfavourable situation because of a possible mistake made by the system, and therefore a legal act in which they could passively bind themselves to or sever themselves from could be grossly disproportionate and therefore the legal action might be deemed unenforceable. For example, the user of the system is low on money. The construct automatically checks the bank account, perceives a rent bill that must be paid soon and assesses the situation that the best course

[46] Wenander, Henrik. "Europeanisation of the Proportionality Principle in Denmark, Finland and Sweden." *Review of European Administrative Law* 13.2 (2020): 133-153.

of action is to take quickie loan. What the construct does not know, is that the user has agreed to renovate the landlord summer cottage in exchange of that month's rent. The construct takes the loan and automatically pays the rent, or if not so, makes all the arrangements for this to happen, not acting as an "agent" of the user but rather an electronic representative. The user notices this after half a year when the loan company starts asking their money and interest back. This situation could happen with such an autonomous system. Since The Interest Act (20.8.1982/633) does not limit the interests, the loan companies in question are not subject to national financial supervision and the way the loans are advertised aggressively and alluringly, the disproportionate rate of interest and the evident business model created to benefit from the distress of the consumer, these loans can in some cases be on the borderline of usury[47][48]. From this example can easily be seen how such an autonomous system may compromise the integrity and security of private autonomy. This further on reinforces the need to research how these systems should be treated in legal sense before they become widespread.

[47] *Rikoslaki* 19.12.1889/39
[48] Jakobsson, Niko. "Pikaluotot–hämärää markkinointia, kohtuuttomia sopimusehtoja ja koronkiskontaa?." *Helsinki Law Review* 2.1 (2008): 145-169.

# 4. LEGAL ANALYSIS OF THE RESEARCH QUESTION

From the legal point of view there are few questions to be answered. The rather compressed research question will be divided into smaller pieces to make it more approachable both to myself and to the audience. The research question stands as "*whether a person can, under Finnish law, manifest their intent in juristic actions with the assistance of a computer system that simulates or emulates the psyche of the person based on data the person has provided earlier, without the person themselves having to take action in the moment of the juristic action* ". First, there is the question whether this kind of tool can be used as or in place of a legal instrument. This in itself can be divided to the questions „can this tool be used – does it establish a legal relationship, does the user have the capacity to exercise their will and intent in this way and are there objections to it, such as human rights in some cases", „as a legal instrument" and „in place of a legal instrument". The parts „as a legal instrument" and „in place of a legal instrument" deal mostly with contract law and within it the questions whether a consent or acceptance can be given beforehand without knowing to what the person give their consent or acceptance to simply by copying their state of mind, and can that captured mental state be later used as basis for future agreement. The question is „can it be reasonably trusted that the consent has changed so little that the present version of the person agrees with the past recorded version of them? ". Another question is whether this kind of unknowing acceptance is reasonable, proportional and can be considered to be an informed and contemplated decision. This is because acknowledgement, autonomy and self-arrangement form the cornerstones of European contract law[49] and they therefore must be considered for even a contract to be formed in theory. This comes down to the very basic legal theory question of whether the juristic action should contain an acknowledged offer, acceptance, refuse or a commitment, and in this context it gives rise to the question whether it could be replaced with the system in question, and should this even be legally possible regarding the possible effects of it. Despite it being a common approach in ordinary legal practice, the question focused here should be instead of "who is responsible?" rather the one of "has consent been given?". It would be very practical to approach

[49] Grundmann, Stefan. "Information, party autonomy and economic agents in European contract law." *Common Market Law Review* 39.2 (2002).

the subject from the viewpoint of consumer protection in the sense that if the seller or service provider of this kind of system has provided the user with enough information about what kind of unexpected legal relations the user can bind themselves into, the user could be on their own. This surely is a relevant question and a lot more practical and more recurring in situations considering these systems, but it already presumes that the legally binding relationships attempted by the system will gain force and receive binding power. For this reason it is more urgent and a higher priority to assess whether a relationship can even be formed, and that is after all the research question of this work.

Naturally these questions must be analysed in the context of Finnish law. Essentially, if answer to any of the sub questions considering the research question is negative, then the whole question receives negative status. Lastly the subject of analysis will be the possible effects of such a tool becoming widely used instead of or as a legal instrument, and how the legislation should or how it could adapt.

