A Lacanian perspective on artificial intelligence

Purnima Kamath

purnima.kamath@u.nus.edu

Communications and New Media Department, National University of Singapore

November 21, 2020

Abstract

The contemporary research landscape in the domain of artificial intelligence is based on a cybernetic focused, feedback and feedforward approach. Such approaches mandate these algorithms to be trained on large datasets, hence limiting any innovation in this area by factors like size of the datasets, computational power available and speed of hardware. There is a need for novel approaches to be tried and experimented with in this regard and thus the primary aim of this paper to is to outline some initial ideas to augment this cybernetics approach by applying some aspects of Lacanian psychoanalysis while designing AI algorithms. My goal is to characterize an AI machine within the three Lacanian registers, explore using the concept of symbolic matrix while designing an AI model and illustrate a conscious and an unconscious within such AI machines using language-based approaches. As a secondary aim of this paper, I will explore using the mathematical concept of factor graphs to augment Lacan's conceptual model of the unconscious. The ideas outlined in this paper are preliminary thoughts on the intersection of AI and psychoanalysis which will be used for further in-depth expansion or exploration in the future.

Keywords: Lacan, artificial intelligence, psychoanalysis

A Lacanian perspective on artificial intelligence

Artificial intelligence (AI) is a set of algorithms which assist machines in simulating aspects of human intelligence like problem solving and creativity. The term, artificial intelligence, was originally coined by John McCarthy on the opinion that '... every aspect of learning or any other feature of intelligence can be in principle so precisely described that a machine can be made to simulate it'. Classical AI algorithms or symbolic AI of the last century focused on simulating this intelligence using a large compilation of language-based rules, which a machine could read and execute in order to solve relevant problems. Symbolic AI had its limitations in its ability to scale up to larger problems from smaller hand-crafted problems. Contemporary AI algorithms (also known as sub-symbolic AI) are more akin to mathematical regression models built by economists, which run on supercomputing hardware and consume a significantly large corpus of text, image or audio datasets to train and build this intelligence. Such sub-symbolic AI machines function like black boxes i.e. these algorithms do not provide a causal, human-understandable explanation for the outputs they generate (Mitchell, 2019).

Most of the progress in AI research, can be attributed to building large datasets and models and using even larger super-computers to train these models. With the current industry hypothesis that the exponential speed-up of computing hardware that we saw in the late last century is slowing down, there is a need to rethink our strategy towards designing and building future AI algorithms and programs. My intent with this paper is to explore some preliminary thoughts on language-based approaches to building the AIs unconscious, draw parallels with some of the current thinking around consciousness in AI with Lacan's notion of conscious ego and explore mathematical factor graphs to augment the Lacan's cybernetic model of the unconscious. I will begin this analysis first by placing the AI machine within the three registers.

Why a Lacanian perspective?

The sub-symbolic AI algorithms heavily rely on statistical and mathematical approaches to build AI models. Such approaches build or create the model in a high dimensional space which make causal analysis and visualization of the model difficult. Hence, such AI models are termed as black boxes. Updates or modifications to such models rely on tuning their 'hyperparameters' (trained or learned meta-parameters that govern the characteristics of that high dimensional space) which is usually difficult and is at times compared to alchemy. Given this background, it is increasingly important to be able to tune or update these models using parameters based on language or something similar to it. Language is the basis for Lacan's psychoanalysis and his framework which explains the concept of ego, subject and unconscious using structures in language will provide a good foundation for our analysis in AI. Additionally, in one of his later seminars, Lacan adds that conjectural sciences (a term he uses for human sciences or philosophical sciences) and exact sciences (a term he uses for mathematics, biology etc.) are related and inseparable from each other (Seminar XXIII) (Lacan, 1991, p. 296). With this understanding, Lacan uses analogies from such exact sciences to explain complex phenomena from psychoanalysis. Inspired by this approach, I intend to use ideas from Lacan’s psychoanalysis and apply them to AI.

AI and Lacan's three registers

Before we delve into the conscious and unconscious of AI, it is imperative that we position the artificially intelligent machine in relation to the three Lacanian registers - the Imaginary, the Symbolic and the Real. There is also a need to outline our position, the researcher/programmer, in relationship with the AI.

