Neuroscience Research 215 (2025) 3–14

Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.sciencedirect.com/journal/neuroscience-research

Review article

Large-scale foundation models and generative AI for BigData neuroscience

Ran Wang [a], Zhe Sage Chen [a,b,c,*]

[a] Department of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA
[b] Department of Neuroscience and Physiology, Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, USA
[c] Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA

ARTICLE INFO

Keywords:
Foundation model
Generative AI
BigData
Transformer
Self-supervised learning
Transfer learning
Representation learning
Embedding
Brain-machine interface

ABSTRACT

Recent advances in machine learning have led to revolutionary breakthroughs in computer games, image and natural language understanding, and scientific discovery. Foundation models and large-scale language models (LLMs) have recently achieved human-like intelligence thanks to BigData. With the help of self-supervised learning (SSL) and transfer learning, these models may potentially reshape the landscapes of neuroscience research and make a significant impact on the future. Here we present a mini-review on recent advances in foundation models and generative AI models as well as their applications in neuroscience, including natural language and speech, semantic memory, brain-machine interfaces (BMIs), and data augmentation. We argue that this paradigm-shift framework will open new avenues for many neuroscience research directions and discuss the accompanying challenges and opportunities.

1. Introduction

Advances in neurotechnology have allowed us to record large-scale, high-throughput neural data through in vivo electrophysiology and brain imaging. These sources of BigData present a challenge for various neural data analyses such as decoding and functional connectivity analysis, as well as closed-loop brain-machine interface (BMI) applications in neuroscience experiments (Chen and Pesaran, 2021. In parallel, machine learning research is also moving very fast. Several reviews on the interplay between AI and neuroscience research have been published (Hassabis et al., 2017; Richards et al., 2019; Saxe et al., 2021; Macpherson et al., 2021. Rapid advances in deep learning and development of large-scale foundation models and large language models (LLMs) have taken the whole world by storm, demonstrating remarkable and revolutionary findings in generating high-resolution synthetic images and yielding human-like natural language understanding (Zhao et al., 2023; Naveed et al., 2023; Singhal et al., 2023. The past few years have witnessed a paradigm shift in AI to foundation models in nearly every aspect of machine learning applications. *How will these technological changes impact neuroscience and what are the implications for the field?* Answering this question is part of our motivation to write this review. However, since the field is relatively new, the number of published studies on neuroscience applications based on foundation models or LLMs is relatively small. Nevertheless, the interest is rapidly growing and many findings derived from this line of research may have a potentially significant impact on neuroscience.

In this mini-review, we first provide a brief overview of foundation models and their building block—transformers, then extend our overview to a broad class of generative AI tools. Next, we review important concepts in representation learning, self-supervised learning (SSL) and transfer learning, which will play important roles in cross-modality applications. Further, we review recent applications of foundation models and generative AI in various neuroscience research areas, including but not limited to large-scale brain imaging data analysis, natural speech and language understanding, memory, emotion, mental state decoding, behavior, BMI, and data augmentation. Finally, we conclude the review with discussions and outlook on future research opportunities and challenges.

2. Foundation models and generative AI

2.1. What are foundation models?

Foundation models have become a new paradigm for building AI systems, in which models trained on a large amount of unlabeled data can be adapted to many other applications. The foundation models are often trained using self-supervision with BigData, and can be adapted to a wide range of tasks (e.g., text, images, speech, structured data, brain

* Corresponding author at: Department of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA.
E-mail address: zhe.chen@nyulangone.org (Z.S. Chen).

https://doi.org/10.1016/j.neures.2024.06.003
Received 9 June 2024; Accepted 9 June 2024
Available online 17 June 2024

signals, and high-dimensional tensor data) (Fig. 1). One of the popular class foundation models is LLMs (Table 1), which take language input and generate synthesized output. In general, foundation models work with multi-modal data types.

In a recent group study conducted at Stanford University, it was concluded that "*foundation models are scientifically interesting due to their impressive performance and capabilities, but what makes them critical to study is the fact that they are quickly being integrated into real-world deployments of AI systems with far-reaching consequences on people*" (Bommasani et al., 2021.

At the very high level, there are two fundamental ideas in the LLM and foundation models: (i) embedding, which aims to convert words or tokens into high-dimensional statistically meaningful numbers; (ii) SSL or contrastive learning.

2.1.1. Embedding

Embedding is a feature extraction technique that nonlinearly transforms the input signal to a representational vector, which enable users to easily index, search, compute, and visualize. In language processing applications, a word embedding is the projection of words onto a meaningful space in which words that "are nearby in meaning" appear nearby in the embedding. Take ChatGPT as an example, the dimensionality of the embedding space can be high-dimensional (hundreds to thousands depending on the specific layer). Therefore, the embedding vectors that contain a string of numbers are located in the coordinates of "linguistic feature space" (Wolfram, 2023). In deep neural networks, embedding layers enable us to learn the relationship between high-dimensional inputs and outputs more efficiently.

2.1.2. SSL learning

In real life, humans and animals can learn efficiently from observation or very few labeled examples, pointing to the limitation of BigData-based supervised learning. SSL is predictive learning in that it aims to predict missing parts of the input. In recent years, SSL techniques have achieved immense successes in natural language processing (NLP) and computer vision by enabling models to learn from BigData at unprecedented scales (Millet et al., 2023; Balestriero et al., 2023. Depending on the objective, SSL can be a generative, contrastive, or generative-contrastive (adversarial) form; a comprehensive survey of SSL is referred to elsewhere (Liu et al., 2023. Under the SSL framework, fine-tuning the pre-trained models with a small percentage of labeled data can achieve comparable results with the supervised training (Eldele et al., 2023. In NLP, pre-training methods like BERT (Bidirectional Encoder Representations from Transformers) have shown strong performance gains using SSL that masks individual words or subword units (Devlin et al., 2019. Recently, Joshi et al. (2020) proposed an extended version of BERT known as SpanBERT, which can mask contiguous random spans instead of random tokens and train the span boundary representations to better predict the entire content of the masked span; by so doing, SpanBERT consistently outperforms BERT, with the largest gains on span selection tasks.

Table 1
A selective list of foundation models and LLMs.

Model	Characterization	Developer
BERT	generative language model	Google
CLIP	language-image pre-training	Open AI
Codex	general-purpose programming model	Open AI
DALE-E, DALL-E2	text-to-image models	Open AI
GPT-3	causal sequence model for NLP	Open AI
GPT-4	multi-modal model	Open AI
SORA	text-to-video model	Open AI
PaLM, PaLM2	multi-lingual pathways language models	Google
LLaMA, LLaMA2	foundational language model, code generation model	Meta
SEER	self-supervised computer vision model	Meta
GATO	multi-modal, multi-task, multi-embodiment policy	DeepMind
DINOv2	foundational models for vision	Meta

2.2. Transformer model

A transformer model is a deep neural network that learns context and thus meaning by tracking relationships in sequential data. Specifically, transformers were developed to solve the problem of sequence transduction that transforms an input sequence to an output sequence, enabling end-to-end learning in machine translation, text generation and sentiment analysis (Vaswani et al., 2017. Transformers are the building blocks in many foundation models, such as BERT and GPT (Generative Pre-trained Transformer). Transformers are computationally efficient in simultaneous sequence processing since model training can be sped up through parallelization, a key feature missing in recurrent neural networks (RNNs) and long short-term memory (LSTM). This feature has also made the creation of LLMs feasible.

