CellPress

Trends in Neurosciences

Opinion

Principles of intensive human neuroimaging

Eline R. Kupers [1,2,*], Tomas Knapen [3,4,5], Elisha P. Merriam [6], and Kendrick N. Kay [1,*]

The rise of large, publicly shared functional magnetic resonance imaging (fMRI) data sets in human neuroscience has focused on acquiring either a few hours of data on many individuals ('wide' fMRI) or many hours of data on a few individuals ('deep' fMRI). In this opinion article, we highlight an emerging approach within deep fMRI, which we refer to as 'intensive' fMRI: one that strives for extensive sampling of cognitive phenomena to support computational modeling and detailed investigation of brain function at the single voxel level. We discuss the fundamental principles, trade-offs, and practical considerations of intensive fMRI. We also emphasize that intensive fMRI does not simply mean collecting more data: it requires careful design of experiments to enable a rich hypothesis space, optimizing data quality, and strategically curating public resources to maximize community impact.

Highlights

A growing number of publicly available human functional magnetic resonance imaging (fMRI) data sets use a 'deep' sampling approach, where many hours of data are acquired from a few individuals. We highlight an emerging approach within deep fMRI, which we refer to as 'intensive' fMRI: the creation of large-scale, publicly shared data sets that extensively sample cognitive phenomena to support the investigation of brain function at the single voxel level.

We discuss key principles of intensive fMRI, its benefits and challenges, and practical considerations in creating intensive fMRI data sets.

We discuss how intensive fMRI can advance scientific discovery in systems neuroscience, machine learning, and artificial intelligence.

The emergence of intensive fMRI

Neuroscience is witnessing a growing interest in collecting and freely distributing large-scale data sets. Here, we focus on fMRI, one of the most popular non-invasive neuroimaging techniques for measuring brain activity. Initial large-scale fMRI data sets focused on wide sampling, maximizing data quantity in terms of sampling many ($n \geq 100$) participants (e.g., the Healthy Brain Network [1], the 1000 Functional Connectomes Project [2], the Human Connectome Project (HCP) [3], the Adolescent Brain Cognitive Development study [4], and the UK BioBank [5]). By dramatically increasing the number of participants relative to conventional fMRI studies, these data sets have supplied the statistical power to study brain–behavior relationships [6] (but see [7,8]) and have revealed fine-scale variability across individuals in the function and organization of cortex [9,10], subcortex [11], and cerebellum [12].

A more recent direction in fMRI is one that emphasizes deep sampling of a small number of participants ($n \leq 20$), each scanned for many hours (e.g., StudyForrest [13], My Connectome [14], Midnight Scan Club [15], and others [16–20]). By collecting many hours of data per participant, these data sets have enabled detailed investigation of individual differences in brain structure and function [21] and their link to perception and behavior [22]. Importantly, this approach enables fine-grained characterization of individual brains in their native space, avoiding the blurring caused by group averaging [23].

In this opinion article, we highlight the emergence of a special form of deep sampling, which we refer to as intensive fMRI. In intensive fMRI, researchers collect massive amounts of data targeting specific cognitive phenomena, with emphasis on maximizing data quality. The goal of intensive fMRI is to generate general-purpose data sets that enable detailed computational characterization of local neural population activity and that help bridge human cognitive neuroscience to machine learning (ML) and artificial intelligence (AI) (Box 1).

Intensive fMRI data sets are resource demanding, but, if done well, they provide tremendous value for the neuroscience community. Intensive fMRI data sets enable a wide range of possible

[1]Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN, USA
[2]Department of Psychology, Stanford University, Stanford, CA, USA
[3]Spinoza Centre for Neuroimaging, Amsterdam, the Netherlands
[4]Netherlands Institute for Neuroscience, Royal Netherlands Academy of Sciences, Amsterdam, the Netherlands
[5]Cognitive Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit, Amsterdam, the Netherlands
[6]Laboratory of Brain and Cognition, National Institute of Mental Health, NIH, Bethesda, MD, USA

*Correspondence: kupers@umn.edu (E.R. Kupers) and kay@umn.edu (K.N. Kay).

Box 1. The marriage between intensive fMRI, machine learning, and artificial intelligence

Over the past two decades, neuroimaging research has shifted from localizations and activations of many fMRI voxels in a brain region [89] toward the use of computational models to understand how information is encoded by local neural populations within a single fMRI voxel [90]. This shift has coincided with the rise of large feedforward convolutional neural networks in ML and AI. These networks have architectures inspired by the human visual system and achieve remarkable performance on a variety of pattern recognition tasks [91]. The impressive performance of these neural networks has sparked a new synergy between neuroscience and ML/AI fields [92–94].