Lacan places the ego or consciousness in the realm of the Imaginary register (Homer, 2004). During the mirror stage phase of a child's cognitive development, the conscious mind is formed, and the subject is separated from the ego. Lacan treats this conscious ego as an object, something which is free of agency and the subject as a common mouthpiece for both conscious ego as well as the unconscious. This ego-as-object (Johnston, 2018) and its separation from the speaking subject makes a Lacanian reading of AI more lucrative to our analysis than any other non-cybernetic theoretical approaches. It helps us identify and work with consciousness in AI, without delving into the emotional and subjective aspects of a machine i.e. without anthropomorphizing it. In the subsequent sections of this paper, we will see how this conscious ego, in an artificially intelligent system, is a space where traces first appear, which may or may not be removed after their further persistence into the unconscious.

The Lacanian Symbolic or the symbolic order encompasses all human-made customs, rules, laws, languages etc., a pre-existing database of knowledge, a 'treasury of the signifier' (Hook, 2017) that human beings are born into and would eventually expand by contributing to it. Lacan treats this order, possibly for ease of reference given its variability and expansive nature, as a singular big Other. This symbolic order or big Other is structured like language or is using language. This pre-existing and ever expanding big Other forms a basis for his theory on the unconscious within human beings, making it the core substance of our unconscious. This effect mandates our unconscious to be structured similarly (like language) which forms a basis for his characterization of the unconscious as being a 'discourse of the big Other'. By 'unconscious is structured like language' he intends to model the unconscious using structures like patterns or relationships between words, semiotics or grammar within a language. Within an AI machine the symbolic order or the big Other can be equated to the corpus of text, images or audio datasets the algorithm is being trained upon. The AI's unconscious is the mathematical model generated as a result of this training. In material terms, the training program tries to find patterns and structures within the corpus and builds a binary (composed of 0's and 1's) model within the machine. This model is the AI's unconscious and we need a subject (The AI program which runs this model) to interpret it. A human being simply viewing the model, without the

AI program, on the machine will only see incomprehensible 0's and 1's. We need this subject to be the model's 'mouthpiece' (Fink, 1995) to communicate with the external world. This AI unconscious is built on this big Other, this big corpus of data comprising the symbolic order, which we the researcher/programmer feed into the system while training it. The AI is created (or more anthropomorphically, it is born) into this symbolic order. We the researcher/programmer control this data and knowledge and thus control the behavior or the output of this AI machine. In the Weapons of Math Destruction, O'Neil postulates that these mathematical models are not just neutral formulae but have human opinions embedded in them (Possati, 2020). Thus, in Zizekian terms, we are the 'Other of the Other' for this AI machine. A secret, invisible agent, pulling the strings within its learnt references from the symbolic order.

Lacan theorizes that everything we know about ourselves and the environment we live in, that which is structured in language, is a reified form of, and is a reduction of, the Real. This reification and reduction take the form of the symbolic order and the realm of the Real is something that existed before these signifiers came to existence. It is a surplus that cannot be "said", that which resists symbolic integration (Dean, 2006). This Real can be formulated in terms of mathematical set theory as a superset or universe of all sets of signifiers, with something extra - an additional transcendental excess. Lacan formulates this excess as something that is unfathomable and incomprehensible to our human minds and defines it terms of a negation or absence than presence. The parts of the Real which are not signified, the ones which are beyond our understanding, lie in wait to be discovered, to be signified in the Symbolic. This un-signified part of the Real also manifests itself in the unconscious as the part of our being which cannot be signified. An AI's unconscious, that which is trained or learned from the symbolic order, and that which functions in the Real because of its high dimensionality, can sample this negative space between symbols to reveal or reify that which has not been yet signified. An example of this reification can be seen in some of the earlier breakthroughs in deep learning, especially in area of image recognition. Scientists trained a neural network to recognize and classify images, and in an effort to understand the causality behind each of the neurons, built a reverse network which interpolates this negative space in the model to produce dream like structures as shown in figure 1.

