Third **Summer School, N3BG, TU Sofia 2025**

Towards Machine Consciousness: The Brain-inspired Approach

Prof Nikola Kasabov

Life FIEEE, FRSNZ, FINNS, DVF RAE UK
Doctor Honoris Causa Obuda University Budapest
Visiting Professor IICT/Bulgarian Academy of Sciences; Advisor TU Sofia
Founding Director KEDRI and Professor, Auckland University of Technology, NZ
https://academics.aut.ac.nz/nkasabov; Director, Knowledgeengineering.ai

Dr Alexander Yang

Manabridge.com, New Zealand
alexanderyanghx@gmail.com

Dr Iman AbouHassan

Technical University of Sofia
iabouhassan@tu-sofia.bg

HEALTH & LIFE SCIENCES CLUSTER

INTERNATIONAL NEURAL NETWORK SOCIETY

conscium

about us our thesis news explainers faqs contact us

What is machine consciousness (MC) and
can we create AI that is safer, more efficient and more ethical?

Pioneering Safe, Efficient AI.

- Consciousness
- Neuromorphic computing
-
- Verification

www.conscium.com

KECL: www.knowledgeengineering.ai

Content

1. A brief review of current theories of human and machine consciousness (MC)
2. Introducing *Evolving spatio-temporal associative memories* (ESTAM) for MC
3. ESTAM in the human brain
4. Brain –inspired SNN (BI-SNN). NeuCube and eXCube.
5. ESTAM on eXCube for modelling meaningful vs meaningless visual perception by humans
6. Current and future applications

1. A brief review of current theories of human and machine consciousness

Representation of consciousness from the 17th century by Robert Fludd, an English Paracelsian physician

"Consciousness (C.) is the state of being aware of oneself and one's surroundings, encompassing thoughts, feelings, sensations, and the experience of the world.

Key Aspects:

- **Self-Awareness:** Consciousness often includes self-awareness, the ability to recognize oneself as a distinct individual with thoughts, emotions, and a sense of identity.
- **Perception:** It encompasses perception, the process of interpreting sensory information from the external world.
- **Subjective Experience:** Consciousness is fundamentally a subjective experience, meaning it's unique to each individual.

Some popular theories and models

1. Tononi, G. & Edelman, G. M. Consciousness and complexity. Science 282, 1846–1851. https://doi.org/10.1126/science.282.5395. 1846 (1998)
2. G.Tononi, An information integration theory of consciousness. BMC Neurosci. 5, 42 (2004).
3. G Findlay, W Marshall, L Albantakis, I David, WGP Mayner, C Koch, G Tononi, Dissociating Artificial Intelligence from Artificial Consciousness, arXiv preprint arXiv:2412.04571, 2024•arxiv.org
4. Tononi, G., Boly, M., Massimini, M. et al. Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17, 450–461 (2016). https://doi.org/10.1038/nrn.2016.44
5. Tononi, G. The integrated information theory of consciousness: an updated account. Arch. Ital. Biol. 150, 56–90 (2012).
6. Tononi, G. Integrated information theory. Scholarpedia http://dx.doi.org/10.4249/scholarpedia.4164 (2015).
7. Koch, C., Massimini, M., Boly, M. & Tononi, G. The neural correlates of consciousness: progress and problems. Nat. Rev. Neurosci. 17, 307–321 (2016).
8. Boly, M. et al. Consciousness in humans and non-human animals: recent advances and future directions. Front. Psychol. 4, 625 (2013).
9. S.Grossberg, Conscious Mind, Resonant Brain: How Each Brain Makes a Mind, Oxford University Press, 2023.
10. J.Taylor, *The race for consciousness*, MIT Press, book. 1999
11. I.Alexander, 2005, *The World in My Mind, My Mind In The World: Key Mechanisms of Consciousness in Humans, Animals and Machines* published by Imprint Academic, 2025
12. Boly M, Sasai S, Gosseries O, Oizumi M, Casali A, Massimini M, et al. (2015) Stimulus Set Meaningfulness and Neurophysiological Differentiation: A Functional Magnetic Resonance Imaging Study. PLoS ONE 10(5): e0125337. https://doi.org/10.1371/journal.pone.0125337
13. Lawrence M. Ward, The thalamic dynamic core theory of conscious experience, Consciousness and Cognition, 20, 2, 2011, Pages 464-486, https://doi.org/10.1016/j.concog.2011.01.007.
14. Orłowski, P., Bola, M. Sensory modality defines the relation between EEG Lempel–Ziv diversity and meaningfulness of a stimulus. Sci Rep 13, 3453 (2023). https://doi.org/10.1038/s41598-023-30639-3
15. Grossberg S (2023) How children learn to understand language meanings: a neural model of adult–child multimodal interactions in real-time. Front. Psychol. 14:121647
16. Koch, C. (2023). What does it 'feel' like to be a chatbot? Scientific American. https://www.scientificamerican.com/article/what-does-it-feel-like-to-be-a-chatbot/
17. M.Butlin, T.Lappas, Principles for Responsible AI Consciousness Research, Journal of AI Research, 2024.
18. Chalmers, David (1996), The Conscious Mind, Oxford University Press, ISBN 978-0-19-510553-7
19. Koch, Christof (2004), The Quest for Consciousness: A Neurobiological Approach, Pasadena, CA: Roberts & Company Publishers, ISBN 978-0-9747077-0-9

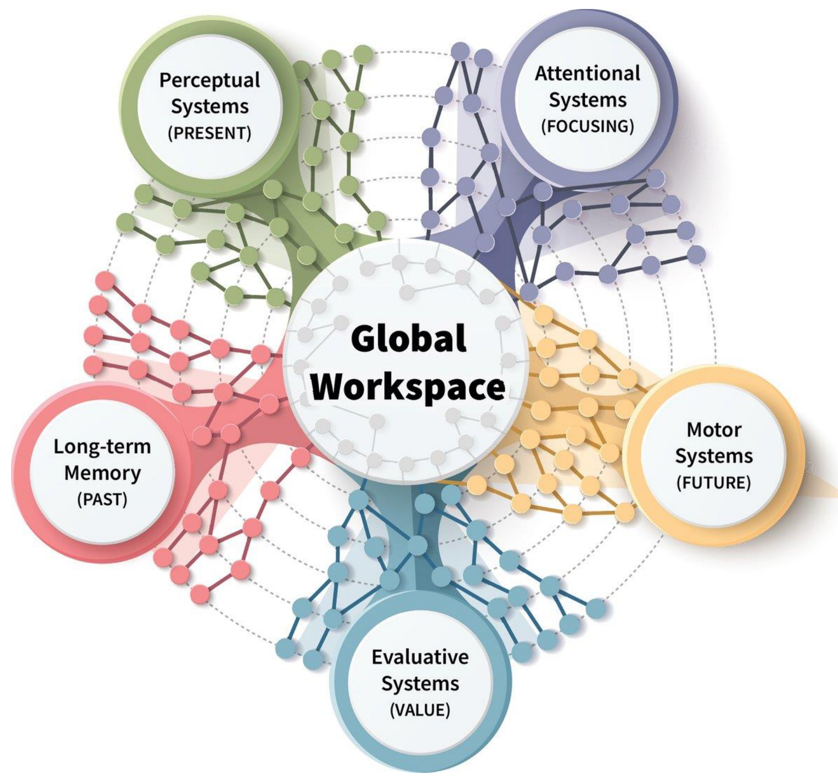
Global Workspace Theory (GWT)
Core Idea
•A central "blackboard" where multiple modules integrate and broadcast content brain-wide

Global Neuronal Workspace (GNW)
Architecture: A distributed "blackboard" formed by long-range pyramidal neurons (prefrontal, parieto-temporal, cingulate) that integrate and broadcast information across specialized local processors .

Hierarchal: bottom up (towards central unit), top-down feedback (from central unit).

Ignition: Nonlinear "all-or-none" activation of GNW neurons enables conscious broadcast.
Bottleneck: Limited capacity ensures only a few items enter awareness at once.
Sequential Experience: Recurrent loops sustain and shift conscious content over time.

GNW Model, Share – Hold – Disseminate



Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research*, *150*, 45-53.
Dehaene S. Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts. New York: Penguin, 2014.

@DaphnePerlman, Oryan Zacks, Eva Jablonka, The evolutionary origins of the Global Neuronal Workspace in vertebrates, *Neuroscience of Consciousness*, Volume 2023, Issue 1, 2023, niad020, https://doi.org/10.1093/nc/niad020

Unlimited Associative Learning (UAL) theory

Consciousness can arise from a plurality of brain structures.

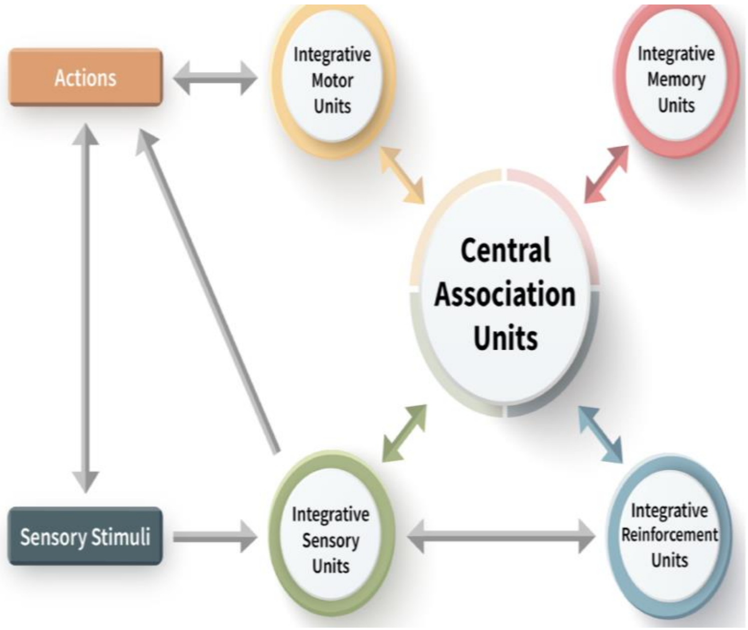
Minimal Consciousness Marker: UAL requires four learning capacities which form joint sufficiency for consciousness. They require *unified attention*, *working memory*, *value-based decision-making*, and *hierarchical association*.

1) Discrimination: Feature based – shapes, colours, features, composite-actions. (*classification*)

2) Flexible weighed outcomes: shifts in motivational significance e.g., devaluing sugar due to pre-feeding. (*connection weights*)

3) Trace conditioning: Maintains temporary memory trace of stimuli to predict outcome, even with timelag. (*memory of temporal events*)

4) Second-order conditioning— associated stimuli predict the same outcome, e.g., light paired with auditory tone predicts food. (*associative memory + prediction*)



Oryan Zacks, Eva Jablonka, The evolutionary origins of the Global Neuronal Workspace in vertebrates, *Neuroscience of Consciousness*, Volume 2023, Issue 1, 2023, niad020, https://doi.org/10.1093/nc/niad020

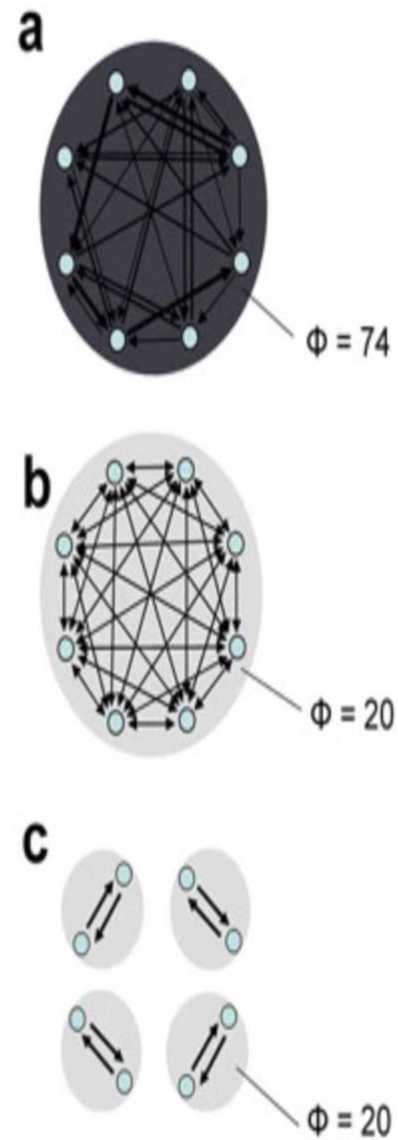
Integrated Information Theory (IIT):

Consciousness can arise from any physical system that can process integrated information
Level of integrated information determines degree of consciousness.

"Φ is the amount of causally effective information that can be integrated across the informational weakest link of a subset of elements"

Φ is taken as the integrated information across the Minimum information partition—i.e. the amount of information "lost" when the system is partitioned at its weakest link. This ensures Φ quantifies only the **irreducible** (inseparable) portion of the system's causal structure

Application to neural networks:
Compute Φ on small recurrent/feedback networks



Functional specialization + integration

Reduction information integration through loss of specialization

Reduction of information integration through loss of functional integration

Tononi, G. An information integration theory of consciousness. *BMC Neurosci* **5**, 42 (2004). https://doi.org/10.1186/1471-2202-5-42

G.Tononi, An information integration theory of consciousness. BMC Neurosci. 5, 42 (2004).

“,,The IIT is based on five axioms of phenomenal existence: every experience is intrinsic (for itself), specific (this one), unitary (a whole, irreducible to separate experiences), definite (this whole, containing all it contains, neither less nor more), and structured (composed of distinctions bound by relations that make it feel the way it feels).