As neuroscientists leverage ML/AI approaches to study the brain, and as ML/AI researchers strive to build neuro-inspired models that reach or exceed human cognitive capabilities, there is a growing need for more and higher quality brain data to support these endeavors. By deeply sampling many experimental conditions within single individuals, intensive fMRI can help meet this need. Indeed, there is an emerging synergy between intensive fMRI data sets and ML/AI [45]. For example, intensive fMRI data sets now play a central role in model prediction competitions, such as the Algonauts Project [95,96] and the Brain-Score platform [97]. Excitingly, there are new intensive fMRI initiatives involving novel types of experiments, such as viewing dynamic videos [33,98], interactive closed-loop behavioral protocols and gameplay [32], and listening to naturalistic spoken stories [31]. Such initiatives expand intensive fMRI into an even wider range of topics in cognitive neuroscience, such as reinforcement learning, and strengthen the synergy between intensive fMRI and ML/AI research.

scientific inferences and, therefore, may be more cost-effective compared with several small-scoped fMRI data sets that test singular hypotheses. In addition, intensive fMRI creates research opportunities for those without the resources to acquire such data sets themselves, improving equity in the scientific community [24].

Here, we lay out the fundamental principles of intensive fMRI. In brief, these principles are: (i) design well (i.e., design a set of carefully controlled experiments that enable a rich hypothesis space); (ii) scan more (i.e., acquire more fMRI data and auxiliary measures per participant than is typical); (iii) optimize quality (i.e., acquire not only more, but also higher quality data), and (iv) share better (i.e., disseminate easily accessible and extensively documented data). Hence, achieving a successful intensive fMRI data set requires simultaneously optimizing many aspects of data creation (Table 1).

We also discuss practical considerations when creating intensive fMRI data sets. Our exposition stems from our experience creating and using the Natural Scenes Dataset (NSD), a large-scale 7 Tesla fMRI data set in which eight participants each viewed 9000–10 000 complex natural scenes

Table 1. Principles, considerations, and practical examples for creating intensive fMRI data sets that provide extensive and high-quality sampling of specific cognitive phenomena

Principles	Considerations	Practical examples
Design well	Elicit rich cognitive phenomena	Sample many distinct experimental conditions Cover a large hypothesis space
Scan more	Have a ‘data-hungry’ mindset	Collect large amounts of data per participant Acquire auxiliary measures (e.g., venograms or quantitative MRI) Acquire physiological data (e.g., heart rate or breathing)
Optimize quality	Optimize data acquisition	Screen, select, and train participants Optimize trial and experimental design Maximize signal-to-noise ratio in single voxels
	Optimize data preparation	Perform frequent and extensive data inspection Tailor preprocessing and analysis steps if needed Develop new analysis techniques to improve data quality
Share better	Create a community-oriented data set	Solicit feedback from community before data collection Release multiple versions of the data to allow for different use cases Include comprehensive documentation

over the course of 30–40 scan sessions while performing a continuous recognition task [25]. In doing so, we do not imply that NSD is the first or the sole intensive fMRI data set. Intensive fMRI is not an all-or-none property and is best viewed as a continuum. Moreover, elements of intensive fMRI can be found in several previous data sets [13–17,19,25–33].

However, we do not rehash issues in big data that have been covered elsewhere, such as the general advantages of collecting large amounts of data from individual participants [34–40], the trade-off between wide versus deep sampling [41,42], and data requirements for establishing robust brain–behavior relationships [43,44].

Principle 1: design well

Intensive fMRI does not simply mean collecting more data. The first principle of intensive fMRI is to create well-designed experiments that provide rich information about the properties, representations, and mechanisms used by neuronal populations. Arguably, this implies that explicit manipulation of sensory, cognitive, and/or motor function (e.g., task-based fMRI) should play a central role in intensive fMRI. While resting-state fMRI has advanced our understanding of functional connectomes and broader brain networks [23], it is limited in its power to reveal insights into neural computations.

'Designing well' involves several considerations. Experimental conditions should be motivated by theory and should seek to elicit specific targeted cognitive processes (e.g., [17,26,33]). Stimuli and trial properties should be carefully designed to optimize statistical power (e.g., to adjudicate different computational models [45]). Stimuli should be high quality (e.g., sharp, high-resolution images [17,27]) and delivered with precise and accurate timing (e.g., as in [25]). Ideally, the stimuli, participant instructions, and experimental protocol provide sufficient constraints such that participants execute the specific intended task [46]. Evidence of task execution can be provided through behavioral responses collected from the participants. Finally, to assist in the analysis and interpretation of the main experiment, researchers should consider conducting additional experiments (e.g., functional localizers [14,31]) or preparing stimulus annotations [30,31,33].