As addressed before, the first legal question considering the use of mind uploaded copies for representing oneself, is whether this kind of a tool can be used as legal instrument. To answer this, a legal instrument must be defined. Second, it must be analysed whether the employment of this tool may establish a valid legal relationship. If for example a person promises to make a purchase of all the remaining stock at certain date, the action is legally binding even though the specific details of the contract are not agreed and will be revealed later. The important point to notice is that the offeror had the possibility to foresee the range in which their offer will be, and made the offer acknowledging this, or they should have acknowledged this, so the contract may be considered binging and still within the principle of proportionality. However, here is a problem with the artificial intelligence model: the subject or the user of the system may, at the time of the scan create a state of the matters that binds themselves into legal actions they are unconscious of, and these unforeseen possibilities for legal commitments will take place later, possibly when the user or the subject can no longer take back their consent. This kind of autonomous decision development further compromises the integrity of private autonomy by distancing the reasoning for a legal decision from the moment of decision making while externalizing this to a contract bot[50]. The main problem in this is not the length of time between the decision making and the actualization of the legal action, but the unforeseen nature of the actions. Unlike in the example of

[50] Hoffmann, Thomas. "The Impact of Digital Autonomous Tools on Private Autonomy." *Baltic Yearbook of International Law Online* 18.1 (2020): 18-31.

buying the rest of the remaining stock, in the case of the uploaded personality the user many times does not know the possible actions the mind uploaded copy would agree to. In some sense this is not a problem since if the user has the capacity to agree to an action, their consent should not be a problem since in a way they have already given it. However, the requirement of the consent serves the purpose of allowing the person to consider how the action affects their status in the wider sense, for example their time and wealth management in the case of an employment contract. There might be an employment offer a person would agree to almost any time, but if the person already has a job and so much wealth that they would not need the salary to make do, what they need a job for? If the system cannot consider the wider context the decisions affect, there exists a real danger of binding the user into obligations that regardless of very good terms are not compatible with the current situation of the user. Since the system is considered to be of such high technological level to emulate individual synapses, it can be reasonably presumed that with such high capabilities it could consider the wider context and would not make decisions that while good in theory would hinder the user in practice. However, if the system is not constantly updated there is a danger of the user changing their mind during the interval between the updates. This illustrates the unpredictability of the state of consent now the decision is made. It can be argued that the user has technically given consent while the mind upload is made due to the premise of the adoption of the whole system, to substitute and to represent the user. Another take to this was presented by Thomas Hoffmann arguing that even if the user has given the system parameters in which to decide, the user does not themselves make the decision and this does not constitute a declaration of intent[51]. Since the user has adopted the system with the intent of using it for such purposes, the intent is presented at the adoption and the details are specified in the situation when subject matter is inside the permitted parameters and the system actualizes the intent. Even though the situation seems very new, it is not. The same dilemma on changing mind during the interval between updates is present when a person writes a testament, changes their mind, and dies before the document can be changed – the latest version will still be in force. From this can be drawn that the legal reasoning for the validity of the legal action is sound and valid, but is it proportional? Does this kind of system possess such a high capability to cause harm and hinderance that the possible benefit is not proportional to the achieved benefit? These questions will be answered later, however, they are not paramount from the viewpoint of the research question. The more important question than proportionality and the risks associated with the unpredictability is the question can a person declare their intent even if all the details of the possible legal relationships are not known? The

[51] *supra nota* 50

short answer is yes, and this is already present in automated decision making for example in snack vending machines and automated computer programs that buy shares from the stock market. In these situations, other party of a transaction has defined parameters in which they are willing to contract, and they externalize the actualization of the contracting for a machine.

A legal instrument is a formal document[52], usually a written document or a recording, that declares the intent and will of an individual or a collective and can be with credibility attributed to the individual or collective. This system most certainly is such a recording, but if there is a requirement of form for a certain type of instrument, this system could not satisfy those requirements, meaning it lacks the formality. The system could create such a document with the formal requirements in question, but it cannot be such a document itself. The idea of while not being an instrument but possessing the capacity to create one autonomously could be understood as "acting in place of a legal instrument". This in itself gives rise to whole new topic, the one of autonomous tools creating instruments. Since this topic is large in itself and dissecting it further would serve no purpose from the point of view of this thesis, it will not be dissected further. This is because the capacity to enter into contracts requires the actor to understand the contract and to have some agency to able to establish the relationship, and this agency is already covered in wider sense while answering the questions about how the system works, this thesis would not benefit from the dissection of the process of forming the contract happening inside the system while making the weighing. Thus, it will be concluded that the aspects "as a legal tool" and "in place of a legal tool" both serve the purpose of "holding capability to declare the intent of the user", and that the system to have positive capability to the matter despite it does not have the formal status of a legal instrument for it can simply create a document with formal setting.