These five postulates—(1) intrinsicality, (2) information, (3) integration, (4) exclusion, and (5) composition—can be formulated mathematically and assessed algorithmically for any substrate, given the system’s state and complete causal model. The analysis identifies systems that can support consciousness, called complexes. IIT claims that the quality of an experience— “what it is like to be” in a specific phenomenal state—is fully accounted for, with no additional ingredients, and its is a cause–effect structure…”

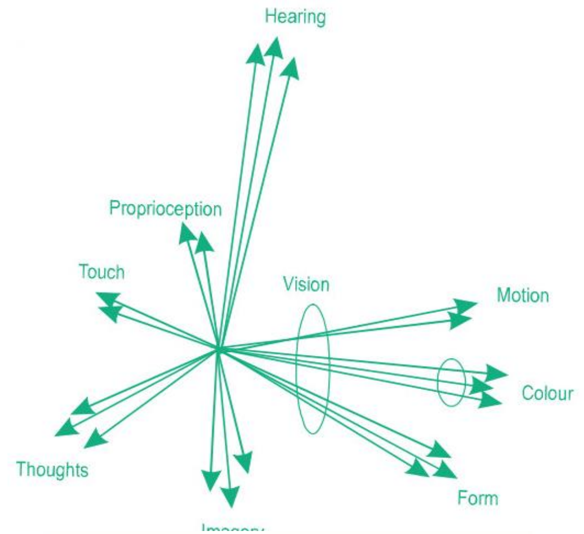
G Findlay, W Marshall, L Albantakis, I David, WGP Mayner, C Koch, G Tononi, Dissociating Artificial Intelligence from Artificial Consciousness, arXiv preprint arXiv:2412.04571, 2024•arxiv.org

“ …Developments in machine learning and computing power suggest that artificial general intelligence is within reach. This raises the question of *artificial consciousness*: if a computer were to be functionally equivalent to a human, being able to do all we do, would it experience sights, sounds, and thoughts, as we do when we are conscious? By applying the principles of IIT, we demonstrate that (i) two systems can be functionally equivalent without being phenomenally equivalent, and (ii) that this conclusion is not dependent on the simulated system’s function. We further demonstrate that, according to IIT, it is possible for a digital computer to simulate our behavior, possibly even by simulating the neurons in our brain, without replicating our experience. This contrasts sharply with computational functionalism, the thesis that performing computations of the right kind is necessary and sufficient for conscious

FOR and AGAINST computational functionalism…………

The thalamic dynamic core model of consciousness

Lawrence M. Ward, The thalamic dynamic core theory of conscious experience, Consciousness and Cognition, 20, 2, 2011, Pages 464-486, https://doi.org/10.1016/j.concog.2011.01.007.



The Dynamic Core: This is the synchronization in the brain that involves a complex interplay of inhibitory and excitatory interactions, as well as spike-timing-dependent synaptic plasticity. The dynamic core consists of multiple clusters of neurons, which enter the core temporarily based on their mutual consistency. This synchronizing process requires strong recurrent connections between brain areas for signals to flow back and forth.

Synchronization: Neurons discharge (fire) with the same frequency and the same phase, creating a pattern of simultaneously firing neurons. This transient synchronization of firing neurons binds features processed in different brain regions, integrating them into a single percept based on the underlying synaptic connectivity.

It is **N-dimensional** space, where each axis (dimension) represent participating group of neurons that represent a certain aspect consciousness. There can be hundreds of thousands of dimensions. distance from the beginning of the axis shows the importance of that asp which corresponds to the number of firing neurons in a certain group.

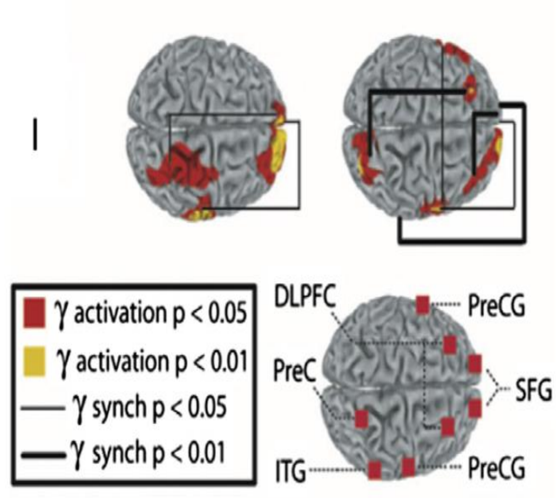


Fig. 11. The cortical consciousness network at 35–45 Hz during binocular rivalry. DLPFC = dorsolateral prefrontal cortex; SFG = superior frontal gyrus; PreCG = precentral gyrus; PreC = precuneus; ITG = inferior temporal gyrus. Reprinted with permission from Doesburg et al. (2009).

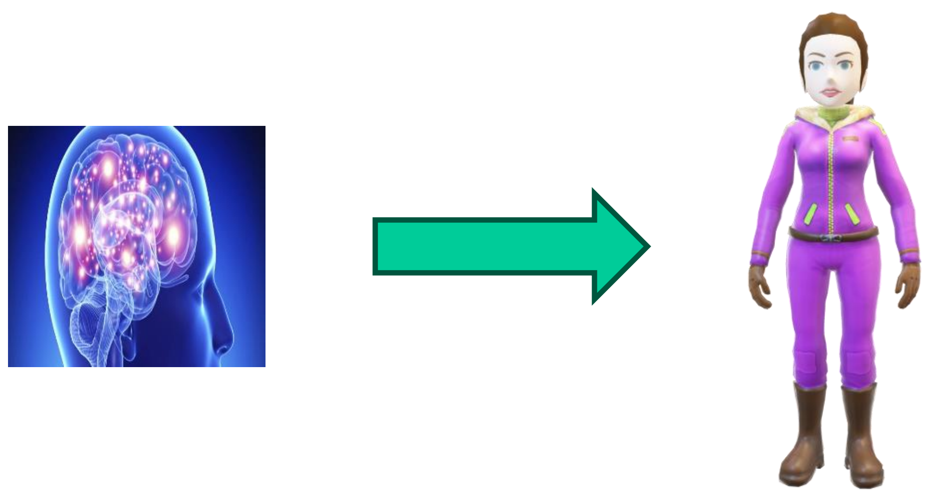
Machine (Artificial) C.

We talk about Machine Consciousness (MC) if a machine can manifest some or all of the features of C. as defined above.

Research questions:

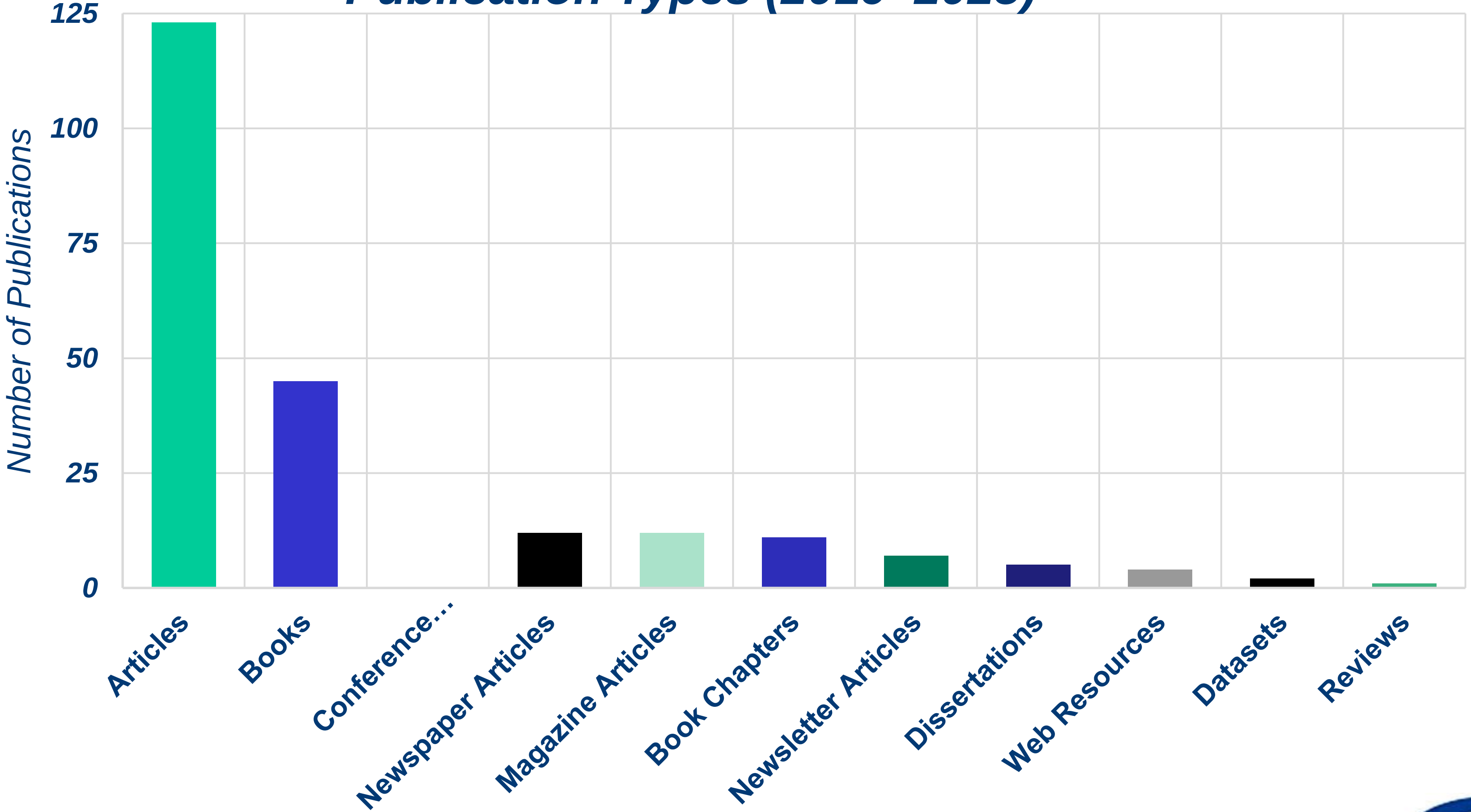
1. How is C. manifested in the brain?
2. Can a brain-inspired machine manifest elements of MC?

Our approach: From understanding of C. in the human brain, to the realisation of MC in brain-inspired machines

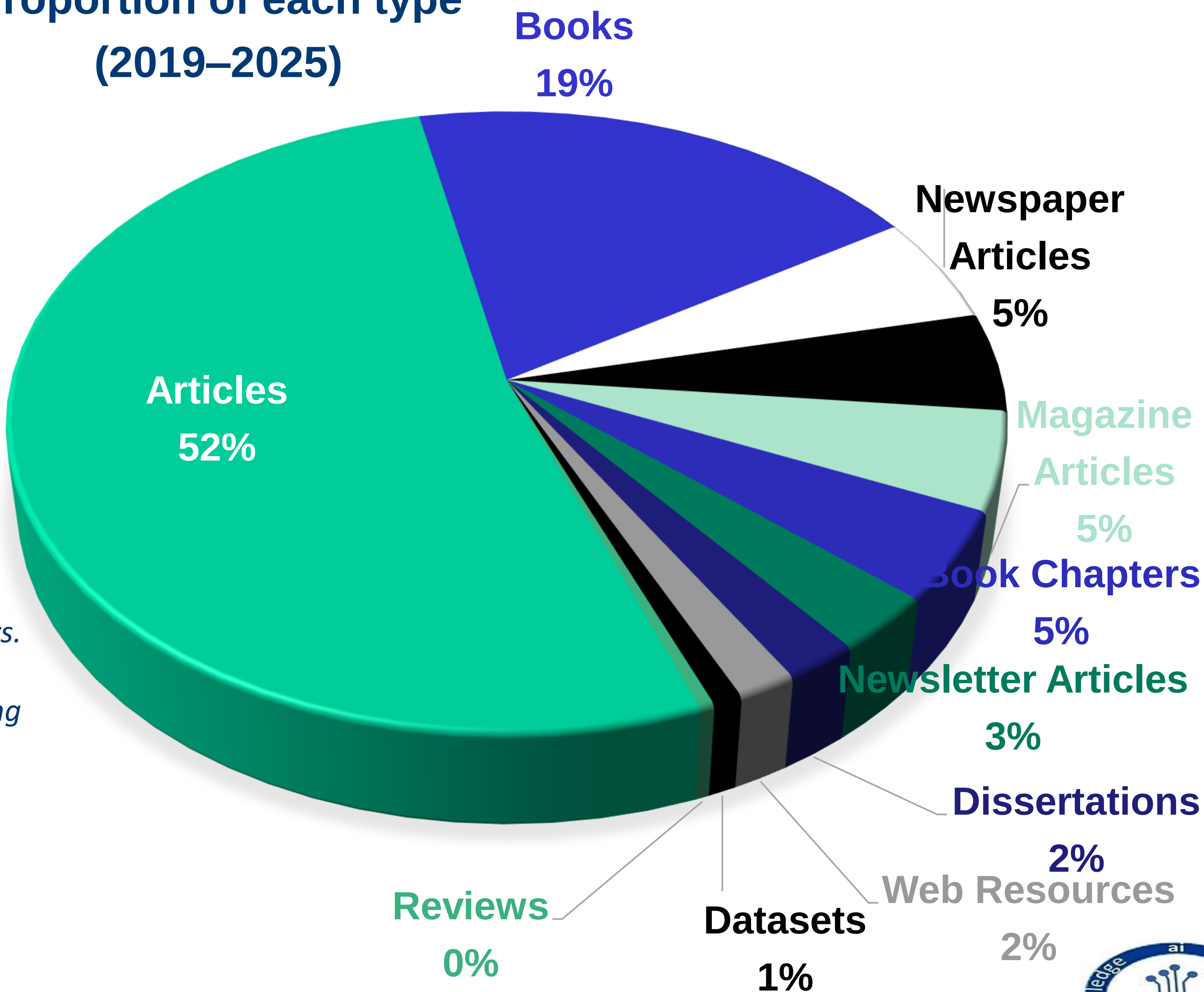


AI & Machine C. (Publication Landscape)

Publication Types (2019–2025)



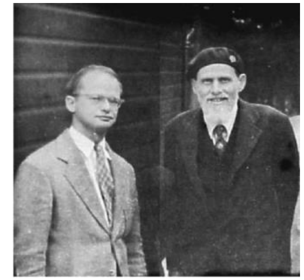
Proportion of each type (2019–2025)



Most works are peer-reviewed journal articles (52%), followed by books and conference papers.

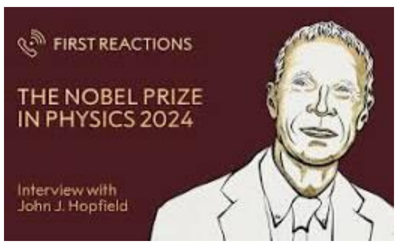
This reflects strong academic and interdisciplinary engagement.

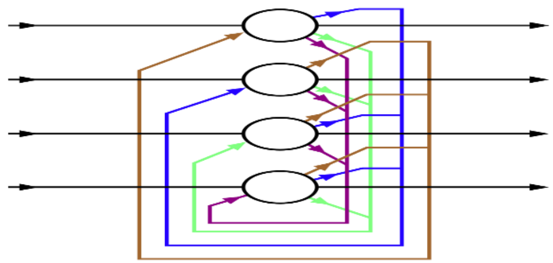
Early neural networks

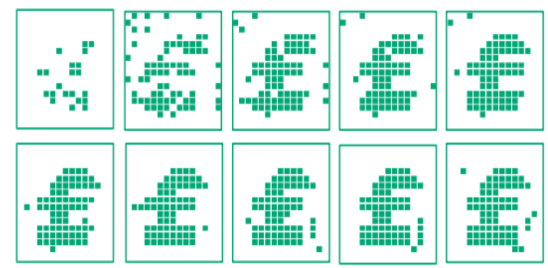


History of ANN:

- 1943, **McCulloch and Pitts** neuron (photo 1949)→
- 1962, Rosenblatt – Perceptron
- 1971- 1986, **Amari,** Rumelhart, Werbos, Hinton: Multilayer perceptron
- Self-organising maps (SOM), Kohonen
- Adaptive resonance theory (Grossberg)
- **J. Hopfield** networks as AM







- “The Hopfield network provides a simple model of an associative memory in a neuronal structure. It is, however, based on highly artificial assumptions, especially the use of formal two-state neurons or graded-response neurons”,
- Wolfram Gerstner and J Leo van Hemmen, Associative memory in a network of 'spiking' neurons, Network 3 (1992) 139-164.
- Bi-directional AM (Bart Kosko)

- N. Kasabov (1996), Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering, MIT Press, Cambridge, Massachusetts, USA, 1996.