When designed well, an intensive fMRI data set can enable rigorous testing of a large space of hypotheses (Figure 1). For example, in the NSD [25], the combination of sampling responses to many naturalistic scene images and the use of sophisticated computational analyses revealed new food-selective networks in ventral temporal cortex [47–49], new mechanistic insights into how the brain preserves temporal memories over long timescales (days, weeks, months, and longer) [50], and new evidence in a long-standing debate regarding orientation tuning in early visual cortex [51].

Principle 2: scan more

The second principle of intensive fMRI is to collect a very large amount of data, including fMRI responses to many distinct experimental conditions, complementary anatomical scan contrasts (e.g., T1-weighted and T2-weighted anatomical scans, venograms, angiograms, diffusion-weighted MRI, quantitative MRI), and multiple auxiliary measures (e.g., functional localizers, behavioral data, eye-tracking data, and physiological data [26,30]). This type of data-hungry mindset was first notable in the field of resting-state fMRI, where data sets included multiple scan contrasts as well as task-based fMRI [13–15,52–55]. Therefore, we consider these resting-state data sets as precursors of intensive fMRI.

Even if the collectors of an intensive fMRI data set do not have plans to exploit a given type of data (e.g., physiological data), they cannot fully predict what types of data users might be interested in,

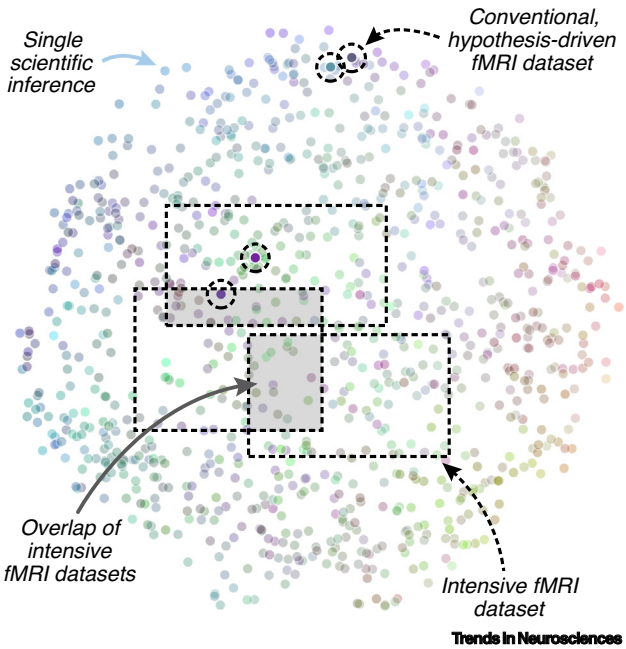

Figure 1. How intensive functional magnetic resonance imaging (fMRI) data sets span the space of possible scientific inferences in cognitive neuroscience research. Consider the space of possible inferences that one might make from experimental data (colored dots). A conventional hypothesis-driven fMRI data set has a small hypothesis space (dashed circle) and allows for a small number of inferences. An intensive fMRI data set has a large hypothesis space (dashed rectangle) and allows for a multitude of different inferences. With inferential overlap across multiple intensive data sets (gray regions), there is the potential to integrate data sets and enable more far-reaching inferences about brain structure and function.

or what analyses users might wish to perform (see Outstanding questions). Hence, we believe it is better to err on the side of collecting more data than one might normally collect, thereby ensuring the longevity of an intensive fMRI data set. For example, saving the phase component of fMRI images can enable certain approaches to denoising [56,57], vascular characterization [58], and distortion compensation [59]. As another example, saving raw *k*-space data from individual receiver coils, although onerous, could allow later capitalization from improved reconstruction methodology. Finally, collecting certain anatomical MRI measures could provide the basis for improved interpretation of fMRI data [60–62]. This includes quantitative MRI, diffusion-weighted MRI, venograms, and anatomical data at the mesoscopic scale [63,64].

Scanning involves inevitable trade-offs between the number of trials, number of experimental conditions, and number of participants. Due to constraints on scan time and the desire to deeply sample individuals, the number of participants in intensive fMRI data sets has historically been relatively low (often less than ten). Hence, one drawback of intensive fMRI is that it is not well suited to answer scientific questions regarding patterns of variability across a population, which is often of interest in clinically oriented fields [65]. For such questions, a wide fMRI design is more appropriate.

