Data Science is Central to Ongoing Advancements in Neuroscience

In 2018, *Nature Methods* published an article stating that "Neuroscience is experiencing a revolution" [1]. The article introduced a novel computational approach, implementing a neural network-based model to infer functional relationships between active brain cells. As developments in neuroscience continue to unfold, that article is one of many suggesting that computing, automation, data analysis, and machine learning will increasingly be at the core of major research achievements and clinical applications in neuroscience.

Since the design of the transformer architecture in 2017 [2] and commercialization of scaled machine learning in recent years, large language models (LLMs) have rapidly become a ubiquitous technology [3]. Though much public attention and industry effort has focused on LLMs and other consumer-facing tools, some of the greatest achievements in machine learning are outside the scope of these applications. For example, machine learning applications have made major contributions to computational biology. In that domain, the AlphaFold algorithms have implemented a model similar to the transformer to increase the number of all known protein structures in the world from less than a few hundred thousand to hundreds of millions [4], [5]. Just as some of the next decade's greatest achievements in biology and neuroscience will require applications of data science and machine learning, some of the greatest opportunities to innovate with data science and machine learning lie in applications to other sciences, including biology and neuroscience.

Artificial neural networks (ANNs) are not the only technology showing promise for innovative applications to neuroscience. As the nervous system contains a network of connected, interacting neurons, known as a connectome, network science is similarly applicable. Network science provides a mathematical foundation for modeling connectomes [6]–[8], including metrics (e.g., centrality, modularity) and algorithms (e.g., page rank and Louvain) for understanding individual nodes (neurons) and their communities within the network [7], [8]. The foundation of network science is combined with ANNs in graph neural networks (GNNs), including graph autoencoders (GAEs), which use convolutions on a product of the graph's adjacency matrix, degree matrix, and node feature embeddings to establish structure-aware embeddings of each node [9], [10]. As modern methodological advances increase capacity for collection of high-resolution connectomes, these network science tools are already beginning to be applied to improve understanding of nervous tissue function (e.g., [11], [12]).

This review considers the intersection of subfields in data science and neuroscience, including connectomics, extracellular electrophysiology, brain computer interfaces, neural signal processing, and spike train analysis. In introducing these subfields, I suggest that they have significant potential for synergy in accelerating understanding of microscale processing and distributed systems in the brain. In addition to providing understanding, I propose that these same technologies will offer significant new clinical interventions for a wide variety of nervous system diseases and disorders in coming decades.

A Connectome Details Neural Connections

The term "connectome" refers to the set of all neural connections in the nervous system, originating in the context of post-human-genome-project bioinformatics -omics research [13]–[15]. The term "biological neural network" (BNN) [16] is also used to reference neural connections, though a BNN may be only a sub-network of the full network–a partial connectome. The field of connectomics is based on a foundational premise that the aggregated structure and function of individual neural connections determines intermediate- and high-level functionality of the brain, such as cognitive, behavioral, and neurological outcomes. These aggregated neural connections can be considered as a network, with neurons as nodes and relationships between nodes as edges. Relationships are directional and may be excitatory, inhibitory, or modulatory, with relationship strength varying with the quantity of neurotransmitter released from one cell and the expression of neurotransmitter receptors on the other. Many factors at systemic and subcellular scales regulate the network as it changes over time [13]. Though a full connectome ideally includes all neurons and all synaptic relationships, connectomes are typically studied at varied scales. A partial connectome includes a sub-network of a complete microscale connectome; a macroscale connectome is a connectome without cellular resolution, resulting in analogs of a network science meta-node and some missing nodes or edges; and a functional connectome is based on functional relationships rather than known physical connections between cells [13], [17]–[20].