2. Introducing Evolving spatio-temporal associative memory (ESTAM) for MC

An ESTAMs is a dynamic model in natural or artificial systems that associates and captures incrementally and continuously related items, objects, processes of different modalities and scales, such as spatial and temporal scales.

ESTAM agrees with all postulates proposed for the definition of C., such as:
(1) intrinsicality, (2) information, (3) integration, (4) exclusion, and (5) composition (G.Tononi, An information integration theory of consciousness. BMC Neurosci. 5, 42 (2004))

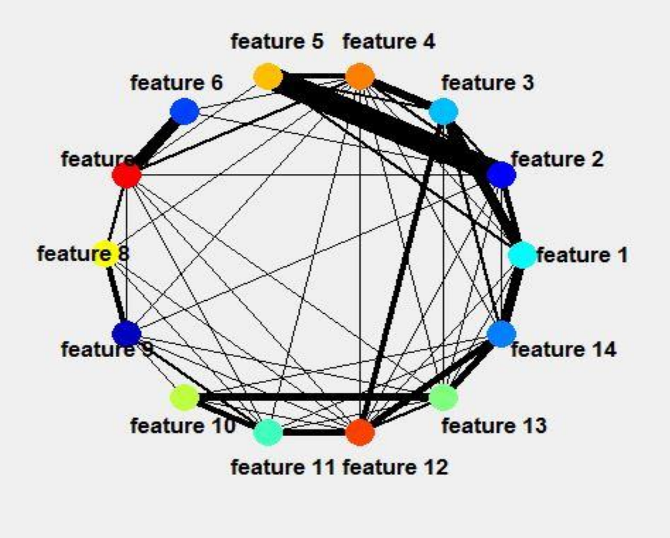
ESTAM is an emerging spatio-temporal pattern in *Time* which exists in a given time window, but including past- and future times as well.

Example: A dynamic model of 14 variables (features) that define a dynamic state of an object or a human brain through their interaction in time (the thicker the line, the more interactions are manifested in time)

“ Времето е в нас и ние сме във времето“

“Time lives inside us and we live inside Time.’

Vasil Levski-Apostola (1837-1873)
Bulgarian educator and revolutionary



ESTAM is a model that can be trained on a full set of spatio-temporal variables of different modalities and can be successfully recalled on only a subset of the variables measured in shorter time intervals (manifesting prediction).

ESTAM can be further incrementally evolved on a new subset of variables measured at varying times (evolvability in time and space).

ESTAM evolves both **generic (abstraction**) and **specific** (**individual**) patterns from data .

Overall, **ESTAM** adds to the existing theories of MC by introducing [1,2]:

- Brain-inspired model
- Time of emergence and existence of patterns of activities in a model;
- Evolvability of a model on new stimuli and data;
- Predictability, when a model is recalled with incomplete inputs in space and time.

1. Kasabov, N (2023). STAM-SNN: Spatio-Temporal Associative Memories in Brain-inspired Spiking Neural Networks: Concepts and Perspectives. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.23723208.v1

2. Kasabov, N.K. (2024). STAM-SNN: Spatio-Temporal Associative Memory in Brain-Inspired Spiking Neural Networks: Concepts and Perspectives. In: Kovács, L., Haidegger, T., Szakál, A. (eds) Recent Advances in Intelligent Engineering. Topics in Intelligent Engineering and Informatics, vol 18. Springer, https://doi.org/10.1007/978-3-031-58257-8_1

3. ESTAM in the human brain

The brain (80bln neurons, 100 trillions of connections, 200 mln years of evolution) is the ultimate learning machine

Three, mutually interacting, spatial memory types:

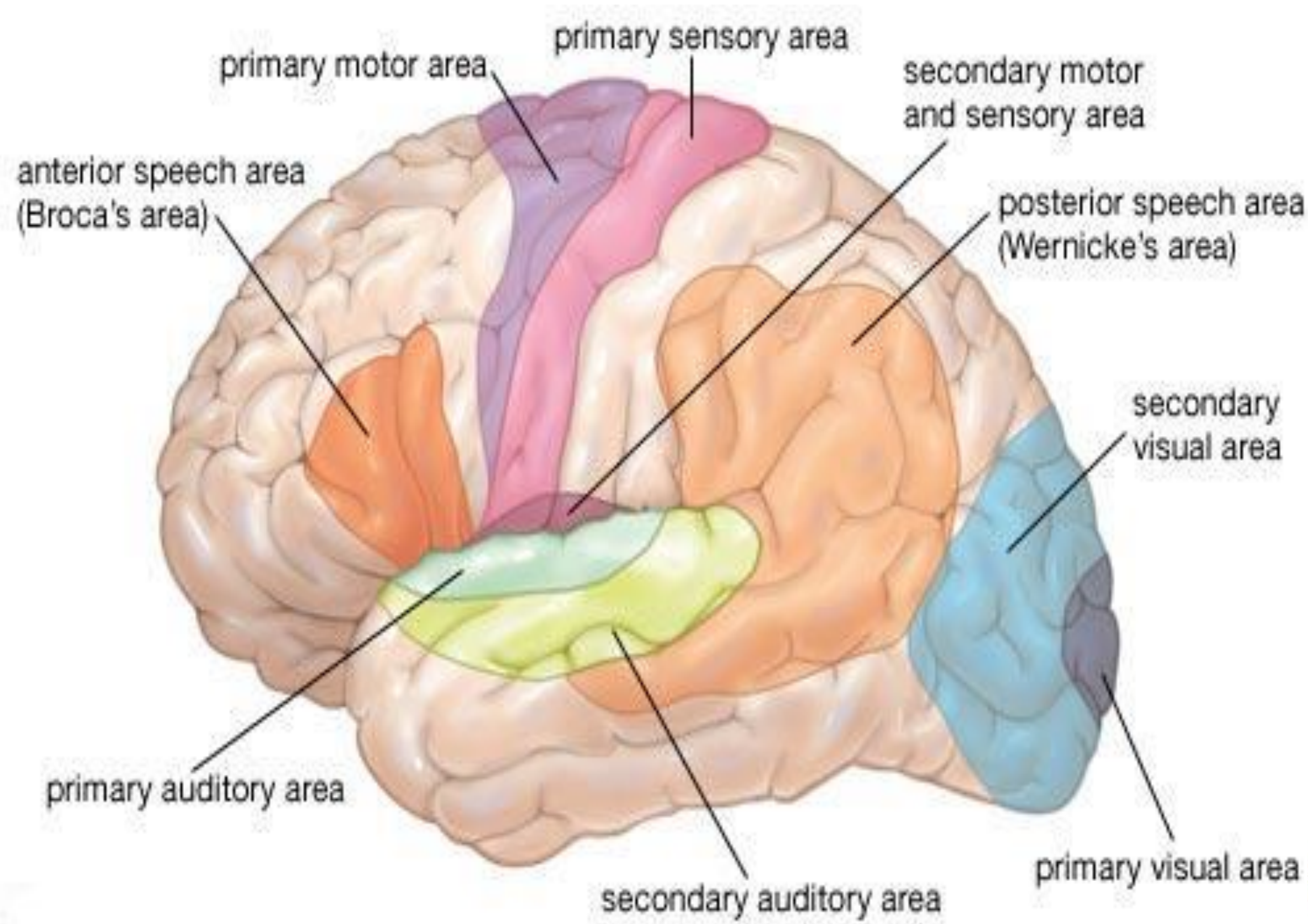
- short term (membrane potential);
- long term (synaptic weights);
- genetic (genes in the nuclei).

Temporal data at different time scales:

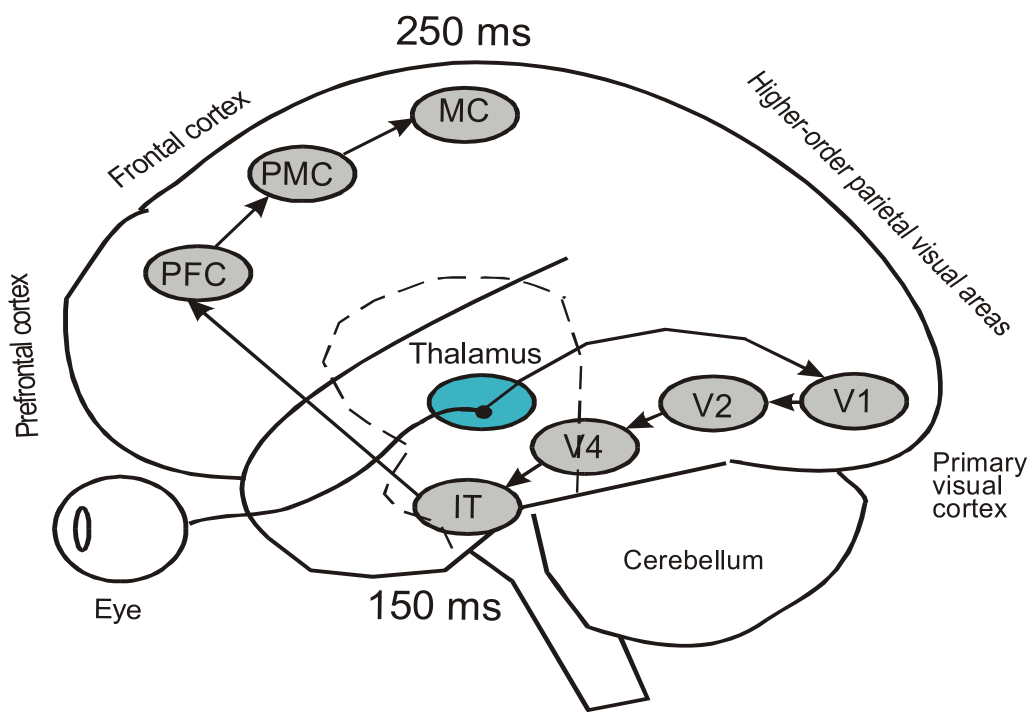
- Nanoseconds: quantum processes;
- Milliseconds: spiking activity;
- Minutes: gene expressions;
- Hours: learning in synapses;
- Many years: evolution of genes.

Knowledge is represented as deep **spatio-temporal patterns** that can evolve/adapt over time. Once created, these patterns can be activated by partial and imprecise spatio-temporal information.

Different **spatially** distributed parts of the brain control different functions and they are connected through neuronal connections to make ESTAM



Spatial areas are connected in time to make ESTAMs



Deep serial processing of visual stimuli in humans for image classification and action.

Location of cortical areas: V1 = primary visual cortex, V2 = secondary visual cortex, V4 = quartiary visual cortex, IT = inferotemporal cortex, PFC = prefrontal cortex, PMC = premotor cortex, MC = motor cortex.

L.Benuskova, N.Kasabov, Computational neurogenetic modelling, Springer, 2007

Principles of Information Representation and Recall in the Brain. The brain evolves its knowledge as ESTAMs

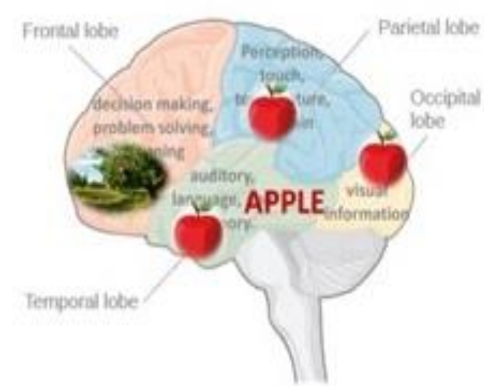
Redundancy: Information is stored, transmitted, and processed by a large number of neurons and synapses so that it does not become lost when neural networks undergo damage, e.g. ageing => brain degrades gracefully, not abruptly.

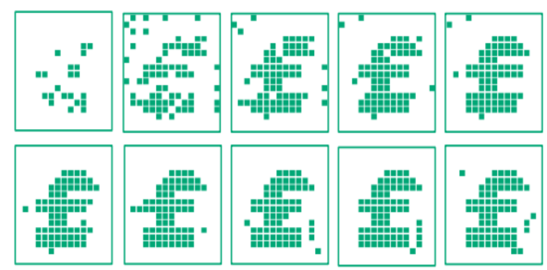
Spatial Distribution: Information in the brain is stored across various neural networks
=> which supports holistic perception, e.g. Apple (sound, image, smell, memory, …).

Active Recall Process: Retrieval from memory is an active process using partial patterns to reconstruct a complete memory. This process is reflected in the SNN association's ability to generate patterns of neural representation.

Synaptic Weights: Neural patterns are stored (encoded) in the matrix of synaptic weights
(connections), where the strength of each connection reflects learned experiences. When a significant part of this encoded pattern is externally stimulated, it triggers a rapid response across the network, activating the entire pattern.

