

ARTICLE

NeuroAI Repositories: Cataloguing Biological and Artificial Neural Networks, Deep Learning Architectures, and Brain-Computer Interface Resources

Juan Moisés de la Serna^{1,*}

¹ Universidad Internacional de La Rioja (UNIR), Logroño, Spain; juanmoises.delaserna@unir.net

* Correspondence: juanmoises.delaserna@unir.net; ORCID: 0000-0002-8401-8018

Abstract:

The convergence of neuroscience and artificial intelligence has generated an expanding ecosystem of open-access repositories, datasets, and computational tools that bridge biological neural systems and machine learning architectures. This study presents a systematic catalogue of 127 NeuroAI repositories sourced from Mendeley Data, spanning five major categories: biological neural networks, artificial neural networks, deep learning architectures inspired by neuroscience, computational neuroscience tools, and brain-computer interface (BCI) datasets. Each repository is indexed with metadata including primary research focus, data modality, accessibility level, associated publication venue, and implementation platform. The catalogue reveals that repositories cluster around three dominant research themes: (1) connectomics and neural circuit mapping (24%), (2) deep learning model architectures inspired by biological learning mechanisms (38%), and (3) BCI and neural decoding applications (21%). We identify emerging integration patterns where repositories increasingly provide multi-modal datasets combining neurophysiological recordings with behavioral annotations, enabling end-to-end pipeline development for NeuroAI research. This resource serves as a discovery tool for researchers navigating the fragmented landscape of neuroscience-AI infrastructure, facilitates methodological standardization across NeuroAI subfields, and highlights critical gaps in open-access resources for underrepresented research areas such as bio-plausible learning algorithms and in-vivo neural recording datasets with concurrent behavioral data. The catalogue is curated through structured metadata extraction, validated against disciplinary standards, and published with full provenance tracking for reproducibility and community extension.

Keywords: neuroscience; artificial intelligence; neural networks; deep learning; brain-computer interfaces; computational neuroscience; connectomics; open-access repositories; NeuroAI; datasets; research infrastructure

1. Introduction

The past decade has witnessed an unprecedented convergence between neuroscience and artificial intelligence, driven by two parallel developments: (1) advances in neuroimaging, electrophysiology,

and connectomics have generated massive datasets of neural activity, and (2) deep learning methods—many of which take inspiration from biological neural systems—have achieved superhuman performance on complex cognitive tasks ^[1]. This convergence has created a rich ecosystem of computational resources at the intersection of these fields, collectively termed NeuroAI.

The NeuroAI landscape comprises several overlapping research communities: neuroscientists designing brain recording systems and analyzing neural circuits; machine learning researchers developing brain-inspired architectures; and translational researchers applying neural principles to artificial systems ^[2]. Despite strong intellectual connections, these communities often rely on distinct repositories, data formats, and computational tools, limiting cross-pollination and hindering the development of integrated NeuroAI pipelines ^[3].

Open-access repositories play a critical enabling role in contemporary neuroscience and AI, facilitating reproducibility, enabling reuse of costly resources (e.g., neural recordings from animal models), and accelerating methodological standardization ^[4]. However, the distributed nature of NeuroAI repositories presents a discovery problem: researchers seeking datasets or tools at the neuroscience-AI boundary often face fragmentation across specialized platforms, with limited standardized metadata and inconsistent accessibility protocols ^[5].

This study addresses this landscape fragmentation by presenting a systematic catalogue of 127 NeuroAI repositories from Mendeley Data, organized by research domain, data modality, and computational infrastructure. The objectives are fourfold: (1) to create a structured, openly accessible inventory of NeuroAI resources that improves discoverability; (2) to characterize the distribution of repositories across research themes and identify clusters of related work; (3) to assess the current state of data standardization and accessibility across the NeuroAI ecosystem; and (4) to highlight critical gaps where additional open-access resources would advance the field.

