A Machine Consciousness architecture based on Deep Learning and Gaussian Processes

Eduardo C. Garrido Merchán[1] and Martín Molina[2]

[1] Universidad Autónoma de Madrid, Madrid, Spain eduardo.garrido@uam.es
[2] Universidad Politécnica de Madrid, Madrid, Spain martin.molina@upm.es

Abstract. Recent developments in machine learning have pushed the tasks that machines can do outside the boundaries of what was thought to be possible years ago. Methodologies such as deep learning or generative models have achieved complex tasks such as generating art pictures or literature automatically. On the other hand, symbolic resources have also been developed further and behave well in problems such as the ones proposed by common sense reasoning. Machine Consciousness is a field that has been deeply studied and several theories based in the functionalism philosophical theory like the global workspace theory or information integration have been proposed that try to explain the ariseness of consciousness in machines. In this work, we propose an architecture that may arise consciousness in a machine based in the global workspace theory and in the assumption that consciousness appear in machines that has cognitive processes and exhibit conscious behaviour. This architecture is based in processes that use the recent developments in artificial intelligence models which output are these correlated activities. For every one of the modules of this architecture, we provide detailed explanations of the models involved and how they communicate with each other to create the cognitive architecture.

Keywords: Machine Consciousness · Machine Learning · Deep Learning · Gaussian Processes · Artificial Intelligence

1 Introduction

Several reviews have been written about machine consciousness [24] [46] [25] that try to sum up all the ideas that literature has proposed about the potential arisal of consciousness in machines [14]. These ideas come from different areas such as artificial intelligence [13], neuroscience [42] or philosophy [49]. Although consciousness can not be measured directly, there exist approaches that have provided potential measures of consciousness in machines [4] [45].

Although the field generates controversy [16] as it lies in the margin of the scientific method it has recently attracted the attention of eminencies of computer science such as Yoshua Bengio, who has provided an approach for how machine consciousness may arise with deep learning [10]. As deep learning [37] have generated machines that implement attention mechanisms [29], a new focus have emerged with the field of machine consciousness based in the astonishing hypothesis [16] that our intelligence and consciousness may arise from very simple principles.

All of the computational approaches for machine consciousness are based in the functionalism theory of consciousness [46]. This theory claims that while mental states correspond to brain states, they are mental states due to their functionality, not due to their physical composition. Hence, consciousness may appear in machines that implement correlated activities that humans do when they are conscious.

Throughout the recent years, there has been amazing advances in the artificial intelligence and machine learning community [41] that does not only include deep learning models. In the machine consciousness literature, it has been hypothesized that consciousness, or phenomenal states [38], may arise from machines that are able to perform tasks that humans are able to do when they are conscious [17] [25]. This is based in the hypothesis that if humans are conscious when producing complex behaviours, then, machines may be conscious when they produce them too [23].

We know, and have measured, that humans are conscious when performing these behaviours thanks to functional magnetic resonance imaging (fMRI) and related techniques [31] [33]. These behaviours can include imagination [55], emotions [51], language communication and social relations [50] or awareness of the environment [34].

Machine learning recent models are able to generate art [21] that deviate from what they are fed to learn, are able to learn how to learn [56], learn from a few examples [52] and are able to transfer knowledge from a different task to behave better in a new one [32]. The applications of these abilities include natural language generation [20], understanding emotions [11] or generating videos [61]. We believe that if the philosophical theory that consciousness arises as a flux of information in any machine [48] is true, if we create a cognitive architecture [12] that is able to produce as much behaviours as possible that are correlated with consciousness in humans, then, the machine may as well be, up to some amount, arise consciousness or phenomenal states.

In this work, we attempt to provide a bridge between the machine learning and the machine consciousness communities by providing the design of a cognitive architecture that produces behaviours that are correlated with machine consciousness through machine learning models. Section 2 will discuss related work done in the past about machine consciousness. Then, in Section 3, we provide a detailed explanation of the modules of our architecture which implement correlate behaviours of consciousness. Section 4 then provides the design of the architecture that unifies these modules. We conclude our work with a section of conclusions and further work.

2 Related Work

Due to different theories explaining the origin of consciousness, several approaches have been proposed to tackle this problem. It is hence critical to organize the work that has been done in the area. In order to do so, we first propose the different processes involving machine consciousness [24] and then, the different approaches that have tackled machine consciousness [46].