Principle 3: optimize quality

The third principle of intensive fMRI is making an effort to achieve the highest quality data. Instead of collecting data that are *good* enough for testing one particular hypothesis (as might be done in conventional fMRI), intensive fMRI seeks to maximize data quality, thereby enabling the testing of many potential hypotheses. Optimization of quality can be performed with respect to all aspects of data creation, including the experimental protocol (stimulus design and trial timing), data

acquisition (scanning parameters, image quality, and participant monitoring), data preparation (preprocessing, quality control, and across-session alignment; as illustrated by [25]), and public release (data curation, data documentation; e.g., [13]).

Improvements in data quality can come from new developments in MRI acquisition methodologies, including ultra-high magnetic field strength, optimized radiofrequency coils, and new pulse sequences and reconstruction methods [66]. A notable example is the Human Connectome Project [3], which developed new multiband-accelerated pulse sequences and customized scanner hardware optimized for functional and diffusion-weighted MRI. However, caution must be exercised. New techniques may be accompanied by a cost–benefit trade-off [67], and 'bleeding edge' methodology, if insufficiently tested, may lead to potential artifacts, scanning instabilities, and other unforeseen issues. For example, when multiband factors are pushed too far, undesirable increases in noise can occur [68].

Another approach for improving data quality is to screen, select, and train participants. Screening could involve finding participants who are highly proficient at staying still, staying awake, and performing instructed tasks; one way to do this is to bring participants in for a test session and quantify the strength and reliability of their fMRI signals (e.g., [25]). Training could involve carefully measuring participant task performance before actual scanning (e.g., [16]). Task performance during training might even be used to select participants. Finally, monitoring participants with performance metrics and providing performance bonuses during data acquisition may be especially effective for achieving high-quality data (as illustrated in [25]). One drawback of restrictive selection of participants is that it may result in a biased demographic sample. For example, the experimenter might end up recruiting mostly individuals from communities who live close to the scanning facility (to facilitate participant retention). As another example, selecting participants who exhibit high reliability of experimentally driven fMRI signals risks undersampling individuals with variable neurocognitive ability. Therefore, selection of participants with the aim of data quality optimization can run counter to diversity and inclusivity goals, and perpetuate marginalization of under-represented communities, a problem endemic in human neuroscience research [69,70]. Thus, it is important for experimenters to assess the broader implications of their recruitment procedures.

A third way to optimize quality is to perform frequent (ideally daily) visual checks of raw and preprocessed data. Such data visualizations are critical for detecting, diagnosing, and remedying any potential problems. Visualizations should be both highly detailed (so that small errors can be detected) and comprehensive (so that the entire data set is checked). Frequent control checks may appear to be a general aspect of creating any fMRI data set. However, we stress that, compared with conventional fMRI, intensive fMRI involves acquiring an unprecedented amount of data. Thoroughly assessing large amounts of data is time-consuming and requires high levels of meticulousness, domain expertise, and computational skills.

Preprocessing pipelines are critical for preparing fMRI data for further use [71–73]. Automated pipelines, especially those that are vetted by the community and conform to accepted standards, may be favorable when analyzing massive amounts of data. Examples include MRIQC [74], which provides comprehensive data quality overviews, and fMRIprep [75], which has shown robust and high-quality preprocessing results. However, we believe that automated tools and quantitative metrics are best complemented with manual visual inspections of preprocessing results (as illustrated in [25]). Furthermore, customization of pipelines (as opposed to off-the-shelf pipelines) may be essential for achieving optimal results when fMRI data have unconventional resolutions or signal-to-noise characteristics [60].

Principle 4: share effectively

Open sharing of data has increasingly become a standard practice throughout neuroscience, especially in the field of human fMRI. Yet, sharing data is not the same as creating data that are designed to be shared. The fourth principle of intensive fMRI is to create data sets that are open by design and serve as community resources.

To be an effective shared resource, an intensive fMRI data set should be accompanied by an accessible and searchable data manual that goes far beyond the level of detail in a traditional methods section of a scientific manuscript. This data manual should be comprehensive, covering the contents of all data files and the procedures used to generate the data files (possibly with links to corresponding code). The manual should also cover experimental protocols, scanning protocols, precise timing logs, known data problems, and so on.