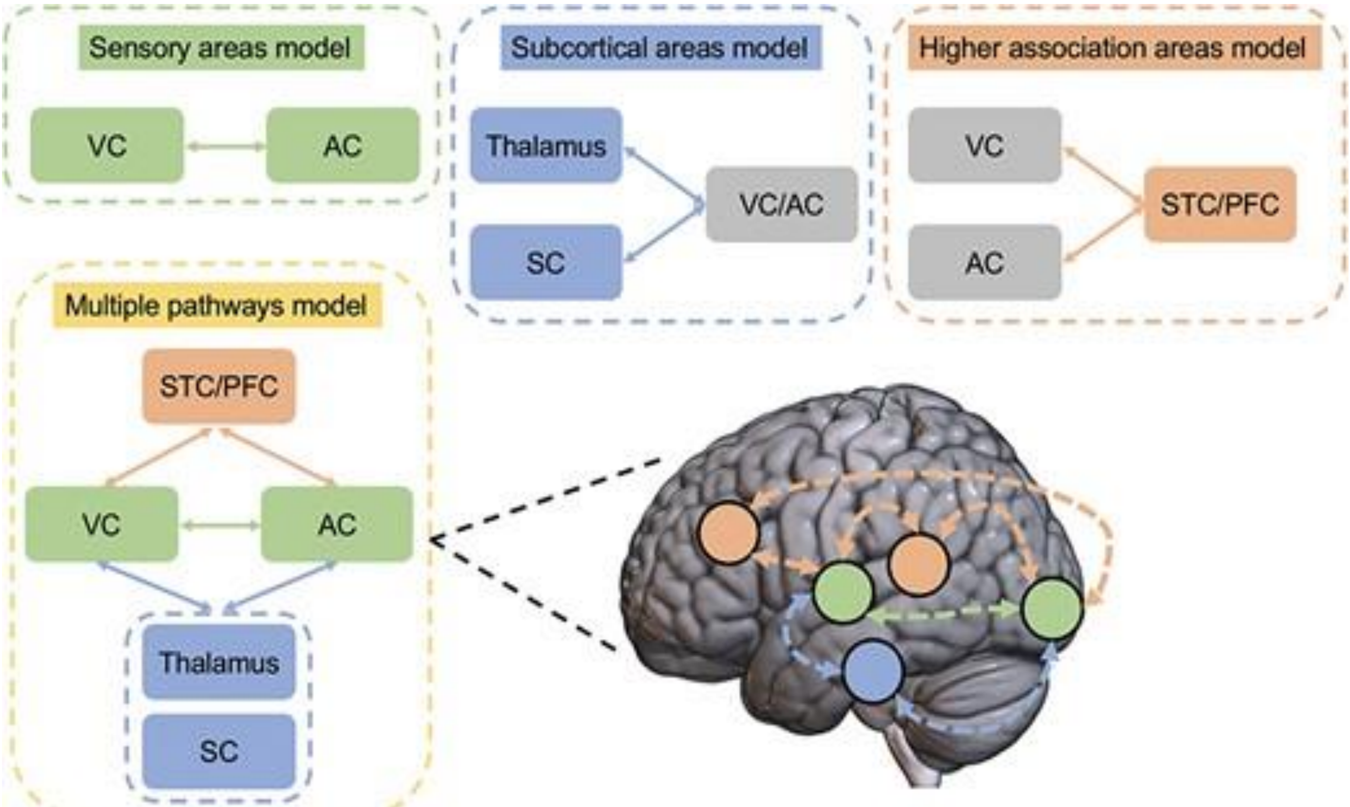
Holistic Nature: Neural patterns representing information are restored and perceived as complete, integrated wholes, consistent with Gestalt principles.





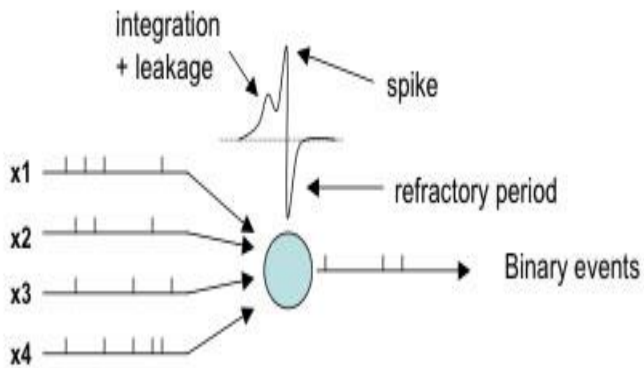
ESTAM of the audio-visual system in the brain

- Sensory areas model posits that direct connections between visual cortex (VC) and auditory cortex (AC) are underlying the audiovisual integration process. Subcortical areas model posits that audiovisual integration takes place at subcortical areas such as thalamus or superior colliculus (SC).



Chuanji Gao, Jessica J Green, Xuan Yang, Sewon Oh, Jongwan Kim, Svetlana V Shinkareva, Audiovisual integration in the human brain: a coordinate-based meta-analysis, *Cerebral Cortex*, Volume 33, Issue 9, 1 May 2023, Pages 5574–5584, https://doi.org/10.1093/cercor/bhac443

4 **Brain –inspired SNN (BI-SNN). NeuCube and eXCube.**

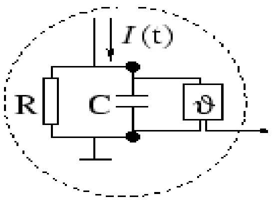


Information processing in Spiking neural networks (SNN)

- Trains of spikes
- Time, frequency and space
- Synchronisation and stochasticity
- Evolvability…

Spiking neuron models

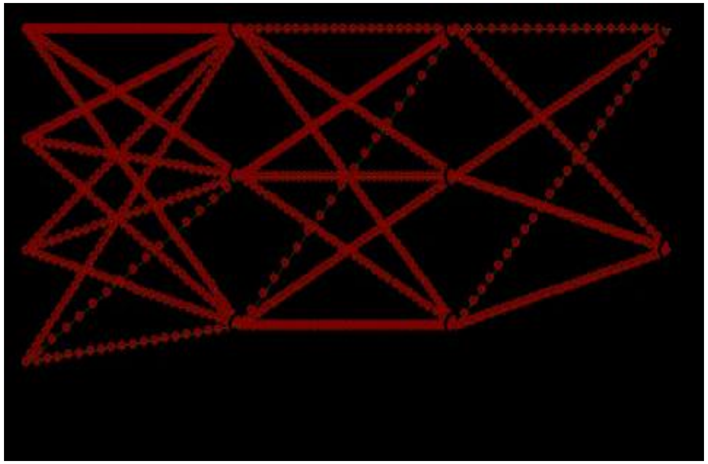
- Leaky Integrate-and-fire
- Probabilistic model
- Neurogenetic model



$$\tau_m \frac{du}{dt} = -u(t) + RI(t)$$

They offer the potential for:

- Spatio-temporal data processing
- Bridging higher level functions and "lower" level genetics
- Integration of modalities



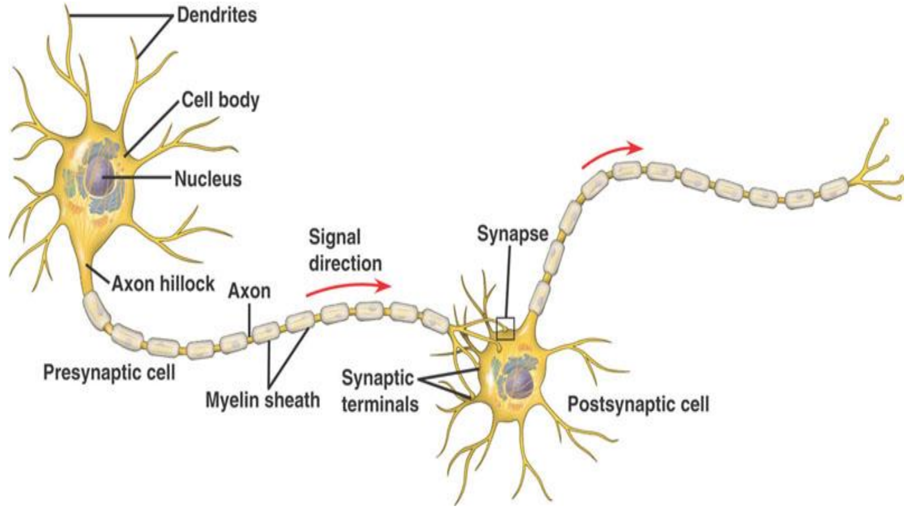
SNN open the field of brain-inspired (cognitive, neuromorphic) computing.

"*The goal of brain-inspired computing is to deliver a scalable neural network substrate while approaching fundamental limits of time, space, and energy,*" IBM Fellow **Dharmendra Modha**, chief scientist of Brain-inspired Computing at IBM R

Methods for unsupervised learning in SNN

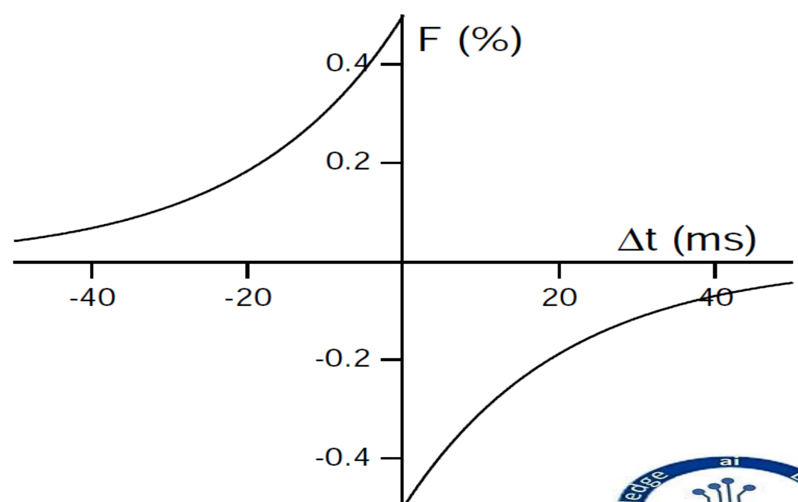
Spike-Time Dependent Plasticity (STDP)
(Abbott and Nelson, 2000).

- Hebbian form of plasticity in the form of long-term potentiation (LTP) and depression (LTD)
- Effect of synapses are strengthened or weakened based on the timing of pre-synaptic spikes and post-synaptic action potential.
- Through STDP connected neurons learn consecutive spatio-temporal associations from data.
- Variations of the STDP



- Pre-synaptic activity that precedes post-synaptic firing can induce LTP, reversing this temporal order causes LTD: $\Delta t = t_{pre} - t_{post}$

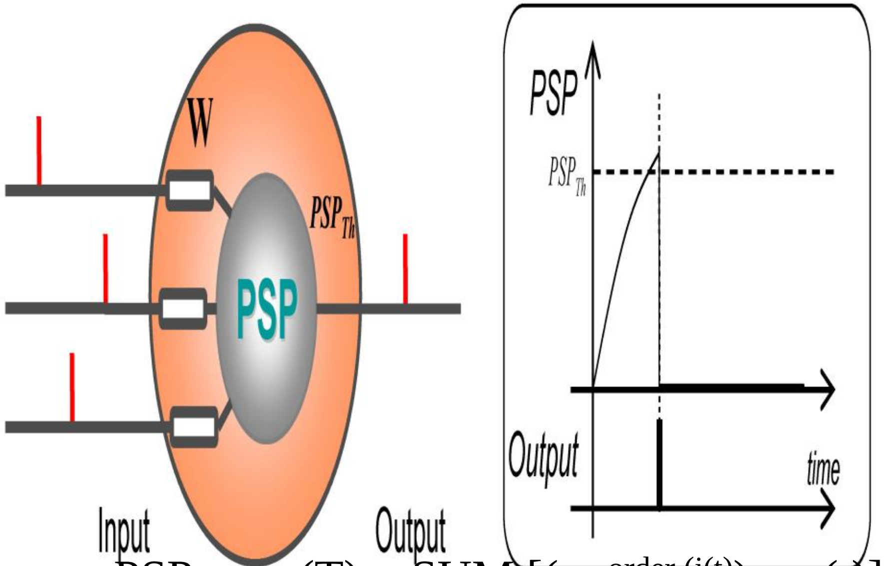
- Predictive spiking activity of neurons



nkasabov@aut.ac.nz

Methods for (predictive) supervised learning in SNN as ESTAM

Rank order (RO) learning rule (Thorpe et al, 1998)



$$\Delta w_{ji} = m^{\text{order}(j)}$$

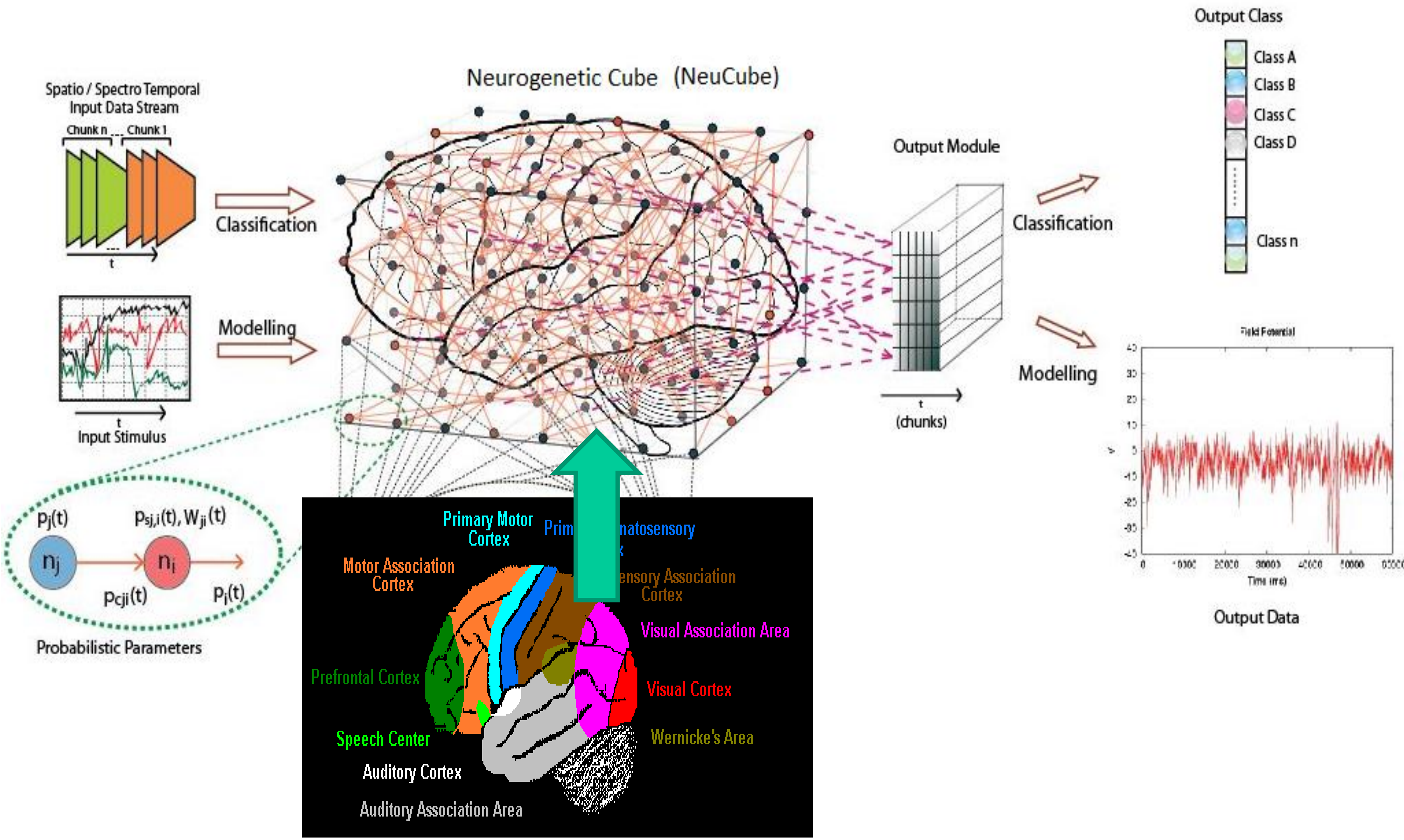
$$u_i(t) = \begin{cases} 0 & \text{if fired} \\ \sum_{j|f(j)<t} w_{ji} m_i^{\text{order}(j)} & \text{else} \end{cases}$$