Machine consciousness processes involve mainly four categories ordered from 1 to 4 in function of how close to generating real awareness they are [25].

Level 1 includes machines that implements external behaviour associated with consciousness. Some of the described behaviours in the introduction section like social interactions implemented in machines would be level 1 and the field of artificial general intelligence [27] lies in this level. Several authors [30] [39] argue that machines implementing these behaviours may produce consciousness, but there is controversy. In this work, we will propose an architecture that contains modules that produces these behaviours. When such an architecture exists, we are talking about level 3 machines, that is, machines with an architecture that is claimed to be a cause or correlate of human consciousness. Several architectures have been proposed before [18] but none of them include both Deep Learning and Gaussian Processes.

Some of the proposed modules, like the imagination one, are cognitive characteristics [3] associated with consciousness. Machines that implement these processes are level 2 machines. Other examples of these processes are attention, emotion, depiction and planning. Lastly, phenomenally conscious machines are level 4 machines , but this category is outside the scope of this work as we could only hypothesize that our level 2-3 design could emerge phenomenal states [2].

Several approaches have tackled the previous categories of machine consciousness. A classification of them all [46] includes five categories: First one are methods based in the global workspace theory [5]. According to this theory, consciousness emerges from a system, like the brain, with a collection of distributed specialized networks with a fleeting memory capacity whose focal contents are widely distributed to many unconscious specialized networks, called contexts. These contexts work together to jointly constrain conscious events and to shape conscious contents [6]. These theory has support of the neuroscience community [7] and the computer science community [10]. We are also inspired by this theory to provide a cognitive architecture [12] with machine learning techniques.

The other categories include methods that suggest that consciousness emerges from a certain amount of information processing and integration [9], from creating an internal self-model [43], from generating higher-level representations [1] and from attention mechanisms [35].

Lastly, machine consciousness has risen as a research topic for the Deep Learning literature as well [10], where the interest resides in learning representations of high-level concepts of the kind humans manipulate with language. This paper tries to give a machine learning perspective, mentioning that elements of a conscious thought are selected through attention mechanisms [8]. We suggest that machine learning and related techniques are able to work as a global workspace, process a high amount of information, can generate internal self-models and higher level representations and have attention mechanisms. Hence, if machine consciousness can be implemented, we argue that machine learning techniques are fundamental to do so as they englobe all the categories of machine consciousness approaches.

3 Machine Consciousness Correlate Processes

We now provide the module design that implement cognitive processes and exhibit external behaviour that are correlated with consciousness [25].

3.1 Simulating dreams

In order to simulate dreams, we first have to record photos $\mathbf{P}_i$ when being awake and store them in a semantic network [54]. Then, dreams will use that information $\mathbf{P} : \mathbf{P}_i \in \mathbf{P}$ to generate a sequence of images $\mathbf{D}_i \in \mathbf{D}$. We define a dream as a function d that converts a subset of a sequence of images $\mathcal{P} \in \mathbf{P}$ and a subset of a sequence of style images $\mathcal{S} \in \mathbf{S}$ in a new set of images $\mathcal{D}$, that is $\mathcal{D} = d(\mathcal{P}, \mathcal{S})$. In order to generate this procedure, we propose two processes for this simulator:

The first process involves the classification of images $\mathcal{P}$ into a semantic network R. In this process, we assume that a previous categorized semantic network R exists and that a robot has already learned to classify images $\mathcal{P}$ into that network R.

An implementation of this process can have ImageNet [19] as the semantic network of images. ImageNet is a resource with more than 14.000.000 images $\mathcal{D}(\mathbf{X})$ and more than 21.000 categories $\mathbf{y}$. ImageNet uses the hierarchy of WordNet [22] to classify photos, having each category y_i a semantic meaning and being organized as a graph $G = V, E$ that can be traversed, where $v \in V$ is the node representing category y_i. Convolutional neural networks [36] or more advance neural models such as Efficient Net L2 [60] or ResNet [59] neural models to classify taken photos into ImageNet. Let NN be the neural model that implements the robot, the robot will classify each input image P to category y_i, inserting it in the graph G through the NN trained on the ImageNet dataset $\mathcal{D}(\mathbf{X}, \mathbf{y})$, that is: $y_i = NN(P|\mathcal{D}(\mathbf{X}, \mathbf{y}))$.

