Artificial Intelligence (AI): Hidden Rules of our Society

Dr. Mika-Petri Laakkonen, University of Lapland

ABSTRACT

AI explanatory research community has become vital in AI, because explanatory analysis of AI opens up and explains the operating principles of the hidden rules of AI. Explanatory AI community expounds the construction role between human-machine in AI modelling, and breaks the gap in understanding the hidden layers of AI.

Instead of Alan Turning's (1936) well-known problem of decidability (*Entscheidungsproblem)* the review approaches the AI hidden rules of our society from them knowledge interest (*Erkenntnisinteresse)* premises coined by German social theorist Jürgen Habermas (1970; 1978).

This review illuminates how our contemporary society is constructed with AI models and hidden rules of artificial intelligence. It shall enlighten the Artificial Intelligence (AI) modelling complexity and illustrate AI hidden rules functionality in our society. AI explanatory research community opens field of wider discursion for socio-technological scholars, where aim is to understand AI role in our society.

INTRODUCTION

Critical ethics in philosophy of information technology argues that artificial intelligence has generated power inequalities and vulnerabilities in our socio-technological world. It is urgent to think and explore AI critically because of their power to define futures by locking-in certain social, political, economic and environmental relations while simultaneously locking-out other modes of relations for long periods of time.

People are entwined with technology, which collect data from us and non-human objects. This big data is calculated with super computers and AI produces new information back into the world we are living in. AI has formed invisible hidden rules inside our everyday life infrastructure and particularly into technological devices we use daily, such as mobile phones, networks, and platforms. Digital data is convenient for AI. Digital data allows AI to optimized, make decision, and predict dynamic of our society.

Artificial intelligence (AI) analyses data structures of the world. For independent skills for AI, it can be found novel applications in our everyday life. For instance, most of the data in the Internet today is in video format. Artificial Intelligence sums up the compression of YouTube videos. File compression saves video distributor costs. The viewer might see this step forward from getting the video downloaded even faster, while they download films from Netflix.

The real world and real life are messy and complicated. No one has made us a rule book on how the world works. Hidden layers of neural networks of AI struggle like humans, and learn from their mistakes based on feedback. The ideas for the latest AI programs are simple. Search algorithms for the latest applications, i.e., a calculation program, such as e.g., Monte Carlo's tree search has long been used in computers decision processes. It is a probabilistic search algorithm, that combines classic tree search heuristic learning patterns, when a tree is in a branch and have a number of options model balances exploration and exploitation bound heuristics and learns from the feedback from the twigs. Futurists think that learning probabilistic artificial intelligence such as Monte Carlo's tree search also achieves superhuman abilities, during decision process. The program is set to achieve the main goal. In pursuit of it,

it achieves qualities that are considered witty, if not intelligent, in human daily life.

AI is considered to be crucial in solving biological problems. Artificial intelligence infers how tens of thousands of proteins in the human body fold in cells and organisms. Proteins keep us alive. If you know their structure, it will help decisively in the development of medicine. Hidden Layers latent variables equal an unobserved quantity that helps to explain observed data for example covid 19 inflection (unobserved) can explain why the person has fever (observed/measurable variable). In probabilistic neural networks, hidden neurons act as latent variables, though their latent interpretation is not explicitly known.

From the perspective of data analysis, all things in the world e.g., in healthcare patients can be presented with a point in time. In addition to time, things can have millions of properties whose value can change – in variables such as health status. These variables can be presented as directions in space so that the location of the point tells the values of the variable e.g., health status. A million-dimensional health status in space cannot be imagined, but mathematically there is nothing strange about it. Once we have knowledge of many similar things, we begin to see the characters and their location and behavior at the latest. The points are not uniform in space but there are connections between the variables, and they form space characters, correlations, clusters, anomalies, signals, and regularities. The description of these characters is called modelling.

AI models are able to understand temporal attributes to detect causal relationships and coincidences (anomalies) from data. AI algorithms are reorganising society, where data, algorithms and predictions are the profound elements of organising. AI models construct hidden invisible layers of society. Hidden layers of deep-learning neural networks require explanation and argumentation capacity before they are mature enough to be used in our contemporary society.

