Positioning spontaneous activity as ‘Adhesive Dots’: Lessons from AI for data integration in neuroscience

openaccessgovernment.org/article/positioning-spontaneous-activity-as-adhesive-dots-lessons-from-ai-for-data-integration-in-neuroscience/199714

Emily Warrender October 13, 2025

In the previous article, I argued that advancing data integration in neuroscience requires incorporating resting-state spontaneous activity into each experiment, framing it as ‘adhesive dots.’ Here, I extend that discussion by drawing strategic lessons from the success of large language models (LLMs) and by concretizing the earlier claims from the perspective of data

What LLMs can teach us about data integration?

The worldwide construction of data centers illustrates how AI development has advanced through scaling – expanding data volume, model size, and computational resources. LLM performance improves according to power-law scaling when all three expand together. [1] Furthermore, scaling model size and dataset size in tandem has been shown to be near-optimal. [2]

Yet progress has required more than scale: cleaning and curation have been equally crucial. GPT-3 demonstrated the power of large-scale training with filtered Common Crawl, [3] and T5 achieved major improvements by building the C4 dataset after aggressively removing duplicates and low-quality text. [4] PaLM 2 also reported the significant impact of data quality. [5] On the other hand, concerns have been raised about the potential exhaustion of high-quality web text, [6] and partly motivated by applications such as edge AI (the concept of running AI on devices or chips), efficiency efforts such as Mixture-of-Experts (MoE) and compact models are also being pursued in parallel. [7,8] In short, AI has advanced through scaling and curation, while efficiency has also evolved along a complementary path.

Current landscape and challenges in neuroscience

By contrast, neuroscience has yet to fully address the ‘limits of data accumulation’ or the ‘optimization of modeling.’ Advances in optical methods, such as two-photon calcium imaging, now enable large-scale simultaneous measurements at single-neuron resolution. Recently, functional data spanning multiple fields of view have been integrated with EM connectomics, yielding analyses on the order of 75,000 neurons in total – marking major progress. [9]

The greater challenge, however, lies in behavioral and environmental diversity. Laboratory experiments still focus largely on ‘screen-based stimuli’ and ‘controlled tasks.’ Although natural scene stimuli are increasingly employed, [10] real-world contexts are far harder to reproduce. Consider sudden crowd surges in a train station, unexpected issues at immigration control, or nighttime evacuation after a major earthquake with power outages and aftershocks. Such scenarios are common in life, but even if reproduced and recorded, the resulting datasets would be rare and highly specialized. Thus, it becomes essential to examine how such data – naturally incorporating individual differences – can be meaningfully connected to others.

This contextual diversity makes integration particularly difficult. Unlike web text, which is relatively static and independent at scale, neural time-series data are strongly influenced by arousal, attention, individuality, apparatus, and surrounding environment. Therefore, standardized and shareable frameworks (NWB, BIDS, DANDI, OpenNeuro), [(11)] together with detailed metadata such as illumination, arousal state, and behavioral logs, are indispensable.

