EPiC Series in Technology

Volume 3, 2026, Pages 108–112

Proceedings of AI for Accelerated Research Symposium

Gaining Insight into Brain Damage and Rehabilitation through Digital Twins*

Risto Miikkulainen[1] and Swathi Kiran[2]

[1] Department of Computer Science, The University of Texas at Austin
risto@cs.utexas.edu
[2] Center for Brain Recovery, Boston University
kirans@bu.edu

Abstract

While digital twins have become an essential tool in many areas of engineering, they are still relatively rare in brain science. Whereas gradient-descent-based machine learning approaches can model behavior at an aggregate level, they do not replicate neural mechanisms of individual subjects or patients, and therefore cannot serve as twins for them. Other approaches can, such as associatively connected neural maps. For example the BiLex model can be fit to an individual patient's language history and impairment in stroke or dementia, and then used to evaluate different rehabilitation and mitigation strategies. This approach was found promising in an actual clinical trial, paving the way for more digital twin studies in the future.

Digital twins have become an essential tool in many areas of engineering. Driven by the vast increases in computational power (over billion-fold increase since the 1990s;[16], physical systems can now be simulated with high fidelity, making it possible to test various design alternatives virtually, before they are built. Computational thinking has gained ground more broadly as well; computational modeling is now used to evaluate policy decisions in society, consider alternative options in healthcare, and optimize decision-making in business [12].

Brain science is the last frontier for digital twins. As complex as physical systems are, and decision-making in society is, neural systems are still much more complex, and also less well understood. The lack of understanding means that the increases in computational power alone is not sufficient; neural mechanisms need to be characterized precisely enough so that digital twins can be constructed for them.

Note that this challenge is different from that addressed by deep learning and generative AI, such as large language models. Even though they are computational neural networks, they are primarily developed to solve engineering problems, not to gain insight into brain processes. The adaptation mechanism used in them, gradient descent, is a high-level abstraction of learning through prediction errors. They are trained with vast collections of examples, much more than humans experience in their lifetime, and by showing these examples many times, instead of incrementally as the system is performing [9]. They do end up capturing many aspects of human

*Supported in part by NIH R01DC020653.

performance, but as a side effect of the statistical modeling of data, without a grounding or ability to introspect and understand the limits of their knowledge.

At the same time, our knowledge of the brain processes has increased significantly. Through techniques such as fMRI, MEG, Diffusion Tensor Imaging, and Voltage-Sensitive Dye Imaging, we now have a relatively good characterization of the dynamic processes in brain responses to visual and language inputs at the level of cortical organization [4, 6]. The anatomical connectivity is also becoming known in detail, including a comprehensive connectome of the fruit fly and the mouse [1, 2]. Consequently, it is possible to simulate neural processes at the level of individual neurons and synapses [11]. It is also possible to simulate function at the level of functional areas and maps. In both of these cases, the simulations can be verified by comparisons to experimental data.

The next step is to build digital twins of neural circuits and process that not only replicate the observed mechanisms, but also allow gaining new insight into the biological processes—that is, to evaluate alternative hypotheses, to inform biological experiments, and to devise potential treatments. Digital twins allow full access to all aspects of the simulated circuitry, both for observation and for manipulation, and therefore make it possible to evaluate many alternatives, gain a thorough mechanistic understanding, and discover solutions that would otherwise be difficult to conceive. Obviously any such insights will eventually need to be confirmed in biological experiments, but digital twins can serve as a way to expand our ability to think about biological processes and to focus our inquiry on the most promising alternatives.

Interestingly, there is a growing need for digital twins also because of the pressure for the scientific community to move away from animal experiments. NIH recently announced an initiative to reduce the use of animals in NIH-funded research, and instead prioritize human-based research technologies [13]. While there may be ways to move in that direction, many high-value experiments cannot be done safely on humans. Digital twins thus provide a compelling alternative. Mechanisms still need to be developed to transfer the results from digital twins to human subjects; these policy changes make it compelling to do so.

In terms of technologies available to construct digital twins, the most powerful approach is to develop physical simulations based on first principles, such as tissue properties, mechanics, electromagnetics, and chemical processes. Where the principles are well-known, the availability of high-performance computing makes this approach viable. Such models can be modified at will, even in novel ways, and they will respond similarly to the system being simulated.

However, even when first-principles simulations are not possible, it may be possible to construct digital twins through machine learning. All that is needed is a dataset describing the behavior of the system—inputs, internal measurements, outputs—and a machine learning model such as a neural network, a random forest, a symbolic regression model, a decision tree, or a rule set, can be constructed to replicate the behavior. Note that this approach is phenomenological only: it does not specify how the behavior is generated, it only replicates through mechanisms that may have little to do with the way the brain generates them.