To feed images in the neural model NN to be classified in the graph G, we would need a robot with an integrated camara to take the photos $\mathbf{P}$ and define a period of being awake T_a and asleep T_s. These parameters can be configured differently for every robot. We suggest to save additional images $\mathcal{S}$ that will represent different styles seen like for example dark places or broad landscapes in a different semantic network R_s.

Second, we need to define the state of being dreaming given by time T_s. We suggest to use a random walk [47] like the one performed in the Metropolis Hastings algorithm [15] to simulate movement into the semantic networks of images R and styles R_s that are related by semantic distance $d_s(y_i, y_j)$ in their graphs G, G_s given by the number of edges that connect each category. At each step, we select two images $\mathbf{P}_i, \mathbf{P}_i^s$ and invoke Deep Style neural networks [40] to generate a new image with the selected photo and an style applied $\mathbf{D}_i = DS_{nn}(\mathbf{P}_i, \mathbf{P}_s^i)$. Similar models such as a Generative Adversarial Network [44] could also be used for this task. The robot then will get attention to that photo and save it. We can observe examples of generated photos using the Deep Dream Generator by this procedure in Figure 1.

Fig. 1. Generated photos representing dreams by the Deep Dream Generator models (`http://deepdreamgenerator.com/`)

The initial categories y_{init}, y_{init}^s are chosen at random. To select the new categories, we perform a random walk in the graph G and G_s given by some uniform distribution with a lower l_l and upper u_l limit, whose sampled value we will call step size ω. For the sake of readability, when we refer to increment or decrement the step size ω, we mean to increment or decrement the lower and upper limits of the uniform distribution $U(l_l, u_l)$. If we set those parameters to a high number, dreams will contain different concepts and if we set it to a low number, they will have similar concepts.

We repeat the mentioned process by performing an iteration of the random walk and generating new images in an iterative fashion after a certain amount of time T_s or a number of iterations n_i. For each generated image $\mathbf{D}_i$, we store it and feed it to the camera of the robot. The sequence of recreated images $\mathbf{D}$ recreates the dream.

After dreaming, the robot is awake and will capture images with the process described by the following module. If we iterate both processes, the robot will acquire images in the first process $\mathcal{P}, \mathcal{S}$ and simulate new images $\mathbf{D}$ based in the acquired ones in the dream process.

3.2 Depiction. Being aware of the environment

In order to perform this task, we suggest to implement a robot that moves autonomously in a given environment E. For the sake of simplicity we are going to assume that E is 2-dimensional $E \in \mathbb{R}^2$. This robot will remember the images $\mathbf{D}$ that has previously dreamed as described in the previous module. The robot, when awake, will try to return to the location or neighbourhood $\mathcal{N} \in E$ where the images that has dreamed are located.

We will generate a 2-dimensional function of location importance with a sample from a Gaussian Process [58] $f_l^r \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')) \in \mathbb{R}^2$ over the environment E , discretized by a grid, for each robot r with interesting places to visit. We can observe examples of such functions at Figure 2 The resolution of the grid r_g can set the size of the environment E. Gaussian Processes models are flexible priors or distributions over functions where inference takes place directly in the functional space $\mathcal{F}$. This functional space will contain every possible environment that can be created $E \in \mathcal{F}$. Given a covariance function $k(\mathbf{x}, \mathbf{x}')$ and its hyperparameters, we control the shape of the generated functions. GPs has been chosen as a probabilistic model of E as it creates stationary smooth functions that a robot can traverse easily by some heuristic. The generated environment by the

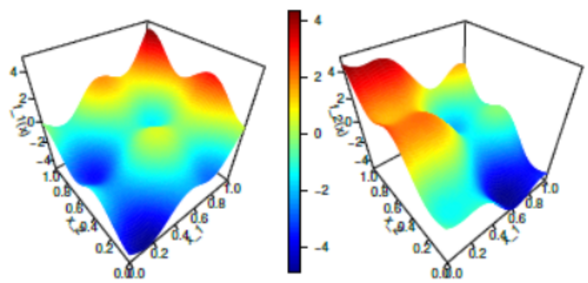

Fig. 2. Sampled functions from a 2-dimensional GP representing the importance value of each location of the environment for the robots.