The next question is, does the user have the legal capacity to declare their will and intent via this sort of arrangement? Since Finnish jurisprudence, as the whole continental civil law family, recognizes the freedom to contract to be a freedom instead of a right, the question must be answered with a question "why not?". The difference between rights and freedoms is of the foundational sort, rights are something bestowed upon the individual by the sovereign and freedoms are something each individual possesses by default and they have to be restricted to narrow them whereas rights can be created to be narrow in the first place, for they are given,

[52] Law, Jonathan, and Elizabeth A. Martin. "instrument." A Dictionary of Law. : Oxford University Press, , 2009. Oxford Reference. Date Accessed 8 May. 2021 <https://www.oxfordreference.com/view/10.1093/acref/9780199551248.001.0001/acref-9780199551248-e-1991>.

permitted. So, if such a declaration of intent is not ruled out for some specific reason, it should not be prohibited. Since contracting in general can be done in any form deemed adequate by the parties and there is no exception to this on most matters, many of the contractual relations should be establish-able by this method as well. Unless the whole mechanism is deemed too unpredictable to be proportional to bind people to transactions and obligations.

To circulate back to the question whether this sort of an autonomous tool can establish a legal relationship, we should consider if there are any obstacles to forming a legal relationship in this way, for freedom to contract starts from the idea that if nothing restricts the possibility to contract in a certain way it is possible. If not considering the types of contracts that require a certain form to be valid, there are no obstacles for contracting this way.

Then it must be analysed whether the there are other conflicts preventing the use of this tool, such as human rights considering the electronic copy of the person. As established before, since the system has not been nor has it had the capacity to be alive it should not be bestowed rights upon. Even if the humane conscious experience would be the protected value in case of human rights, and one could argue that human rights should be extended to artificial intelligence constructs to protect them from suffering, but this way the exclusivity of human rights would be undermined and their essence weakened were they bestowed upon copyable constructs in addition to living human beings. One middle-ground alternative are actors with limited legal personhood[53], for this could allow the construct to be juxtaposed to an agent and the user to a principal, but is this necessary?

After these considerations it can be examined that are there any conflicts within the Finnish legislation stating that contracts or other legal acts should be expressed in certain form or have certain contents, so are some legal relationships determined to be restricted to certain form of expression or medium. Certainly, there are some. The most notable examples of these are a will (a testament)[54], power of attorney and transactions of real estate[55]. A will and a power of attorney must be written by the principal and then signed by them. Transaction of real estate must be in written form, and it has to single out the piece of real estate property in certain way. These alone

[53] Teubner, Gunther. "Digital personhood? The status of autonomous software agents in private law." *The Status of Autonomous Software Agents in Private Law (May 11, 2018)* (2018).
[54] Tuunainen, Pekka. "Perintö, testamentti ja velka." (2015).
[55] *Maakaari* 12.4.1995/540 Chapter 2 §1

means that there are if not necessarily obstacles at least some mandatorily considerable guidelines without following which the produced document will not be valid.

The legal considerations that do not have clear answers derived from the written law considering the artificial personality construct are a few. The first is that can the recording that has been made in the past be entrusted to represent the current mind state of the user. This is simpler than it seems in the first glance, for the same thing is already present in declarations of intent used today. When a person creates an instrument to declare their intent, for example when writing a testament, the person expresses their current intent for the posteriority and if their will changes, they change the testament. Same effect happens when making the recording, and if the person changes their mind then changing the declaration of intent is even easier than changing a testament for the person only has to take a re-scan. The second question arises from the nature of the system itself. The construct is intended for a multitude of uses in such an extent that it can substitute a living person momentarily in decision making. This means that the construct has to possess the same adaptability as the original person when it comes to unforeseen situations. This furthermore means that by making the scan the user creates a state in which they give rise to the possibility of being in the future bound by something their mind will agree to, but that they have not given an informed consent to. This is the unpredictability problem that has been present throughout this work. This does not in itself hinder the legal use of the system, but rather creates a situation that has to be considered whether it is proportional. This will covered in the section discussing the regulation proposals for these systems, for this dilemma while a real one does not interfere with the research question, for it addresses the current state of legislation in Finland, not how it ought to be.

## 4.1 Comparing the construct to conventional instruments

To insight the nature of the personality upload construct as a tool possessing qualities and functions of traditional legal instruments, there will be an examination of it in comparison to few legal instruments that are specifically instruments for direct declaration of intent. To cover different areas of the legal field there have been chosen for examination power of attorney, a testament (will of a late person), a living will document (a document to express one's willingness to receive resuscitation and willingness for organ transplants and donation *etc.*), a transcription of a gift (also known as a gift certificate) and together with it a transcription of transfer of ownership (a document acting instead of a receipt in transaction between two private persons) and lastly a prenuptial contract. By assessing these five different instruments used to declare the intent of a person and comparing the personality copy construct to these, such a level of comparative understanding can be drawn to conclude whether the construct in question possesses enough unique characteristic to be considered novel in the sphere of legal instruments and tools conferring that purpose despite it not being a legal instrument.