To maximize the utility of a data set to the larger community, the needs of users should be considered at multiple stages of data-set creation. Before data collection, creators should consider soliciting feedback from the community. This can provide valuable insights into what types of data are needed and could inform critical aspects of experimental design. After data collection, creators should prepare data in ways that accommodate different users' needs. For instance, researchers should consider releasing not only preprocessed data to enable the quickest uptake, but also the raw data in a standardized file format and file organization (e.g., following naming conventions of the Brain Imaging Data Structure [76]). In addition, data sets should include code or software tools if necessary for easy data access and manipulation (for illustrative examples, see [3,13,25,27,31]). Finally, with the emergence of very large data sets, proper data management (storage, analysis, and sharing) becomes crucial, and researchers may consider software tools to streamline data release and version control (e.g., [28]). Example of such tools include the Extensible Neuroimaging Archive Toolkit [77] and DataLad [78].

Public sharing of intensive fMRI data comes with the responsibility to safeguard the privacy of participants and prevent misuse [79,80]. As ML and AI algorithms become more powerful, researchers should consider the risk that deidentified data can be reidentified [81]. To navigate the ethical and legal landscape of public data, researchers may consider resources from international data governance efforts, such as the Brain Research International Data Governance & Exchange project [82].

Challenges of intensive fMRI

Intensive fMRI faces multiple challenges that need to be addressed to secure its future in the scientific process (see Outstanding questions). Here, we identify three major challenges. The first is concentration of power. Only a limited number of groups have the expertise, capacity, institutional infrastructure, and financial resources to create intensive fMRI data sets. These groups ultimately decide what types of data set to create and the boundaries of the hypothesis spaces, and there is a risk of concentrating power in too few hands.

A second challenge is data set staleness or overfitting. There is a growing community of researchers who are productively using publicly available data. Interestingly, despite the thousands of data sets hosted by online data repositories, such as OpenNeuro [83], most users take advantage of only a handful of data sets [84]. A potential pitfall of mining a small number of data sets is overfitting. For example, a given data set may be unrepresentative of the broader population or of the cognitive phenomena being targeted, which may lead to biased scientific inferences. Intensive fMRI may very well exacerbate this effect. Data mining can also result in situations where results from intensive

fMRI data sets are used to generate hypotheses that are then inadvertently tested by other groups on the same data sets. To avoid such circular inferences, users should carefully separate exploratory versus confirmatory analyses and be aware of the specific data sets used by others.

The third challenge is that computational models fitted on intensive fMRI data sets may inherit biases present in the data. For instance, the stimulus set used in a given experiment may contain idiosyncratic distributions of image classes [85] or categories [86] inherited from large image databases. Careful curation of the stimulus set can combat these idiosyncrasies and achieve more balanced distributions [30]. Furthermore, data set users can mitigate biases by analyzing multiple data sets and by exploiting small data sets that are not commonly used, but which may provide valuable insights [87].

Concluding remarks and future perspectives

Several labs around the world are now creating large-scale, publicly-available intensive fMRI data sets. These data sets provide unprecedentedly deep sampling of rich cognitive phenomena and are being widely used by researchers to gain new insights into brain function and structure. The continued rise of intensive fMRI initiatives may lead to fundamental changes in the way in which science is conducted. For example, intensive fMRI data sets could become benchmark data sets for exploration of scientific hypotheses. Furthermore, such data sets could be routinely combined with smaller conventional fMRI experiments that provide more direct tests of hypotheses. This integration of large- and small-scale data sets will require effective methods for translating large data sets to smaller data sets [88] as well as methods for bridging big data sets (see Outstanding questions). In the future, we hope that routine integration of a wealth of publicly shared small and large neuroimaging data sets will help answer new questions in cognitive neuroscience, ML, and AI research.

Outstanding questions

To ensure the longevity of intensive fMRI data sets, creators may seek to overshoot data targets. What specific aspects of data acquisition should researchers overshoot? How can researchers future-proof their data set against new acquisition technologies?

How can the research community exert influence on the design and preparation of intensive fMRI data sets? How can data set creators avoid collisions and capitalize on the relative strengths of different data sets?

How can researchers integrate information from multiple intensive fMRI data sets? One possibility is to design shared experiments that could serve as 'Rosetta Stones' to bridge different data sets. What stimuli and tasks would be optimal for these shared experiments? What analysis techniques would be optimal for mapping neural activity from one data set to another?

When generating new hypotheses from prior work involving large-scale publicly available fMRI data sets, researchers may wish to track exactly how these data sets were used to prevent circular scientific inferences. What are effective ways to track the specific use of intensive fMRI data sets by other researchers?