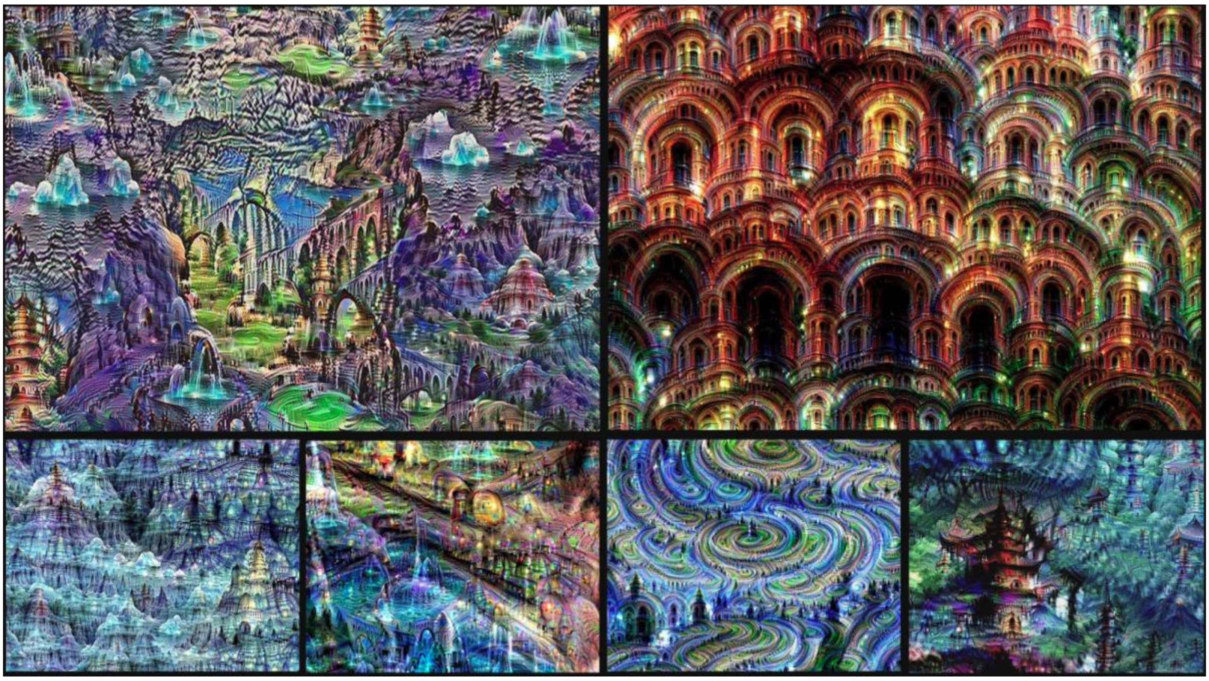

Figure 1: Image from Google AI's blog on 'Inceptionism' titled 'Neural net dreams' (Mordvintsev et al., 2015). A neural network trained for image classification to recognize places is run in reverse with emphasis on few outer layers of the neurons to produce dream like images

Unconscious in AI

Most current AI machines implement computations that reflect unconscious processing (Dehaene et al., 2017). An AI's model, its unconscious, is a high dimensional mathematical space, which at often times is difficult to visualize for us as human beings. As seen in the previous section, this model is backed by and created based on the symbolic order, but it functions in the Real and manifests in the Imaginary, implying that the model's outputs need not limit nor conform to the symbolic representations within its learning corpus. Lacan positions exact sciences (like physics and mathematics) very close to the Real (Seminar XXIII, Book II) (Lacan, 1991, p. 297). For example, he elucidates the concept of time, as being based on the rotation of Earth around the Sun and is signified in the Symbolic by its units of measure. It commonly exists for everyone, even when we are consciously not aware of it. This measure of time, which is now signified, is sampled from the Real. This notion of sampling from the Real, into the Symbolic, that which did not exist before, provides an excellent basis for AI to exhibit creativity or build new symbols or images, by sampling its model in the Real.