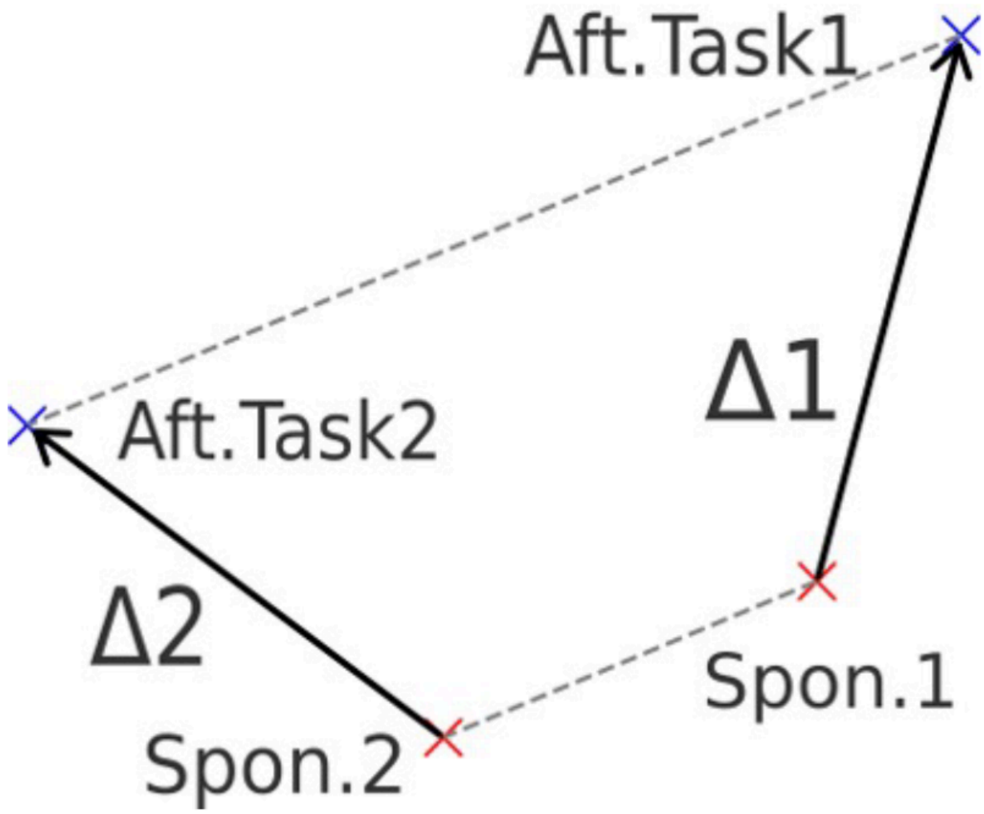

Figure 1. Relationship between spontaneous activity and task-related activity
Spontaneous activity states (Spon.1, Spon.2) represent the baseline states before the task. For simplicity, they are depicted as points in this figure, but in reality they are temporally fluctuating dynamics. Conventionally, analyses have been limited to quantifying the changes Δ1 and Δ2 in post-task activity (Aft.Task1, Aft.Task2) relative to each spontaneous state, without considering the relationship between Spon.1 and Spon.2. If the two differ substantially, comparing only Δ1 and Δ2 is insufficient to properly discuss task effects. Therefore, understanding the relative relationship between spontaneous states is essential, and this figure illustrates the necessity of comparing baseline states in addition to observing differences.

The Idea of a ‘ten-minute spontaneous activity’ baseline

As a realistic step, I have proposed adding a ‘ten-minute spontaneous activity’ segment to each experiment. Spontaneous activity provides a statistical foundation less constrained by specific tasks or environments, reflecting arousal, attention, and individuality while serving as the substrate for task-evoked activity. This has been supported by findings from both human fMRI and mouse research. [12]

Moreover, resting brain activity exhibits scale-free long-range correlations lasting minutes to tens of minutes. [13] A ten-minute window thus captures the key temporal scales while remaining feasible as a unifying standard across laboratories. Longer recordings are, of course, preferable, but a two-step strategy – first establishing a ten-minute baseline and then extending it for refinement – is the most pragmatic approach.

Attaching this “ten-minute spontaneous activity” baseline forms the “adhesive dots” (as described in a previous article), enabling cross-comparison across studies (Fig.1).

This is not an abstract ideal: the role of resting-state structure as a foundation for interpreting and predicting task responses has been empirically demonstrated. [12]

Definitions and non-stationarity of spontaneous activity

The definition of spontaneous activity differs across species and paradigms. In humans, it is typically defined as an ‘eyes-open, fixation-rest state,’ whereas in animals it is categorized as ‘head-fixed, task-free’ or ‘freely moving without tasks.’ Importantly, spontaneous activity is not a static point, but a fluctuating dynamic influenced by arousal and microenvironmental factors. Thus, detailed metadata are indispensable. Notably, the diversity within spontaneous activity is far smaller than the vast diversity of tasks and environments.

This – the “Principle of External Complexity” – highlights that in situations like crowded trains or large gatherings, where one brain is surrounded by dozens or even thousands of other brains, environmental complexity can easily exceed an individual’s internal complexity, making neural data integration difficult. Focusing first on the limited variability of spontaneous activity provides AI with a practical intermediate target for translation and alignment.

A bridge to the next article

The key lesson from LLMs is that breakthroughs emerged not from scaling alone but at the convergence of scaling, curation, and efficiency. In neuroscience, progress likewise requires addressing not only the expansion of neuron counts but also the challenge of behavioral and environmental diversity. As a preparatory step, standardizing the inclusion of a “ten-minute spontaneous activity” segment in each experiment – curated and shared as adhesive dots – would provide a common foundation for integration. This article has emphasized data-side strategies; the next will examine how AI can serve as the glue, through representational mapping and transformation learning, to connect fragmented datasets into a coherent understanding.

CLICK HERE for references

We gratefully acknowledge support from JSPS KAKENHI Grant Number 23K18493, ‘3R Innovation via InVitro2InVivo.’

CLICK HERE previous article

Primary Contributor

Masanori Shimono
Keio University School of Medicine