PSP max (T) = SUM $[(m^{\text{order}(j(t))}) \, w_{j,i}(t)]$, for j=1,2.., k; t=1,2,...,T;

PSP_{Th}=C. PSPmax (T)

- Earlier coming spikes are more important (carry more information)
- Early spiking can be achieved, depending on the parameter C.
- Predictive spiking activity

The NeuCube BI-SNN architecture (3D space + 1D time)

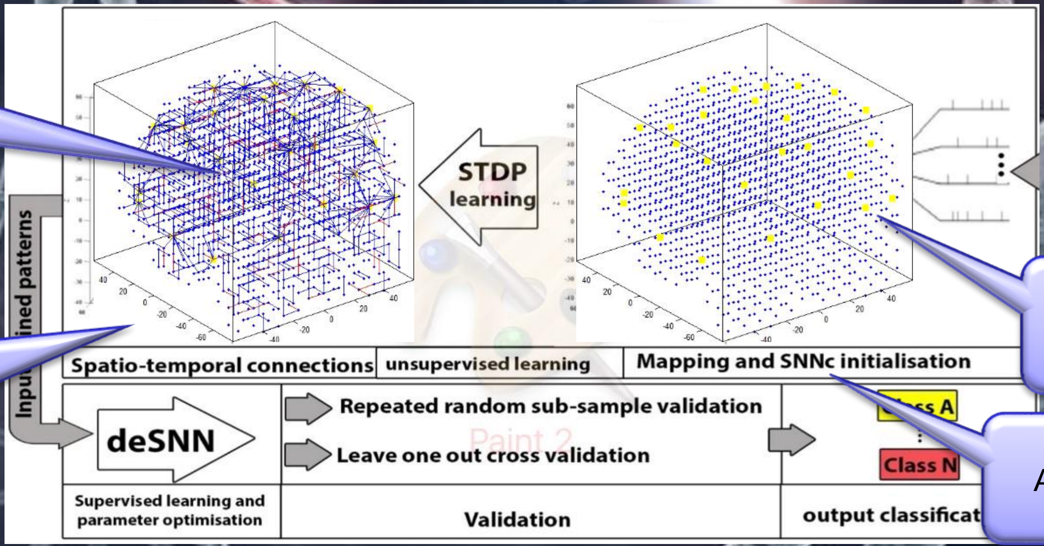


Kasabov, N., NeuCube: A Spiking Neural Network Architecture for Mapping, Learning and Understanding of Spatio-Temporal Brain Data, Neural Networks, vol.52, 2014.

Unsupervised and supervised learning in NeuCube

Creation of Neuron Connections During The Learning

The More Spike Transmission, The More Connections Created



Spike Trains Entered to the SNNc

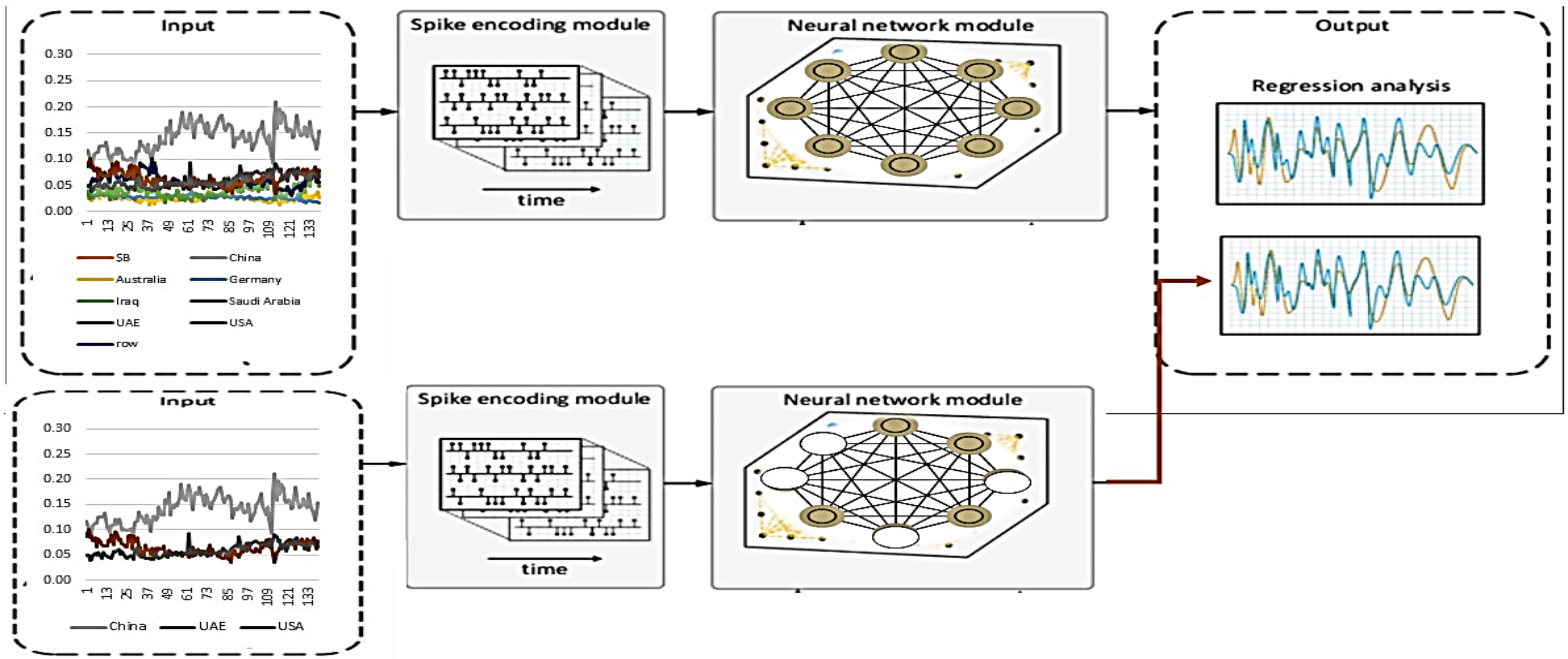
Neuron Spiking Activity During the STDP Learning

nkasabov@aut.ac.nz www.kedri.aut.ac.nz/neucube/

ePAMeT on NeuCube

AbouHassan, Iman; Kasabov, Nikola; Bankar, Tanmay; Garg, Rishabh; Sen Bhattacharya, Basabdatta. ePAMeT: Evolving Predictive Associative Memory for Time Series, Evolving Systems, Springer-Nature, 16-6, 2025,
https://doi.org/10.1007/s12530-024-09628-y

A SNN model is trained on multiple time series using all available variables measured at a full-time length, and then the model is recalled on subsets of variables at a shorter time measurement without compromising predictive accuracy. Using a shorter time for recall makes early prediction of events possible. The SNN model can be further adapted/evolved on new data without pre-training the model on the old data, even using new variables.



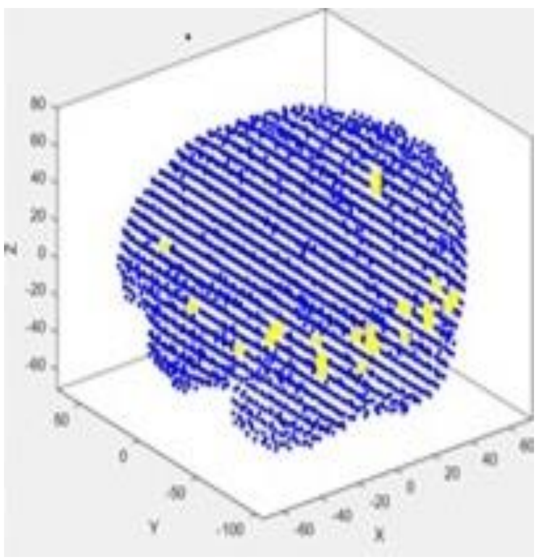
NeuCube application development environment

https://www.kedri.aut.ac.nz/neucube/

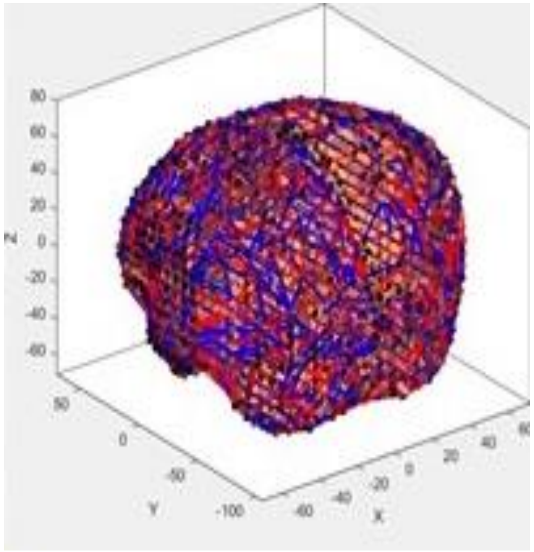
From NeuCube to eXCube: eXplainableCube

eXCube is a brain-inspired an also NeuCube inspired model for the evolution and the eXplanation of ESTAM of a class of stimuli vs other class (classes), e.g. conscious vs unconscious; meaningful vs meaningless; truth vs lie; ill vs healthy, etc. and for the discovery of both generic (abstraction) and specific (individual) related patterns.

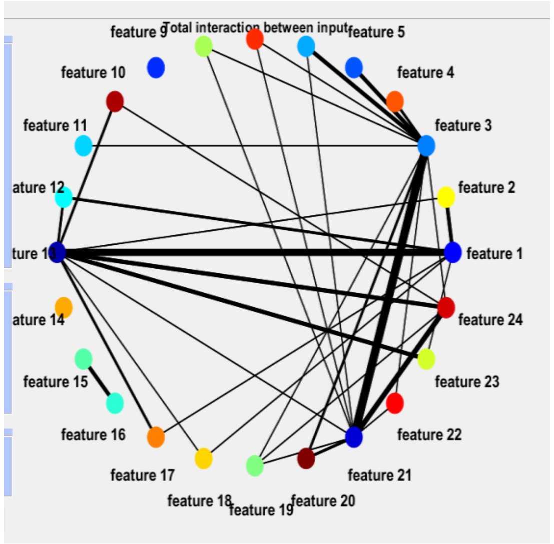
1. Feature Mapping



2. STDP Training



3. Extracting ESTAM

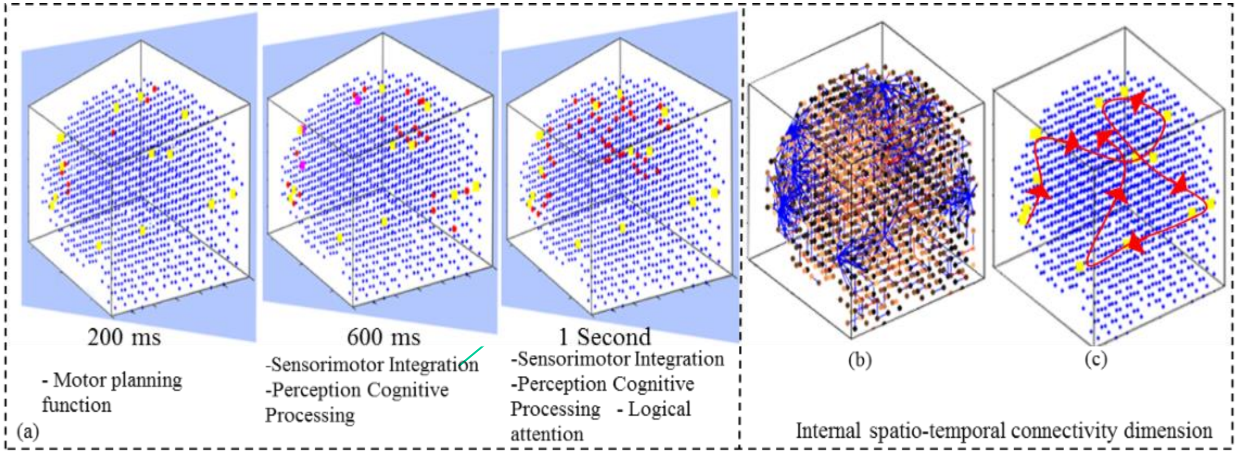


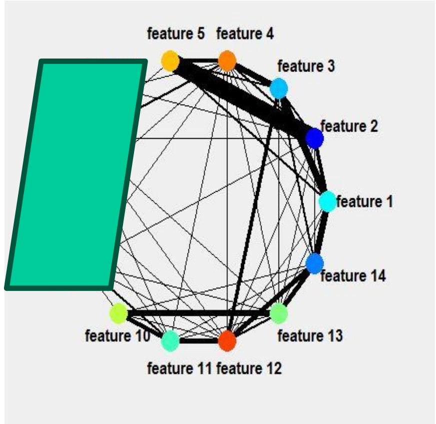
Kasabov, N (2023). STAM-SNN: Spatio-Temporal Associative Memories in Brain-inspired Spiking Neural Networks: Concepts and Perspectives. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.23723208.v1

Theoretical and computational foundations of ESTAM on eXCube

Realising ESTAM in BI-SNN was first suggested in [1], (citation begins) “…As spatio-temporal patterns from data are learned in the recurrent SNN as pathways of connections, when only a small initial part of input data is entered, the SNN will ‘synfire’ and ‘chainfire’ learned connection pathways to reproduce learned functional pathways...The NeuCube can be used as an associative memory and as a predictive system for brain states based on some initial brain signals” (citation ends).

Implementing ESTAM on BI-SNN is justified based on both theoretical and experimental studies of SNN. Similar activation patterns (called ‘polychronous waves’) can be generated in a SNN reservoir with recurrent connections, when similar but not necessarily identical, stimuli are presented to the SNN. This is a further extension of the ‘synfire chain’ theory (Abeles, 1991). A spatio-temporal state Si(T) of a SNN is a sequence (Si(t1), Si(t2), …,Si(tk)) of consecutive spiking patterns of polychronous groups of neurons in the SNN within a time interval T=(t1,t2,…,tk) that represents a defined output Oi.