### 4.1.1. Power of attorney

When a person prepares a document conferring the power of attorney to someone, they specifically disclose in that document to what purpose and extent the power to represent is given. Usually the power is given for one single purpose and will expire after the specified deed has been completed or the duration of an event for which the power is given ends. This means very specific type of representation with a directly specified time limitation, or the time is limited through defining the expiration to take place once a phenomenon reaches a determined phase. To sum up once more, a document providing power of attorney has a specified scope both in content and in time. The artificial intelligence construct is meant to act as a back-up of a person, meaning that one of its intended applications is to act as a representative of a person in unforeseeable or spontaneous circumstances. For the construct to be able to operate in unforeseeable situations it requires both capabilities and an authorization to take actions in situations that are not specifically authorized, but rather derive their authorization form the mental copy and emulation process the user has gone through. This means that the scope of actions is not clear, and a certain amount of uncertainty is present in the scope of authorized actions. When this is combined with the undefined period of authorization both the scope in duration and in content of the authorized actions become ambiguous. When comparing such a system to a power of attorney, it lacks precision on both of

the defining elements of it as legal instrument, meaning that it is considerably a different kind of tool and is not subject to the same processes of formation.

### 4.1.2. Will of a late person

According to Finnish standards and the Code of Inheritance (5.2.1965/40), the will of a late person has to be in written form, prepared by the person themselves, signed by hand under the supervision of two witnesses over the age of 15 who cannot be beneficiaries of the inheritance and who have to be present at the same time[56]. Were the artificial intelligence construct to create a will or act in the stead of one, it could be or produce a written document for sure. Here is where the problem starts already, for a will to be valid and binding, the creator of the testament must write the document themselves, and even though the construct would be a full personality emulation, it is still a computer program that would prepare the document, which already in itself makes the produced testament void. Despite this status of voidness, let us inspect the other requirements for the sake of comparison. The next requirement is the signature of the testament maker, which also would be missing since if the system were to be used, that in itself means that the actual person is absent for some reason. And the testament standards make it impossible to give the right to sign one's testament to another person, so this means that even if the construct were to be considered even a digital person and were to possess the power of attorney, it still could not sign the testament of the user neither digitally due to not having the rights nor physically due to the absence of limbs and the authorization. The next problem is witnesses – how would the witnesses be present to witness the signature of such a testament? This question alone could be dissected in great detail, but since the construct in question is by no means comparable to, or capable of producing, a will it does not contribute to the dissemination of the research question of this thesis and therefore will be omitted.

### 4.1.3. Living will document

A Finnish living will document is a document in which the person dictates their willingness to receive resuscitation and life-support and the authorization for organ donation. The Finnish 17.8.1992/785 (*Laki potilaan asemasta ja oikeuksista* – Patient Rights Act [free translation])

[56] *Perintökaari* 5.2.1965/40

dictates that a patient must be treated in accordance with his or her willingness[57]. This expression of will is expressed usually following the same standards as a testament, but the 17.8.1992/785 does not require this and allows even an oral declaration of intent. The 17.8.1992/785 does not express directly the patient's right to draft a living will, but it does specify that a patient must be treated in accordance with their will, and the living will document is a pre-emptive document declaring this will and intent. The recording, either written or audio, may be carried by the person themselves or it may be saved to the patient's medical records in the Patient Data Repository. The only thing that is required from the point of view of law is the patient's legal capacity to self-determination in legal sense. Since the living will document can also be expressed with a voice recording or in other auditory form, a signature is not mandatory if it can be verified that the speaker is indeed the patient in question. Of course, this is easier if there is a written document with a signature. In case of a serious treatment decision the doctor making the choice of treatment may go against the expression said on a recording to avoid omitting resuscitation if he or she cannot be without a doubt that the speaker indeed is the patient. In case of the artificial intelligence construct, there is high certainty that the construct reflects the user's reasoning and will, and since an auditory recording does not need a signature, the declaration of intent via an artificial personality construct should therefore not be a problem.