The transformer model has a seq2seq neural network architecture, consisting of encoding, decoding and self-attention modules (Fig. 2a). There are several concepts fundamental to computations in the transformer:

- word embeddings: computing vector representations of words.
- positional embeddings: encoding the position of each token in a sequence and adding the positional information to the word embeddings
- attention: understanding the context of a word by considering the words that come before or after it. In other words, if the meaning is a result of the relationship between words, then self-attention is a general way of learning the meaning underlying a sentence (Vaswani et al., 2017.
- self-attention: weighing the importance of different parts of the input sequence against each other.

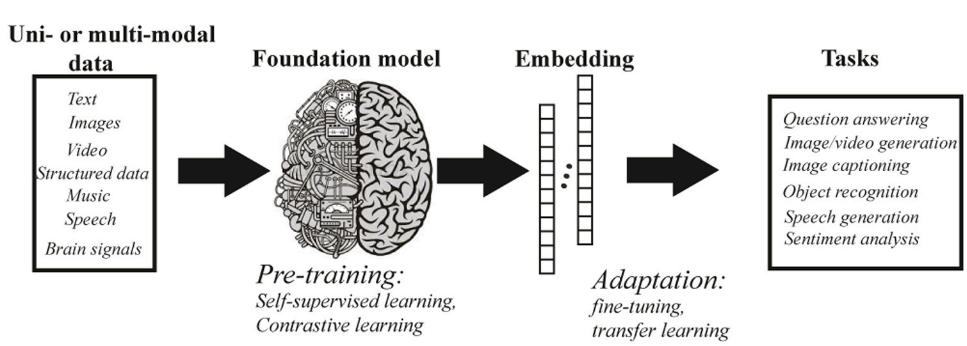

Fig. 1. A schematic diagram of foundation models. Unlike traditional machine learning models are designed for a specific task, foundation models are trained on a wide range of unlabeled data and can also perform other tasks.

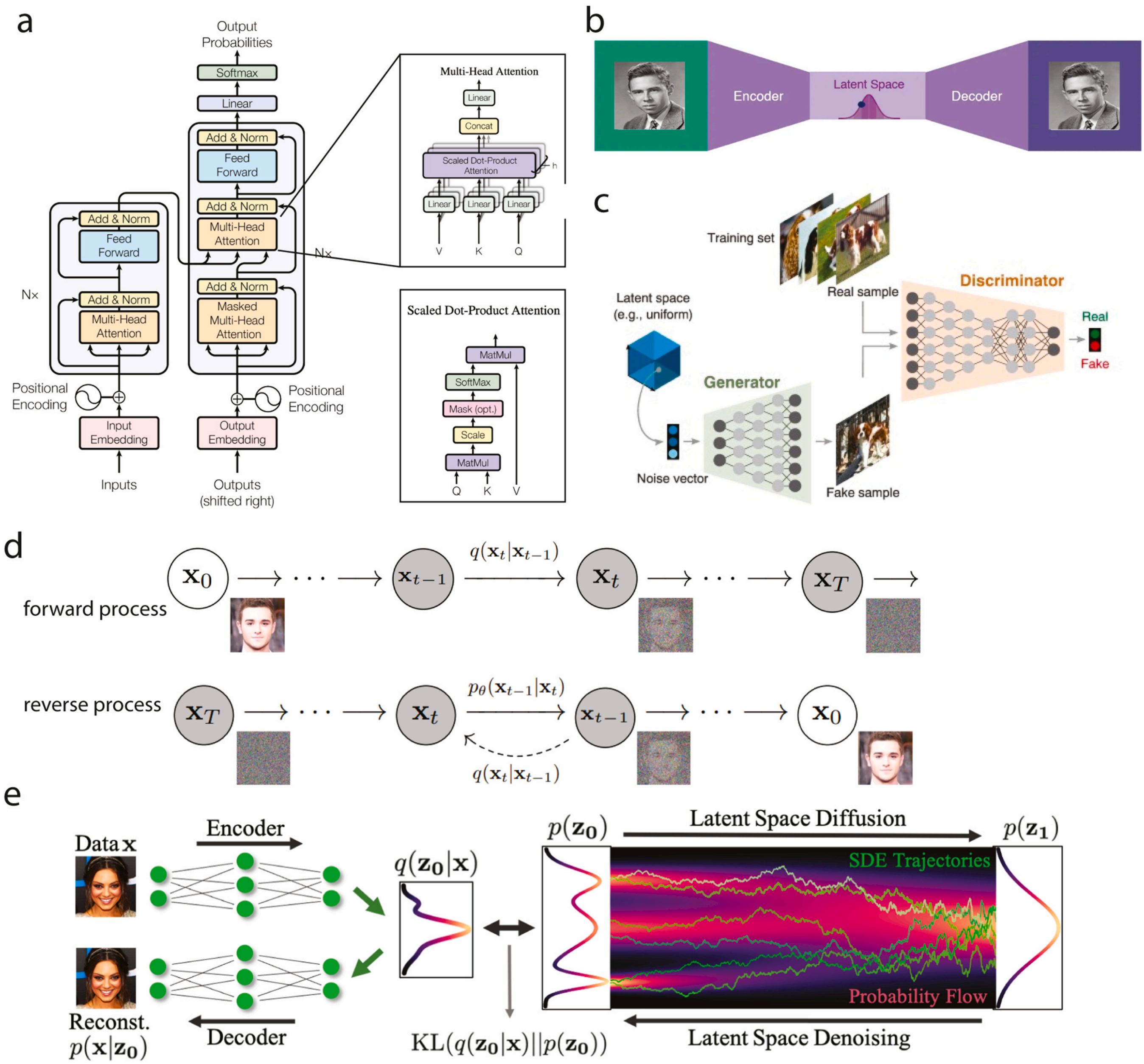

Fig. 2. Schematics of several generative AI models: (a) the Transformer architecture, (b) VAE, (c) GAN, (d) diffusion model. (e) LSGM. The multi-head attention in the decoder implements several masked, single-attention functions, where [Q, K, V] represent three embedding vectors for queries, keys, and values, respectively. To minimize the reconstruction error characterized by the distribution $p(\mathbf{x}|\mathbf{z}_0)$, the encoder and decoder are trained to minimize the KL divergence between a proposal distribution $q(\mathbf{z}_0|\mathbf{x})$ and the distribution $p(\mathbf{z}_0)$.
(a) Panel a is adapted with permission from (Vaswani et al., 2017, where the architecture follows an encoder-decoder structure without the use of convolution and recurrence. (b) Panel c is adapted with permission from (Goetschalckx et al., 2021, Elsevier, where the GAN was used to generate synthetic images by simultaneously training a discriminator and generator. (c) Panel e is adapted with permission from (Vahdat et al., 2021, where $\mathbf{z}_0$ and $\mathbf{z}_1$ represent two latent variables following a diffusion-based stochastic differential equation (SDE). (d) Panel d is adapted with permission from (Ho et al., 2020, in which the forward process is sampled from a Markov chain characterized by the transition distribution $q(\mathbf{x}_t|\mathbf{x}_{t-1})$, and the reverse process is sampled from a parameterized distribution $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$.