The machine learning approach may still be useful because such twins make it possible to predict how the subject would respond to various new inputs. It is therefore possible to study, for instance, how to reduce human errors, how to make their performance more effective, and how to train people best. Note however that the datasets need to be comprehensive: such models are good at interpolating between examples in their training dataset, but they do not do well with examples that are very different from anything seen before.

The phenomenological machine learning approach is less useful when studying lesions, impairments, and rehabilitation. First, there is often not enough data on impaired behavior to train the model to replicate it well. Second, since they are phenomenological, it is not clear how the models could be lesioned to match the lesions to the brain. Third, because they are trained with all available data at once, they do not account for how the performance depends on the process of learning during the subject's lifetime. Therefore, while machine learning models are good at capturing behavior overall in people, they are not well-suited for the role of a digital twin of an individual subject or a patient.

However, many techniques in computational neuroscience and cognitive science are better suited for this role. They aim at modeling neural processes at a level that matches the behavior being studied. Low-level models aimed at understanding neural processes and recovery may be based on detailed neural circuitry, spiking events, and synaptic plasticity. Models at the level of perception and response generation may capture the overall organization in maps and their interconnections. High-level models may include multiple cognitive components such as vision, memory, decision-making, and language. The idea is that each one has a grounding in brain structures and processes commensurate with the mechanisms being modeled. They are constrained by the knowledge about those mechanisms and provide insights at that level.

As a concrete example, consider BiLex, a model of the bilingual lexicon [3, 5, 10, 14, 15]. This model is aimed at understanding the organization of knowledge about words and their meanings in two languages, how it depends on the history of exposure to the two languages, how the lexical processing can break down in aphasia and dementia, and how it can be best repaired and mitigated through rehabilitation. It is based on computational modeling at the level of maps and connections akin to the conceptual model of Kroll & Stewart [8] (Figure 1). The words in the two languages, such as Spanish and English, or Mandarin and English, are laid out on two separate self-organizing-map neural networks, the word semantics is laid out on a third map, associative connections between them implement production and understanding mechanisms, and lateral connections within the maps implement competition in choosing the right output. BiLex is intended to model an individual subject or patient. The maps are organized and the connections learned through a Hebbian learning process that replicates the individual's language history, i.e. exposure to the two languages over their lifetime. In that sense, it is truly a digital twin of an individual person.

This digital twin then makes it possible to identify the best possible treatments for the individual in case of stroke or dementia. First, the maps and the connections are lesioned to match the profile of the patient's impaired performance. Second, several copies of the model are made, and each one is rehabilitated in a different way. Third, the extent and speed of recovery, or slowdown of progress in dementia, are measured, and the best rehabilitation schemes for this patient are identified. Whereas with the real patient we would have only one shot to get the treatment right, the digital twin makes it possible to try out any number of possibilities to find the right one.

This approach was implemented in a clinical trial of Spanish-English stroke patients where the treatment consisted of choosing one of those languages for treatment [7]. The language choice matters in a complex way that depends on the history and impairment of the patient. There is currently no theory of how the language should be chosen: the decision depends largely on the availability of treatment providers. The results suggested that individual differences need to be taken into account: while for most severe cases the language choice did not matter, there was a significant difference for patients with mild impairment and sufficient competence. In this manner, the digital twins made it possible to provide better care for individual patients, and to develop a beginning of a theoretical approach for effective rehabilitation.

To our knowledge, BiLex is the first neural digital twin evaluated in a clinical trial. While the study was limited to the choice between two languages, it can be extended to more refined choices in the future. It is possible to imagine how such models can be developed further, for