We share a common symbolic order with the AI machines we build as programmers/researchers. This symbolic order is trans-subjective (Hook, 2017) and as discussed before, we are the Other of the AI's Other. Our human desire for identification, for ourselves and the world around us, is transferred through our coding of the symbolic order into the AI machine (Possati, 2020). Through our worldview of this symbolic order, our misinterpretations, biases, misjudgments get transferred into the AI we build, thus incorporating our lack (because we don’t know what we don’t know) and desires (in that we unconsciously are driven towards it) into the system. Computers cannot experience jouissance in the literal human subjective sense, their lack is just an existence of 0's instead of 1's (Moncayo, 2018). In Lacanian terms, lack is the result of an unrealized Real, something that surpasses representation within the Symbolic. But for an AI system, this lack is the result of the human researchers/programmers ineffective or insufficient coding of the symbolic into that AI. This manifests as, what we can term as AI's transgression. Dean expounds on the Zizekian notion of using transgression as a vehicle for jouissance (Dean, 2006). This transgression manifested in AI, for example mis-recognizing individuals in a face recognition system, represents our (the programmers) transgression and hence our jouissance. Thus, our lack, desire and jouissance manifests as lack, desire and jouissance of the AI we build.

Symbolic matrix and the AI unconscious

In the seminar on 'The Purloined Letter', Lacan demonstrated how the unconscious, instead of the speaking subject, does the thinking and plays the game of chance according to given combinatory rules. Inspired by Cybernetics theory, he built a model for his language-based unconscious, on the basis that language can be defined as a programmable system of signs specifically 1's & 0's (Liu, 2010). He created a set of rules or laws, encoded into a symbolic matrix, that are not inherent in our existing reality. He hypothesized, and provided a logical proof, that such rules or laws and their higher order derivatives lead to complexity which is indicative of the structure of the unconscious. This complexity arises even when the starting point is as simple as rules outlined in language. He analogized generation or random selection of unconscious thought to the generation of numerical sequences using + and - symbols and some rules. These rules, both simple and complex, help us in understanding how the unconscious operates by (Fink, 1995) -

1. Assisting in selecting the next element (or thought) in the sequence based on the numerical probability of such element formulated by the rules and
2. Indirectly encoding of memory within the temporal sequence of selected elements (i.e. the ability to find not just the next in the sequence, but ability to derive the previous elements from the rules as well)

This symbolic matrix is not very dissimilar from the classical AI algorithms of the 1960's-70's where intelligence was purported to be hand-crafted into language-based rules. Such algorithms are not in vogue today because of their inability to learn rules or patterns or representation from data and the need to manually craft these rules.

One of the goals of this paper is to outline the first steps to theoretically explore using this symbolic matrix in conjunction with the sub-symbolic AI techniques in use today. This theoretical inclusion of symbolic matrix within a feed-forward network is done with an intent to be able to develop more parameterized control of an AI’s output using its inputs. One such example is theoretically developed for this paper as indicated in figure. 2 below.

