A Computer Science Perspective on Models of the Mind

Teresa Nicole Brooks Abu Kamruzzaman Avery Leider and Charles C. Tappert
Seidenberg School of Computer Science and Information Systems
Pace University, Pleasantville NY

Abstract

Recent findings about the human mind add to our understanding of the working brain. Pattern recognition algorithms are expressed by the neurons firing together in that part of the brain most active during thinking. These algorithms are variation-tolerant, self-organizing, and continuously adapting. There are three seminal models of biological inspiration in Artificial Intelligence that must be included in the development of a standard model of the mind. These are Rosenblatt's perceptrons, Kutzweil's Pattern Recognition Theory of Mind and Hawkins's Hierarchical Temporal Memory. In particular, the Hierarchical Temporal Memory includes groundbreaking insights into how the brain simplifies, reducing data complexity, how it accommodates variations in input, and how it organizes data. We survey the research in the range of academic areas engaged in the standard model of the mind and find the common unifying ideas of Rosenblatt, Kurzweil and Hawkins.

Introduction

Artificial Intelligence (AI) research is unique as it often requires cross-discipline knowledge from fields such as linguistics, cognitive science and neuroscience. As our understanding of what constitutes intelligence and the fundamental principals that govern it grows, researchers have been trying to develop systems that can reason, learn and solve problems like humans. This desire has fueled decades of contentious debates, regarding the vastly different approaches to build intelligence systems, but in recent years there appears to be an emerging consensus regarding the current state of the field among the AI research community (E, Lebiere, and Rosenbloom 2017). This consensus was observed during the 2013 AAAI Fall Symposium on Integrated Cognition and ultimately led to an initial proposal for a standard model of human-like minds.

Since the 2013 AAAI Fall Symposium, there has been a call to engage the international research community to further develop a standard model of the mind. In this paper we will not propose changes or critiques of the current proposed model. We will however enrich the discussion by presenting a computer science perspective on developing models of the mind.

What We Currently Know About The Human Mind

Only the human brain has intelligence, meaning the ability to evaluate itself and change (Kurzweil 2013). Mind is the core functional object for any mammal. There is no actual visibility of mind. We understand the brain as the visible entity for mind. Our understanding of the mind has improved through studying artificial intelligence, cognitive science, neuroscience, and robotics (E, Lebiere, and Rosenbloom 2017).

We understand the mind to be "the faculty of consciousness and thought." (Dictionary 2007). The mind is capable of perception, judgment, imagination, recognition, memory, feelings, attitudes, and emotions. All of these things determine our choice of action. Various structures of the brain are responsible for specific mental processes. For example, our consciousness is affected by the prefrontal cortex, using collections of neurons with parallel connections across other regions in the brain (Opris and Casanova 2014) Opris (Opris et al. 2013) suggests that interactions between inter laminar prefrontal microcircuits, the posterior parietal cortex, and cortico-striatal-thalamo-cortical circuits are responsible for making our decisions. The neocortex of the brain is responsible for sensory perception, recognition of everything from visual objects to abstract concepts, controlling movement, reasoning from spatial orientation, rational thought and language in what we regard as "thinking" (Kurzweil 2013). The neocortex may also recognize patterns. According to Kurzweil, this is possible because the neocortex has a columnar organization, as first discovered by Vernon Mountcastle.

We can use the brain's architecture as a blueprint for designing a digital counterpart. Kurzweil (Kurzweil 2013) estimates the mind contains "30-100 million bytes of compressed code," and artificial intelligence, if created based on this design and using hidden Markov models and genetic algorithms, could surpass the human mind in its capabilities. But an artificial brain of that ability will require massive computational power that will not be reached for another decade (Jolivet et al. 2015). The Blue Brain Project (Markram 2006) has thus far only managed to replicate a rodent brain (Markram et al. 2011). Figure 1 below shows the time line of the actual and projected progress of the Blue

Brain simulation project by year in a 45 degree line on a graph. The progression in the X axis represents computer speed(FLOPS) and in the Y axis represents computer memory in bytes needed to run the project.