What are ways to incentivize the re-preprocessing of existing data sets using better data preparation methods? To what extent would it be fruitful for the neuroscience community to hold preprocessing 'challenges', similar to model prediction 'challenges'?

Acknowledgment

This work was supported by NIH grant R01EY034118 (K.K.).

Declaration of interests

The authors declare no competing interests in relation to this work.

References

1. Alexander, L.M. *et al.* (2017) An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* 4, 170181
2. Mennes, M. *et al.* (2013) Making data sharing work: the FCP/INDI experience. *NeuroImage* 82, 683–691
3. Van Essen, D.C. *et al.* (2012) The Human Connectome Project: a data acquisition perspective. *NeuroImage* 62, 2222–2231
4. Casey, B.J. *et al.* (2018) The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* 32, 43–54
5. Miller, K.L. *et al.* (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536
6. Marek, S. *et al.* (2022) Reproducible brain-wide association studies require thousands of individuals. *Nature* 603, 654–660
7. Spisak, T. *et al.* (2023) Multivariate BWAS can be replicable with moderate sample sizes. *Nature* 615, E4–E7
8. Wu, J. *et al.* (2023) The challenges and prospects of brain-based prediction of behaviour. *Nat. Hum. Behav.* 7, 1255–1264
9. Benson, N.C. *et al.* (2022) Variability of the surface area of the V1, V2, and V3 maps in a large sample of human observers. *J. Neurosci.* 42, 8629–8646
10. Ribeiro, F.L. *et al.* (2023) Variability of visual field maps in human early extrastriate cortex challenges the canonical model of organization of V2 and V3. *Elife* 12, e86439
11. Tian, Y. *et al.* (2020) Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* 23, 1421–1432
12. van Es, D.M. *et al.* (2019) Topographic Maps of visual space in the human cerebellum. *Curr. Biol.* 29, 1689–1694
13. Hanke, M. *et al.* (2014) A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci. Data* 1, 140003
14. Poldrack, R.A. *et al.* (2015) Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* 6, 8885
15. Gordon, E.M. *et al.* (2017) Precision functional mapping of individual human brains. *Neuron* 95, 791–807
16. Pinho, A.L. *et al.* (2018) Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci. Data* 5, 180105
17. Chang, N. *et al.* (2019) BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Sci. Data* 6, 49
18. Seeliger, K. *et al.* (2019) A large single-participant fMRI dataset for probing brain responses to naturalistic stimuli in space and time. *bioRxiv*, Published online July 2, 2019. https://doi.org/10.1101/687681
19. Boyle, J.A. *et al.* (2020) The Courtois project on neuronal modelling - 2020 data release. In *The 26th Annual Meeting of the Organization for Human Brain Mapping*, Organization for Human Brain Mapping, Poster 1939.