- multi-head attention: allowing the network to learn multiple ways of weighing the input sequence against itself.

In addition to NLP applications, the transformer architecture has been applied in other domains such as computer vision (Dosovitskiy et al., 2020, visual stimulus classification (Bagchi and Bathula, 2022, neural data analysis (Ye and Pandarinath, 2021, and reinforcement learning (RL) (Li et al., 2023a.

2.3. Generative AI

Generative AI describes a class of algorithms that can be used to create new content, including audio, code, images, text, simulations, and videos. Several representative generative AI algorithms are summarized below.

- Variational Autoencoder (VAE): VAE is a generative AI algorithm that uses deep learning to generate new content, detect anomalies and remove noise (Kingma and Welling, 2013. VAE consists of an encoder and a decoder, separated by the latent space (Fig. 2b). The latent space contains an abstract representation of the data containing only the most meaningful information (i.e., dimensionality reduction). The model can learn the data distribution, so that a corresponding output can be reconstructed based on a new sample input.

- Generative Adversarial Network (GAN): A GAN is a class of deep learning framework that uses two neural networks, the generator and the discriminator (Fig. 2c), to generate new and realistic synthetic data that are similar to the samples among the training set. Specifically, the generator network takes random noise as input and generates synthetic data, and aims to produce data that are indistinguishable from the real data in the training set. The generator tries to create realistic samples and follow the patterns present in the original dataset. On the other hand, the discriminator network evaluates the data it receives and tries to distinguish between real data from the training set and the synthetic data produced by the generator. Its goal is to correctly classify whether the input data is real or generated by the generator. The discriminator provides feedback to the generator, helping it improve its generated samples. To date, the GAN and many of its variants have numerous applications in image generation, image-to-image translation, super-resolution imaging, text-to-image synthesis, and video generation (Goodfellow et al., 2014; Gui et al., 2020; Goetschalckx et al., 2021.
- Generative Pre-trained Transformer (GPT): GPT is specifically referred to as a series of language models that use the transformer architecture to understand and generate coherent and contextually relevant text. Because of powerful predictive ability, GPT is effective for a variety of NLP tasks, including text generation, translation, and summarization. The basic idea behind GPT is to apply SSL and train the large-scale model based on big datasets containing a diverse range of text from various sources. Upon the completion of learning, the model takes the sequence of tokens that corresponds to the text in the past and finds an embedding that represents them, and further generates a large number of values that turn into probabilities for predicting possible next tokens (Wolfram, 2023. The newer GPT developments, such as GPT-3 (Brown et al.), GPT-4, and SORA, represent a landmark in this line of technology due to their impressive generative power and being trained on increasingly complex large-scale models.
- Diffusion Model: Diffusion models refer to a class of latent generative models that are used to model the distribution of data based on Markov chains and variational inference (Fig. 2d) (Ho et al., 2020; Rombach et al., 2022. These models are designed to capture the underlying data distribution by iteratively transforming a simple distribution into a complex one. Diffusion models offer a promising avenue for deep generative modeling owing to robust expressive capacity, and ability to generate data via ancestral sampling without the prerequisite of a posterior distribution. Unlike other deep generative models such as VAE and GAN, training diffusion models is relatively simple. To date, diffusion models have been used in image generation, NLP, and time series analysis.
- Latent Score-based Generative Model (LSGM): The LSGM generalizes the ideas of VAE and the diffusion model, maps the input onto a latent space and applies the diffusion model in the latent embeddings of the data (Fig. 2e) (Vahdat et al., 2021. As an extension of score-based generative models (Song and Ermon, 2019; Song et al., 2021, the LSGM has several key computational advantages: synthesis speed, expressivity, and tailored encoders and decoders.

Foundation models can serve as a basis for generative AI. BERT and GPT models have already been used as the building blocks for developing more sophisticated generative AI models. For instance, Fei et al. (2022) developed a self-supervised pre-trained foundation model on vision-language multi-modal input, which only requires weak semantic correlated image-text training pairs; specifically, they demonstrated that the foundation model not only can generate high-level concepts and describe complicated scenes, but also can synthesize new images or samples, which represents a critical step towards artificial general intelligence (AGI).

Furthermore, foundation models may provide a starting point for developing more advanced generative AI systems. Researchers and developers often fine-tune or extend the foundation models to create specialized generative models tailored to specific tasks or domains. More importantly, foundation models may facilitate transfer learning, which is vital for generative AI as it allows models to leverage the knowledge and representations learned by foundation models to generate diverse and contextually appropriate content across different domains. One exciting application of generative AI is to decode brain signals and transform them into text or images, which may have a clinical impact on the lives of individuals with traumatic brain injury (TBI) or severe paralysis who cannot communicate through speech, typing, or gestures (Metzger et al., 2023,2022; Défossez et al., 2023; VanRullen and Reddy, 2019; Tang et al., 2023. Recently, GAN-based (Dado et al., 2022 and diffusion model-based (Takagi and Nishimoto, 2023 approaches have been developed to reconstruct human faces or visual images from fMRI recordings. See (Gong et al., 2023 for a short review on generative AI for brain imaging applications.

3. Representation learning and transfer learning

3.1. Representation learning

Representation Learning refers to a class of machine learning algorithms that extract meaningful patterns from raw data to create representations easily understood or processed (Bengio et al., 2014. During this process, dimensionality reduction, regularization, invariance, and sparsity play important roles. Current LLMs heavily rely on effective representation learning algorithms. Representation learning can be achieved by unsupervised, supervised, and self-supervised frameworks. For instance, as a special case of SSL paradigm, contrastive learning can learn an embedding space such that similar instances have close representations while dissimilar instances stay far apart from each other. In addition to computer vision and NLP tasks, contrastive learning has been used to extract meaningful representations from neural data, including data from electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and other neuroscience modalities (Kostas et al., 2021. For instance, contrastive learning has enabled researchers to uncover patterns in brain connectivity data, providing insight into the organization and communication between different brain regions, and identifying connectivity-based biomarkers between healthy and pathological brains (Tong et al., 2021. Contrastive learning can also learn representations in the latent feature space based on dimensionality reduction. One such an example is contrastive PCA (cPCA), which can identify the dominant subspace that distinguishes two datasets collected from different conditions (Abid et al., 2018. Additionally, contrastive variational autoencoder (cVAE) (Aglinskas et al., 2022, as an extension of cPCA, offers a more flexible approach capable of modeling nonlinear relationships between the inputs and latent features. Finally, another contrastive learning paradigm, contrastive predictive coding (CPC) (van den Oord et al., 2019, learns self-supervised representations by predicting the future in latent space by using autoregressive models and VAE; the model uses a probabilistic contrastive loss, which induces the latent space to capture information that is maximally useful to predict future data.