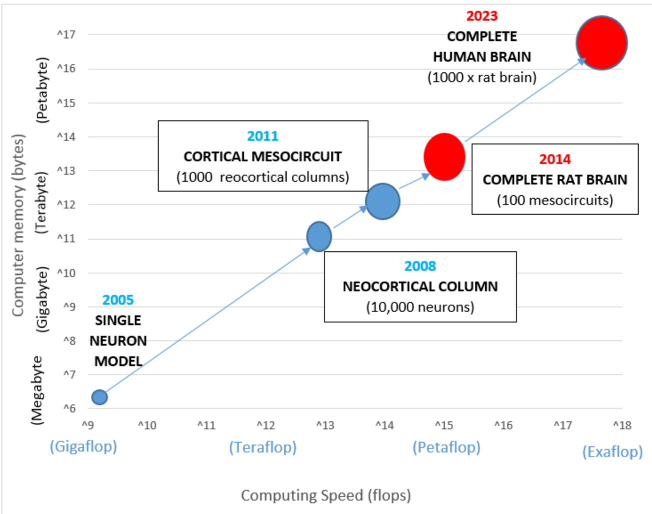

Figure 1: Actual and projected progress of the Blue Brain brain simulation project (adapted) (Kurzweil 2013).

Dharmendra Modha and others simulated a digital brain model, cell-by-cell, of a partial human visual neocortex that contains 1.6 billion virtual neurons and 9 trillion synapses equivalent to a cat neocortex (Kurzweil 2013). Hierarchical Hidden Markov Models(HHMMs) (Fine, Singer, and Tishby 1998) are used for speech recognition and natural language texts.

Theories For Developing Artificial Minds

The neocortex is the part of the brain where researchers believe thinking occurs. It is this belief along with discoveries regarding the structure and functionality of the neocortex that has motivated researchers to develop biologically inspired models of artificial minds. In this section we will briefly describe key attributes of two exemplar theories for modeling artificial minds.

Pattern Recognition Theory of Mind (PRTM)

Ray Kurzweil's Pattern Recognition Theory of Mind is a theory for describing the basic algorithms of the neocortex. It is based on the hypothesis that the neocortex is a homogeneous, recursive structure that is composed of a large number of basic structural units called pattern recognizers (Kurzweil 2013), hence making the neocortex itself a pattern recognizer. This bottom-up, hierarchical organization of patterns and pattern recognizers is a key attribute of this theory, as it allows for the expression, matching and storage of complex and abstract concepts.

Pattern recognizers are defined by a hierarchy of self-organizing connections that link together. When a new pattern is learned, new connections are formed between the pattern recognizers that were involved in recognizing the given input pattern. Each pattern recognizer is responsible for identifying a single input pattern. There is also redundancy among pattern recognizers to identity the same input pattern. This redundancy allows for generalization of pattern identification, which allows a system to learn and tolerate variations in input patterns. The *self-organization* of pattern recognizers is an important feature of PRTM, as it enables a system to create new connections and remove obsolete connections while it learns new input patterns over time.

Kurzweil proposes using hierarchical hidden Markov models (HHMM) to implement "self-organizing hierarchical pattern recognition" (Kurzweil 2013). Hierarchical hidden Markov models is a statically model where each state is its own self-contained probabilistic model and each state yields a sequence of observations symbols instead of an individual symbol (Park 1998). In PRTM based systems hierarchical hidden Markov models each internal state on each level represents a single neocoritcal pattern recognition module as depicted in Figure 2. States on each level can identity redundant but similar patterns.

In the example show in Figure 3, Level 1 represents a low-level input pattern. Levels 2 and 3 represent higher-level input patterns or concepts that contains input patterns from one level below it. Each white circle labeled $\mathrm{PR}_{i,j}$ represents a pattern recognition module. The gray lines represent vertical state transitions and the black lines represent horizontal state transitions, where which transition has a calculated probability.