GP $f_l \in \mathbb{R}^2$ will contain high-valued locations of interest to take photos from and low-valued locations to ignore. When the robot reaches these places, it will take photos of the environment and save them for the dream module. Once visited, these places will be penalized by a local penalization procedure [28]. These kind of procedures get a neighbourhood $N \in f_l$ centered in the place of interest $\mathbf{r} \in N$ and penalizes this zone by, for example, a multivariate gaussian distribution centered in the place that the robot takes the photo with an identity matrix as covariance to penalize equally all dimensions $f_l(N) = f_l(N) - MVG(\mathbf{r}, \mathbf{I})$. The dreamed places will be given high rewards in the Gaussian Process prior of the scenario and retrieved if taken photos from.

The robot will end navigation when it is exhausted after its time awake T_a. We can simulate fatigue through a non deterministic function $p(r,t) \in [0,1]$ of time since it has last slept. Each time that the robot takes a photo, fatigue will be incremented or decremented depending on the reward given by the photo. We can assign a threshold $\phi \in [0,1]$ for fatigue. When the robot is awake the threshold takes value 0 and it is incremented as a function of time or by ϵ when taking a photo. The robot will fall asleep after the maximum time awake or if the non deterministic function samples a value higher than the threshold $p(r,t) > \phi$. Specific parametric forms for the function can be configured for each robot.

The action of taking or not the photo t_p in each place g of the grid will be non deterministic and dependant on the value of f_l, that is $t_p(f_l, g)$. Higher values will incur in a higher probability of taking a photo. The robot can take a photo my sampling t_p periodically after an amount of steps s in f_l or after a certain amount of time. The specific parametric function of t_p can be configured as well differently for each robot.

The Gaussian process sample f_l will be contaminated by i.i.d. gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_{gj})$ in each position of the grid g and dimension j in the period of being awake to favour exploration over unknown places. Higher values of σ will enforce exploration. The robot will navigate through the interesting places through a local search metaheuristic [57] mechanism similar that favours exploration and exploitation. We can also favour exploitation through the computation of the gradient of the Gaussian process [53]. Random rewards will be put in the scenario. In order to exhibit a human behaviour it is recommended to find a good tradeoff between exploration and exploitation.

Through these mechanism we expose external human behaviour of interest in places based in subjective experiences, that are the dreams of the robot. We also reinforce the previous module by feeding it with different images $\mathcal{P}$ of the environment and creating new dreams $\mathcal{D}$, that are cognitive processes of the robot.

3.3 Emotion simulation

In this section, we define a process that models emotions through objective functions $e(t) \in [0, 1]$ of time. The main reason why we implement emotions in these robots is because they are going to influence the Gaussian Process prior f_l of the environment E. If the robots feels confident and happy $e_h(t) \approx 1$, unknown near areas of E to the position of the robot g will be rewarded to be explored. If R is a neighbourhood of f_l containing a reward, we can reward its value by sampling from a multivariate gaussian distribution centered in the reward $f_l(R) = f_l(R) - MVG(\mathbf{r}, \mathbf{I})$. By doing this process, the robot will enter a positive cycle and take photos of interesting places. By performing this action, we increment $e_h(t)$ by a uniform distribution which limits $[l, u]$ can be parametrized. If, in contrast, the robot feels sad and fear $e_h(t) \approx 0$, movement across the grid will be penalized by incrementing fatigue and decrementing the step size ω of the random walk, entering a negative loop.

These cognitive processes will exhibit external behaviour that will show if a robot is happy or sad by its activity on the grid. We provide an exit of the cycles by images of dreams $\mathcal{D}$. Dreams can also influence emotions $e(t)$ and make the robot behave differently. If an image resembles a visited area that had got high value of f_l, happiness will be incremented by a parametrizable amount $e(t) = e(t-1) + \delta \sim U[l, u]$, where t represents time. If images of places with low value of f_l are displayed, the opposite operation will be performed $e(t) = e(t-1) - \delta \sim U[l, u]$. By implementing these emotions, the robot exhibits human behaviour, which is a reflect of potential consciousness. Happiness could also affect the fatigue function, by alleviating it if the robot is happy or increasing it in the other case.

Other emotions that may be optimized are curiosity and boredom $e_c(t) \in [0,1]$, that would affect the Gaussian Process sampled function f_l by penalizing already saw places by an $MVG(\mathbf{0},\mathbf{I})$ and rewarding unknown places also by an $MVG(\mathbf{0},\mathbf{I})$. A last example can be friendship and solitude $e_f(t)$, based in relations with other robots that are going to be described further or courage and fear $e_c(t)$ that will condition the movements across the environment by incrementing the step size ω of the random walk. The described fatigue function can also be seen as an emotion. Particular parametric forms of the functions are open for the robot developer to be implemented.