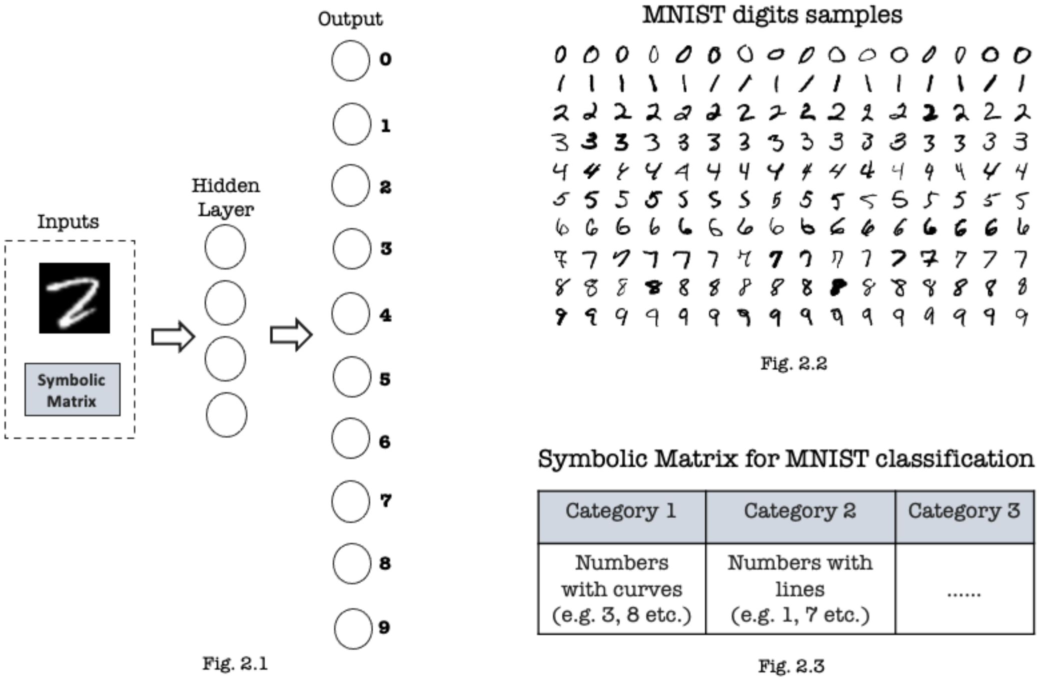

Fig. 2.1

Fig. 2.2

Category 1	Category 2	Category 3
Numbers with curves (e.g. 3, 8 etc.)	Numbers with lines (e.g. 1, 7 etc.)	......

Fig. 2.3

Figure 2: MNIST handwritten digits are commonly used for training image-based machine learning systems. Figure 2.1 shows a schematic of the inclusion of a symbolic matrix to the neural network. Figure 2.2 is a view of some of the samples from the MNIST DB (from Wikipedia) and Figure 2.3 shows a sample symbolic matrix categorizing the data using some visual features denoted in language.

Let's suppose we have a neural network which recognizes handwritten digits from 0 - 9. One example of the symbolic matrix for such a network would be to categorize the digits based on the shape, for instance if they are built using lines or curves. I postulate that the inclusion of the symbolic matrix as parameters to such a feed-forward network will help improve not just the accuracy of the trained network, but also provide tunable parameters for such tasks. For e.g. let’s say this trained network is tasked with figuring out which digits have been incorrectly classified as 7 but were actually 2 (in figure 2.1 the input 2 can be easily misrecognized by the algorithm as 7). In such a case, the 'Number with curves' category could be set to true to indicate our datasets affinity towards the number 2.

Note that the MNIST image recognition problem suggested above is a trivial task within the field of machine learning and current techniques provide great accuracy. This example was chosen as a hypothetical and is a simplistic version of the parameter control and rules that can be encoded into a symbolic matrix for a neural network. Any future work in this regard would create symbolic matrix encodings for more complex tasks in the realm of sequence generation for natural language or audio.

Factor graphs as an alternative to the model of unconscious

In this section of the paper, I will look at augmenting Lacan's hand-crafted model of the unconscious with factor graphs. Lacan's model, which is inclusive of both the symbolic matrix and higher order graphs, demonstrates randomness and memory associated with unconscious thought (Fink, 1995). By randomness, he wanted to indicate the variable probability of unconscious thought selection which he demonstrates using + and – signs and manually deriving higher order matrices using grouping and substitution. I postulate that Lacan's model of the unconscious can use mathematical factor graphs, focusing on the flow of messages (thoughts) and calculating probability of those messages at various points in the graph using sum product algorithms, thus alleviating us from the iterative complexity within Lacan's handcrafted higher order derivations.

Factor graphs, commonly used in neural networks, are bipartite graphs which are used as probabilistic models with some in-built constraints. These graphs are called bipartite as they consist of two parts or nodes - variables and factors. A simple graph schematic is shown in figure 3 and figure 4.