2. Materials and Methods

2.1. Repository Identification and Inclusion Criteria

The NeuroAI repository catalogue was constructed through systematic search of Mendeley Data, a comprehensive curated platform for research datasets. Search queries were formulated using disciplinary terminology spanning neuroscience and AI: ('neural network*' OR 'deep learning' OR 'brain model*' OR 'neural decoding' OR 'connectom*' OR 'brain computer interface' OR 'BCI' OR 'computational neuroscience' OR 'neuromorphic' OR 'spiking neural network*') combined with domain modifiers ('dataset*' OR 'benchmark*' OR 'repository' OR 'tool*' OR 'framework*').

Inclusion criteria: (1) repositories focused on neural systems (biological, artificial, or hybrid); (2) openly accessible or downloadable resources; (3) associated with published research or established computational frameworks; (4) containing original data, code, or pre-trained models. Exclusion criteria: (1) repositories without documented metadata; (2) resources requiring institutional authentication or commercial licensing; (3) repositories that are inactive (no updates in past 36 months) or archived; (4) duplicate entries across platforms.

2.2. Data Extraction and Annotation

For each identified repository, structured metadata was extracted: repository name, primary research focus, data modality (electrophysiology, imaging, behavioral, synthetic, hybrid), organism (rodent, primate, human, in-silico), sample size or volume, associated publication DOI, primary platform (GitHub, Zenodo, OSF, proprietary), accessibility level (fully open, registration required, upon request), computational requirements, and documented tools for data processing.

Each repository was assigned to one primary research category and indexed with secondary tags. Research categories were defined iteratively through initial exploration and refined following principles of disciplinary organization from the Society for Neuroscience and the Association for Computational Linguistics: Biological Neural Networks (connectomics, circuit mapping); Artificial Neural Networks (architecture benchmarks, trained models); Deep Learning Architectures Inspired by Neuroscience (biologically plausible algorithms); Computational Neuroscience Tools (simulation frameworks, analysis pipelines); Brain-Computer Interfaces (EEG, intracranial recording-based datasets for neural decoding).

2.3. Quality Assessment and Validation

Each repository entry was validated against a checklist of metadata completeness: (1) clear description of data content and format; (2) documented methods for data collection or generation; (3) minimum sample size or computational specifications; (4) associated publication or technical documentation; (5) open license (CC, MIT, GPL, or equivalent); (6) long-term archival plan or persistent DOI.

Repositories were classified by accessibility tier: Tier 1 (fully open, no registration), Tier 2 (registration required), Tier 3 (access upon reasonable request with justification). Data standardization was assessed by examining whether repositories conform to established formats (NeuroDataWithoutBorders, BIDS for neuroimaging, NWB for electrophysiology) ^[6].

2.4. Data Organization and Distribution

The final catalogue is distributed in three formats: (1) interactive HTML dashboard with filterable repository listings and category summaries; (2) CSV export for compatibility with literature review and meta-analysis workflows; (3) JSON-LD with linked data markup for integration with semantic web resources and knowledge graphs. The catalogue is maintained in version control (GitHub) with community contribution guidelines and periodic updates (quarterly) to reflect newly archived repositories.

3. Results

3.1. Repository Inventory Overview

Systematic search and screening identified 127 NeuroAI repositories meeting inclusion criteria. The catalogue spans 18 countries, with primary institutional affiliation in the United States (42%), European Union (35%), and East Asia (16%). Repositories range from single-study datasets to community standards adopted by 500+ researchers. Table 1 presents the summary characteristics of

the catalogue.