3.4 Social relationships with other robots

If we want to simulate emotions $e(t)$ like the ones felt with humans to show behaviour correlated with consciousness, we need to model these emotions to be not only a function of the environment interaction f_l but also of relationships with other robots. For this reason, we consider that an essential component for the cognitive processes of the robots must be the interaction with other robots to share experiences, in the form of photos $\mathcal{P}$ in this setting, and influence the emotions $e(t)$.

Emotions like friendship or solitude $e_s(t)$ are dependant on social interactions. We define here a social interaction $\alpha(\beta_x,\beta_y)$ as the change of a photo $\mathcal{P}_x$ of a robot β_x with a photo $\mathcal{P}_y$ of a robot β_y when both robots share the same location g in the environment E.

Each robot β_i has a different function sampled from the GP prior $f_l^i \sim \mathcal{GP}(0,k(\mathbf{x},\mathbf{x}')) \in \mathbb{R}^2$ of the environment E. As each photo $\mathcal{P}_i$ related to a position of the grid g_i, it will have, for every robot β_i a different value $f_l^i(g_i)$, conditioning the rest of the emotions. If the photo refers to a location that the robot likes according to its prior f_l^i, emotions will make the robot more active. Although, if this is not the case, the robot may enter a negative cycle.

By interacting with each other, robots β will share images $\mathcal{P}$ or dreamed images $\mathcal{D}$ of the environment E that will modify their Gaussian Process sampled function f_l^i and the other emotions of the robot. Specific parametric forms are again free for the programmer of the robot to be set. By implementing these social simulation, robots exhibit interactions between them that affect their emotions, which is as well a correlated effect of consciousness.

4 An Unified Architecture for the Models

In the previous section, we have described how can we implement external behaviours correlated with consciousness in machines. All the described processes can be implemented in a certain amount of robots β with an environment E that they can traverse and get photos $\mathcal{P}$ from. In this section, we provide a diagram with all the modules described to illustrate how the information flows in our architecture. We can see the diagram in Figure 3. We can observe how

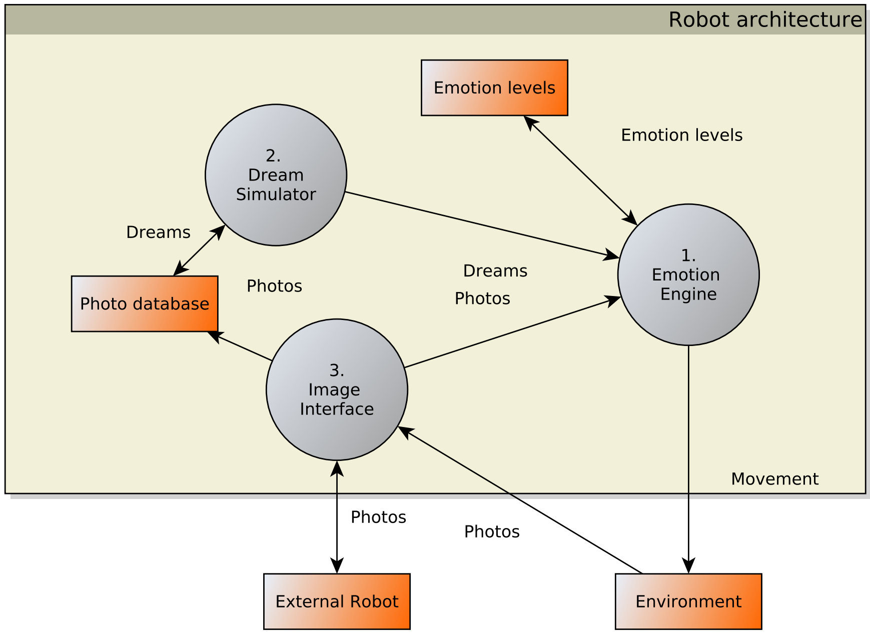

Fig. 3. Diagram flow with all the modules of the system.

robots share images, showing social behaviour. These interactions affect their emotions and incur in a different movement across the environment, reflecting emotions, commonly correlated with consciousness. Cognitive interior processes include dreaming images that are function of the perceived images, affecting emotions. These behaviours could, according to the cited theories, be a correlation of consciousness in robots.