20. Siegel, J.S. *et al.* (2024) Psilocybin desynchronizes the human brain. *Nature* 632, 131–138
21. Yang, E. *et al.* (2023) The default network dominates neural responses to evolving movie stories. *Nat. Commun.* 14, 4197
22. Pinho, A.L. *et al.* (2021) Subject-specific segregation of functional territories based on deep phenotyping. *Hum. Brain Mapp.* 42, 841–870
23. Finn, E.S. *et al.* (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671
24. Eickhoff, S. *et al.* (2016) Sharing the wealth: neuroimaging data repositories. *NeuroImage* 124, 1065–1068
25. Allen, E.J. *et al.* (2022) A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* 25, 116–126
26. Van Essen, D.C. *et al.* (2013) The WU-Minn Human Connectome Project: an overview. *NeuroImage* 80, 62–79
27. Benson, N.C. *et al.* (2018) The Human Connectome Project 7 Tesla retinotopy dataset: description and population receptive field analysis. *J. Vis.* 18, 23
28. Nastase, S.A. *et al.* (2021) The 'Narratives' fMRI dataset for evaluating models of naturalistic language comprehension. *Sci. Data* 8, 250
29. Gong, Z. *et al.* (2023) A large-scale fMRI dataset for the visual processing of naturalistic scenes. *Sci. Data* 10, 559
30. Hebart, M.N. *et al.* (2023) THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *Elife* 12, e82580
31. LeBel, A. *et al.* (2023) A natural language fMRI dataset for voxelwise encoding models. *Sci. Data* 10, 555
32. Shim, W.M. *et al.* (2023) Exploring the broad cognitive landscape with the 7T Naturalistic Perception, Action & Cognition (NatPAC) dataset. In *ISMRM Workshop on Current Issues in Brain Function*, The International Society for Magnetic Resonance in Medicine, Talk 2, Session 3: Big/Deep Data + Artificial Intelligence in fMRI.
33. Lahner, B. *et al.* (2024) Modeling short visual events through the BOLD moments video fMRI dataset and metadata. *Nat. Commun.* 15, 6241
34. Turk-Browne, N.B. (2013) Functional interactions as big data in the human brain. *Science* 342, 580–584
35. Van Horn, J.D. and Toga, A.W. (2014) Human neuroimaging as a 'Big Data' science. *Brain Imaging Behav.* 8, 323–331
36. Sejnowski, T.J. *et al.* (2014) Putting big data to good use in neuroscience. *Nat. Neurosci.* 17, 1440–1441
37. Ascoli, G.A. *et al.* (2017) Win-win data sharing in neuroscience. *Nat. Methods* 14, 112–116
38. Xia, M. and He, Y. (2017) Functional connectomics from a 'big data' perspective. *NeuroImage* 160, 152–167
39. Fair, D. *et al.* (2021) Developmental cognitive neuroscience in the era of networks and big data: strengths, weaknesses, opportunities, and threats. *Annu. Rev. Dev. Psychol.* 3, 249–275
40. Horien, C. *et al.* (2021) A hitchhiker's guide to working with large, open-source neuroimaging datasets. *Nat. Hum. Behav.* 5, 185–193
41. Michon, K.J. *et al.* (2022) Person-specific and precision neuroimaging: current methods and future directions. *NeuroImage* 263, 119589
42. Tibon, R. *et al.* (2022) Bridging the big (data) gap: levels of control in small- and large-scale cognitive neuroscience research. *Trends Neurosci.* 45, 507–516
43. Baker, D.H. *et al.* (2021) Power contours: optimising sample size and precision in experimental psychology and human neuroscience. *Psychol. Methods* 26, 295–314
44. Rosenberg, M.D. and Finn, E.S. (2022) How to establish robust brain–behavior relationships without thousands of individuals. *Nat. Neurosci.* 25, 835–837
45. Naselaris, T. *et al.* (2021) Extensive sampling for complete models of individual brains. *Curr. Opin. Behav. Sci.* 40, 45–51
46. Kay, K. *et al.* (2023) Tasks and their role in visual neuroscience. *Neuron* 111, 1697–1713
47. Khosla, M. *et al.* (2022) A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Curr. Biol.* 32, 4159–4171
48. Jain, N. *et al.* (2023) Selectivity for food in human ventral visual cortex. *Commun. Biol.* 6, 175
49. Pennock, I.M.L. *et al.* (2023) Color-biased regions in the ventral visual pathway are food selective. *Curr. Biol.* 33, 134–146
50. Zou, F. *et al.* (2023) Re-expression of CA1 and entorhinal activity patterns preserves temporal context memory at long timescales. *Nat. Commun.* 14, 4350
51. Roth, Z.N. *et al.* (2022) Natural scene sampling reveals reliable coarse-scale orientation tuning in human V1. *Nat. Commun.* 13, 6469
52. Choe, A.S. *et al.* (2015) Reproducibility and temporal structure in weekly resting-state fMRI over a period of 3.5 years. *PLoS One* 10, e0140134
53. Laumann, T.O. *et al.* (2015) Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87, 657–670
54. Braga, R.M. and Buckner, R.L. (2017) Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* 95, 457–471
55. O'Connor, D. *et al.* (2017) The Healthy Brain Network Serial Scanning Initiative: a resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *Gigascience* 6, 1–14
56. Moeller, S. *et al.* (2021) NOise reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing. *NeuroImage* 226, 117539
57. Dowdle, L.T. *et al.* (2023) Evaluating increases in sensitivity from NORDIC for diverse fMRI acquisition strategies. *NeuroImage* 270, 119949
58. Stanley, O.W. *et al.* (2021) Effects of phase regression on high-resolution functional MRI of the primary visual cortex. *NeuroImage* 227, 117631
59. Van, A.N. *et al.* (2023) Framewise multi-echo distortion correction for superior functional MRI. *bioRxiv*, Published online November 29, 2023. https://doi.org/10.1101/2023.11.28.568744
60. Kay, K. *et al.* (2019) A critical assessment of data quality and venous effects in sub-millimeter fMRI. *NeuroImage* 189, 847–869
61. Weiskopf, N. *et al.* (2021) Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nat. Rev. Phys.* 3, 570–588
62. Kurzawski, J.W. *et al.* (2022) Non-neural factors influencing BOLD response magnitudes within individual subjects. *J. Neurosci.* 42, 7256–7266
63. Lusebrink, F. *et al.* (2021) Comprehensive ultrahigh resolution whole brain in vivo MRI dataset as a human phantom. *Sci. Data* 8, 138
64. Gulban, O.F. *et al.* (2022) Mesoscopic in vivo human T(2)(*) dataset acquired using quantitative MRI at 7 Tesla. *NeuroImage* 264, 119733
65. Thompson, P.M. *et al.* (2020) ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* 10, 100
66. Feinberg, D.A. *et al.* (2023) Next-generation MRI scanner designed for ultra-high-resolution human brain imaging at 7 Tesla. *Nat. Methods* 20, 2048–2057
67. Viessmann, O. and Polimeni, J.R. (2021) High-resolution fMRI at 7 Tesla: challenges, promises and recent developments for individual-focused fMRI studies. *Curr. Opin. Behav. Sci.* 40, 96–104
68. Demetriou, L. *et al.* (2018) A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *NeuroImage* 176, 404–416
69. Henrich, J. *et al.* (2010) Most people are not WEIRD. *Nature* 466, 29
70. Webb, E.K. *et al.* (2022) Addressing racial and phenotypic bias in human neuroscience methods. *Nat. Neurosci.* 25, 410–414
71. Carp, J. (2012) On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. *Front. Neurosci.* 6, 149
72. Bowring, A. *et al.* (2019) Exploring the impact of analysis software on task fMRI results. *Hum. Brain Mapp.* 40, 3362–3384
73. Luppi, A.I. *et al.* (2024) Systematic evaluation of fMRI data-processing pipelines for consistent functional connectomics. *Nat. Commun.* 15, 4745