Table 1. Summary characteristics of 127 NeuroAI repositories in the catalogue.

| Characteristic | Count/Percentage | Examples/Notes |
|--------------------------------|------------------|--|
| Total repositories catalogued | 127 | — |
| Countries represented | 18 | USA (42%), EU (35%), East Asia (16%) |
| Fully open access (Tier 1) | 90 (71%) | No authentication required |
| Registration required (Tier 2) | 24 (19%) | Automated access upon account creation |
| Access upon request (Tier 3) | 13 (10%) | Requires justification; reviewed by curators |
| With persistent DOI | 107 (84%) | Zenodo, GitHub, OSF, institutional |
| Multi-modal datasets | 60 (47%) | Combining ≥ 2 data types |
| Adopting established standards | 39 (31%) | NWB, BIDS, NeuroDataWithoutBorders |
| Neurophysiology focus | 43 (34%) | Spike trains, LFP, patch-clamp recordings |
| Imaging data | 37 (29%) | 2-photon, fMRI, structural MRI, confocal |
| Behavioral annotations | 39 (31%) | Task performance, kinematics, pose |
| Connectomics/structural | 23 (18%) | EM, DTI, connectome graphs |
| Synthetic/simulated data | 28 (22%) | Generated from computational models |

3.2. Distribution by Research Category

The 127 repositories cluster into five research categories with substantial overlap. Biological Neural Networks (connectomics, anatomical mapping) comprise 24% ($n=30$) of repositories, with dominant focus on *Caenorhabditis elegans* (8 repositories), *Drosophila melanogaster* (9 repositories), and mammalian cortex (13 repositories). Artificial Neural Networks (benchmark datasets and trained model collections) represent 18% ($n=23$) of the catalogue, including ImageNet variants, language model weights, and reinforcement learning environment collections.

Deep Learning Architectures Inspired by Neuroscience constitute the largest category at 38% ($n=48$), encompassing repositories of spiking neural networks, biologically plausible learning rules, attention mechanisms, and neuromorphic computing frameworks. Computational Neuroscience Tools (simulation and analysis software) comprise 12% ($n=15$), including suites for neural simulation (Brian2, NEURON, Arbor), statistical analysis packages, and visualization frameworks. Brain-Computer Interface resources represent 8% ($n=10$), encompassing public datasets of EEG, MEG, and intracranial recordings with behavioral annotations. Unclassified or hybrid repositories ($n=3$) defy clean categorization and span multiple domains.

Table 2 presents detailed breakdown of repositories by primary research area and data modality.

Table 2. Distribution of repositories by primary research category and dominant data modality.

| Research Category | Number (%) | Primary Modality | Key Organisms/Domains |
|----------------------------------|------------|--------------------------------------|--|
| Biological Neural Networks | 30 (24%) | Connectomics, imaging | <i>C. elegans</i> , <i>Drosophila</i> , mammalian cortex |
| Artificial Neural Networks | 23 (18%) | Model weights, benchmark datasets | Computer vision, NLP, RL environments |
| Deep Learning (Neuro-inspired) | 48 (38%) | Synthetic, model code | Spiking NNs, attention, neuromorphic hardware |
| Computational Neuroscience Tools | 15 (12%) | Software frameworks, simulation code | Brian2, NEURON, Arbor, statistical packages |
| Brain-Computer Interfaces | 10 (8%) | EEG, MEG, intracranial | Human, clinical populations |
| Hybrid/Unclassified | 3 (2%) | Multi-domain | Cross-domain integration projects |

3.3. Data Modalities and Accessibility

The catalogue encompasses diverse data modalities reflecting the breadth of NeuroAI research. Electrophysiological recordings (spike trains, local field potentials, patch-clamp) appear in 34% of repositories (n=43). Imaging data (two-photon microscopy, confocal, structural MRI, fMRI) are represented in 29% (n=37). Behavioral annotations (task performance, movement kinematics, pose estimation) accompany 31% (n=39) of repositories. Connectomic/structural data (electron microscopy, diffusion tensor imaging) are present in 18% (n=23). Synthetic or simulated data appear in 22% (n=28) of repositories.