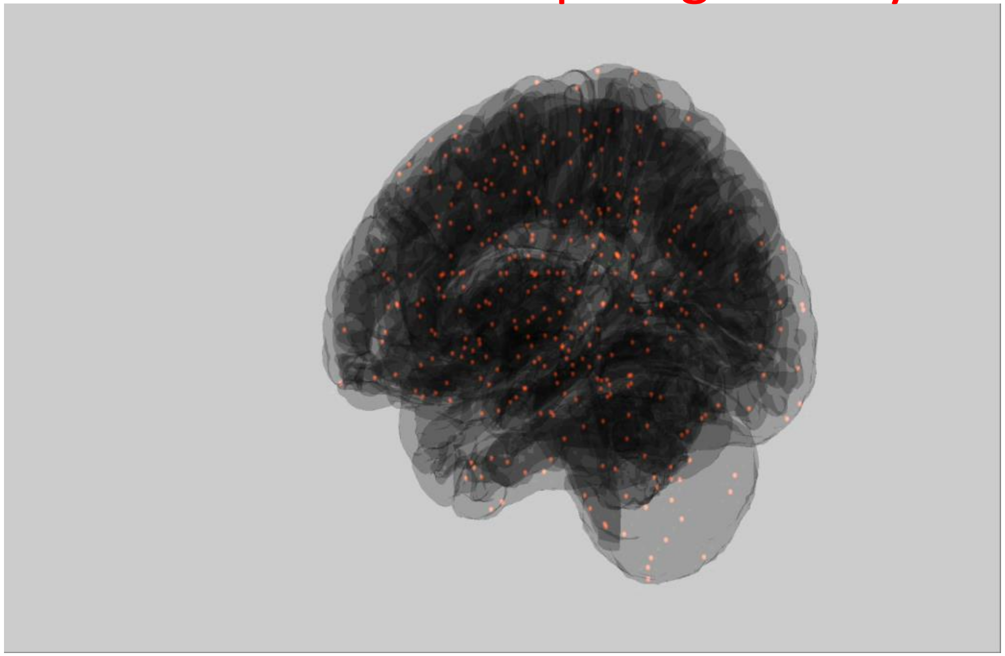


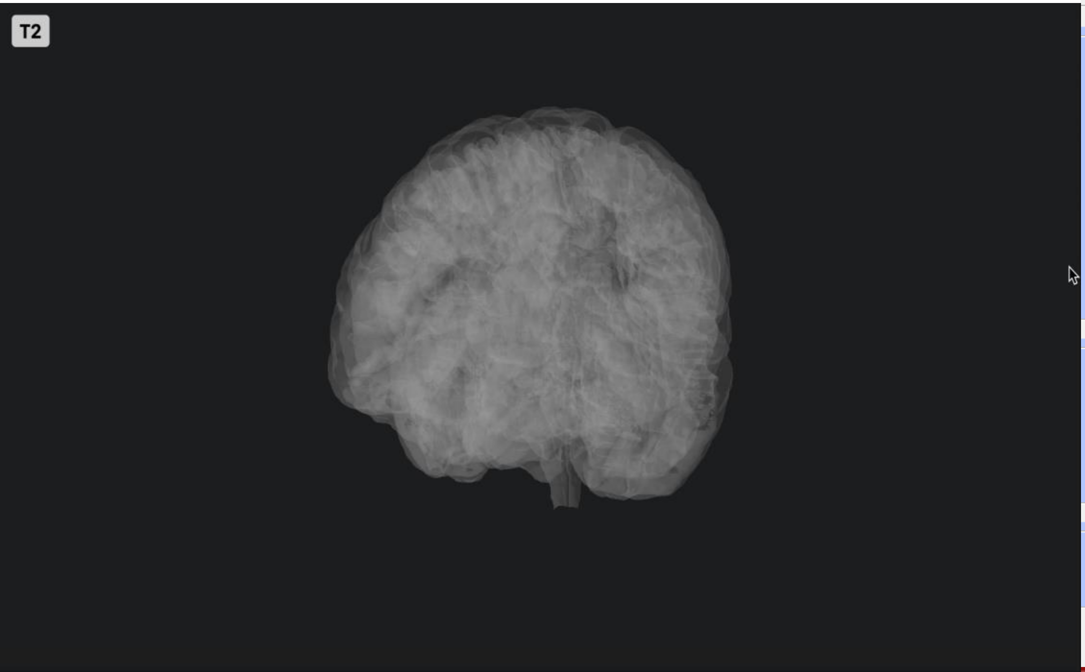


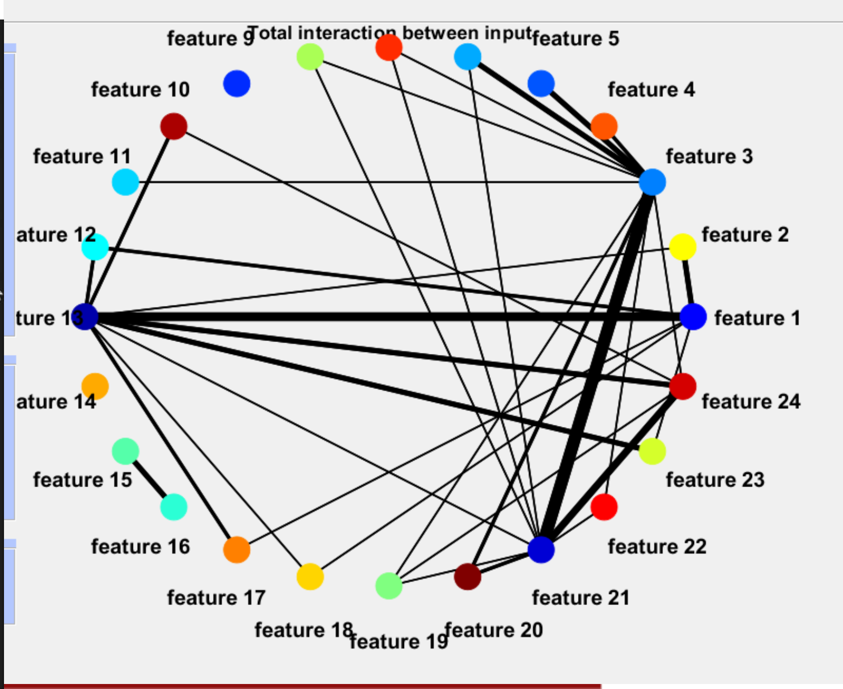
Kasabov, N., NeuCube: A Spiking Neural Network Architecture for Mapping, Learning and Understanding of Spatio-Temporal Brain Data, Neural Networks, vol.52, 2014.

Kasabov, N (2023). STAM-SNN: Spatio-Temporal Associative Memories in Brain-inspired Spiking Neural Networks: Concept Perspectives. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.23723208.v1

eXCube: From spiking activity to ESTAM







5. ESTAM on eXCube for modelling meaningful vs meaningless visual perception in humans

Understanding how the brain differentiates between **meaningful** and **meaningless** stimuli is a cornerstone of neuroscience and artificial intelligence.

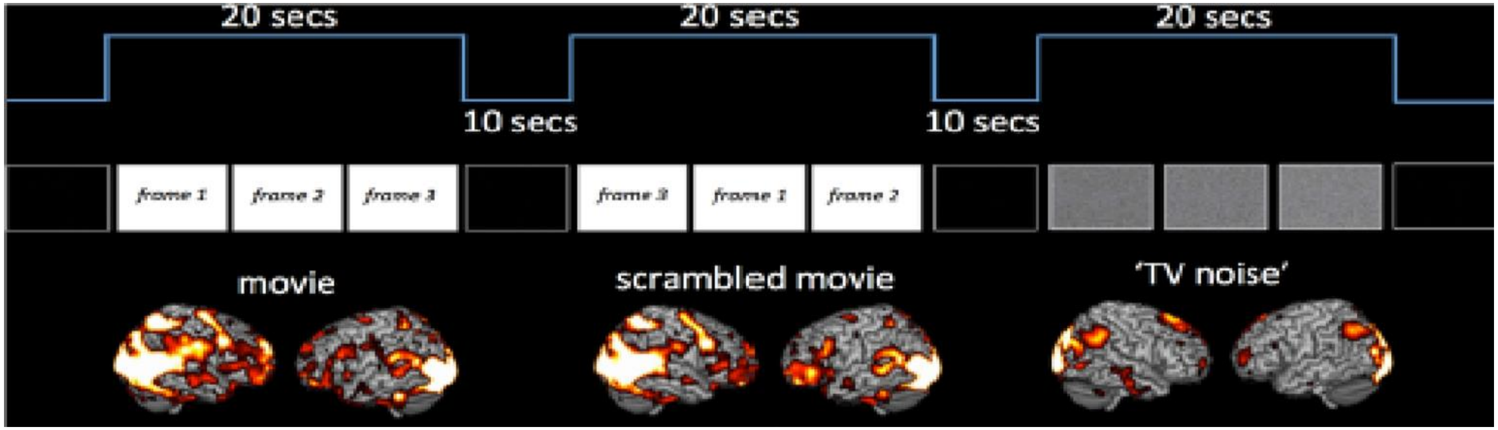
From neuroscience and neural networks to AI machine consciousness, our aim is to uncover the dynamic brain patterns that define meaningful experience. To model this, we need computational systems that mirror the brain's temporal complexity, not just its static snapshots.

Limitations of traditional methods

Traditional methods treats brain activity as static snapshots. They might offer high accuracy but lack temporal dynamics and interpretability and no insight into how decisions are made in time.

Experimental Case study

Boly M, Sasai S, Gosseries O, Oizumi M, Casali A, Massimini M, et al. (2015) Stimulus Set Meaningfulness and Neurophysiological Differentiation: A Functional Magnetic Resonance Imaging Study. PLoS ONE 10(5): e0125337.doi:10.1371/journal.pone.0125337



Settings for the preliminary proof of concept experiments

Goal

1. To analyse fMRI data using eXCube to identify spatio-temporal brain signatures in time (ESTAMs) that distinguish between:

   - Class 1: Watching a meaningful video
   - Class 2: Watching meaningless noise

2. Construct ESTAM model, a 4D spiking trajectories of spatio-temporal activity (3D features × time of activity).

3. Map spike paths over time to find the spatio-temporal signature that discriminates the two classes.

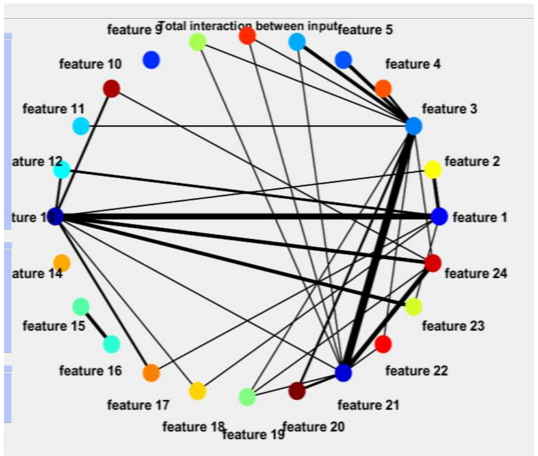
A small data set for proof of concept

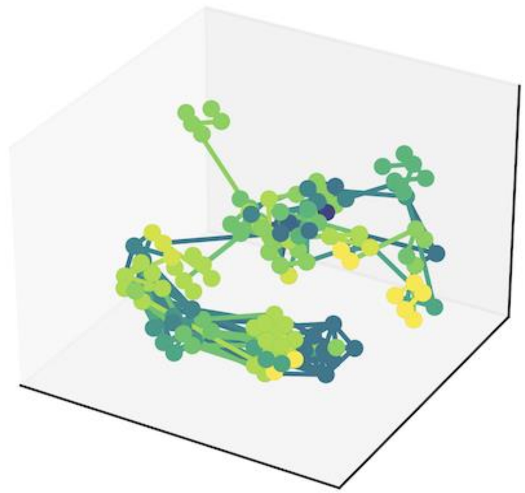
fMRI data from 24 feature voxels, representing areas of visual cortex, parietal cortex; temporal lobe; amygdala; hippocampus; frontotemporal lobe.

- 40 samples (20 per class).
- 16 time points per sample.
- Two classes: meaningful vs meaningless
- Single-subject for now (scalable to 6 subjects)

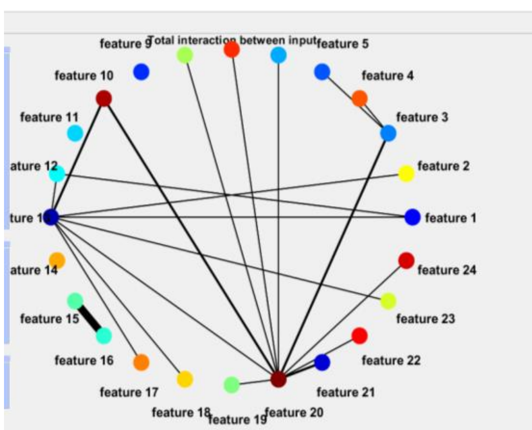
Discovering and visualisation of ESTAMs of meaningful (top) vs meaningless (bottom) visual perception (24 features – voxel areas)

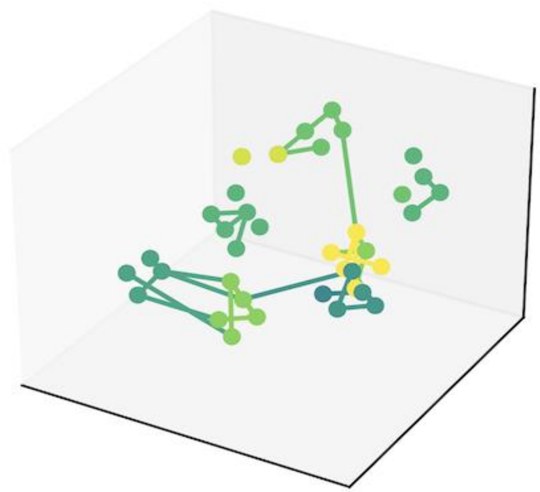
meaning*ful*





meaning*less*

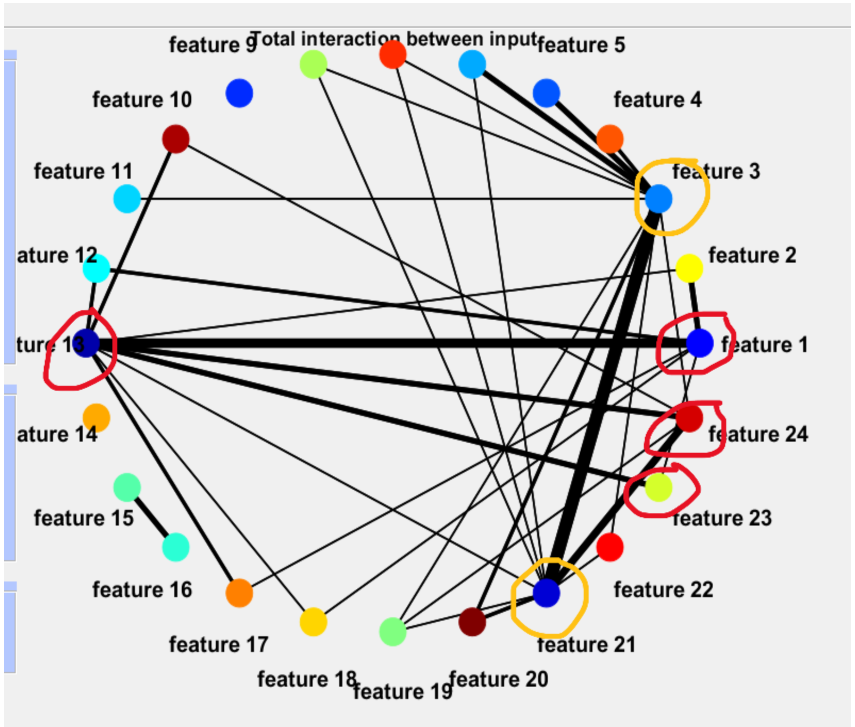




1: right middle occipital gyrus
7: left cuneus
17: right cuneus
16: right postcentral gyrus
13: right fusiform gyrus
14: left insula
15: right postcentral gyrus
16: right postcentral gyrus
12: right fusiform gyrus
9: left precentral gyrus
3: left middle temporal gyrus
15: right postcentral gyrus
14: left insula
12: right fusiform gyrus
9: left precentral gyrus
3: left middle temporal gyrus
23: right cuneus
21: left middle occipital gyrus
24: left lingual gyrus