Both, the dynamic of the world and AI data cycles are continuous, and algorithms are constantly being applied to benefit and to harm our society. In other words, the world and the data in the world are the dynamics with which algorithms operate. The world, man, data and artificial intelligence algorithms are not separated. People make moral choices in the moment they live. For instance: the goal was to launch a resistance vaccine first. However, AI is utilized in armaments and in cyber wars whether we wanted it or not. We need to find the right ethical and moral balance in the development of AI and we need to understand how continuous AI data cycles effect the dynamics of our society.

In this editorial I shall elucidate, how to model AI and how AI models and neural networks hidden layers influence and reconstruct the several areas in our society, where AI already exists. Lastly, I argue that there are profound philosophical problematization, argumentation and explication problems that need to be critically analysed before allowing the AI models and hidden layers to make independent decision in our society.

DEFINING ARTIFICIAL INTELLIGENCE (AI) TOWARDS MORE JUST, TRUST AND SUSTAINABLE SOCIETY

Artificial intelligence (AI) is the intelligence demonstrated by machines (Turning 1936, 1948, 1950, 1956), unlike the natural intelligence of humans and animals, which involves consciousness and emotionality. The beginning of the birth of intelligent machines (Turing 1936, 1950) and the decidability problem known in arithmetic (Entscheidungsproblem) help to understand how artificial intelligence has emerged in our society. The historical background of its context and motivation of research, helps us to understanding AI evolution from its birth up to now.

Instead of well-known problem of decidability this review approach the AI hidden rules of our society from the premises coined by German social theorist Jürgen Habermas (1970; 1978) to understand, how AI present knowledge and how the hidden layers of neural network reorganize our society. In his work, Habermas considered that natural sciences as empirical–analytical sciences aiming to control through prediction, while understanding and emancipation characterize humanities and social sciences.

Colloquially, the term "artificial intelligence" is often used to describe machines that mimic

"cognitive" functions that humans associate with the human mind such as "learning" and "problem solving", this is analogy how "cognitive" functions are learned by machines.

Leading AI textbooks define (see Flasiński, 2006) the field as the study of "intelligence agents": any device that perceives its environment and takes actions that maximize its chance of achieving its goals. 'Strong' AI is usually labelled as artificial general intelligence (AGI) while attempts to emulate 'natural' intelligence have been called artificial biological intelligence (ABI). The newest AI models imitates the laws of physics and physical phenomena. (de Bézenac, Pajot & Gallinari, 2019).

Current definition of AI and "intelligent agents" illustrates looming socio-technological catastrophe, and we need to consider how the current artificial general intelligence (AGI) and "intelligence agents" technological definitions and goals must be redefined, rearranged, retrofitted, or even decommissioned that support and generate more just and sustainable society.

What is moral, ethical and sustainable future for AI? How do they learn the norms and rules of our society? The most modern AI algorithms (deep learning) learn the rules of the world in a similar way as a child does (tabula rasa), without innate processing rules from the world of perception. An understanding of forms, colours, causality, even morality arises purely from experience and observation. Thus, deep learning artificial intelligence has no congenital a priori knowledge. Later in the social sciences, the idea of a "tabula rasa" (Locke, 1689) has also been understood to mean that the environment determines the construction of a human's identity, and now artificial intelligence hidden rules interfere the constructions of our contemporary society.

Artificial intelligence (AI) is already learning the rules themselves. They can start from scratch and find rules in the world. Now, artificial intelligence can operate in the everyday life, which is complicated, and not exactly known. For the first time, we have a program or system that builds its own understanding of how the world works and uses this understanding for design.

AI MODELS PRESENT KNOWLEDGE TO OUR SOCIETY

IoT machines (sensors), computers, mobile devices, video surveillance cameras (gestures, movement, behavioural), internet, social media (text, voice, pictures, video) are the data sources for AI models to learn. AI models constructs, manipulate, optimize and present processes data in visual form to our society.