3.2. Transfer learning

Transfer learning represents a class of machine learning technique where knowledge learned from a task is reused in order to boost performance on a related task or generalize out-of-distribution via targeted re-training (Pan and Yang, 2010. In deep learning models, transfer learning has been widely used in computer vision, image classification, and NLP tasks (Yosinski et al., 2014; Goodfellow et al., 2016.

Transfer learning has found many applications in neuroscience. In neuroimaging data analysis, pre-trained models from NLP or computer vision domains, can be fine-tuned or used to extract features from raw neural data, facilitating out-of-domain tasks such as classification, segmentation, and decoding of neural activity. For instance, pre-trained models from related medical imaging tasks can be adapted to process and interpret neuroimaging data, leading to a more accurate and efficient analysis. Additionally, since the relationship between cognitive tasks is usually represented by similarity of neural representations or activated brain regions, transfer learning may perform better in task decoding with fMRI data if the source and the target cognitive tasks activate similar brain regions (Qu et al., 2022).

In BMI research, transfer learning can improve the performance and adaptability of BMI systems by leveraging knowledge from related tasks. Pre-trained models may help enhance the decoding of neural signals for controlling external devices or for interpreting brain activity associated with specific motor or cognitive tasks.

Transfer learning can assist in the early detection and diagnosis of neurological or psychiatric disorders by leveraging knowledge from related medical domains. Pre-trained models from medical imaging or clinical data analysis can be adapted to identify biomarkers associated with specific pathological conditions, aiding in early intervention and personalized treatment strategies. Notably, transfer learning can work well where the data sample size is small in neuroimaging-based prediction (Ardalan and Subbian, 2022; Malik and Bzdok, 2022 and ECoG/EEG decoding analysis (Zhang et al., 2021; Peterson et al., 2021.

4. Foundation models and generative AI for neuroscience applications

4.1. Context-dependent embedding mapping

As discussed earlier, representation learning can identify context-dependent embeddings for a broad class of input signals. For instance, if the input is a speech signal, the embedding mapping for speech representation may be produced by "wave2vec" (Schneider et al., 2019; Baevski et al., 2020; Millet et al., 2023, HuBERT (Hsu et al., 2021, and "data2vec" (Baevski et al., 2022,2023. If the input is a neural time series such as an EEG signal, the embedding mapping for EEG may include "EEG2vec" (Zhu et al., 2023 or other representation learning methods (Kostas et al., 2021; Rafiei et al., 2022; Wagh et al., 2021. Such methods have been demonstrated in neuroscience applications such as automatic sleep staging (Banville et al., 2020; Yang et al., 2023 and seizure detection (Tang et al., 2022.

In neural data analysis, embeddings have been widely adopted in unsupervised or supervised representation learning. For instance, automated neuron reconstruction and annotation of volume electron microscopy (VEM) datasets of three-dimensional images of brain tissue are computationally intensive and challenging. Schubert et al. (2019) first used unsupervised training to infer morphology embeddings ("neuron2vec") of neuron reconstructions, and then trained cellular morphology neural networks (CMNs) to identify glia cells via supervised classification; they also demonstrated how CMNs can be used to identify subcellular compartments and the cell types of neuron reconstructions.

Embeddings are useful for revealing low-dimensional neural dynamics and modeling naturalistic behaviors (Wang and Guet, 2022; Schneider et al., 2023. Although traditional latent variable models have been used for analyzing neural and behavioral data (Chen, 2015; Latimer et al., 2015; Calhoun et al., 2019; Bolkan et al., 2022; Ashwood et al., 2022; Lakshminarasimhan et al., 2023, most of them are limited in encoding the context dependence. Incorporating task-relevant embedding vectors to form a context-relevant embedding would allow us to perform end-to-end learning efficiently. Recently, Ye and Pandarinath (2021) have proposed a non-recurrent, BERT encoder-based neural data transformer (NDT) model to explicitly model autonomous neural population activity and reported comparable performance between the NDT model and other RNN models. In their NDT model, inputs to transformer layers were first normalized and enriched through contextual information ("self-attention" blocks), and passed through a feed-forward module.

4.2. Brain imaging

Human neuroimaging provides a window to examine healthy and diseased brains, in terms of both structural and functional forms, including EEG, MEG, fMRI, diffusion tensor imaging (DTI), and positron emission tomography (PET). See (Gong et al., 2023 for a review of generative AI for brain imaging, covering co-registration, super--resolution, enhancement, classification, segmentation, cross-modality, brain network analysis, and decoding analysis.

Several lines of work have proposed generative AI approaches to reconstruct visual images based on fMRI data (Seeliger et al., 2018; VanRullen and Reddy, 2019; Ferrante et al., 2023a. For instance, VanRullen and Reddy (2019) first trained a VAE network using a GAN unsupervised procedure over a large dataset of celebrity faces, where the VAE latent space provided a topologically organized 1024-dimensional embedding of each image. Next, they presented thousands of face images to human subjects and learned a linear mapping between multi-voxel fMRI activation patterns and latent embeddings. Finally, they applied this mapping to novel face images, translating fMRI patterns into reconstructed faces.

Lu et al. (2022) developed a self-supervised pre-trained image-text multi-modal foundation model which outperformed CLIP (Contrastive Language-Image Pre-Training) model even with a small percentage ($\sim$ 3.75 %) of training pairs. The image and text were first encoded individually by pre-trained uni-modal large-scale models, vision transformer (ViT) and BERT. The output of BERT was then projected to a trained mapping layer that aligns with ViT features. By comparing the encoded image encoding feature with fMRI imaging of the human visual cortex, their results showed that the proposed multi-modal model has higher prediction accuracy than the uni-modal image encoder.

4.3. Natural language and speech

Speech and language understanding involves a deep comprehension of their generation and processing (in both sound and text), enabling computers to perform tasks such as speech recognition, language translation, sentiment analysis, and text summarization.

Representing human speech from brain signals (such as ECoG and fMRI) may include decoding neural activity associated with speech production, perception, or comprehension. It has been known that natural speech reveals a semantic map that tiles the human cerebral cortex (Huth et al., 2016a, and the semantic space is continuously distributed across the brain describing representations of thousands of object and action categories (Huth et al., 2012,2016b.

On the one hand, the rich features extracted from the foundation models provide a new hypothesis when studying brain representations during specific speech and language tasks. For example, the ECoG activity in the superior temporal gyrus (STG) and inferior frontal gyrus (IFG) of the human brain was found to be correlated with features extracted by the GPT model (Goldstein et al., 2022). Since predictive pre-training of the GPT model was capable of encoding contextual information, word onset, and word surprisal, this finding suggests that the human auditory cortex may encode speech in a similar manner. The contextual encoding phenomenon was also found when correlating neural representations in the human auditory cortex with the HuBERT model's embeddings (Li et al., 2023b).