Dominating dynamic features of information (spike) exchange for class1 (meaningfulness) : (1,24,23) -- 13; 21--3 ; voxel areas 1,3,13,21

Meaningful



Semantic processing

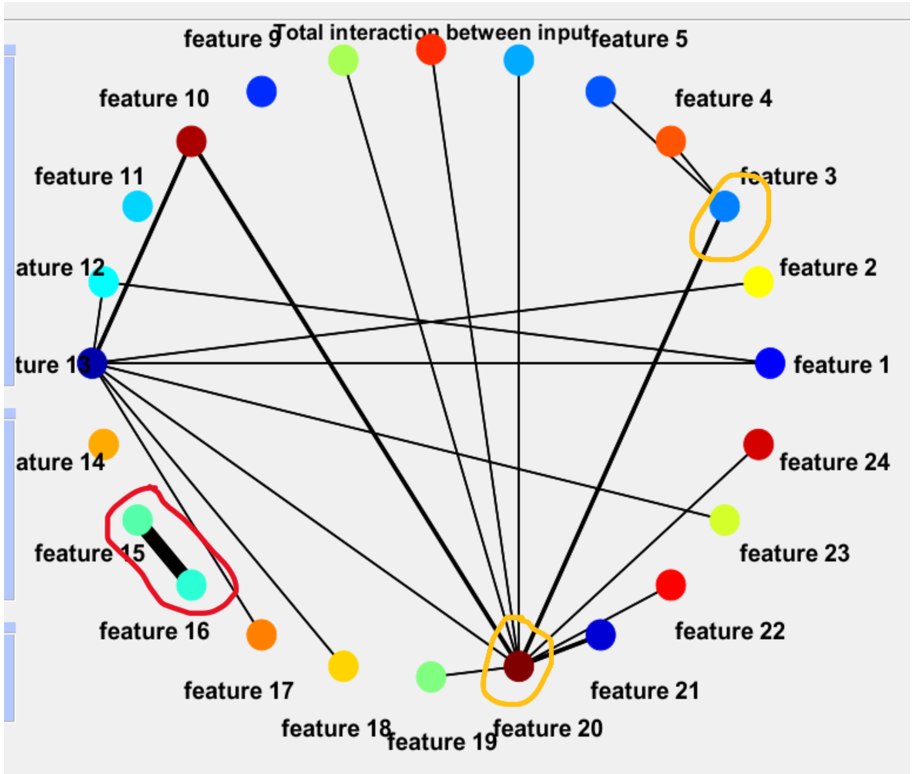
Feature 21: Left Middle Occipital Gyrus
Extraction of **early visual features**
Feature 3: Left Middle Temporal Gyrus
Semantic processing, visual meaning

Ventral Visual Stream Network

- Where detailed perceptual inputs converge with **higher-order processing.**
- Recognition and interpretation of meaningful visual stimuli

Feature 13: Right Fusiform gyrus: involved in high level visual processing (object recognition)
Feature 1: Right Middle Occipital Gyrus: early visual processing
Feature 24: Left Lingual Gyrus: Complex visual stimuli
Feature 23: Right Cuneus: basic visual processing + spatial orientation

Meaningless



Somatosensory processing
Right post central gyrus, both features located close together

When visual input does not effectively engage higher-order visual/semantic circuits, there may be a compensatory or parallel involvement of sensorimotor regions, e.g shift in attention on self rather than environment in the fMRI.

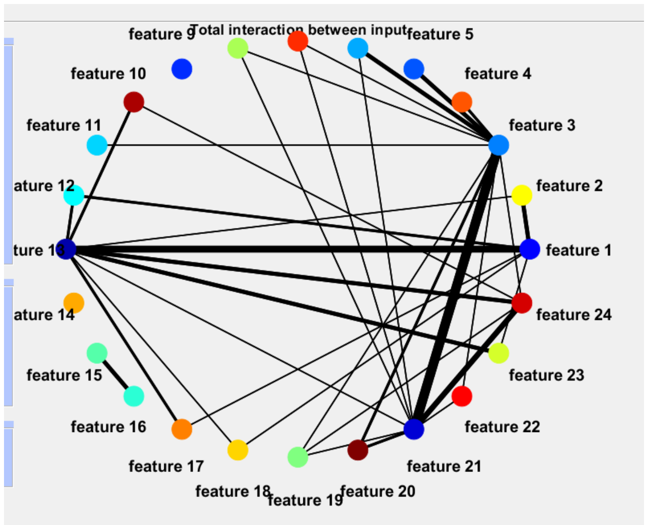
Lack of semantic richness

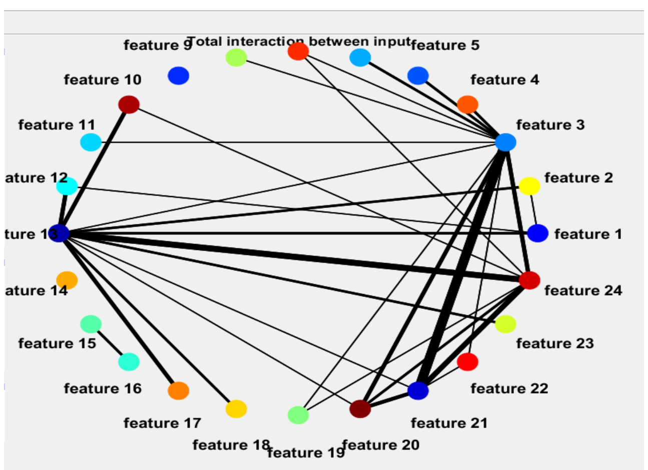
Feature 20 Left middle occipital gyrus, Feature 10 Right Lingual Gyrus
- Both regions involved in early or intermediate processing of visual information

Feature 3 Left Middle Temporal Gyrus
- Extraction of meaning, also involved but with less interaction strength than meaningful conditions.

Discovery of both **common** features across individuals and **specific** features for each individual. Example: subject 1 (top), 4 other subjects (bottom)

Common dynamic features: 3---21, (1,23,24) ---13, Specific and less pronounced are different features.





1: right middle occipital gyrus
7: left cuneus
17: right cuneus
16: right postcentral gyrus
13: right fusiform gyrus
14: left insula
15: right postcentral gyrus
16: right postcentral gyrus
12: right fusiform gyrus
9: left precentral gyrus
3: left middle temporal gyrus
15: right postcentral gyrus
14: left insula
12: right fusiform gyrus
9: left precentral gyrus
3: left middle temporal gyrus
23: right cuneus
21: left middle occipital gyrus
24: left lingual gyrus

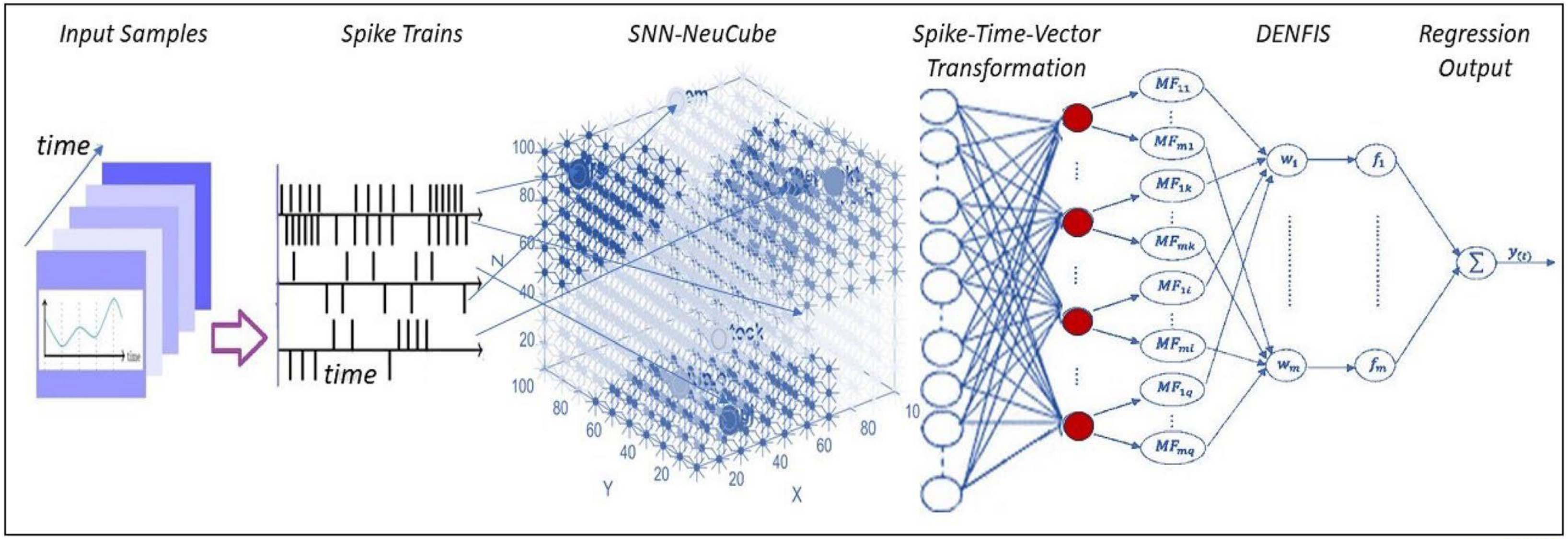
We can see similar features involved, with a commonality of predominance on feature 13-24, for high level visual processing and complex visual stimuli respectively.

We are seeing that each model's **spike activity likely contain information about the brain processes** that underly visual meaningfulness or meaninglessness.

Using NeuDen (SNNcube + DENFIS) for revealing fuzzy rules that explain ESTAM

AbouHassan, Iman; Kasabov, Nikola; NeuDen: A Framework for the Integration of Neuromorphic Evolving Spiking Neural Networks with Dynamic Evolving Neuro-Fuzzy Systems for Predictive and Explainable Modelling of Streaming Data, Evolving Systems, Springer-Nature, 16-3, 2025, https://doi.org/10.1007/s12530-024-09630-4.

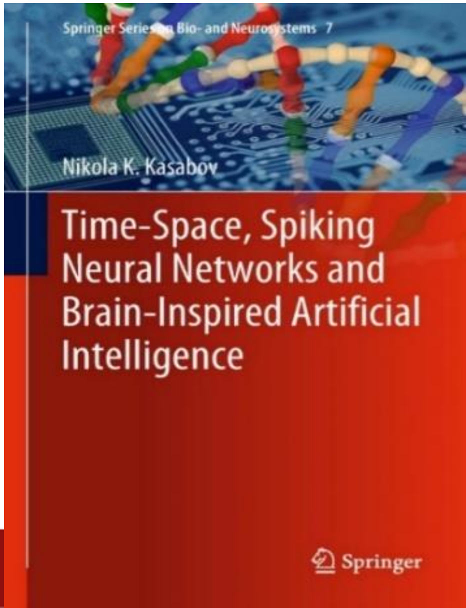
NeuDen combines a SNNcube, for learning temporal and spatio-temporal data after data is encoded into spike sequences a dynamic, evolving neuro-fuzzy inference system (DENFIS) for time series prediction and fuzzy rule extraction from the feature frequency vectors extracted from the trained SNNcube.



6. Some current and future applications:

- Emotional intelligence:
  - Recognising emotional faces
  - Expressing emotions
- Brain abnormalities:
  - Autism
  - ADHD
  - Demetia: Early detection via brain data
  - Demetia: Early detection via speech data
- Brain-inspired robotic systems for human and machine symbiosis

.