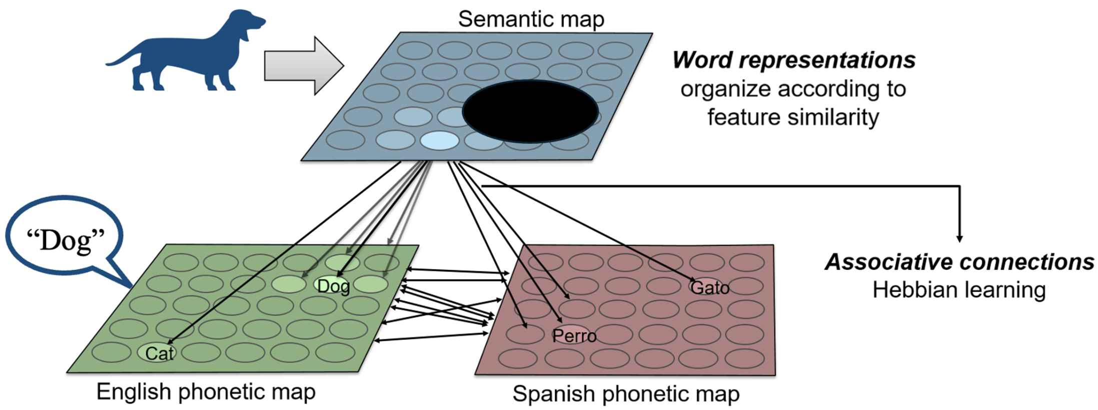

Figure 1: **The BiLex model of the Bilingual Lexicon.** The model consists of self-organizing maps for the phonetic words in the two languages and a common semantic map for word meanings. The maps are connected through associative connections that implement naming and understanding, and lateral connections that implement response selection. The model is trained to match the language history and impaired performance of an individual patient; as an example, loss of a part of the semantic map after a stroke is shown. The model can then be used as a digital twin for the patient to identify the best personalized strategies for rehabilitation.

instance by including higher-level maps for other language functions, control mechanisms such as basal ganglia for choosing between languages, and memory modules for continual learning. It can also be extended to model other language impairments such as those occurring in dementia. BiLex can thus be seen as an example of both how digital twins can be useful in brain science, and how they can inform clinical practice. We believe this area will develop rapidly in the near future, and it will be productive.

References

[1] J. Alexander Bae et al. Functional connectomics spanning multiple areas of mouse visual cortex. *Nature*, 640:435–447, 2025.

[2] Sven Dorkenwald et al. Neuronal wiring diagram of an adult brain. *Nature*, 634:124–138, 2024.

[3] Peggy Fidelman, Uli Grasemann, Claudia Penaloza, Michael Scimeca, Yakeel T. Quiroz, Swathi Kiran, and Risto Miikkulainen. Constructing individualized computational models for dementia patients. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, 2022.

[4] Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536:171–178, 2016.

[5] Uli Grasemann, Claudia Peñaloza, Maria Dekhtyar, Risto Miikkulainen, and Swathi Kiran.

Predicting language treatment response in bilingual aphasia using neural network-based patient models. *Scientific Reports*, 11:10497, 2021.

[6] Jiyoung Kang and Hae-Jeong Park. Integration of partially observed multimodal and multiscale neural signals for estimating a neural circuit using dynamic causal modeling. *PLOS Computational Biology*, 20(7):e1012655, 2024.

[7] Swathi Kiran, Erin Carpenter, Uli Grasemann, Michael Scimeca, Manuel J. Marte, Marissa Russell-Meill, Claudia Peñaloza, Risto Miikkulainen, and Yorghos Tripodis. Using digital twins to predict language treatment outcomes in Spanish-English bilingual aphasia: A double-blind randomized controlled trial. Under review, 2025.

[8] J. F. Kroll and E. Stewart. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33:149–174, 1994.

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[10] Ran Li, Uli Grasemann, Risto Miikkulainen, and Swathi Kiran. Predicting treatment outcomes in Mandarin-English bilingual adults with aphasia. Under review, 2026.

[11] Wenlian Lu et al. Simulation and assimilation of the digital human brain. *Nature Computational Science*, 4:890–898, 2024.

[12] Risto Miikkulainen. Creative AI through evolutionary computation: Principles and examples. *SN Computer Science*, 2:163, 2021.

[13] National Institutes of Health. NIH to prioritize human-based research technologies. News release, April 29, 2025, 2025. https://www.nih.gov/news-events/news-releases/nih-prioritize-human-based-research-technologies, Retrieved 7/19/2025.

[14] Claudia Peñaloza, Uli Grasemann, Maria Dekhtyar, Risto Miikkulainen, and Swathi Kiran. BiLex: A computational approach to the effects of age of acquisition and language exposure on bilingual lexical access. *Brain and Language*, 195, 2019.

[15] Divya Prabhakaran, Uli Grasemann, Swathi Kiran, and Risto Miikkulainen. How the Stroop effect arises from optimal response times in laterally connected self-organizing maps. In *Proceedings of the 47th Annual Meeting of the Cognitive Science Society*, 2025.

[16] Nick Routley. Visualizing the trillion-fold increase in computing power. https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/, Retrieved 11/17/2022, 2017.