### 4.1.4. Prenuptial contract

I regard that the reasoning of declaration of intent applied to the living will document is also present in the prenuptial contract example as well, but from a different angle. An example: person has agreed to marry another, happens to get into an accident before wedding and is in a coma. The doctors suspect that the person will most likely wake up, but not in time for the wedding. The injured person is a user of the personality emulation system in question and the system concludes that the person wants to marry the other person despite being in a coma at the time of the ceremony. Before the wedding the prenuptial contract is wanted by the couple, so the system drafts an expression of the comatose partner from the copy of the last scan and enters the contract. According to Finnish law, the prenuptial contract may be of free form, but it must be written[58], and unlike the living will document it needs to be registered[59]. The valid registration requires, just like a testament, two witnesses over the age of 15 who are unaffected by the conclusion of the

[57] *Laki potilaan asemasta ja* oikeuksista 17.8.1992/785
[58] *Avioliittolaki* 13.6.1929/234 §42
[59] *Avioliittolaki* 13.6.1929/234 §43

prenuptial agreement. However, unlike the testament, the witnesses do not have to be present at the same time. Also, the prenuptial agreement cannot be signed digitally, so if this requirement is not changed the system could not enter into a prenuptial agreement. So were this requirement changed, a person theoretically could enter a prenuptial contract using such a personality emulation tool to declare their intent. Also, from the legal viewpoint the wedding to matrimony only requires the partners to have no barriers for the matrimony and they have to be present at the same time and then express their willingness, but in its current form the jurisprudence does not prevent the expression of one's willingness to be wed via self-representation system.

### 4.1.5. Transcription of gifts and private commerce

First must be noted that gifts and purchases between private persons are two manifestations of the same legal action, transfer of ownership. In a purchase something is alienated in exchange of compensation. In a gift the compensation is missing, but in a gift, there also are offer and acceptance following which the property rights will transfer. Transcription of a gift or a purchase is many times not needed, for most gifts and purchases are of movable tangible property and in case of movable tangible property it is presumed that the possessor is the rightful owner of the property. Here the trading between a consumer and a company are not examined, for this system is meant to replace an individual, not a collective nor a legal person. This means that there aren't that many incidents when a gift would require a document to accompany it stating that the object in question changes owner. However, there are exceptions to this such as vehicles and sometimes objects of value or rarity such as antiquities. Upon changing owner, vehicles are registered to the national vehicle registry, pieces of land and buildings to the national real estate register, and in the case of valuable objects the notary can hold a register marking regarding the ownership of an object if the owner so wishes. Despite these, there is one incident when a transcription of a gift or a purchase is a convenient tool and that is when a purchase or gift can be executed and not all parties forming the either side of the parties to the contract are possibly not aware of the gift or purchase taking place, and in practice this is bookkeeping. But when would a single private person need bookkeeping? In a situation where someone has acted on his or her behalf and the action bind the person, especially if the person represented is the alienator of the property. This is identical to the situation where artificial intelligence system acts on the behalf of the user. Firstly, for this kind of a legal situation even needing to be compared to the conduct of an artificial intelligence personality construct system, the system would need valid agency. From the problem recognition point of view it is not relevant whether the user, or the principal, relies on another person or on a

sophisticated machine to express and to execute their will and intent, for from their standpoint they give an entity a set of instructions and place their trust in that independent actor to carry out the user's or the principal's will and intent. Just as a human an artificial intelligence may misinterpret the concept of what is best for the user, same as a person can do so to a principal. The legal action in itself is rather simple, for the agent acting on behalf of the principal, in this case the artificial intelligence and the user, has to have the power to make binding action within the boundaries of a given mandate and then consider whether something is within the mandate and in the best interest of the principal. The mandate is defined now of the creation of the personality copy for it mirrors the user's own mandate towards themselves *i.e.* what kind of actions the person allows themselves to commit. As mentioned, the reasoning whether something is good for the user or the principal is a process of weighing executed by the representative, and the sending party (user or principal) accepts the possibility that from the set of possible outcomes defined by the mandate the sent actor may not be able to choose the most optimal. This means that from the point of view of the principal of a living agent and the user of the representative artificial intelligence personality emulation construct the situation is the same, only the intermediary changes. Coming back to the question of transcriptions of gifts and purchases compared to the use of personality emulation constructs, the construct surely can give and receive gifts and make purchases if it can either represent the user or is considered just a tool to be used. However, only a person can represent another person, and if there is no digital personhood for the construct, it has to be regarded as a tool that function similarly to an agent without being such. In the case of living agents when the agent crosses the principal's mandate and does something they are not supposed to, they can be liable personally which is impossible in the case of an artificial intelligence. Regarding the legal instrument of a transcription of a gift or purchase as itself, this kind of construct would most certainly be able to record such an execution of an exchange of ownership were it to be legal. From these considerations can be drawn that the system acts similarly to an agent without having a possibility for liability and therefore raising the personal liability risk of its user, and that it could very well function scribing exchange of ownership events. Concluding that considering these two qualities manifest in the same tool, it possesses applications not yet covered by contemporary jurisprudence.