But how do AI models present knowledge? In AI models, we can see four main ways, or language, to present knowledge: 1) observations, 2) rules, 3) hyper-levels, or 4) probability distributions. Observation based findings can represent knowledge in the sense that when a program predicts, it directly uses the comparative findings available to it for its work. The predictive principle, therefore, is that some units with similar characteristics are more likely to be similar in many respects. These have been called instance-based, local method and analogizations.

Logical rules are a natural way for a person to present knowledge. Rules use logic tools: and or, if z then y, etc. Rules can only be assumed to be either true or false. Or let the value vary between these. These methods have been called rule-based methods and symbolic artificial intelligence. Hyper-levels, on the other hand, use the tools of arithmetic, algebraic calculus, summation, multiplication, derivatives of various functions consisting of these. These methods have been called hyper-surface-based global methods.

Uncertainty can be described by probability distributions, coefficients, or noise. Uncertainty differs from soft truth value. In uncertainty we assume that some exact truth exists, but we are just not sure about it. The role of artificial intelligence, is to make uncertainty visible and to learn more from uncertainties to present knowledge.

AI models can be assembled like Legos. Distributions can give the probabilities of any variable for every possible value. Which can be updated as more information about the actual value becomes available. The measure and the predictor are in this case input variables and the benefit is an out-put variable. Variables between

the input and output data can be observed or latent.

Latent means that the values of the variable are not detected at any stage, but they are only assumed to exist, and their values are calculated according to the model. With the help of latent variables, one can learn from the input data, completely new variables can be added to the model, control measurement errors can be made, the model can be made more flexible, etc.

The relationship between variables means that the variables are not independent of each other. Independence means that information about the values of one variable does not provide any information about the values of the other variable. Or in other words, does not help predict the value of another variable. Connections can be set up in countless ways. The connections can have a certain direction, or it can be symmetrical. The connection may or may not be causal. Furthermore, connections can be set randomly by experts, or each variable can be linked to each other.

Many of the most advanced AI modelling techniques can be seen as variations, intermediate forms, combinations, and collections of these. For example, the neural network is only a nested regression equation, and the decision trees are just concatenated rules. Only the search process is done. From the metaphorical rule knobs, a specific program is selected, tested how well the program predicts, and the program's rule knobs are corrected until it is good enough.

The visual way of thinking about data will once again help us. Learning is about colouring the data space. Colour can mean the probability of each possible value or the predicted value itself. Simply where there are a lot of points based on the data, i.e., there are likely to be future observations of observation units. For example, in healthcare, we measure the value of one unit of a patient in one unit. We can now throw most of the space away because we know they are impossible and now we have a better prediction of the value of the second variable. A more compact way of thinking about such a forecasting situation is to present the forecast directly as a colour. Now the purpose of learning is to colour the space correctly.

MYSTERIOUS METAPHORICAL KNOBS

But what exactly are these metaphorical knobs? The adjustment knob is any interchangeable feature of the model you want to set as such. If you do not want the search process to change something in the template, it means that you restrict some templates beyond the search.

Five types of adjustment knobs can be distinguished. 1) We can try completely different models in the search. Such as those based on rules, hybrid levels, or observations. 2) In the absence of a single modelling method, we can experiment with different model structures, such as the numbers of latent variables (hidden layers). 3) Models usually have many other settings in addition to the structure that define different properties of the model, hybrid parameters. 4) The models have many different numbers that represent the actual knowledge of the model, parameters such as coefficients and constants. 5) We can include in the search not only the AI model but also the performance value and the optimization algorithm itself. All we must do is present these programs as models and go through all the previous four points.

Why do we make assumptions and delimitations at all? Because whenever alternatives are added to a search, the number of alternatives to try grow explosively and the search becomes more difficult. If we can narrow down our search based on our own knowledge, it's often worth doing.

AI HIDDEN LAYERS LEARNS WITH OR WITHOUT THE HUMAN

Human is involved in building AI models at the beginning and end by adjusting the deep learning algorithm model to match reality. Human makes the first decision when setting a basic goal for AI based on his/her view of knowledge, reality, or values. Through critical suspicion, man attempts to bring out ambiguous expressions or question assumptions related to the subject perception. Human decides the number of the hidden layers of deep-learning neural networks. By adjusting the deep learning algorithm model to match reality, human and artificial intelligence algorithms share the

explication decision-making of the solved problem.