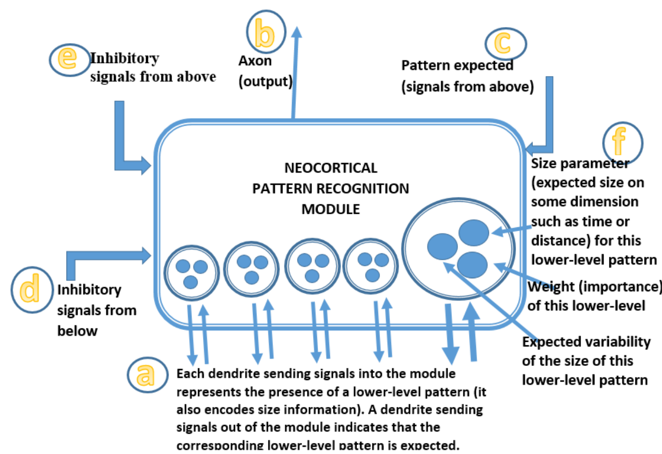

Figure 2: Diagram of A Single Neocortical Pattern Recognition Module (adapted) (Kurzweil 2013).

When a pattern recognizer receives an input pattern it calculates a score (probability). This score is calculated by matching observed magnitudes of each feature of an input pattern against learned size and size variability parameters that are associated with this features. This score is used to determine if a pattern recognizer was successful at identifying the single input pattern it is assigned to recognize. If the score exceeds a learned threshold, it will signal a neighboring pattern recognizer one level higher in the hierarchy that

Table 1: Summary of biologically inspired theories for modeling an artificial neocortex.

Model	Implementation	Key Components
Pattern Recognition Theory of The Mind (PRTM)	Hierarchical Hidden Markov Models (HHMM)	Proposes a model of an artificial neocortex that is a self-organizing, recursive, pattern matching structure where the relationships between the pattern matchers are defined by their self-organizing connections (Kurzweil 2013). In addition, patterns are also modeled as recursive structures where a hierarchy of patterns enables the construction of complex patterns and abstract concepts.
Hierarchical Temporal Memory (HTM)	Sparse Distributed Representations (SDRs)	Proposes a "biologically constrained" (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016) theory that describes the fundamental principals of the neocortex. Under this theory the neocortex is defined as a hierarchical memory system, where memory (sensory patterns) are modeled as "time changing (temporal) patterns" (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016). HTM is the core technology for building intelligent machines whose functionality is constrained by the fundamental principles of the neocortex.

it successfully identified it's input pattern, which means the recognizer above it can now process its input pattern. Parameters such as the matching threshold, size and size variability are learned by running genetic algorithms (Kurzweil 2013). Pattern recognizers have the ability to send signals to other pattern recognizers below and above them in the hierarchy. This enables the higher-level pattern recognizers to signal lower-level recognizers to lower their matching thresholds because most of a given input pattern has been identified.

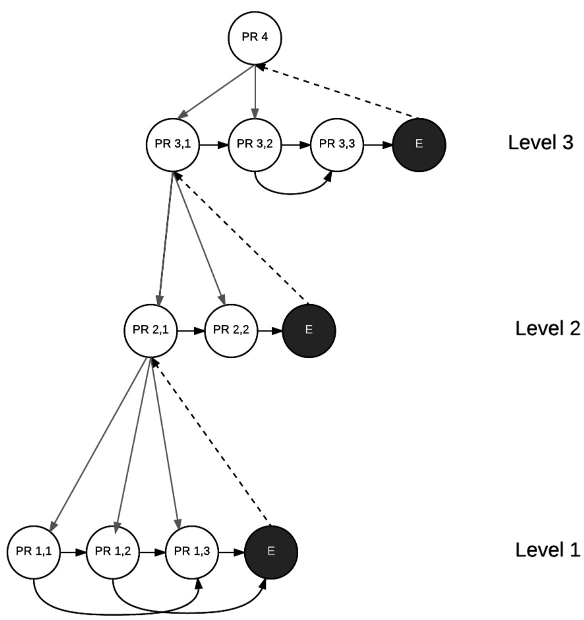

Figure 3: Example of a hierarchical hidden Markov model implementing a hierarchy of redundant pattern recognition modules in a PRTM system.

