# The Global Workspace Theory: A Step Towards Artificial General Intelligence

1st Mohamed Abdelwahab
*Electrical and Computer Engineering Department*
*University of Toronto*
Toronto, Canada
mo.abdelwahab@mail.utoronto.ca

2nd Parham Aarabi
*Electrical and Computer Engineering Department*
*University of Toronto*
Toronto, Canada
p@arh.am

***Abstract*—Global Workspace Theory (GWT) and Artificial General Intelligence (AGI) are two concepts in cognitive science and Artificial Intelligence, respectively. This paper discusses the possibility of achieving AGI using a deep learning implementation of GWT. The shared latent space is trained using the latent spaces of the connected deep learning modules. This implementation aims to enhance the performance of specialized models in their specified tasks and achieve more general functions from single-task/specialized modules. The paper also discusses the possible applications of this implementation in healthcare.**

***Index Terms*—Artificial General Intelligence, Deep Learning, Global Workspace Theory**

## I. Introduction

In the past few years, there has been a continuous development for deep learning methods which has led to solving more diverse problems. Most of the developed methods, however, focus on solving single tasks, or what is termed as Narrow Artificial Intelligence, with less research being done on reaching Artificial General Intelligence (AGI) in which more general-purpose models are developed to perform a variety of tasks.

In another area of research related to defining consciousness and simulating or even creating conscious machines, the Global Workspace Theory (GWT) is one of the widely known theories that hypothesize how the conscious brain works. Some evidence for GWT is provided in neuroscience research [1]. Since it is one of the few theories that validate the possibility of artificial consciousness, some research is being done into incorporating GWT in artificial systems in an attempt to reach consciousness.

In this paper, one of the proposed implementations for GWT is reviewed. This proposed implementation is based on the combination of pretrained deep learning models' latent spaces to form a higher-level amodal latent space within a greater system. It uses unsupervised learning to build this latent space and uses an attention mechanism to give focus to certain modules depending on the context of the task. Rather than focusing on giving rise to artificial consciousness, we are more interested in the possibility of this implementation achieving AGI. We also highlight how such an implementation can be used in healthcare systems in the future.

## II. Background

GWT, illustrated in Fig. 1, is one of the most prominent functional theories explaining how consciousness emerges. It was first proposed in [2] and later updated in [3]. The theory posits that the brain is composed of multiple specialized unconscious modules that perform specific functions. These modules are interconnected through a global workspace, and they continuously compete to gain access to that workspace. Depending on the inputs or on the task requirement, one module will be able to access the global workspace and broadcast or share its contents with other distinct specialized modules. According to the theory, conscious awareness arises from this sharing of information between the different modules.

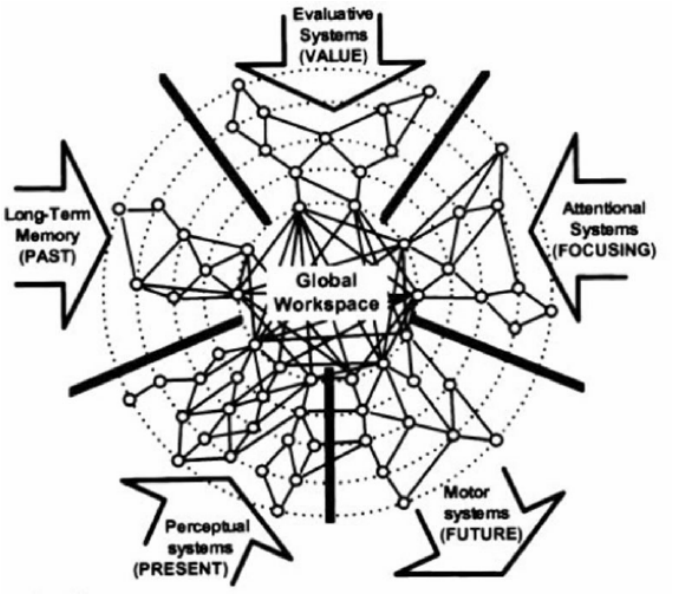


Fig. 1. Visualization of GWT

## III. Artificial GWT Implementation

There are a few papers linking GWT with deep learning [4], [5]. A possible deep learning implementation for GWT is illustrated in Fig. 2. Each deep learning model (also called module in this context) in this setup is pretrained on inputs

of different modalities. The global latent workspace (GLW) is a shared latent space which is trained using unsupervised cycle consistency [6]. Internal copies of the latent spaces of each module is used in the training of GLW. Assuming that the GLW is attending to a certain module (attention will be explained further ahead), the latent activity of this module is coupled with its internal copy and is translated via the GLW to the different latent spaces of the other modules. This is known as the broadcasting phenomenon. The translation fed to each internal latent space copy leads to their activation and thus they send back translations through the GLW. The translations and back-translations are used to measure the cycle consistency error for the GLW which is then modified using backpropagation. This translation and back-translation occurs continuously in a recurrent manner leading to the continuous learning and enhancement of the GLW. In this case, the GLW is an amodal latent space which learns higher-level concepts. The backpropagation here can also lead to the modification of the latent spaces and weight parameters of the generative models used in this system, leading to better information grounding of these modules i.e. the meaning of a certain representation in one domain is better acquired when related to other domains. The more modules are connected to the GLW, the better the GLW generalizes over the data and the more novel tasks can be solved by the system.