5 Conclusions and further work

In this paper, we have illustrated an architecture of processes that, if are implemented in robots being executed in an environment, they will exhibit external human behaviour and have cognitive interior processes. According to machine consciousness theory [25], both characteristics could be correlated with machine consciousness in robots [24].

We plan to implement all the described processes in robots to get empirical evidence of their behaviour and execute machine consciousness tests to extract useful conclusions. Futher work will also include optimizing the emotions described in Section 3 by some mechanism such as constrained Multi-objective Bayesian Optimization [26] in order to create a global and dynamical policy for the behaviour of the robots, that we believe that is more flexible than the proposed one.

Acknowledgments

The authors gratefully acknowledge the use of the facilities of Centro de Computación Científica (CCC) at Universidad Autónoma de Madrid. The authors also acknowledge financial support from Spanish Plan Nacional I+D+i, grants TIN2016-76406-P and TEC2016-81900-REDT.

References

1. Aleksander, I. *Impossible minds: my neurons, my consciousness.* World Scientific, 1996.
2. Aleksander, I. Why axiomatic models of being conscious? *Journal of Consciousness Studies 14*, 7 (2007), 15–27.
3. Aleksander, I., and Dunmall, B. Axioms and tests for the presence of minimal consciousness in agents i: preamble. *Journal of Consciousness Studies 10*, 4-5 (2003), 7–18.
4. Arrabales, R., Ledezma, A., and Sanchis, A. Consscale: A pragmatic scale for measuring the level of consciousness in artificial agents. *Journal of Consciousness Studies 17*, 3-4 (2010), 131–164.
5. Baars, B. J. In the theatre of consciousness. global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies 4*, 4 (1997), 292–309.
6. Baars, B. J. The global workspace theory of consciousness. *The Blackwell companion to consciousness* (2007), 236–246.
7. Baars, B. J., and Newman, J. A neurobiological interpretation of global workspace theory. *Consciousness in philosophy and cognitive neuroscience* (1994), 211–226.
8. Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
9. Balduzzi, D., and Tononi, G. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS computational biology 4*, 6 (2008).
10. Bengio, Y. The consciousness prior. *arXiv preprint arXiv:1709.08568* (2017).
11. Cambria, E. Affective computing and sentiment analysis. *IEEE Intelligent Systems 31*, 2 (2016), 102–107.
12. Chella, A., Frixione, M., and Gaglio, S. A cognitive architecture for robot self-consciousness. *Artificial intelligence in medicine 44*, 2 (2008), 147–154.
13. Chella, A., and Manzotti, R. Artificial intelligence and consciousness. In *Association for the advancement of Artificial Intelligence Fall Symposium* (2007), pp. 1–8.
14. Chella, A., and Manzotti, R. Artificial consciousness. In *Perception-Action Cycle*. Springer, 2011, pp. 637–671.
15. Chib, S., and Greenberg, E. Understanding the metropolis-hastings algorithm. *The american statistician 49*, 4 (1995), 327–335.
16. Crick, F., and Clark, J. The astonishing hypothesis. *Journal of Consciousness Studies 1*, 1 (1994), 10–16.
17. Damasio, A., and Dolan, R. J. The feeling of what happens. *Nature 401*, 6756 (1999), 847–847.