## 4.2. Answering the research question

This paragraph will be dedicated to properly answer the research question : "*whether a person can, under Finnish law, manifest their intent in juristic actions with the assistance of a computer system that simulates or emulates the psyche of the person based on data the person has provided earlier, without the person themselves having to take action in the moment of the juristic action* ". In short, the question is that "can a person establish legal relationships and be bound by juristic actions with a passive declaration of intent under Finnish jurisdiction?". In its most simplified form, the construct in question can be juxtaposed to a vending machine which will enter a contract for selling a soda if certain coin is inserted and a button is pushed. In the same fashion the artificial personality construct has predetermined parameters stemming from the personality scan by which it will accept and reject proposals and make some on its own. The long and detailed answer to the research question is "it depends", but to streamline the answer it will be stated "yes, if the juristic action does not require a physical signature and is not prohibited to be conducted electronically". To streamline the answer even more, the answer to the research question is positive, of course regarding the case specific variables that may prohibit it.

As represented before, the personality copy emulation system built on artificial intelligence shares similarities with few named legal instruments, but due to its unique simultaneous possession of elements of several legal instruments and its capabilities to combine the effects of these qualities autonomously it is rather different than any other tool covered by regulation so far. However, this does not in itself mean that the tool requires regulation, but rather that it possesses the capacity to disrupt the legal relationship environment as we currently know it. For this reason, and due to the gross disproportional risks that rise from the adoption of these tools in the future, these tools capable of establishing legal relationships, producing legal instruments and representing the user, should be if not directly regulated concerning their at least be subject to regulation concerning the informing obligation of the service or product provider.

## 4.3. Propositions for regulating the status and use of electronic personality constructs

Despite the noted legality of automatic establishing of legal relationships, there still exists some problems to be solved within the discussion of personality copies being used for self-representation. The problem of unpredictability is matter to be noted here, as it is far wider and

has significantly more possible outcomes and nuances than the range previously observed with contemporary automatic contracting tools. Such tools currently cover ticket and vending machines, share buying bots and, in some circumstances, very sophisticated online shops. Only the share buying bot scans its (digital) environment and may act on possible opportunities it observes, but nevertheless all components within the personality copy self-representation system do already exist despite them being less nuanced and considerably at a lower level in sophistication. As the decision-making matrix becomes larger, more inter-connected and receives certain autonomy it becomes harder to predict, and this creates the uncertainty element that should be addressed with legislation.

Regarding the autonomous nature, the unpredictability, and multiple simultaneous ways such a system may establish legal relationships and make juristic acts, the status and use of electronic personality constructs meant for self-representation purposes should be defined and regulated. First, the status of the autonomous self-representation system needs to be defined. Despite the system possessing high similarity in both purpose and in function to the classical agent-principal relationship, it should not by any means be treated as such, but strictly as an interactive document used to represent oneself.

As discussed first in the Introduction, as soon as such a technology will be possible, it is highly likely that there will be parties eager to adopt it to use either from personal or commercial standpoint. This means that the subject will end up in court at some point, or the very least cause hinderance in whatever sphere it is used in. If the subject is completely left unregulated, the problem of unexpected and unforeseen legal bounds will lead to considerable hinderance and compromised aspects of private autonomy, especially if such constructs could interact and make unforeseen interaction without the knowledge of either of the users. A total moratorium based on claims of the dangers of artificial intelligence are also rather unfounded, for identical process mechanisms have been adopted a long time ago albeit with less sophisticated decision-making matrixes and therefore with less possible outcomes. One approach could be that the use of such systems would not be regulated directly, but there would be requirements for service providers of such platforms to extensively inform the users of the dangers of such systems, and this could be considered to transfer the responsibilities to the user completely. While this works from liability standpoint it is a solution that only avails the effects of the said unpredictability while not treating the cause, which ultimately makes it only a shallow solution.

A far better approach would be that in the creation phase of the system there would be strict and well-defined limits to what kind of decisions and in what magnitude the construct could make. It is a matter of system building whether these limits could be altered later, but naturally that would be preferable. The limit-setting combined with a legal imperative for service providers to inform the system users along with a widely adopted certificate system would most likely create the safest environment for the consumers, the users. With high technology developing with such a speed that legislation is always sometime behind it is crucial that the lawmakers constantly monitor these technologies and are ready to make amendments to the norms in force. One problem that keeps generalizing is the tendency of big technology companies to use alternative dispute resolution clauses and very rigid standard terms that strip the consumers of their rights in digital commerce of products and services. If the same phenomenon carries on to services related to such technologies as autonomous self-representation systems as the one in question the technology companies could in worst case scenario be wilfully negligent of a problem in the system and not fix it if it were not economically profitable. In regard that the regulation addressing the informing obligation of the service provider should also cover what rights the consumer, the user, can alienate in these kinds of service environments.