On the other hand, a growing number of studies have focused on decoding human speech from invasive brain recordings, using either intracranial ECoG or intracortical spiking activity (Metzger et al., 2022; Moses et al., 2021; Willett et al., 2023) (see the review of BMI applications below). Recently, Défossez et al. (2023) have developed a contrastive learning approach to decode speech based on non-invasive magneto- or electro-encephalography (MEG/EEG). They first employed a large-scale pre-trained speech encoding model ("wave2vec 2.0" (Baevski et al., 2020) to extract semantic features from speech, and then trained a decoding model to extract features that converged to the speech features of the corresponding trail while diverging from speech features of other trails. The model was capable of identifying the speech segment with features that best matched the decoded neural features. This work represents a large step forward in clinical practice without putting patients at the risk of brain surgery.

Furthermore, EEG signals can be leveraged to augment multi-modal NLP models while using less training data (Hollenstein et al., 2021; in combination with EEG data, BERT embeddings have shown consistently improved performance for NLP tasks.

4.4. Memory and semantic reconstruction

In the traditional episodic memory paradigm, subjects are usually required to memorize arbitrary items (words or images), lacking the fundamental components in real-life naturalistic events occurring over a longer timescale. Multimedia stimuli such as music and film, however, may provide rich contextual and naturalistic memory behaviors (Groussard et al., 2009).

In neuroscience experiments, recollection of short audiovisual segments from movies can be viewed as a proxy to real-life memory that consists of a stream of continuous sensory experiences. In contrast to pure reconstruction of static images from brain imaging (Shen et al., 2019; Horikawa et al., 2013), reconstructing high-quality images with correct semantics from brain recordings is more challenging due to the complex underlying representations of brain signals and the scarcity of data annotations. In the literature, neural decoders have been developed for semantic reconstruction of movie or visual experiences (Huth et al., 2016b; Nishimoto et al., 2011). Extension of this framework using generative AI would represent a promising research direction.

Recently, Chen et al. (2023) proposed a conditional diffusion model with sparse masked modeling for human visual decoding. Inspired by sparse coding in the primary visual cortex, they first applied SSL and mask modeling in a large latent space for fMRI data; then they augmented a latent diffusion model (LDM) to reconstruct highly plausible images with semantically matching details from fMRI recordings using very few paired annotations.

4.5. Mental state and emotion

Decoding brain states and mental processes based on brain imaging data has been an active research area (Poldrack et al., 2012; Rubin et al., 2017). However, the common challenge is that sample sizes are relatively small and models are prone to overfitting. Recently, to decode mental states, Thomas et al. (2022) proposed to leverage publicly shared fMRI data (https://openneuro.org/) to pretrain a foundation model. Their procedure consisted of two steps. In the first step, they performed self-supervised learning on fMRI time series using various model strategies: seq-to-seq autoencoder, casual sequence modeling (similar to GPT-3), sequence-BERT, and network-BERT. In the second step, they applied a plug-in and adaptation to decoding mental states. In so doing, the mental states can be viewed as a high-dimensional neural embedding, and NLP-inspired architectures were able to learn useful representations of fMRI time series; more importantly, the pre-trained model also improved the decoding accuracy of mental states (compared to several baseline models).

Decoding emotions from brain activity is one fundamental task in human-computer interaction, yet most decoding methods are limited by the number of emotion categories or have ignored the discrepancy of emotion expression between two brain hemispheres. Recently, Fu et al. (2022) proposed a multi-view multi-label hybrid model for fine-grained emotion decoding: the generative component is a multi-view VAE that learns the brain activity of left and right hemispheres, as well as their differences, and the discriminative component is a multi-label classification network. Furthermore, they used a label-aware module for emotion-specific neural representation learning and modeled the dependency of emotional states by a masked self-attention mechanisms.

4.6. Naturalistic behavior

An important goal in neuroscience is to uncover the circuit mechanisms underlying cognitive processes and behavior, for which quantitative behavioral descriptions may play a vital role in linking brain activity and behavior (Krakauer et al., 2017; Pereira et al., 2020). Unlike constrained behaviors (such as head-fixed tasks or planar reach-and-grasp movement), naturalistic behavior refers to the behavior that animals have a tendency to exhibit under natural or realistic conditions, which is often beneficial to biological functioning.

Given the success of sequence modeling in NLP, it is tempting to frame behavior analysis as a sequence modeling problem and apply this idea to context-relevant behavioral embedding and attention computation. Recently, Reed et al. (2022) have proposed a generalist agent (GATO) model for multi-modal, multi-task learning. Specifically, they encoded various modalities into a single vector space of "tokens" that can be ingested by a large sequence model such as transformers; they also proposed various "tokenization" approaches to capture the large amount of multi-modal data that include standard vision and language datasets and some RL benchmarks.

4.7. Brain-machine interfaces

A BMI is a system that establishes a direct communication pathway between the brain's electrical activity and an external device, reading out the encoded stimuli (e.g., speech, vision, location) or translating thought into action (i.e., neuroprosthetics) (Gilja et al., 2012; Lebedev and Nicolelis, 2017; Willett et al., 2021). Such mind-reading devices can be used not only for clinical applications (Shanechi, 2019; Moses et al., 2021; Zhang et al., 2023; Sun et al., 2022, but also for scientific inquiry in basic science questions (Sadtler et al., 2014.

Data sources in different BMIs have a varying degree of signal-to-noise ratio (SNR). For instance, while sharing the same temporal resolution, ECoG has a higher SNR than the scalp EEG. On the other hand, calcium imaging or fMRI data have a much lower temporal resolution than ECoG or EEG. Because of this variability, directly mapping neural signals onto decoding targets (e.g., text, speech, and music) is not optimal. Pre-trained foundation models can mitigate this by incorporating prior knowledge about the decoding targets, aligning them more closely with the neural signals.

To date, LLMs have been incorporated into BMI systems to enhance text decoding. A wide range of machine learning techniques have been employed to increase the efficiency and accuracy of EEG-based spelling systems (Speier et al., 2016. In practice, these language models can either auto-complete decoded words or be integrated into classifiers to refine the probability estimates of potential letters based on previously decoded ones. Leveraging language models has proven to significantly reduce word-error-rates, especially when decoding text from intracranial ECoG or Utah array during speech attempts (Moses et al., 2021; Metzger et al., 2022,2023; Willett et al., 2021,2023. A notable recent study (Tang et al., 2023 utilized a pre-trained GPT-2 model to interpret perceived speech from fMRI scans, converting neural patterns into text. This research, which involved over 16 h of fMRI data from participants listening to stories, has showcased the potential of BMI in decoding imagined speech and even in cross-modal decoding, such as interpreting text representations of mental states during silent film viewing.