The gold standard for collection of a connectome is ex-vivo electron microscopy, a method that began to be used in the 1980's [6], [21]. The process includes delicate slicing of the tissue and imaging of each slice, followed by stitching visualizations together to identify all neurons and connections. Current methods restrict the scale of this exercise to only small animal models such as C. elegans and drosophila melanogaster [6], [21], [22]. In larger mammals, ex-vivo microscopy has been used to produce partial connectomes from pieces of neural tissue, and to produce macroscale connectomes from the full brain [23]–[25]. Continued methodological advances will likely continue to expand the scope and scale of these processes. While these developments are promising, microscopy is limited to only ex-vivo research, therefore microscopy-based connectomics must be combined with in-vivo neural recording to understand functional

relationships and work toward clinical interventions.

A major milestone in connectomics was the Human Connectome Project (HCP) [20], which pioneered methods for collection of human connectomes in-vivo via magnetic resonance imaging (MRI), including functional MRI (fMRI) and diffusion MRI (dMRI). From 2010 to 2019, teams collected a large initial dataset of fMRI-based macroscale connectomes and standardized many processes for large-scale study with MRI. Many projects today continue to iterate and build on the HCP with expanded scope, considering additional populations and analyses. While these methods have proven extremely useful in many research cases, they observe neural activity only indirectly by measuring fluid flow through tissue. Measurement resolution is also restricted by the physical limitations of radiowaves used in MRI sensors. As a result, cellular-level activity cannot be measured, and determination of causative relationships between nuclei is limited. For example, excitatory vs. inhibitory relationships typically cannot be identified [26].

Extracellular Electrophysiology is Fundamental to Brain Computer Interfaces

In order to improve on current microscopy and MRI methods, existing data should be combined with in-vivo neural recording. Such recording would assist in determining functional relationships between cells, eventually at scale, for a given human or animal during a given period in time. This would improve understanding of cellular underpinnings of high-level function of the nervous system, and build toward improved clinical interventions for many neurological and psychiatric disorders [27], [28]. Ideally, neural recording will have sufficient resolution in time and space to differentiate precise firing activity of individual cells. While whole-brain recording at this scale may be unlikely, recording from a maximal spatial area would likely give a useful partial functional connectome [29].

========= PICK UP HERE =========

Neural recording may only provide functional data, as no methods currently exist for visualizing solid tissue to micron resolution, but functional data is likely sufficient for many applications, or can be combined with macroscale connectomic data (e.g., anatomical location of major nerve tracts [TODO: check spelling]) for additional insights [CITE: TODO]. Last, any recorded data must be organized to enable scalable processing to automate recognition of functional connectomes.

Scaled Electrophysiology Relies on Strong Neural Signal Processing

The primary output of electrophysiological recording is a set of voltages, each associated with a time and a recording site (channel). These values are typically organized into a matrix, with rows corresponding to channels and columns corresponding to timepoints. With hundreds of channels and thousands of timepoints per second (devices typically work at 1,000-10,000Hz), these matrices can become quite large. This raw data can be used directly in some cases, but it does not provide a set of identified neurons, nor information about the neurons, such as when they fired an action potential. The process of converting the raw voltage data into this more useful information is known as spike sorting, a type of neural signal processing [CITE: TODO, notes]. There are many spike sorting algorithms [CITE: TODO, notes], but we will focus on one algorithm that leads in the field of spike sorting for large sets of parallel-collected signals or channels called KiloSort [CITE: TODO, kilosort 2016 and 2022].

A key initial step in neural signal processing for algorithms like kilosort is to separate the signals by

Processing Spike Trains Can Produce a Functional Connectome

Conclusion

References

[1] C. Pandarinath, D. J. O'Shea, J. Collins, *et al.*, "Inferring single-trial neural population dynamics using sequential auto-encoders," *Nature Methods*, vol. 15, no. 10, pp. 805–815, 2018, ISSN: 1548-7105. DOI: 10.1038/s41592-018-0109-9. [Online]. Available: https://doi.org/10.1038/s41592-018-0109-9.

[2] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention is all you need*, 2023. arXiv: 1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762.