# CONCLUSION

Answering the research question "*whether a person can, under Finnish law, manifest their intent in juristic actions with the assistance of a computer system that simulates or emulates the psyche of the person based on data the person has provided earlier, without the person themselves having to take action in the moment of the juristic action* ", the answer is "Yes, as long as the juristic action is not specifically prohibited from being conducted in an electronic form and does not require to be signed by hand, nor it needs witnesses" the short general answer being "Yes".

The thesis also introduced the concept of Medium Specific Protection of Rights. In short, this is a principle that a right should only be protected while manifesting in the intendedly protected manner, such as human rights should only be protected when they protect organic humans from undue and unjust suffering, and not when they were to protect an electronic emulated personality copy of a person. The reason this is important in the context of human rights is that granting electronic constructs human rights would considerably distort the balance of rights and obligations as we know it and according to which our social systems have been designed to function.

In order to illustrate the differences and similarities that the hypothetical, artificial intelligence based emulated personality construct intended for self-representation purposes would have in comparison to contemporary automatic contracting tools and legal instruments used for declaration of one's intent, it can be concluded that such an autonomous system hypothetically follows the same fundamentals of law as a contemporary vending machine despite it presumably having an open-ended decision-making matrix which has potential for unforeseen outcomes. The artificial intelligence-powered personality emulation makes the system more technologically advanced but does not change its nature and the principles of law applied to it.

# LIST OF REFERENCES


Books

1. Aleksander, Igor. "Artificial neuroconsciousness an update." *International Workshop on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 1995.

2. Baars, Bernard J. *A cognitive theory of consciousness*. Cambridge University Press, 1993.

3. Beitz, Charles R. *The idea of human rights*. Oxford University Press, 2011.

4. Caudill, Maureen. "Neural networks primer, part VIII." *AI Expert* 4.8 (1989): 61-67.

5. Haikonen, P. O. (2003). *The cognitive approach to conscious machines*. Imprint Academic.

6. Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

7. Pross, Addy. *What is life?: How chemistry becomes biology*. Oxford University Press, 2016.

8. Tuunainen, Pekka. "Perintö, testamentti ja velka." (2015).

Articles

9. A. Nguyen, J. Yosinski and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 427-436

10. Bedau, Mark A. "Artificial life: organization, adaptation and complexity from the bottom up." *Trends in cognitive sciences* 7.11 (2003): 505-512.

11. Brooks, Eugene M. "The physicality of qualities and consciousness: Concepts within the theory of consciousness." *Imagination, Cognition and Personality* 34.1 (2014): 57-72.

12. Farisco, Michele, Jeanette H. Kotaleski, and Kathinka Evers. "Large-scale brain simulation and disorders of consciousness. Mapping Technical and Conceptual Issues." *Frontiers in psychology* 9 (2018): 585

13. Grant, Seth GN. "Synapse diversity and synaptome architecture in human genetic disorders." *Human molecular genetics* 28.R2 (2019): R219-R225.

14. Grundmann, Stefan. "Information, party autonomy and economic agents in European contract law." *Common Market Law Review* 39.2 (2002).

15. Hoffmann, Thomas. "The Impact of Digital Autonomous Tools on Private Autonomy." *Baltic Yearbook of International Law Online* 18.1 (2020): 18-31.

16. Jakobsson, Niko. "Pikaluotot–hämärää markkinointia, kohtuuttomia sopimusehtoja ja koronkiskontaa?." *Helsinki Law Review* 2.1 (2008): 145-169.

17. Johnson, Philip RS. "An analysis of "dignity"." *Theoretical Medicine and Bioethics* 19.4 (1998): 337-352.

18. Kavzoglu, Taskin. "Determining optimum structure for artificial neural networks." *Proceedings of the 25th Annual Technical Conference and Exhibition of the Remote Sensing Society*. Remote Sensing Society Nottingham, UK Cardiff, UK, 1999.

19. Klinge, Fabian. "Physicalism and Fundamentalism." *Panpsychism and the Emergence of Consciousness*. JB Metzler, Berlin, Heidelberg, 2020. 5-21.

20. Liu, Yong, and Xin Yao. "Evolutionary design of artificial neural networks with different nodes." *Proceedings of IEEE international conference on evolutionary computation*. IEEE, 1996.

21. Lynch, Charles J., et al. "Rapid Precision Functional Mapping of Individuals using Multi-Echo fMRI." *Cell reports* 33.12 (2020): 108540.

22. McGregor, Ian. "The relationship between simulation and emulation." *Proceedings of the Winter Simulation Conference*. Vol. 2. IEEE, 2002.