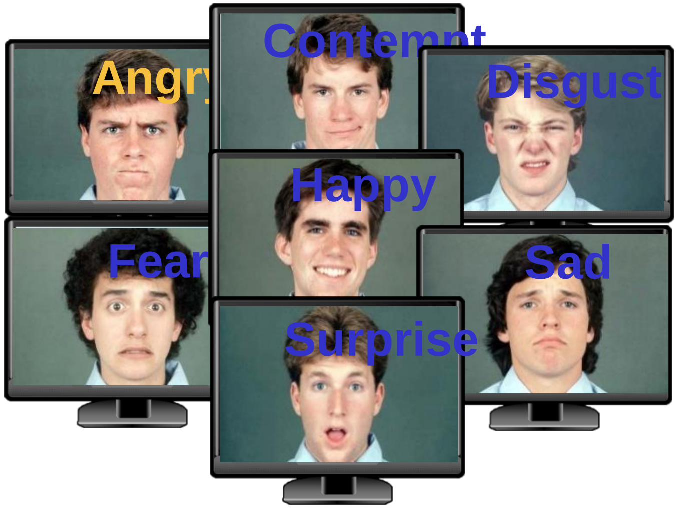


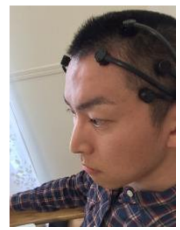
Kasabov, N., Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence, Springer (2019), https://www.springer.com/gp/book/9783662577134

MC for emotion recognition and emotion expression by machines

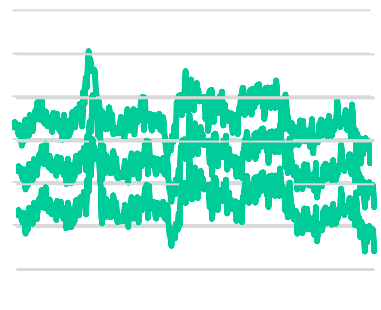
(H.Kawano, Z.Doborjeh, N.Kasabov, Proc. ICONIP 2016, Kyoto, 2016)

Facial Expression Perception Task

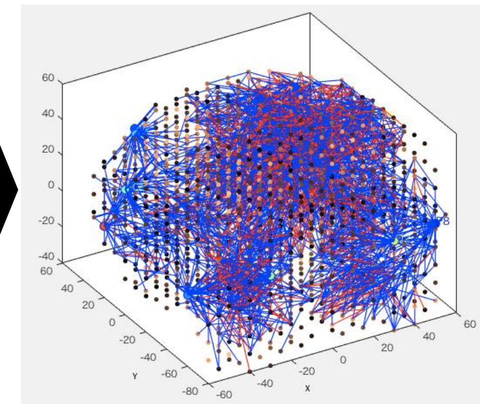




14ch EEG



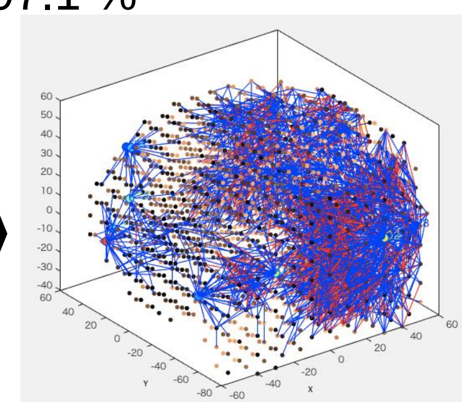
94.3 %



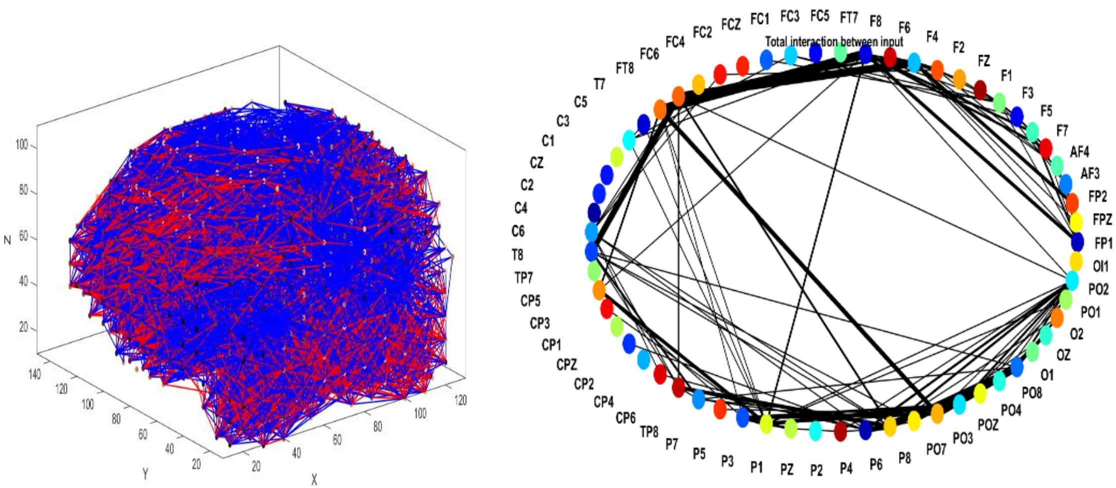
Face Expression Production Task

14ch EEG

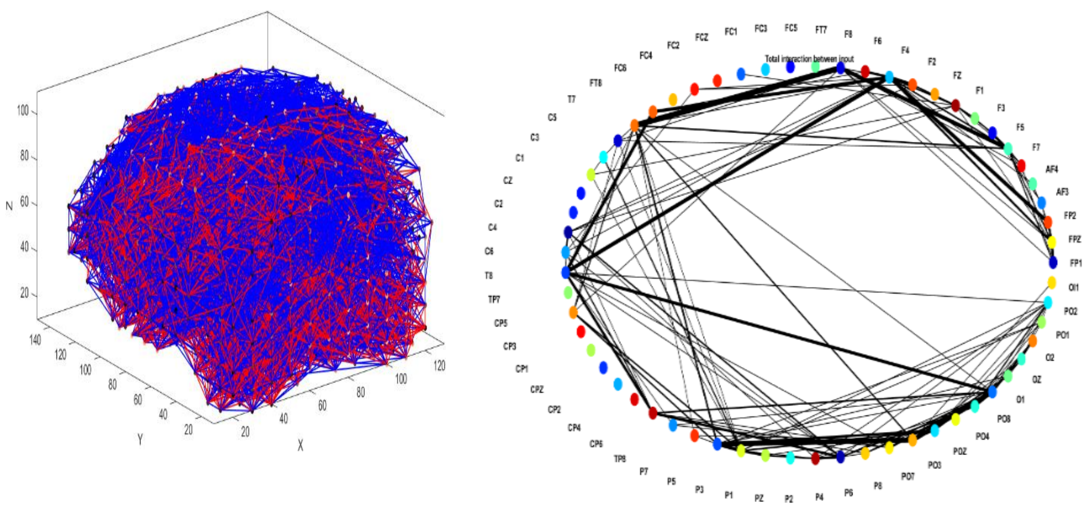
97.1 % NeuCube



Understanding depression through ESTAM



(a) ESTAM of eXCube trained on depressed subjects EEG data;



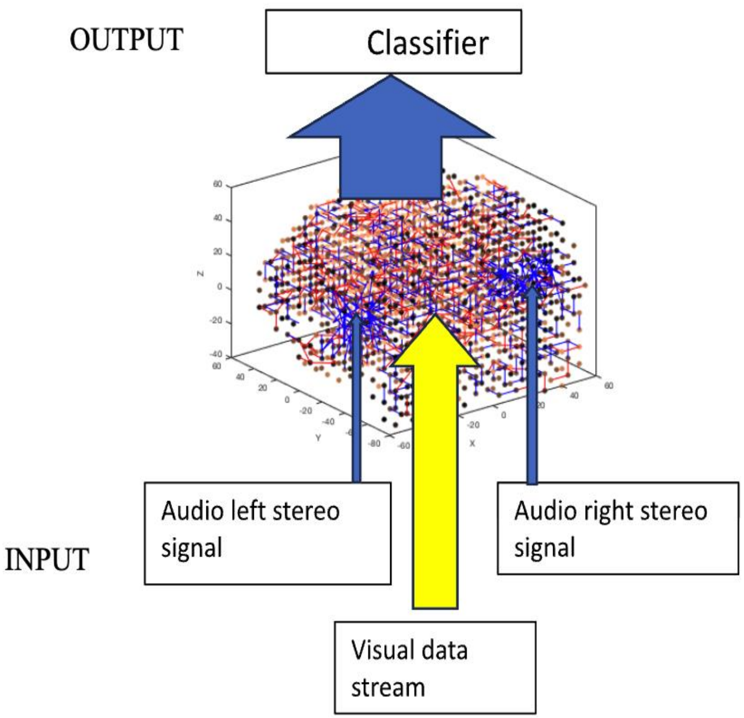
(b) ESTAM of eXCube trained on healthy subjects EEG data;

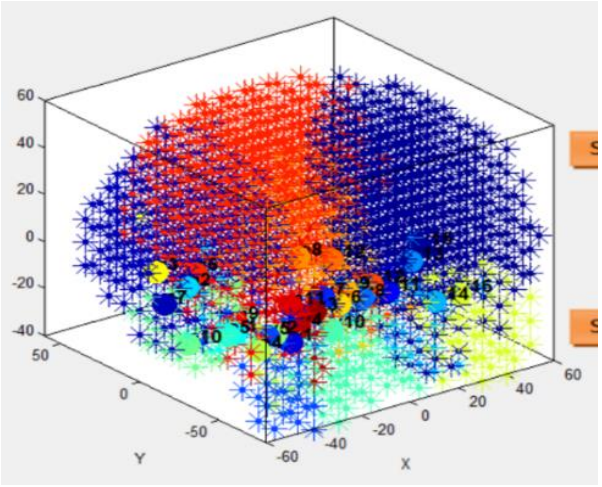
nkasabov@aut.ac.nz

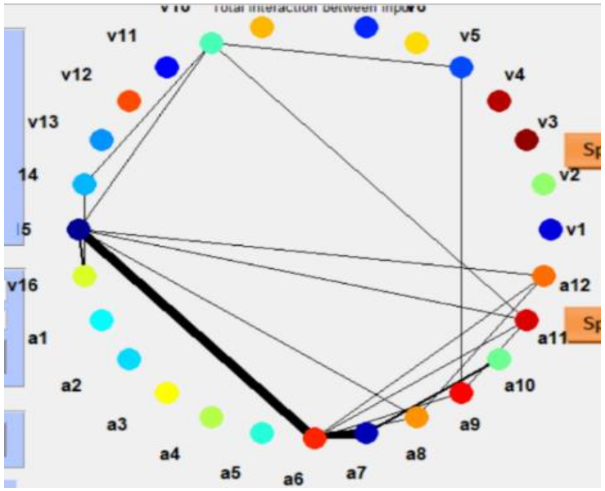
Multimodal audio-visual ESTAM

Kasabov, N., B.Bhattacharya, I.Sabousaan et al, AViAM-SNN: A Framework for Audio-Visual Associative Memories using Brain-inspired Spiking Neural Networks, Neural Computation and Applications (NCAA) 2025 (under review)

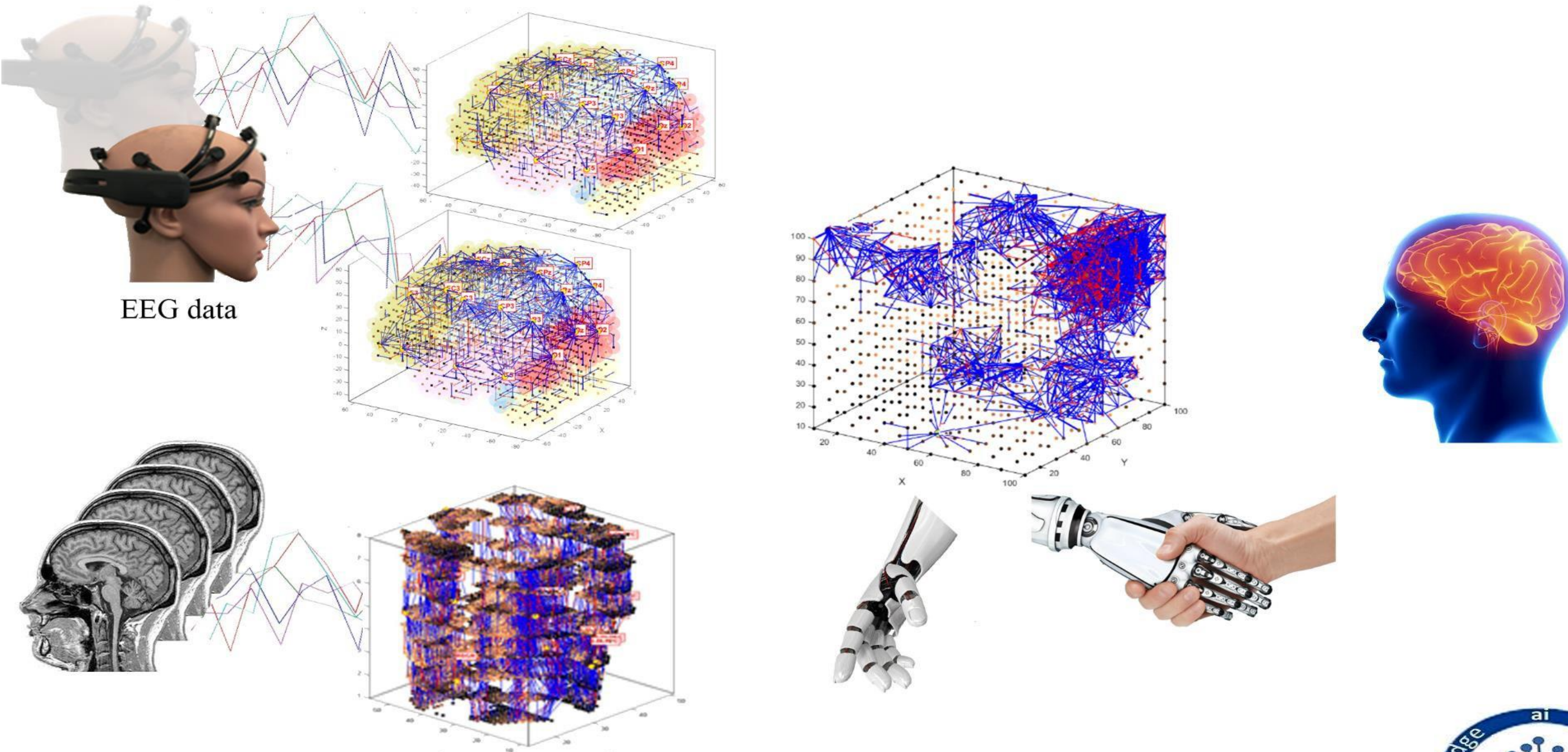
- Training on both modalities
- Extract ESTAM
- Recall on only one modality





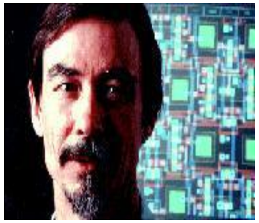


Human and machine symbiosis for a better communication and a better understanding of what machines can do and if they manifest MC.



Realisation of MC on neuromorphic hardware platforms

High speed and low power consumption!!!



Carver Mead (1989): A hardware model of an IF neuron:
The Axon-Hillock circuit.

SpiNNaker (*Furber, S., To Build a Brain, IEEE Spectrum, vol.49, Number 8, 39-41, 2012).*

INI Zurich SNN chips (Giacomo Indiveri)

Silicon retina (the DVS) and silicon cochlea (ETH, Zurich, Toby Delbruck))

The IBM True North (D.Modha et al, 2016): 1mln neurons and 1 billion of synapses

Loihi by Intel

In conclusion:
Can MC make AI safer, more efficient and more ethical?

- Safer:
  - Make decisions based on meaningful rather then meaningless information
  - Discriminate harm from no harm

- More efficient
  - Aware of it can and what it cannot do
  - Much, much less energy consuming (brain-like, neuromoprhic)

- More ethical
  - Teaching AI to evolve moral values
  - Discriminate truth from lie (misinformation)
  - Explainable
  - Manifesting sentience and emotions

nkasabov@aut.ac.nz

Thank you

- Conscium Ltd for funding this research to Knowledge Engineering Cons. Ltd
- Assia Kassabova for helping with the presentation

Questions?