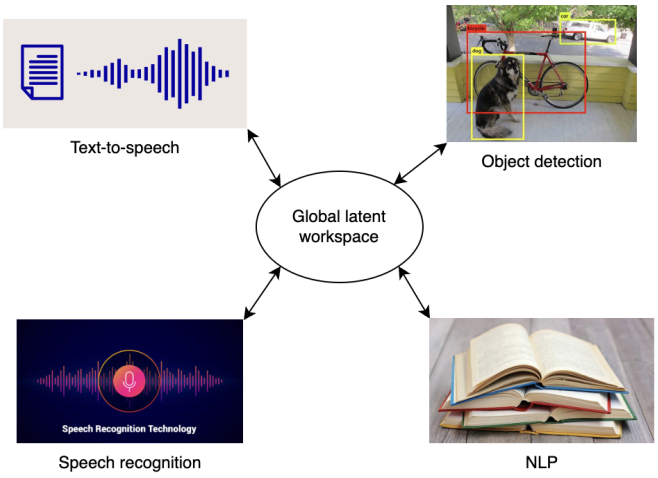


Fig. 2. Proposed implementation for GWT

The Global Workspace allows only certain modules to broadcast their content to the other modules. This is managed by an attention mechanism inside the Global Workspace. There are two forms of attention in this case, top-down attention, and bottom-up attention. Top-down attention occurs when a certain task signal is fed to the GLW which thus produces a corresponding attention-query that is continuously emitted to every module, and is compared with each key produced from the modules. The modules that have a query-key match are granted access to broadcast information in the GLW and thus have their latent spaces being coupled to their internal copies in the GLW. Bottom-up attention occurs when a module receives salient inputs, and would thus emit a master key that supersedes the query produced from the GLW. This is equivalent to having our brains bringing conscious awareness to exceptionally strong sensory inputs.

This implementation of GWT is advantageous in that, due to the grounding effect accomplished in each module, these modules should perform better in their specified tasks than they would have if they were operating independently. Furthermore, the ability to connect different combinations of modules to the GLW entails its ability to achieve more general functions from these single-task/specialized modules.

## IV. Applications of GWT to Healthcare

The mentioned implementation of GWT can be applied to many fields, including healthcare. The authors of [7] mention the success of current deep learning models in tasks such as clinical imaging and processing electronic health records. They predict that deep learning models should, in the future, be able to make disease risk prediction, produce personalized prescriptions and treatment recommendations. The multi-modal aspect of the mentioned GWT implementation allows it to combine different types of data from which decisions can be made. It can thus combine the information from the medical scans of a patient along with their medical history, to produce a prescription or a treatment plan for that patient.

## V. Conclusion

This paper reviews a possible implementation for the GWT. The implementation demonstrates how the architecture can continuously learn from the input stimuli and is able to give awareness to certain stimuli. It also shows how such an implementation can be used in healthcare applications. One limitation to this proposed implementation would be the huge computational power required to execute the continuous recurrent learning that should occur in the GLW. A future step that can be taken towards achieving this implementation would be to use small modules and train the GLW with them. This would enable the preliminary testing of the capabilities of such an implementation.

## References


[1] B. J. Baars, N. Geld, and R. Kozma, "Global Workspace Theory (GWT) and Prefrontal Cortex: Recent Developments," Frontiers in Psychology, vol. 12, Nov. 2021.

[2] B. J. Baars, "A Cognitive Theory of Consciousness," Cambridge University Press, 1988.

[3] B. J. Baars, "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience," Progress in Brain Research, vol. 150, pp. 45-53, 2005.

[4] Y. Bengio, "The Consciousness Prior," arXiv preprint arXiv:1709.08568, Sep. 2017. [Online]. Available: https://arxiv.org/abs/1709.08568

[5] A. Juliani, R. Kanai, and S. Sasai, "The Perceiver Architecture is a Functional Global Workspace," Proceedings of the 44th Annual Conference of the Cognitive Science Society, vol. 44, 2022.

[6] R. VanRullen and R. Kanai, "Deep learning and the global workspace theory," Trends in Neurosciences, vol. 44, no. 9, pp. 692-704, 2021.

[7] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236-1246, Nov. 2018.