[3] H. Naveed, A. U. Khan, S. Qiu, *et al.*, "A comprehensive overview of large language models," *ACM Trans. Intell. Syst. Technol.*, Jun. 2025, Just Accepted, ISSN: 2157-6904. DOI: 10.1145/3744746. [Online]. Available: https://doi.org/10.1145/3744746.

[4] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583–589, 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: https://doi.org/10.1038/s41586-021-03819-2.

[5] M. Varadi, D. Bertoni, P. Magana, *et al.*, "Alphafold protein structure database in 2024: Providing structure coverage for over 214 million protein sequences," *Nucleic Acids Research*, vol. 52, no. D1, pp. D368–D375, Jan. 2024, Published on behalf of Nucleic Acids Research, ISSN: 1362-4962. DOI: 10.1093/nar/gkad1011. [Online]. Available: https://doi.org/10.1093/nar/gkad1011.

[6] S. W. Emmons, "The beginning of connectomics: A commentary on white et al. (1986) 'the structure of the nervous system of the nematode caenorhabditis elegans'," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1666, p. 20 140 309, Apr. 2015, PMCID: PMC4360118, ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0309. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4360118/.

[7] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, UK: Cambridge University Press, 2010, ISBN: 9780521195331.

[8] M. E. J. Newman, *Networks: An Introduction*. Oxford, UK: Oxford University Press, 2010, ISBN: 9780199206650.

[9] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2018. arXiv: 1710.10903 [stat.ML]. [Online]. Available: https://arxiv.org/abs/1710.10903.

[10] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *arXiv preprint arXiv:1611.07308*, 2016.

[11] A. Srinivasan, R. Raja, J. O. Glass, *et al.*, "Graph neural network learning on the pediatric structural connectome," *Tomography*, vol. 11, no. 2, p. 14, 2025, Epub ahead of print: 2025-01-29, ISSN: 2379-139X. DOI: 10.3390/tomography11020014. [Online]. Available: https://doi.org/10.3390/tomography11020014.

[12] J. Neudorf, S. Kress, and R. Borowsky, "Structure can predict function in the human brain: A graph neural network deep learning model of functional connectivity and centrality based on structural connectivity," *Brain Structure and Function*, vol. 227, no. 1, pp. 331–343, 2022, Epub 2021-10-11, ISSN: 1863-2661. DOI: 10.1007/s00429-021-02403-8. [Online]. Available: https://doi.org/10.1007/s00429-021-02403-8.

[13] J. Ciarrusta and T. Arichi, "Chapter 1 - neurobiology and the connectome," in *Connectome Analysis*, M. D. Schirmer, T. Arichi, and A. W. Chung, Eds., Academic Press, 2023, pp. 3–23, ISBN: 978-0-323-85280-7. DOI: https://doi.org/10.1016/B978-0-323-85280-7.00012-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323852807000129.

[14] A. Mahapatra, "Omics in the postgenomic era," *ACS Chemical Biology*, vol. 5, no. 4, pp. 343–344, 2010, ISSN: 1554-8929. DOI: 10.1021/cb1000873. [Online]. Available: https://doi.org/10.1021/cb1000873.

[15] E. D. Green, J. D. Watson, and F. S. Collins, "Human genome project: Twenty-five years of big biology.," *Nature*, vol. 526, no. 7571, pp. 29–31, Oct. 2015, ISSN: 0028-0836. DOI: 10.1038/526029a. [Online]. Available: https://www.nature.com/articles/526029a.

[16] K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, and N. Le, "Spiking neural networks and their applications: A review," *Brain Sciences*, vol. 12, no. 7, p. 863, Jun. 2022, E-published 2022 Jun 30. DOI: 10.3390/brainsci12070863. [Online]. Available: https://doi.org/10.3390/brainsci12070863.

[17] L. Baxter, "Chapter 3 - functional network construction using functional mri," in *Connectome Analysis*, M. D. Schirmer, T. Arichi, and A. W. Chung, Eds., Academic Press, 2023, pp. 45–69, ISBN: 978-0-323-85280-7. DOI: https://doi.org/10.1016/B978-0-323-85280-7.00002-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323852807000026.