Humans are not logical thinkers. Artificial intelligence models human brain functions and the hidden layers of neural networks can be equated to human memory. Humans manipulate AI models with the mathematical activation functions (e.g. sigmoid function, gradient descent) and change the gradients and weights of neurons in the hidden layers and find down the levels of abstraction. AI models are not purely based on logic, models are influenced by humans impulsive, manipulative and intentional actions. The intellectual rationality of artificial intelligence vs. humans' irrational functions and external random stimuluses are constantly shaping both AI models' and society.

Deep learning algorithms are constructed with logical and precise conceptualization methods and techniques. At the beginning of deep learning algorithms, new outputs are sought, the stages of new output postures are problematization, expulsion, and argumentation. At the input and output ends of the problematization, explication, and argumentation control knobs is the human, and between these are the hidden layers of deep-learning neural networks that make decisions based on human regulation.

HOW DO AI LEARN?

Learning methods once again have a myriad of different properties. There are four main types of guidance. Supervised learning means that the program is given complete example observations, i.e., model performances. For example, we could give the program radiological images of the names and locations of the medical findings shown in the images. Semi-supervised learning means that in addition to model performance, the program is also given observations that do not have ready-made solutions. So, in the example, these would be just radiological images.

Reinforcement learning can be thought of as a slightly weaker indirect guidance. Reinforcement learning is about giving the program only good or bad feedback about its operation instead of examples. The program will be rewarded and punished. In principle, the program learns to predict the feedback it receives based on

its observations and actions and, of course, the desired action. In a radiological example, reinforcement-learning artificial intelligence could, for example, prescribe treatments based on a pattern to obtain feedback on the well-being of its patients.

Unsupervised learning means that no model performance or feedback is given, only observations. The program can still learn the characters formed by the observations and the new variables. Although the variables do not have human-given names (labels). In practice, the program can learn the same task again. In the example, the program could learn from the images a new variable with values ranging from 0 to 1. If we look at this visually, we might find that the variable does reflect the probability of a tumour.

The result of AI models forms an association, or even assimilations in our society. Humans understand the decisions of AI models in delay, when the consequence of the AI action has been seen. At the same time, by observing the consequences associated with the rules, deep-learning algorithms learn. Is there the adversarial neural network, kind of competition on-going between human vs. hidden layers of complicated AI models? Only an explanatory analysis of deep-learning neural networks enlightens us about what is the process of problematisation, reasoning and argumentation that lay behind the decisions AI models.

AI EXPLANATORY (ERKENNTNISINTERESSE) PROBLEM

AI explanatory research provides novel field of research for AI socio-technological scholars, where aim is to understand AI (deep learning algorithms) role in our society. AI explanatory *(Erkenntnisinteresse)* research community has become vital in AI. The explanatory analysis opens and explains the operating principles of the deep-learning algorithm, divides the construction role between human-machine in AI modelling, and breaks the gap in understanding the hidden layers of deep-learning algorithms.

In democratic societies decisions must be reasoned. In other words, present arguments in support of or against decision-making. Relevant questions in this context include: Is

the proposed solution consistent, clear, justified, and compatible within and between cases? Is the AI solution the answer to the problem posed? Is it able to deal with other similar problems in an interesting way? The arguments put forward, in turn, can lead to the emergence of new philosophical problems, attempts to solve them, etc., which give the philosophical debate the same internal continuity as is observed in the development of other disciplines.

The decisions must be reasoned, and answers must be worded in such a way that they are clear and comprehensible. This philosophical problem is to formulate a response attempt. Formulating a response to a decision includes refining questions, analysing and defining concepts, presenting new ideas, and formulating philosophical arguments and perceptions. In some cases, the result may be the elimination of the problem by displaying it as a false problem due to unclear use of the language.

The reasoning of artificial intelligence arises in a computer network that is roughly reminiscent of the human brain. Networks have been tuned to different layers in artificial intelligence. The layers are often so complex that even program writers can't predict how well artificial intelligence works and what it can do. Only experiments show its power. The neural network uses the feedback it receives to deduce how to implement the strategy to remodify the AI model.