Input patterns are one-dimensional vectors and like the pattern recognizers, they are organized in a hierarchy of lower and higher level patterns. Multidimensional input pattern features are reduced to one-dimension vectors by a sparse coding technique called vector quantization, which Kurzweil first used in speech recognition software (Kurzweil 2013). This process assigns a vector to be a point in a cluster and each cluster is labeled using an integer (Kurzweil 2013). For example, when a new input pattern arrives in a system it is assigned to a cluster and can then be identified by the integer that represents its cluster. Vector quantization is an optimization that reduces data complexity while retaining the key features that are important for recognizing a pattern.

Hierarchical Temporal Memory (HTM)

Hierarchical Temporal Memory (HTM) is a "biologically constrained" theoretical framework that describes the fundamental principals of the neocortex (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016). It is the successor of George and Hawkin's memory-prediction theory (George and Hawkins 2005) which defined a theoretical model of the human neocortex. Like Kurzweil's Pattern Recognition Theory of Mind it defines a the neocortex as a recursive, homogeneous structure which a hierarchical organization.

Hierarchical Temporal Memory is based on three key principals: common algorithms of cortical regions, hierarchy, and sparsely connected neurons (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016).

For the first principal, Hawkins, J. et al, propose that because of the neocortex's structural uniformity and the fact that we now know that regions of the neocortex perform very similar actions there must be some fundamental algorithms that can generate behaviors for all sensory perceptions such as hearing, vision and language (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016). They also define all sensory actions as "temporal inference problems" where by patterns are a hierarchy of lower and higher level time changing patterns.

The second principal is arguably the most important principal of HTM, as it asserts that cortical regions are defined by a logical hierarchy of connections whereby higher-level perceptions are derived from lower-level sensory patterns. As depicted in Figure 4, these sensory input patterns are processed and passed up the hierarchy as beliefs, which are then used by the highest levels to make predictions. Like Kurzweil's PRTM theory, this hierarchical organization of cortical regions and patterns allows for the expression, matching and storage of complex and abstract concepts. In

HTM based systems memory is organized as a hierarchy of lower and higher levels of memory, where lower-level sensory patterns are taken in as input. These input patterns are processed and passed up the hierarchy as beliefs, which are then used by the highest levels to make predictions. These predictions are then passed down and can be used for example to trigger motor functions.

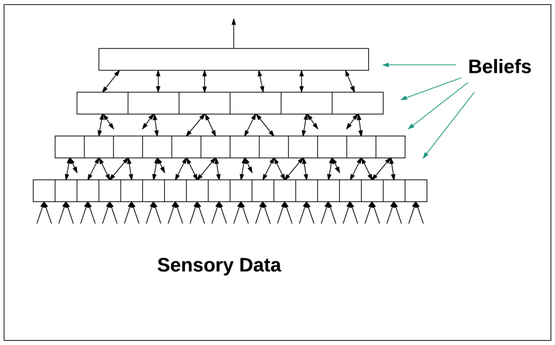

Figure 4: HTM's hierarchy of memory nodes (adapted) (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016).

The third principal defines what HTM considers the foundations of biological intelligence (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016). It asserts that the sparse connections (activation) of neurons is what allows the mind to efficiently learn new sequences as well as make predictions. It is these "moment-to-moment" thoughts and perceptions that define which neurons are active at any given point in time. In HTM systems information is represented as Spare Distributed Representations (SDRs). SDRs are vectors of thousands of bits, where active neurons are represented by 1s and 'off' neurons are represented by 0s, where only a small percentage of neurons are actually active. This representation not only allows the system to represent sparsely connected neurons but acts as a means of data encoding as each bit encodes not only data but contextual information as well. This form of representation also offers interesting mathematical properties such as it's union property which allow systems to perform efficient pattern matching by only comparing a small number of number of features (Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin 2016). For example, when comparing SDRs a system can easily determine how semantically different or similar they are. Such differences and similarities are defined by the number of active bits they share in the same position of the vector.