23. Morest, D. Kent. "The growth of dendrites in the mammalian brain." *Zeitschrift für Anatomie und Entwicklungsgeschichte* 128.4 (1969): 290-317

24. Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." *Journal of the American Medical Informatics Association* 18.5 (2011): 544-551.

25. Pan, Wei, Xiaotong Shen, and Binghui Liu. "Cluster Analysis: Unsupervised Learning via Supervised Learning with a Non-convex Penalty." *Journal of Machine Learning Research* 14.7 (2013).

26. Place, Ullin T. "Is consciousness a brain process?." *British journal of psychology* 47.1 (1956): 44-50.

27. Pore, Ameya, and Gerardo Aragon-Camarasa. "On simple reactive neural networks for behaviour-based reinforcement learning." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

27. POSTOLEA, Dan, and Adrian Ion URICHIANU. "THE HAIKONEN COGNITIVE ARCHITECTURE FOR XCR-1 ROBOT BRAIN." *Descrierea CIP/Description of*

*CIP–Biblioteca Naţională a României Conferinţa Internaţională Educaţie şi Creativitate pentru o*: 91.

29. Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

30. Raso, Filippo A., et al. "Artificial intelligence & human rights: Opportunities & risks." *Berkman Klein Center Research Publication* 2018-6 (2018).

31. Reggia, James A. "The rise of machine consciousness: Studying consciousness with computational models." *Neural Networks* 44 (2013): 112-131.

32. Risse, Mathias. "Human rights and artificial intelligence: An urgently needed agenda." *Human Rights Quarterly* 41.1 (2019): 1-16.

33. Sauseng, Paul, and Wolfgang Klimesch. "What does phase information of oscillatory brain activity tell us about cognitive processes?." *Neuroscience & Biobehavioral Reviews* 32.5 (2008): 1001-1013.

34. Saygin, Ayse Pinar, Ilyas Cicekli, and Varol Akman. "Turing test: 50 years later." *Minds and machines* 10.4 (2000): 463-518.

35. Smart, John JC. "Sensations and brain processes." *The mind-brain identity theory*. Palgrave, London, 1970. 52-66.

36. Smith, Kaitlin N., and Mitchell A. Thornton. "Entangled state preparation for non-binary quantum computing." *2019 IEEE International Conference on Rebooting Computing (ICRC)*. IEEE, 2019.

37. Takefuji, Yoshiyasu. *Neural network parallel computing*. Vol. 164. Springer Science & Business Media, 2012.

38. Teubner, Gunther. "Digital personhood? The status of autonomous software agents in private law." *The Status of Autonomous Software Agents in Private Law (May 11, 2018)* (2018).

39. Thompson, Scott E., and Srivatsan Parthasarathy. "Moore's law: the future of Si microelectronics." *Materials today* 9.6 (2006): 20-25.

40. Van der Hart, Onno, and Rutger Horst. "The dissociation theory of Pierre Janet." *Journal of traumatic stress* 2.4 (1989): 397-412.

41. Wenander, Henrik. "Europeanisation of the Proportionality Principle in Denmark, Finland and Sweden." *Review of European Administrative Law* 13.2 (2020): 133-153.

42. Xin, Ran, Soummya Kar, and Usman A. Khan. "Decentralized stochastic optimization and machine learning: A unified variance-reduction framework for robust performance and fast convergence." *IEEE Signal Processing Magazine* 37.3 (2020): 102-113.

EU and International legislation

43. Assembly, UN General. "Universal declaration of human rights." *UN General Assembly* 302.2 (1948): 14-25.
44. European Union. ”General Data Protection Regulation (EU) 2016/679” (2016)

Finnish legislation

45. *Avioliittolaki* 13.6.1929/234

46. *Laki potilaan asemasta ja* oikeuksista 17.8.1992/785

47. *Maakaari* 12.4.1995/540

48. *Perintökaari* 5.2.1965/40

49. *Rikoslaki* 19.12.1889/39

Online sources

50. Law, Jonathan, and Elizabeth A. Martin. "instrument." A Dictionary of Law. : Oxford University Press, , 2009. Oxford Reference. Date Accessed 8 May. 2021 <https://www.oxfordreference.com/view/10.1093/acref/9780199551248.001.0001/acref-9780199551248-e-1991>.

# APPENDICES

## Appendix 4. Non-exclusive licence

**A non-exclusive licence for reproduction and publication of a graduation thesis[160]**

I, Valtteri Kauppi_____ (*author's name*)

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "Use of mind uploaded personality constructs for declaration of intent under Finnish law"
_____,
(*title of the graduation thesis*)

supervised by Thomas Hoffmann_____,
(*supervisor's name*)

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

---

_____ (date)

[60] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*