The consequences of deep learning rules change the system but the change of system is understood only when the consequence of the action can be seen. By observing the consequences associated with the rules, deep-learning algorithms learn. The explanatory analysis of deep-learning neural networks tells what the processes of reason are, the focus of AI explanatory analysis here are: problematization, explication and argumentation in the context of AI modelling and neural networks hidden layers.

Artificial intelligence in Mathematics are "truths" that make assumptions and draw conclusions based on them. AI algorithms imitate the process of reality and automates functions. Mathematics can be thought of as a language modelled by quantity, form, change, uncertainty, and the relationships between them. AI models in Mathematics are "truths"

that make assumptions and draw conclusions based on data, but do they represent the truth of reality of our societies? Do they have enough the explanatory power to be trusted, when we make critical decisions in our society?

The hidden layers of deep learning algorithms create rules and organise our society. However, they do not justify the decisions they make based on their activities and options. There are immense explanatory problems in AI models and hidden-layer architecture on deep-learning neural networks. AI models do not justify decisions made based on their functions and alternatives. Only explanatory analysis of deep-learning neural networks tells us what the processes of reason are.

AI HIDDEN LAYERS (LATEO "LIE HIDDEN") OF OUR SOCIETY

Hidden rules of our society are drawn from digit-virtual reality. AI models counteracts with digit-virtual reality. The world is becoming more and more entwined with virtual hidden data. For instance, we seldom use the physical cash anymore as credit cards are disappearing and instead of cash, we have digital wallet applications (such as Google Pay, WeChat etc.). The monetary systems have reconstructed governance model, where states and banks don't have monetary power anymore. Monetary exchange is directly and globally operated with and within citizens, where hidden layers of AI secure (hash functions), optimize and predict the human-to-human transactions.

Artificial intelligence models human brain functions and the hidden layers of neural networks can be equated to human memory. A deep neural network (DNN) has multiple hidden layers between the input and output layers. What are the hidden layers of neural networks? What is human's role in the complex hidden layered of artificial intelligence neural networks? First question is easy to answer, but the second question has moral aspect in it. Similar way as humans can manipulate their own "consciousness" with memory, the hidden layers of neural network can be manipulated with mathematical models. The hidden layers are between input and output of data from our society. The hidden layers of neural network

can, like us humans are able to, search, process, manage, form, build, retrieve, restore, predict, optimize, decide and control part of information in our society.

In deep neural network latent variables are variables that are not directly observed but are rather inferred (through mathematical models such as activation functions or iterative optimization algorithms such as gradient descent) from other variables that are observed (directly measured). Mathematical models that aim to explain observed variables in terms of latent variables are called latent variable models. Latent variable models are used in many areas of our society, including psychology, demography, economics, engineering, medicine, management and now ever more often in social science. In neural networks, a hidden layer is located between the input and output of the algorithm, in which the function applies weights to the inputs and directs them through an activation function as the output. In short, the hidden layers perform nonlinear transformations of the inputs entered into the network. Hidden layers vary depending on the function of the neural network, and similarly, the layers may vary depending on their associated weights.

Latent variables may correspond to aspects of physical reality. These could in principle be measured but may not be for practical reasons. In this situation, the term hidden variables are commonly used (reflecting the fact that the variables are meaningful, but not observable). Other latent variables correspond to abstract concepts, like categories, behavioural or mental states, or data structures. The terms hypothetical variables or hypothetical constructs may be used in these situations.

The use of latent variables can serve to reduce the dimensionality of data. Many observable variables can be aggregated in a model to represent an underlying concept, making it easier to understand the data. In this sense, they serve a function like that of scientific theories. At the same time, latent variables link observable "sub-symbolic" data in the real world.

CONCLUSION

AI models and hidden layers of deep-learning algorithms (Black-box AI) reorganize our society, and we are ignorant to its future consistencies. The society is in the middle of the complex and layered environment of artificial intelligence neural networks. The coincidence events in the time series sequences of AI models' shape the reality and, contingency consequences of deep learning neural networks organise our society. In other words, coincidences and unexpected events occur simultaneously in both artificial neural networks and the real world. Coincidence is present in both habitats, artificial intelligence neural networks, and society. Which produces a coincidence when the society surrounding the neural networks are hidden layers?