Common Approaches For Developing Artificial Minds Biologically inspired theories for developing artificial minds share common approaches. Many employ methods for sparsely coding input patterns in order to reduce data complexity, while retaining key features of the pattern and making pattern matching and storage more efficient. The neocortex as well as input patterns are often organized as a hierarchy of lower and higher level structures. This hierarchical organization is key allows the learning and representation of complex and abstract concepts. Lastly, these models share common origins with early models in machine learning such as deep learning, such as Rosenblatt's (1962) topological model of the nervous system depicted in Figure 5. Like their predecessors, modern theories for developing artificial minds build complex networks where weights that describe the relationships between units in the network are learned, however unlike deep learning models where the connections are fixed these models employ online-learning, which enables them to learn concepts by making new connections over time as well as pruning obsolete connections when the relationships are no longer valid.

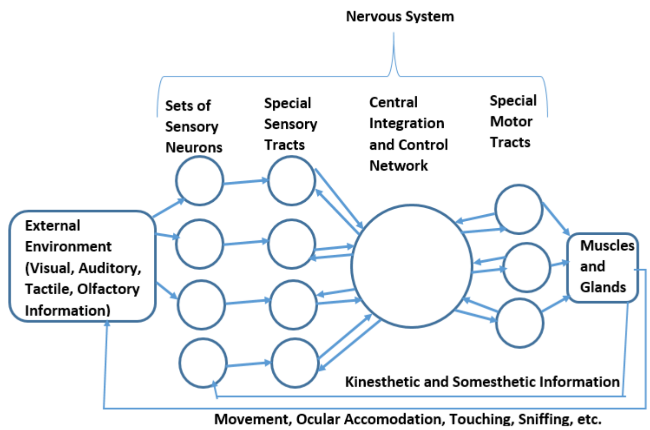

Figure 5: Topological model of the nervous system (adapted) (Rosenblatt 1962).

Relationships Between Deep Learning and Brain Architecture

Six key concepts of Deep Learning in Artificial Intelligence (AI) and the functioning of the human brain reflect the evolving merge of the standard model of the mind and AI's deep learning tools.

First in our model is the concept of brain memory, analogous to computer storage. Memory or storage is key to the function of both the human brain and in deep learning. Memory in the brain, like a deep learning network, stores input data, weighs parameters, and acts on computed data - however dynamically and nomadically the patterns of neurons and synapses accomplish these actions - and machine deep learning uses dynamic RAM (DRAM), static RAM (SRAM) internally and externally, as all computers are designed to save new data to storage in order to function (Nahal 2017).

Second in our model is learning. Both deep learning and the brain learn from their respective datasets. Both of them use the stored data to execute their intelligent actions. Neural networks require frequent access to data to learn from data and stores the entire dataset in computer memory, just as the brain stores the the "brain dataset" of pattern recognizers in the hippocampus, learning from the frequency of the high-level features from cortical neurons (Fontana 2017).

Third and fundamental is the circuit diagram that represents the electrical foundations of the current electronic computers and its similarity to the biological connectome structure (NIH 2017) that is the focus of intense research investment by the United States government (Van Essen et al. 2013). The complex brain connectome structure of two connected brain neurons (each of which can be connected to up to 10,000 other neurons) can be built using deep learning architecture. To predict the connectome between the brain neurons the architecture is built using convolutional layers, max-pooling layers and recurrent layers from the bottom up and a dynamically programmed layer on top to align the output sequences of salient temporal patterns from the two recurrent layers as depicted in figure 6 (Veeriah, Durvasula, and Qi 2015).