Foundation models have also been instrumental in enhancing the performance of BMI systems, especially in decoding audio and visual signals (Metzger et al., 2023; Anumanchipalli et al., 2019; Wang et al., 2020,2023; Takagi and Nishimoto, 2023; Denk et al., 2023; Bellier et al., 2023; Benchetrit et al., 2023). For instance, Metzger et al. (2023) utilized a pre-trained speech generative model to decode clear speech from neural signals. Specifically, they used a sophisticated transformer-based speech encoding model ("HuBERT") to learn a compact representation of speech, which was then transformed into high-quality speech using a pre-trained synthesizer. Beyond speech, music decoding has also seen progresses with the aid of generative AI. Multiple lines of recent research (Denk et al., 2023; Bellier et al., 2023) have demonstrated the feasibility of decoding music from neural signals using deep learning, with pre-trained models such as musicLM (Agostinelli et al., 2023, to produce high-quality outputs. Similarly, image reconstruction from fMRI scans has achieved remarkable accuracy with the help of image generative models such as the VAE, GAN, and diffusion models (Takagi and Nishimoto, 2023; Ferrante et al., 2023b; VanRullen and Reddy, 2019; Huang et al., 2021; Ozcelik and VanRullen, 2023). In these studies, neural signals were first converted into latent representations, and then used to produce images through various generative models (Table 2). For instance, a two-stage scene reconstruction framework called "Brain--Diffuser" has been proposed: in the first stage, a low-level image is first reconstructed via a very deep VAE, and in the second stage, a latent diffusion model conditioned on predicted multi-modal (text and visual) features is used to reconstruct high-quality images (Ozcelik and VanRullen, 2023).

Remarkably, Benchetrit et al. (2023) developed an real-time visual decoding strategy from MEG recordings using a foundation model. The model consists of three modules: (i) pre-trained embedding obtained from images, (ii) an MEG module trained end-to-end, and (iii) a pre-trained image generator. The brain-to-image readout was decoded with a foundational image model known as DINOv2. The authors reported that MEG-based decoding can recover high-level visual features compared to fMRI-based decoding, offering a real-time BMI paradigm ($\sim$ 250 ms delay) for the human brain.

To date, most of brain decoding applications have been reported in human research since data format and acquisition are relatively universal, which may not be the case in animal studies. Recently, built upon a foundation model known as Perceiver IO (Jaegle et al., 2022, Azabou et al. (2023) developed a new framework called POYO (Pre-training On manY NeurOns) for large-scale training transformer models end-to-end on multi-session and across-individual electrophysiology datasets. POYO introduces innovative spike-based tokenization strategies and uses pre-trained models (with possible fine tuning) for neural population decoding. Using a transformer architecture, POYO applies both cross-attention and self-attention in the latent space after latent embeddings of neural events. Their work demonstrates the power of transfer learning and transformer to achieve rapid and scalable neural decoding.

Table 2
A representative list of recent BMI and neural decoding studies based on generative AI.

Study	Data	Model	Application
(Anumanchipalli et al., 2019	ECoG	bidirectional LSTM	brain2speech
(Wang et al., 2020	ECoG	GAN, transfer learning	brain2speech
(Wang et al., 2023	ECoG	ResNet	brain2speech
(Willett et al., 2021	ECoG	RNN, language model	brain2text
(Willett et al., 2023	microelectrode arrays	RNN	brain2speech2text
(Metzger et al., 2022	ECoG	neural network	brain2text
(Metzger et al., 2023	ECoG	HuBERT, bidirectional RNN	brain2speech
(Liu et al., 2023	ECoG	sequential CNN-LSTM	brain2speech
(Tang et al., 2023	fMRI	GPT-2	brain2text
(Takagi and Nishimoto, 2023	fMRI	diffusion model	brain2image
(VanRullen and Reddy, 2019	fMRI	VAE, GAN	brain2face
(Ferrante et al., 2023a	fMRI	CNN	brain2feature
(Ferrante et al., 2023b	fMRI	generative image-to-text transformer	brain2image&text
(Huang et al., 2021	fMRI	deep VAE, LSTM	brain2image
(Ozcelik and VanRullen, 2023	fMRI	very deep VAE, diffusion model	brain2image
(Défossez et al., 2023	MEG, EEG	Contrastive Language-Image Pre-Training	brain2speech
(Benchetrit et al., 2023	MEG	DINOv2	brain2image
(Bellier et al., 2023	ECoG	feedforward neural network	brain2music
(Denk et al., 2023	fMRI	MusicLM	brain2music
(Azabou et al., 2023	spikes	PerceiverIO	brain2behavior

4.8. Data augmentation

Machine learning-driven data augmentation techniques are beneficial to alleviate the sample imbalance or insufficiency problem (Chawla et al., 2002; He and Garcia, 2009. This is particularly important for improving the generalization ability of deep learning. Recently, data-centric deep learning or generative AI strategies (e.g., data regeneration and synthetic data generation) have been proposed to improve the consistency between the existing and augmented data, especially in clinical applications where labeled samples may be scarce or the data privacy is a concern (Zhang et al., 2022. For instance, combining RNN and GAN may help construct generative models of synthetic time series and impute missing data (Yoon et al., 2019; Lee et al., 2021; Habashi et al., 2023. In one example, combined GAN and VAE models utilized three-dimensional convolution to model high-dimensional fMRI sensors with structured spatial correlations and the synthesized datasets were then used to augment classifiers designed to predict cognitive and behavioral outcomes (Zhuang et al., 2019. In another example, an auxiliary classifier GAN (AC-GAN) was used to generate synthetic interictal epileptiform discharges (IED) from EEG recordings of epileptic patients (Geng et al., 2021; Geng and Chen, 2021.

Bird et al. (2021) employed an LLM (based on GPT-2) to augment the EEG/MEG dataset for a classification task. After initial training, the GPT model was used to generate realistic synthetic neural signals as the augmented data; a marginal improvement was reported in classification performance.

Soingern et al. (2023) investigated a diffusion model-based EEG data augmentation strategy known as "WaveGrad" and compared it with other commonly used data augmentation techniques; their results showed a consistent improvement and better efficacy in the EEG motor imagery classification task.

Recently, a text data augmentation approach based on ChatGPT (named AugGPT) (Dai et al., 2023 has been developed to overcome the challenge of limited sample sizes in NLP tasks (Pellicer et al., 2023. Specifically, sentences in the training set were rephrased into conceptually similar variations as the augmented data with the same label of the original sample. The results showed that data augmentation based on such a large-scale pre-trained model increased the classification accuracy by a big margin in comparison with standard data augmentation methods. However, more research is still needed to see whether similar techniques can apply to neural data augmentation.

5. Discussion and conclusion

5.1. Crosstalk between AI and neuroscience

AI and neuroscience have been driving each other forward. Not only has neuroscience inspired the development of deep learning and AI technologies (Hassabis et al., 2017, but explainable AI and deep learning have also generated opportunities for in-depth neuroscience investigations (Richards et al., 2019; Saxe et al., 2021. For instance, biologically constrained CNN models have enabled neuroscientists to directly compare data in the visual cortex and uncover the underlying computational principles (Yamins and Hong, 2014; Yamins and DiCarlo, 2016; Shi et al., 2022. Recently, Schneider et al. (2023) proposed a contrastive learning-based neural network model for jointly modeling neural and behavioral dynamics. The SSL algorithm, known as CERBA, which combines ideas from nonlinear independent component analysis (ICA) with contrastive learning, may identify interpretable and consistent neural embeddings of high-dimensional neural recordings using auxiliary variables (such as time or behavioral measures). Importantly, it can generate embeddings across multiple subjects and cope with distribution shifts among experimental sessions, subjects, and recording modalities. In another example, Caucheteux et al. (2023) applied deep language algorithms (based on GPT-2) to predict nearby words and discovered that the activations of language models linearly map onto the brain responses to speech, and these predictions are organized hierarchically in frontoparietal and temporal cortices. These findings illustrate that the synergy between neuroscience and AI can largely improve our understanding of human cognition.