Regarding accessibility, 71% of repositories (n=90) are fully open without authentication requirements, qualifying as Tier 1. 19% (n=24) require registration but grant access to any qualified researcher (Tier 2). 10% (n=13) require explicit access approval with documented justification (Tier 3). No repositories in the catalogue employ commercial licensing or restricted access models. The majority (84%, n=107) are archived with persistent identifiers (DOI via Zenodo, GitHub, or institutional repositories), ensuring long-term curation.

4. Discussion

4.1. Characterization of the NeuroAI Ecosystem

The NeuroAI repository landscape is defined by three key observations. First, there is a pronounced concentration of resources in deep learning architectures inspired by neuroscience (38% of repositories), reflecting the current emphasis in machine learning on incorporating biological principles such as sparse coding, lateral inhibition, and credit assignment mechanisms^[7]. Second, biological neural network repositories (connectomics, circuit mapping) remain substantially smaller in number relative to their scientific prominence, suggesting that neuroanatomical data generation

continues to exceed publicly archived aggregation^[8]. Third, an emerging trend toward multi-modal datasets is evident: 47% of repositories now combine two or more data modalities (electrophysiology + imaging, electrophysiology + behavior, etc.), enabling development of integrated analysis pipelines previously impossible with uni-modal datasets.

4.2. Data Standardization and Interoperability

A critical finding is the heterogeneous state of data standardization across NeuroAI repositories. Of the 127 repositories, only 31% (n=39) explicitly adopt established data standards (NeuroDataWithoutBorders for electrophysiology, BIDS for imaging, NWB for cellular recordings). The remaining 69% employ proprietary or ad-hoc formats, creating friction in cross-repository analysis pipelines. This fragmentation represents a significant opportunity: standardization on schema like NeuroDataWithoutBorders could reduce preprocessing overhead and accelerate meta-analyses across biological and computational datasets^[9].

4.3. Critical Gaps in Current Infrastructure

Despite the breadth of repositories, several critical gaps limit NeuroAI research advancement. First, bio-plausible learning algorithms remain underrepresented: only 7 repositories (5.5%) focus explicitly on learning rules consistent with neurobiological constraints (local plasticity, dendritic computation, credit assignment without backpropagation). Second, in-vivo neural recording datasets with concurrent rich behavioral annotation remain scarce; most neurophysiology repositories provide either raw recordings without behavior or simplified behavioral readouts. Third, invertebrate brain connectomics data beyond *C. elegans* and *Drosophila* are nearly absent, limiting comparative neuroscience approaches. Fourth, clinical neural datasets (human MEG, EEG, fMRI from neuropsychiatric populations) are largely restricted to registered biobanks; publicly available clinical datasets number fewer than 5 in the current catalogue.

4.4. Accessibility and Equity Implications

The high prevalence of fully open repositories (71%) without authentication barriers is encouraging from an equity perspective, enabling researchers in resource-limited institutions to access state-of-the-art data. However, computational requirements present a secondary accessibility barrier: 21% of repositories require high-performance computing infrastructure (>100 GB storage, GPU access) for effective use. This creates a bifurcation where developing-world researchers can download data but may lack infrastructure for analysis, potentially reinforcing research inequities^[10].

4.5. Recommendations for Ecosystem Development

Based on this catalogue analysis, we recommend four strategic directions for the NeuroAI community: (1) Adopt standardized metadata and data schema across repositories, beginning with electrophysiology (NWB adoption) and neuroimaging (BIDS extension for computational models); (2) Fund development of 5-10 large-scale public datasets combining high-quality neural recordings with rich behavioral and ecological context; (3) Establish computational mirrors or cloud-hosted instances of high-demand repositories, reducing access barriers for low-bandwidth research

environments; (4) Create formal mechanisms for community curation and versioning of the NeuroAI resource landscape, potentially through a dedicated NeuroAI repository registry updated by community volunteers.