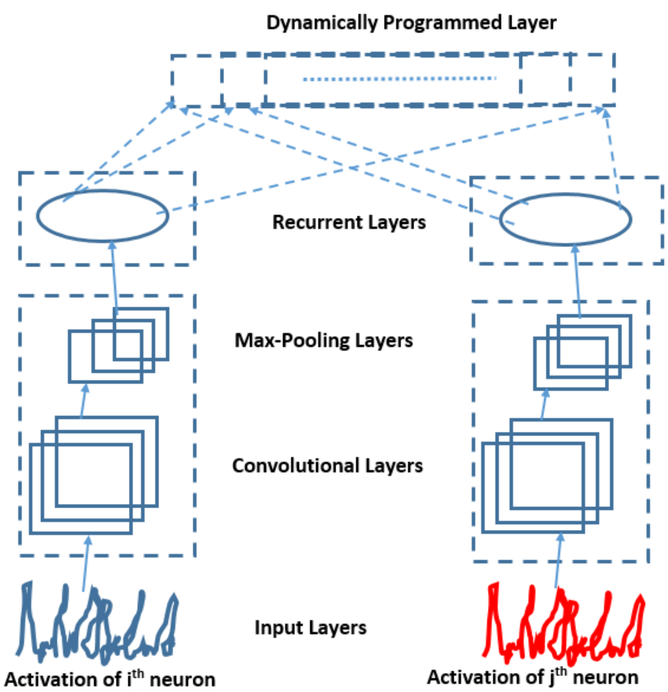

Figure 6: Dynamically programmed proposed deep learning architecture for brain connectome (adapted) (Veeriah, Durvasula, and Qi 2015).

Fourth in our model is emotion. Using feature representations such as Convolutional Neural Networks (CNN), the brains multimodal emotions information such as visual stimuli, past experiences, context, motion, face expressions can be simulated in computer systems by deep learning neural network models (Ghayoumi and Bansal 2016). CNN basic structure depicted in figure 7.

Behavior is the fifth concept in our model. The hypothesis of the integration of Deep Learning and Neuroscience behavior says that (1) neurons can gradually mature their synapses, (2) neurons in different brain areas can optimize different sensors to improve over time, and (3) different brain areas are pre-structured to solve identical computational problems posed by behavior. This hypothesis is supported by simulations of the implementation of multiple layers of neurons from neural circuitry deep learning (Marblestone, Wayne, and Kording 2016).

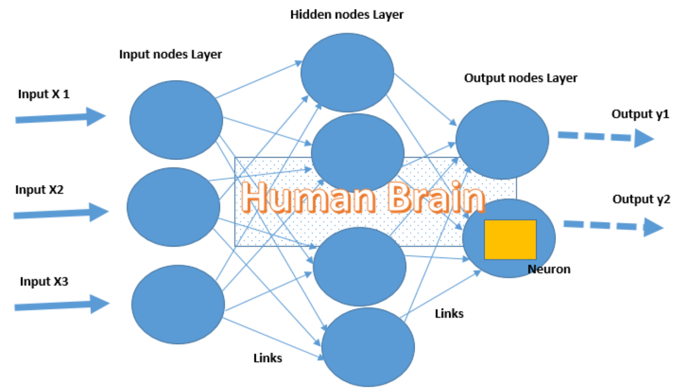

Figure 7: Basic structure of an Convolutional Neural Network (adapted) (Kurzweil 2013).

The sixth concept in the model has the potential to unify all the other concepts, which is the China Brain Project (Poo et al. 2016) that is the work in China that competes with the Human Connectome Project of the U.S. and also explores the broad relationship between deep learning and human brain. The aim of the China Brain Project includes creating an artificial brain, to better understand the working of the brain at different levels, and promote deep learning projects with the collaborative assistance of neuroscientists working with AI researchers. The goals of this project are cognitive computational models, such as those that will support the emerging standard brain model, and brain-inspired chips.

Conclusion

We are engaged in a rapidly accelerating area of research that is deeply fundamental to our awareness of the design of the brain and how its biology uses pattern recognition, data structure self-organization, and simplification of complex data, which, if we can incorporate these findings into our research model of the mind, promise to accelerate our capacity for creating artificial learners that can expand and improve on the design of biology. This area is engaging significant state-sponsored investment, both in the China Brain Project in China and also in the Human Connectome Project in the US, which will continue to fuel research. The promise of further exploration is bringing in greater collaboration across the academic disciplines that illuminate our understanding of the brain: artificial intelligence, linguistics, cognitive science and neuroscience. As we work to link together our research by using the new standard model of the mind, we need to include the fundamental ideas of computer science that come from the works of Rosenblatt, Kurzweil and Hawkins.