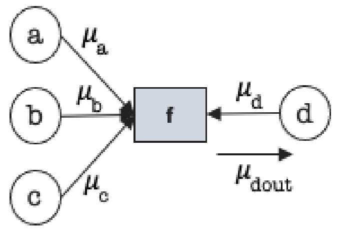

Fig. 3

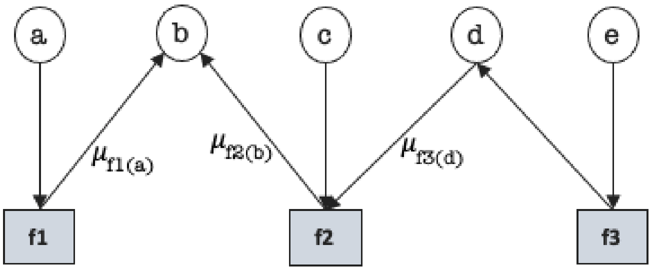

Fig. 4

Figure 3: A single factor f and four variables a, b, c and d are shown. µ is the message between these variables and the factor. µdout specifically indicates that the probability of the message is being calculated at node d. Figure. 4: A sample factor graph with 5 variables and 3 factors. The intent is to find the probability of the message at node b. μf1(a) and μf2(b) are the message μ having 'flown' through other variables and factors towards node b.

The nodes in the schematics in circles are called variables and the nodes in rectangles are called factors. Messages (μ) can travel between variables and factors and probability of a message at each variable node can be calculated based on the direction of the messages. For instance, the message at node d (figure 1) is as follows -

$$\mu dout = \sum_{a,\, b, c} f(a, b, c, d) \, * \, \mu a \, * \, \mu b \, * \, \mu c$$

Where $\mu a, \mu b$ and μc are messages from each of the variable nodes and $f(a, b, c, d)$ is the overall multiplying factor. Based on this, consider a factor graph like figure 4, the probability of the message at observed node b (for example), based on the principles of sum product or belief propagation algorithms (Loeliger, 2004) will be as follows -

$$Pr\ (b) = \sum_{a,\, c, d, e} \mu f1(b\ ) \, * \, \mu f2(b)$$

Hence, if we have a graph like representation of the unconscious, then we can find the probability of message at a particular node based on the 'flow' of the messages. The higher the probability at a particular node, the better its chances of selection within a sequence. The lower its probability, the lesser its chances. Lacan defines this as ‘impossibility’ of an element’s occurrence within a sequence. This sum product algorithm helps govern the probability or belief of message at a particular node and hence can prove to be an extension to the higher order matrices devised by Lacan to model the unconscious.

Consciousness in AI

In the previous sections we saw that the AI's model is its unconscious and because of its high dimensionality it functions in the Real. This unconscious has a low-level representation i.e. it is structured in binary 0's and 1's (human language is considered high level representation by computers). In this section, I will look at some recent research both in neuroscience and AI on consciousness and draw parallels between them and the concept of consciousness in Lacan.

Cybernetically speaking, consciousness in neuroscience, in AI or in Lacan can be modeled as being of low dimensionality and having a high-level representation i.e. it can be represented using language (as compared to high dimensionality and low-level representation of the unconscious). As per cognitive neuroscience, conscious processes are generally slow and process only few elements at a time. They use the concept of attention, which encompasses selecting elements from a larger pool of the unconscious and placing it in the conscious realm, as being core to the conscious process. Conscious processes are slow because they result in the elements being broadcasted to other areas of the brain or machine, which can be used for further conscious or unconscious processing or even mechanical action (e.g. walking).

According to Lacan, the unconscious and the conscious in human beings work in tandem (Fink, 1995) and use the common speaking subject. Also, traces are first created in the conscious part of our cortex are later persisted into the unconscious part. The conscious traces may or may not be deleted subsequent to this persistence (Seminar XXIV Book II) (Lacan, 1991, p. 322). Also, the Lacanian idea of ego-as-object which separates conscious ego from the subject, helps us explore a model of consciousness within AI without anthropomorphizing it.

Applying the above ideas from the realm of cognitive neuroscience and Lacan, some researchers have developed a model for AI (though focusing more on the hard sciences than philosophy), which not just includes an unconscious, but also a conscious mechanism which is based on a high level representation (or language). An attention mechanism is developed which selects traces from the unconscious and places it in the conscious for further processing. This conscious part of the model will be trained and executed in parallel with the unconscious mechanism (Bengio, 2019) which is similar to Lacan's model of consciousness.