18. Dehaene, S., and Changeux, J.-P. Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS biology 3*, 5 (2005).
19. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (2009), Ieee, pp. 248–255.
20. Deng, L., and Liu, Y. *Deep learning in natural language processing*. Springer, 2018.
21. Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068* (2017).
22. Fellbaum, C. Wordnet. *The encyclopedia of applied linguistics* (2012).
23. Gamez, D. *The development and analysis of conscious machines*. PhD thesis, University of Essex Colchester, 2008.
24. Gamez, D. Progress in machine consciousness. *Consciousness and cognition 17*, 3 (2008), 887–910.
25. Gamez, D. *Human and machine consciousness*. Open Book Publishers, 2018.
26. Garrido-Merchán, E. C., and Hernández-Lobato, D. Predictive entropy search for multi-objective bayesian optimization with constraints. *Neurocomputing 361* (2019), 50–68.
27. Goertzel, B., and Pennachin, C. *Artificial general intelligence*, vol. 2. Springer, 2007.
28. González, J., Dai, Z., Hennig, P., and Lawrence, N. Batch bayesian optimization via local penalization. In *Artificial intelligence and statistics* (2016), pp. 648–657.
29. Graziano, M. S. The attention schema theory: A foundation for engineering artificial consciousness. *Frontiers in Robotics and AI 4* (2017), 60.
30. Harnad, S. Can a machine be conscious? how? *Journal of Consciousness Studies 10*, 4-5 (2003), 69–75.
31. He, B. J., and Raichle, M. E. The fmri signal, slow cortical potential and consciousness. *Trends in cognitive sciences 13*, 7 (2009), 302–309.
32. Kaiser, L., Gomez, A. N., Shazeer, N., Vaswani, A., Parmar, N., Jones, L., and Uszkoreit, J. One model to learn them all. *arXiv preprint arXiv:1706.05137* (2017).
33. Kamitani, Y., and Tong, F. Decoding the visual and subjective contents of the human brain. *Nature neuroscience 8*, 5 (2005), 679–685.
34. Kitamura, T., Tahara, T., and Asami, K.-I. How can a robot have consciousness? *Advanced Robotics 14*, 4 (2000), 263–275.
35. Koch, C., and Tsuchiya, N. Attention and consciousness: two distinct brain processes. *Trends in cognitive sciences 11*, 1 (2007), 16–22.
36. Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.
37. LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *nature 521*, 7553 (2015), 436–444.
38. Loar, B. Phenomenal states. *Philosophical perspectives 4* (1990), 81–108.
39. Moor, J. H. Testing robots for qualia. In *Perspectives on mind*. Springer, 1988, pp. 107–118.
40. Mordvintsev, A., Olah, C., and Tyka, M. Inceptionism: Going deeper into neural networks.

41. Murphy, K. P. *Machine learning: a probabilistic perspective.* MIT press, 2012.
42. Pennartz, C., Farisco, M., and Evers, K. Indicators and criteria of consciousness in animals and intelligent machines: an inside-out approach. *Frontiers in systems neuroscience 13* (2019), 25.
43. Perlis, D. Consciousness as self-function. *Journal of Consciousness Studies 4*, 5-6 (1997), 509–525.
44. Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
45. Raoult, A., and Yampolskiy, R. Reviewing tests for machine consciousness. *Retrieved from ResearchGate* (2015).
46. Reggia, J. A. The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks 44* (2013), 112–131.
47. Roberts, G. O., Gelman, A., Gilks, W. R., et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability 7*, 1 (1997), 110–120.
48. Rose, D. Consciousness: philosophical, psychological, and neural theories.
49. Searle, J. R. *Mind: A brief introduction.* oxford university press, 2004.
50. Searle, J. R., Willis, S., et al. *Consciousness and language.* Cambridge University Press, 2002.
51. Shanahan, M. Consciousness, emotion, and imagination: a brain-inspired architecture for cognitive robotics. In *In Proceedings of the AISB05 Workshop: Next Generation Approaches to Machine Consciousness* (2005), Citeseer.
52. Snell, J., Swersky, K., and Zemel, R. Prototypical networks for few-shot learning. In *Advances in neural information processing systems* (2017), pp. 4077–4087.
53. Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D. J., and Rasmussen, C. E. Derivative observations in gaussian process models of dynamic systems. In *Advances in neural information processing systems* (2003), pp. 1057–1064.
54. Sowa, J. F. Semantic networks.
55. Stuart, S. A. Machine consciousness: Cognitive and kinaesthetic imagination. *Journal of Consciousness Studies 14*, 7 (2007), 141–153.
56. Thrun, S., and Pratt, L. *Learning to learn.* Springer Science & Business Media, 2012.
57. Voss, S., Martello, S., Osman, I. H., and Roucairol, C. *Meta-heuristics: Advances and trends in local search paradigms for optimization.* Springer Science & Business Media, 2012.
58. Williams, C. K., and Rasmussen, C. E. *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
59. Wu, Z., Shen, C., and Van Den Hengel, A. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition 90* (2019), 119–133.
60. Xie, Q., Hovy, E., Luong, M.-T., and Le, Q. V. Self-training with noisy student improves imagenet classification. *arXiv preprint arXiv:1911.04252* (2019).
61. Xiong, W., Luo, W., Ma, L., Liu, W., and Luo, J. Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 2364–2373.