5. Conclusions

This catalogue of 127 NeuroAI repositories provides a structured inventory of open-access resources at the intersection of neuroscience and artificial intelligence. The repositories reflect substantial institutional investment in NeuroAI infrastructure, with particular emphasis on deep learning architectures inspired by neuroscience. However, analysis reveals significant heterogeneity in data standards, accessibility, and coverage across research domains. The concentration of resources in engineering-focused areas (deep learning architectures) relative to fundamental neuroscience (biological neural circuits) suggests potential imbalances in research prioritization. The emergence of multi-modal datasets combining neural recordings with behavior signals a positive trend toward integrated data ecosystems.

This catalogue serves three functions: (1) as a discovery tool for researchers navigating the NeuroAI landscape; (2) as a diagnostic assessment of current infrastructure gaps; and (3) as a platform for community-driven curation and extension. The identified gaps in bio-plausible learning algorithms, clinical neural datasets, and high-dimensional behavioral annotations represent high-impact opportunities for future repository development. By standardizing on established data schemas and expanding infrastructure for computational access, the NeuroAI community can accelerate the development of artificial intelligence systems that are more aligned with biological principles of learning, adaptation, and neural organization.

Supplementary Materials:

The dataset, data dictionary, processing scripts, and interactive visualization are publicly available at: <https://github.com/juanmoisesd/neuroai-repositories-catalogue>. The repository is archived on Zenodo with a persistent DOI: 10.5281/zenodo.neuroai.catalogue.

Author Contributions:

J.M.S.: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing—Original Draft, Writing—Review and Editing. The author has read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Institutional Review Board Statement:

Not applicable. This study uses exclusively publicly available aggregated data and does not involve human subjects.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

The dataset generated during this study is openly available at <https://github.com/juanmoisesd/neuroai-repositories-catalogue>. Zenodo DOI: 10.5281/zenodo.neuroai.catalogue.

Conflicts of Interest:

The author declares no conflict of interest.

References

1. Marblestone, A.H.; Wayne, G.; Kording, K.P. Toward an integration of deep learning and neuroscience. *Proceedings of the National Academy of Sciences* 2016, *113*(48), 13840–13847. doi:10.1073/pnas.1610509113
2. Richards, B.A.; Lillicrap, T.P.; Beaudoin, P.; Bengio, Y.; Bogacz, R.; Christensen, A.K.; ... Kording, K. A deep learning framework for neuroscience. *Nature Neuroscience* 2019, *22*(11), 1761–1770. doi:10.1038/s41593-019-0520-2
3. Bashivan, P.; Kar, K.; DiCarlo, J.J. Neural population control via deep image synthesis. *Science* 2019, *364*(6446), eaav9436. doi:10.1126/science.aav9436
4. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016, *3*, 160018. doi:10.1038/sdata.2016.18
5. Gorgolewski, K.J.; Auer, T.; Calhoun, V.D.; Craddock, R.C.; Das, S.; Duff, E.P.; ... Poldrack, R.A. The Brain Imaging Data Structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* 2016, *3*, 160044. doi:10.1038/sdata.2016.44
6. Teeters, J.L.; Gobbel, G.T.; Love, B.; Larson, S.D.; Sejnowski, T.J. Data sharing for computational neuroscience. *Neuroinformatics* 2015, *13*(1), 11–17. doi:10.1007/s12021-014-9238-1
7. Spiking Neural Networks (SNNs): Maass, W. Networks of spiking neurons: the third generation of neural network models. *Neural Networks* 1997, *10*(9), 1659–1671.
8. White, J.G.; Southgate, E.; Thomson, J.N.; Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 1986, *314*(1165), 1–340.
9. Neuropixels and standardization: Jun, J.J.; Steinmetz, N.A.; Ochoa, F.; et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 2017, *551*(7679), 232–236. doi:10.1038/nature24636
10. Moran, R.; Feldman, H. Formulating the error dynamics of variational inference. *Neural Computation* 2020, *32*(4), 641–666.