[18] J. Blommaert and D. Christiaens, "Chapter 2 - structural network construction using diffusion mri," in *Connectome Analysis*, M. D. Schirmer, T. Arichi, and A. W. Chung, Eds., Academic Press, 2023, pp. 25–44, ISBN: 978-0-323-85280-7. DOI: https://doi.org/10.1016/B978-0-323-85280-7.00007-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323852807000075.

[19] T. J. Sejnowski, "Nanoconnectomics," in *Micro-, Meso- and Macro-Connectomics of the Brain*, H. Kennedy, D. C. V. Essen, and Y. Christen, Eds., Accessed via NCBI Bookshelf, Cham (CH): Springer, Mar. 2016, ch. 1, pp. 1–10. DOI: 10.1007/978-3-319-27777-6_1. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK435767/.

[20] J. S. Elam, M. F. Glasser, M. P. Harms, *et al.*, "The human connectome project: A retrospective," *NeuroImage*, vol. 244, p. 118 543, Dec. 2021, Epub 2021 Sep 8, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118543.

[21] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode caenorhabditis elegans," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 314, no. 1165, pp. 1–340, Nov. 1986, ISSN: 0962-8436. DOI: 10.1098/rstb.1986.0056.

[22] L. K. Scheffer, C. S. Xu, M. Januszewski, *et al.*, "A connectome and analysis of the adult drosophila central brain," *eLife*, vol. 9, e57443, 2020, ISSN: 2050-084X. DOI: 10.7554/eLife.57443. [Online]. Available: https://doi.org/10.7554/eLife.57443.

[23] A. Motta, M. Berning, K. M. Boergens, *et al.*, "Dense connectomic reconstruction in layer 4 of the somatosensory cortex," *Science*, vol. 366, no. 6469, eaay3134, 2019. DOI: 10.1126/science.aay3134. eprint: https://www.science.org/doi/pdf/10.1126/science.aay3134. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aay3134.

[24] M. Helmstaedter, K. L. Briggman, and W. Denk, "High-accuracy neurite reconstruction for high-throughput neuroanatomy," *Nature Neuroscience*, vol. 14, no. 8, pp. 1081–1088, 2011, ISSN: 1546-1726. DOI: 10.1038/nn.2868. [Online]. Available: https://doi.org/10.1038/nn.2868.

[25] K. Amunts, C. Lepage, L. Borgeat, *et al.*, "Bigbrain: An ultrahigh-resolution 3d human brain model," *Science*, vol. 340, no. 6139, pp. 1472–1475, 2013.

[26] P. Hagmann, L. Cammoun, X. Gigandet, *et al.*, "Mr connectomics: Principles and challenges," *Journal of Neuroscience Methods*, vol. 194, no. 1, pp. 34–45, 2010, Proceedings of the Workshop "Neuroanatomical Tracing and Systems Neuroscience: The State of the Art", ISSN: 0165-0270. DOI: https://doi.org/10.1016/j.jneumeth.2010.01.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027010000361.

[27] R. Kobayashi and S. Shinomoto, "Inference of monosynaptic connections from parallel spike trains: A review," *Neuroscience Research*, vol. 215, pp. 37–46, 2025, Will BigData change Neuroscience? ISSN: 0168-0102. DOI: https://doi.org/10.1016/j.neures.2024.07.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016801022400097X.

[28] Y. Zhang, Y. Wang, M. Azabou, *et al.*, "Neural encoding and decoding at scale," *arXiv preprint arXiv:2504.08201*, 2025. arXiv: 2504.08201 [q-bio.NC]. [Online]. Available: https://arxiv.org/abs/2504.08201.

[29] J. J. Jun, N. A. Steinmetz, J. H. Siegle, *et al.*, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, 2017, ISSN: 1476-4687. DOI: 10.1038/nature24636. [Online]. Available: https://doi.org/10.1038/nature24636.