It is also worth mentioning that current AI technologies have relied on oversimplified models of neural systems. First and foremost, the standard artificial neurons in deep neural networks are “point neurons” that focus on somatic computation, yet the importance of nonlinear dendritic computation has been ignored. However, it has been known that the dendrite also plays an important role in neuronal computations and biological learning, such as enhancing expressivity of single neurons, improving neuronal resources and generalization abilities, utilizing internal learning signals, and enabling continual learning, contextual representation, and predictive coding (Acharya et al., 2022; Hodassman et al., 2022; Hawkins and Ahmad, 2016. Deep learning models have the potential to reproduce the computational complexity of biologically realistic neurons’ I/O properties (Beniaguev et al., 2021. Second, brain oscillations are important hallmarks in representing neural dynamics for a wide range of tasks in cognition, attention, memory, decision-making, and sensorimotor integration. Future development of next-generation neuroAI models and biologically plausible learning algorithms remains a central research direction to transform a “black-box” to “glass box” model while achieving a good trade-off between performance and interpretability.

5.2. Outlook and outstanding questions

Looking ahead, foundation models and generative AI will likely see a rapid research growth in method development and applications, especially in brain imaging and large-scale neural and behavioral data analyses (Moor et al., 2023. In clinical applications, foundation models and generative AI may have a great impact on personalized medicine. A growing number of ChatBots, such as ChatGPT and Bard, can play an active role in mitigating the worldwide crisis in mental health (Chen et al., 2022. In multi-modal BMI systems, generative AI will help combine speech, vision and motor modalities to improve the functionality and decoding accuracy. Future developments of brain-to-content neurotechnologies may have promising applications in immersive virtual reality, video games, marketing, and personalizied education.

In computer vision and NLP, BigData-empowered machine learning technologies have improved in performance steadily over the years, assessed by quantitative benchmarks (Deng et al., 2009; Russakovsky et al., 2015. However, there is a major gap in basic and clinical neuroscience applications because of the lack of high-quality training samples and benchmarks. However, this issue has been noticed in some research communities (e.g., AI for Epilepsy, (Chung et al., 2022), and the status quo may change soon. Several large-scale public datasets (Table 3) have become increasingly popular in open-source research.

Finally, we present several outstanding questions that might motivate future research in the intersection of AI and neuroscience.

- Since the majority of foundation models have been trained on single-modal data, it is unclear whether the model would benefit from training based on multi-modal or cross-modal data when the decoding domain is on single modality. For instance, in simultaneous EEG-fMRI recordings, can we train a foundation model based on their joint measurements, and then apply the pre-trained model in EEG-alone or fMRI-alone decoding analysis? While the prior knowledge of the cross-modal relationship may be beneficial, the variability in SNR and spatiotemporal resolution between the two modalities may create practical barriers. Furthermore, it remains an open question how we should apply SSL to identify an optimal analysis pipeline for multi-modal neuroimaging data.
- Representation learning and foundation models have great potential in RL, including end-to-end policy learning (Bahl et al., 2020 and multi-agent communications (Foerster et al., 2016. However, it remains unclear how well the foundation models and learned embedding representations can generalize across tasks in RL. For instance, RL algorithms have been developed in BMI applications, enabling individuals with motor disabilities to control external devices using neural signals. It still needs to be thoroughly tested whether the pre-trained policy can generalize across subjects, tasks, and environments. Identifying common as well as individualized decision-making or control policies under the new representation learning paradigm will continue to be an active research topic.
- While ChatGPT can be used as an interface between users and external systems, serving as a bridge between individuals with limited mobility and the external world, it is vital to revolutionize the communication capability of BMIs by translating thoughts into text-based information and refining the dynamics of human-machine interaction. However, it remains unclear how ChatGPT or GPT-like models can be optimally integrated into the BMI systems. Furthermore, can we adapt these models or generative AI to interpret and produce text that syncs flawlessly with a user’s intentions while abiding by ethical and privacy mandates? The recurrent engagement of users with ChatGPT offers prospects to transform the lives of those with disabilities and to develop personalized and adaptable BMI systems, escalating user gratification and optimizing system outputs.
- Ongoing research has continued producing new frontiers in foundation models and generative AI, such as the new autonomous AI agent tools (AutoGPT, MetaGPT and AutoGen) (see a compiled list at https://github.com/steven2358/awesome-generative-ai).

Table 3
Public datasets for open-source neuroscience research.

Creator	Data modality	Website
Allen Brain Observatory	electrophysiology, calcium imaging	https://observatory.brain-map.org/
CRCNS	electrophysiology, neuroimaging	https://crcns.org/data-sets
MIT Lab for Computational Physiology	medical research data	https://physionet.org
Harvard EEG Dataset	EEG	https://bdsp.io/content/harvard-eeg-db/
OpenNeuro	human neuroimaging	OpenNeuro.org
Human Connectome Project	neuroimaging	https://www.humanconnectome.org/

Integration of these emerging AI technologies into neuroscience applications presents more challenges and opportunities.

In conclusion, many research areas in neuroscience have greatly benefited from BigData-empowered machine learning. Exploitation of large-scale foundation models, generative AI, and transfer learning tools will enable us to potentially probe neuroscience questions and brain-to-content technology in new dimensions. The landscape of neuroscience research is rapidly changing, and our imagination is only the limit for creativity. We hope this mini-review will inspire more exciting work in the near future.

Funding

The work was partially supported by grants MH118928, DA056394, NS123928, NS121776, MH132642, and NS135170 from the US National Institutes of Health. We thank Katherine Eisert for careful proofreading.

CRediT authorship contribution statement

Zhe Sage Chen: Conceptualization, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Ran Wang:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing interests.

Data Availability

No data was used for the research described in the current article.

Acknowledgments

The authors thank Dr. Ryota Kobayashi and Dr. Ken Nakae for the invitation to participate in a Special Session at the Annual Japanese Neuroscience Meeting held in Sendai, Japan on August 3, 2023, which motivated the writing of this review.

References

Abid, A., Zhang, M.J., Bagaria, V.K., Zou, J., 2018. Exploring patterns enriched in a dataset with contrastive principal component analysis. Nat. Commun. 9, 2134. https://doi.org/10.1038/s41467-018-04608-8.

Acharya, J., Basu, A., Legenstein, R., Limbacher, T., Poirazi, P., Wu, X., 2022. Dendritic computing: branching deeper into machine learning. Neuroscience 489, 275–289. https://doi.org/10.1016/j.neuroscience.2021.10.001.

Aglinskas, A., Hartshorne, J.K., Anzellotti, S., 2022. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 376, 1074–1076. https://doi.org/10.1126/science.abm2461.

Agostinelli, A., Denk, T.I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., et al., 2023.MusicLM: Generating music from text.10.48550/arXiv.2301.11325.

Anumanchipalli, G.K., Chartier, J., Chang, E.F., 2019. Speech synthesis from neural decoding of spoken sentences. Nature 568, 493–498. https://doi.org/10.1038/s41586-019-1119-1.