AI models are complicated, with multiple and multilayer parameters in the relational relationships in our society. The hidden layers in AI models are key structural and learning elements to comprehend the process of input and output of our society. AI hidden layers are invisible, but they continuedly influence our every day of life, what we see, what we do and how we behave. Thus, AI's hidden layer functions in our socio-technological world need to be understood in a deeper level to build more just, trust and sustainable society. AI is deeply entangled with and withing our society.

AI is not apart from our society and its dynamics cannot be understood purely by mathematical premises. How to study this complicated dynamic relationship, where the AI and society is seen as a one phenomenon? Even for experts the complexity of AI models is hard to understand as phenomena of their own. In close mathematical frame of "truth" without the society impact, it is much easier to answer the questions how AI models present knowledge, how AI learn, what are the activation functions and what are hidden layers of AI models. But much more demanding task is to answer why AI models function as they do in our society. Also, it is easy to see the influences AI have in our society afterwards. But how to predict AI behavioural coincidences in our society and how AI could help us to predict catastrophes and build safer, resilience and sustainable society.

Nevertheless, prediction of future is hard for humans and AI models, because prediction is often made relaying on historical data. For instance, Markov chain AI model is based on the idea that what happened in the past, is going to happen also in the future, and the future is predicted based on the present moment. Thus, our society cannot forget the past if we want to know where we are now.

Current AI research is dominantly natural science oriented. Natural science accepts only mathematical "knowledge" with empirical evidence as the "truth". Experimental knowledge has no place in AI. Even in arising field of artificial emotional intelligence (AEI) computers read emotions by analysing data, including facial expressions, gestures, tone of voice, force of keystrokes etc. to determine human's emotional stages and then reach to it.

We need to understand how the current artificial intelligence algorithms formulate, construct, predict, decide, manipulate, control, and manage our society and how to create future AI models that support and generate more just,

trust and sustainable future. There does not exist the common AI interdisciplinary language, which could provide us the opportunity to form a common understanding of how AI models and deep learning neural network hidden layers work in our society. In theory of social science, the society constructs the identity of a human. What is constructing identity of human being in contemporary society? When AI models and hidden layers dominate, manipulate and re-organized our everyday of life. AI models do not only provide recommendations but also make decisions based on our behavioural and organize the vital daily operations in our life.

To conclude, AI hidden rules in our digit-virtual world has changed the dynamics of our society. Intelligent computers and machines justify the decisions purely based on data analysis, without historical experiences, emotions, and consciousness of the world, and without understanding the social relationship between humans. AI models have taken permanent invisible position to interfere in society-human-society interaction.

REFERENCES

de Bézenac E, Pajot, A. & Gallinari, P. (2019). Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge. *Journal of Statistics.*

Flasiński, Mariusz (2016). *Introduction to Artificial Intelligence*, Springer

Habermas J. (1970) Knowledge and Interest. In. Emmet D., MacIntyre A. (eds) *Sociological Theory and Philosophical Analysis.* London: Palgrave Macmillan.

Habermas, J. (1978). *Knowledge and Human Interests* (2nd ed.). London: Heinemann.

Locke, J. (1689). *An Essay Concerning Human Understanding II.i*, edited by K.P. Winkler. Indianapolis: Hackett Publishing Company.

Turning, A. M. (1936). *On Computable Numbers, with an Application to the Entscheidungsproblem.* Proceedings of the London Mathematical Society, Serie 2, 42 (1936–37), pp. 230–265

Turning, A. M. (1948). *Intelligence machinery, report for National Physical Laboratory*, in Machine Intelligence 7, eds. B. Meltzer and D. Michie

Turning, A. M. (1950) Computing machinery and intelligence, Mind 49: pp 433-460

Turning, A. M. (1956) *Can a Machine Think*, in volume 4 of The World of Mathematics, ed. James R. Newman, pp 2099–2123, Simon & Schuster