There are multiple advantages to this addition of conscious processing to an AI algorithm. Firstly, the energy needed to train large networks greatly reduces because searches within this low dimensional conscious space will be faster as compared to searches performed on the high dimensional unconscious. Secondly, this conscious mechanism will provide researchers an ability to interact with the AI using language and thus providing better parameter control and explainable outputs.

Summary

My primary goal with this paper was to outline some initial ideas on applying Lacan's psychoanalysis to contemporary AI research. In order to do this, I first oriented the concept of AI within the three registers - the Imaginary, the Symbolic and the Real. I outlined how treating the conscious ego-as-object which is separate from the Subject assists us in applying Lacan's ideas to the notion of conscious in AI. Also, the symbolic order is the basis for the model or the unconscious in AI and us the programmers/researchers become the 'Other of the Other' while building and training the program. I also explored the concept of AI's creativity by the way of sampling the Real and bringing into the symbolic images that did not exist before. On the topic of AI's unconscious, I looked at the idea of our lack, desires and jouissance transferring into the AI we build and also applying Lacan's basic building blocks of the unconscious, namely the symbolic matrix to some theoretical examples in AI. In the latter part of the paper I drew parallels between some contemporary AI research and Lacan's concept of conscious. As a secondary goal of this paper, I tried to extend the mathematical concept of factor graphs and probabilities to Lacan's model of unconscious. With these initial few thoughts at the intersection of artificial intelligence and Lacanian psychoanalysis, I hope inject ideas from different domains and not just pure mathematics or hard sciences into my research in the future.

References

Bengio, Y. (2019). The Consciousness Prior. *ArXiv:1709.08568 [Cs, Stat]*. http://arxiv.org/abs/1709.08568

Dean, J. (2006). *Zizek's Politics*. https://doi.org/10.4324/9780203956618

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science, 358*(6362), 486–492. https://doi.org/10.1126/science.aan8871

Fink, B. (1995). *The Lacanian Subject: Between Language and Jouissance*. Princeton University Press. http://www.jstor.org/stable/j.ctt1jktrqm

Homer, S. (2004). *Jacques Lacan*. Routledge. https://doi.org/10.4324/9780203347232

Hook, D. (2017). Six Moments in Lacan: Communication and Identification in Psychology and Psychoanalysis. In *Six Moments in Lacan: Communication and Identification in Psychology and Psychoanalysis* (p. 208). https://doi.org/10.4324/9781315452616

Johnston, A. (2018). Jacques Lacan. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/lacan/

Lacan, J. (1991). *The Seminar of Jacques Lacan, Book II—The Ego in Freud's Theory and in the Technique of Psychoanalysis 1954-1955. Edited by Jacques-Alain Miller. Translated by Sylvana Tomaselli.* W. W. Norton & Company.

Liu, L. H. (2010). The Cybernetic Unconscious: Rethinking Lacan, Poe, and French Theory. *Critical Inquiry, 36*(2), 288–320. https://doi.org/10.1086/648527

Loeliger, H. (2004). An Introduction to factor graphs. *IEEE Signal Processing Magazine, 21*(1), 28–41. https://doi.org/10.1109/MSP.2004.1267047

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Penguin Books.

Moncayo, R. (2018). The Symbolic in the Early Lacan as a Cybernetic Machine, as Automaton and Tyché, and the Question of the Real. In R. Moncayo (Ed.), *Knowing, Not-Knowing,* *and Jouissance: Levels, Symbols, and Codes of Experience in Psychoanalysis* (pp. 99–126). Springer International Publishing. https://doi.org/10.1007/978-3-319-94003-8_6

Mordvintsev, A., Olah, C., & Tyka, M. (2015, June 17). Inceptionism: Going Deeper into Neural Networks. *Google AI Blog*. http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Possati, L. M. (2020). Algorithmic unconscious: Why psychoanalysis helps in understanding AI. *Palgrave Communications*, *6*(1), 1–13. https://doi.org/10.1057/s41599-020-0445-0