References

Dictionary, O. E. 2007. Oxford english dictionary online.

E, J.; Lebiere, C. L.; and Rosenbloom, P. S. 2017. A Standard Model of the Mind: Toward a Common Computational

Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *to appear in AI Magazine*.

Fine, S.; Singer, Y.; and Tishby, N. 1998. The hierarchical hidden markov model: Analysis and applications. *Machine learning* 32(1):41–62.

Fontana, A. 2017. A deep learning-inspired model of the hippocampus as storage device of the brain extended dataset. *arXiv preprint arXiv:1706.05932*.

George, D., and Hawkins, J. 2005. A Hierarchical Bayesian model of invariant pattern recognition in the visual cortex. *Proc. 2005 IEEE Int. Jt. Conf. Neural Networks, IJCNN'05.* 1812–1817.

Ghayoumi, M., and Bansal, A. K. 2016. Multimodal architecture for emotion in robots using deep learning. In *Future Technologies Conference (FTC)*, 901–907. IEEE.

Hawkins, J. and Ahmad, S. and Purdy, S. and Lavin, A. 2016. Biological and Machine Intelligence (BAMI). Initial online release 0.4.

Jolivet, R.; Coggan, J. S.; Allaman, I.; and Magistretti, P. J. 2015. Multi-timescale Modeling of Activity-Dependent Metabolic Coupling in the Neuron-Glia-Vasculature Ensemble. *PLoS Comput. Biol.* 11(2).

Kurzweil, R. 2013. *How to Create a Mind: The Secret of Human Thought Revealed*. New York, NY, USA: Penguin Books.

Marblestone, A. H.; Wayne, G.; and Kording, K. P. 2016. Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience* 10.

Markram, H.; Meier, K.; Lippert, T.; Grillner, S.; Frackowiak, R.; Dehaene, S.; Knoll, A.; Sompolinsky, H.; Verstreken, K.; DeFelipe, J.; Grant, S.; Changeux, J. P.; and Sariam, A. 2011. Introducing the Human Brain Project. In *Procedia Comput. Sci.*, volume 7, 39–42.

Markram, H. 2006. The blue brain project. *Nature Reviews Neuroscience* 7(2):153–160.

Nahal, S. 2017. *The Relationship Between Deep Learning and Brain Function: Proceedings of Student-Faculty Research Day*, volume May 5. Pleasantville, NY, USA: CSIS, Pace University.

2017. National institute of health human connectome project.

Opris, I., and Casanova, M. F. 2014. Prefrontal cortical minicolumn: From executive control to disrupted cognitive processing.

Opris, I.; Santos, L.; Gerhardt, G. A.; Song, D.; Berger, T. W.; Hampson, R. E.; and Deadwyler, S. A. 2013. Prefrontal cortical microcircuits bind perception to executive control. *Scientific reports* 3.

Park, F. 1998. The Hierarchical Hidden Markov Model : Analysis and Applications. *Machine Learning* 62:41–62.

Poo, M.-m.; Du, J.-l.; Ip, N. Y.; Xiong, Z.-Q.; Xu, B.; and Tan, T. 2016. China brain project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron* 92(3):591–596.

Rosenblatt, F. 1962. Principles of neurodynamics: perceptrons and the theory of brain mechanisms.

Van Essen, D. C.; Smith, S. M.; Barch, D. M.; Behrens, T. E.; Yacoub, E.; Ugurbil, K.; Consortium, W.-M. H.; et al. 2013. The wu-minn human connectome project: an overview. *Neuroimage* 80:62–79.

Veeriah, V.; Durvasula, R.; and Qi, G.-J. 2015. Deep learning architecture with dynamically programmed layers for brain connectome prediction. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1205–1214. ACM.