74. Esteban, O. *et al.* (2017) MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* 12, e0184661
75. Esteban, O. *et al.* (2019) fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116
76. Gorgolewski, K. *et al.* (2016) The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3, 160044
77. Marcus, D.S. *et al.* (2007) The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 11–34
78. Halchenko, Y.O. *et al.* (2021) DataLad: distributed system for joint management of code, data, and their relationship. *J. Open Source Softw.* 6, 3262
79. Yuste, R. *et al.* (2017) Four ethical priorities for neurotechnologies and AI. *Nature* 551, 159–163
80. Ochang, P. *et al.* (2023) Towards an understanding of global brain data governance: ethical positions that underpin global brain data governance discourse. *Front. Big Data* 6, 1240660
81. Rocher, L. *et al.* (2019) Estimating the success of re-identifications in incomplete datasets using generative models. *Nat. Commun.* 10, 3069
82. Eke, D.O. *et al.* (2022) International data governance for neuroscience. *Neuron* 110, 600–612
83. Markiewicz, C.J. *et al.* (2021) The OpenNeuro resource for sharing of neuroscience data. *Elife* 10, e71774
84. Wallis, J.C. *et al.* (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One* 8, e67332
85. Yang, K. *et al.* (2020) Towards fairer datasets: filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. In *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Hildebrandt, M. *et al.*, eds), pp. 547–588, ACM
86. Oksuz, K. *et al.* (2021) Imbalance problems in object detection: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3388–3415
87. Ferguson, A.R. *et al.* (2014) Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nat. Neurosci.* 17, 1442–1447
88. He, T. *et al.* (2022) Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. *Nat. Neurosci.* 25, 795–804
89. Fedorenko, E. (2021) The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Curr. Opin. Behav. Sci.* 40, 105–112
90. Gardner, J.L. and Merriam, E.P. (2021) Population models, not analyses, of human neuroscience measurements. *Annu. Rev. Vis. Sci.* 7, 225–255
91. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
92. Krizhevsky, A. *et al.* (2012) ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Proces. Syst.* 25, 1097–1105
93. Yamins, D.L. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8619–8624
94. Güçlü, U. and van Gerven, M.A.J. (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35, 10005–10014
95. Cichy, R.M. *et al.* (2019) The Algonauts Project. *Nature. Mach. Intell.* 1, 613
96. Gifford, A.T. *et al.* (2023) The Algonauts Project 2023 challenge: how the human brain makes sense of natural scenes. *arXiv*, Published online January 9, 2023. http://dx.doi.org/10.48550/arXiv.2301.03198
97. Schrimpf, M. *et al.* (2018) Brain-Score: which artificial neural network for object recognition is most brain-like? *bioRxiv*, Published online September 5, 2018. https://doi.org/10.1101/407007
98. Zhou, M. *et al.* (2023) A large-scale fMRI dataset for human action recognition. *Sci. Data* 10, 415