Ardalan, Z., Subbian, V., 2022. Transfer learning approaches for neuroimaging analysis: a scoping review. Front. Artif. Intell. 5, 780405 https://doi.org/10.3389/frai.2022.780405.

Ashwood, Z.C., Roy, N.A., Stone, I.R., Laboratory, I.B., Urai, A.E., Churchland, A.K., Pouget, A., Pillow, J.W., 2022. Mice alternate between discrete strategies during perceptual decision-making. Nat. Neurosci. 25, 201–212. https://doi.org/10.1038/s41593-021-01007-z.

Azabou, M., Arora, V., Ganesh, V., Mao, X., Nachimuthu, S., Mendelson, M., Richards, B., Perich, M., Lajoie, G., Dyer, E.L., 2023.A Unified, Scalable Framework for Neural Population Decoding, In: Thirty-seventh Conference on Neural Information Processing Systems.

Baevski, A., Babu, A., Hsu, W.N., Auli, M., 2023.Efficient Self-supervised Learning with Contextualized Target Representations for Vision, Speech and Language, arXiv: 2212.07525.

Baevski, A., Hsu, W.N., Xu, Q., Babu, A., Gu, J., Auli, M., 2022.Data2vec: A general framework for self-supervised learning in speech, vision and language.arXiv: 2202.03555.

Baevski, A., Zhou, H., Mohamed, A., Auli, M., 2020.wav2vec 2.0: A Framework for Self-supervised Learning of Speech Representations, arXiv:2006.11477.

Bagchi, S., Bathula, D.R., 2022. EEG-ConvTransformer for single-trial EEG-based visual stimulus classification. Pattern Recognit. 129, 108757 https://doi.org/10.1016/j.patcog.2022.108757.

Bahl, S., Mukadam, M., Gupta, A., Pathak, D., 2020.Neural Dynamic Policies for End-to-end Sensorimotor Learning, In: NeurIPS.

Balestriero, R., Ibrahim, M., Sobal, V., Morcos, A., Shekhar, S., Goldstein, T., Bordes, F., Bardes, A., Mialon, G., Tian, Y., Schwarzschild, A., Wilson, A.G., Geiping, J., Garrido, Q., Fernandez, P., Bar, A., Pirsiavash, H., LeCun, Y., Goldblum, M., 2023.A Cookbook of Self-supervised Learning.10.48550/arXiv.2304.12210.

Banville, H., Chehab, O., Hyvärinen, A., Engemann, D.A., Gramfort, A., 2020.Uncovering the structure of clinical EEG signals with self-supervised learning.10.48550/arXiv.2007.16104.

Bellier, L., Llorens, A., Marciano, D., Gunduz, A., Schalk, G., Brunner, P., Knight, R.T., 2023. Music can be reconstructed from human auditory cortex activity using nonlinear decoding models. PLoS Biol. 21, e3002176 https://doi.org/10.1371/journal.pbio.3002176.

Benchetrit, Y., Banville1, H., King, J.R., 2023.Brain decoding: toward real-time reconstruction of visual perception.⟨https://ai.meta.com/static-resource/image-decoding⟩.

Bengio, Y., Courville, A., Vincent, P., 2014.Representation Learning: A Review and New Perspectives, arXiv:1206.5538.

Beniaguev, D., Segev, I., London, M., 2021. Single cortical neurons as deep artificial neural networks. Neuron 109, 2727–2739. https://doi.org/10.1016/j.neuron.2021.07.002.

Bird, J.J., Pritchard, M., Fratini, A., Ekárt, A., Faria, D.R., 2021. Synthetic biological signals machine-generated by GPT-2 improve the classification of EEG and EMG through data augmentation. IEEE Robot. Autom. Lett. 6, 3498–3504. https://doi.org/10.1109/LRA.2021.3056355.

Bolkan, S.S., Stone, I.R., Pinto, L., Ashwood, Z.C., Garcia, J.M.I., Herman, A.L., Singh, P., Bandi, A., Cox, J., Zimmerman, C.A., Cho, J.R., Engelhard, B., Pillow, J.W., Witten, I. B., 2022. Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. Nat. Neurosci. 25, 345–357. https://doi.org/10.1038/s41593-022-01021-9.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., vonArx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N.S., Chen, A.S., Creel, K.A., Davis, J., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L.E., Goel, K., Goodman, N.D., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D.E., Hong, J., Hsu, K., Huang, J., Icard, T.F., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P.W., Krass, M.S., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X.L., Li, X., Ma, T., Malik, A., Manning, C.D., Mirchandani, S.P., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J.C., Nilforoshan, H., Nyarko, J.F., Ogut, G., Orr, L., Papadimitriou, I., Park, J.S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y.H., Ruiz, C., Ryan, J., R'e, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K.P., Tamkin, A., Taori, R., Thomas, A.W., Tramèr, F., Wang, R.E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S.M., Yasunaga, M., You, J., Zaharia, M.A., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., Liang, P., 2021.On the Opportunities and Risks of Foundation Models, ArXiv ⟨https://crfm.stanford.edu/assets/report.pdf⟩.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language models are few-shot learners. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 1877–1901

Calhoun, A.J., Pillow, J.W., Murthy, M., 2019. Unsupervised identification of the internal states that shape natural behavior. Nat. Neurosci. 22, 2040–2049. https://doi.org/10.1371/journal.pbio.3002176.

Caucheteux, C., Gramfort, A., King, J.R., 2023. Decoding speech from non-invasive brain recordings. Nat. Hum. Behav. 7, 430–441. https://doi.org/10.1038/s41562-022-01516-2.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, Z., 2015. Advanced State Space Methods for Neural and Clinical Data (Ed.). Cambridge University Press.

Chen, Z.S., Kulkarni, P., Galatzer-Levy, I., Bigio, B., Nasca, C., Zhang, Y., 2022. Modern views of machine learning for future precision psychiatry. Patterns 3, 100602. https://doi.org/10.1016/j.patter.2022.100602.

Chen, Z.S., Pesaran, B., 2021. Improving scalability in systems neuroscience. Neuron 109, 1776–1790. https://doi.org/10.1016/j.neuron.2021.03.025.

Chen, Z., Qing, J., Xiang, T., Yue, W.L., Zhou, J.H., 2023.Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding.10.48550/arXiv.2211.06956.

Chung, Y.G., Jeon, Y., Yoo, S., Kim, H., Hwang, H., 2022. Big data analysis and artificial intelligence in epilepsy—common data model analysis and machine learning-based

seizure detection and forecasting. Clin. Exp. Pedia 65, 272–282. https://doi.org/10.3345/cep.2021.00766.
Dado, T., Güçlütürk, Y., Ambrogioni, L., Ras, G., Bosch, S., van Gerven, M., Güçlü, U., 2022. Hyperrealistic neural decoding for reconstructing faces from fMRI activations via the GAN latent space. Sci. Rep. 12, 14. https://doi.org/10.1038/s41598-021-03938-w.
Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., Zhao, L., Xu, S., Liu, W., Liu, N., et al., 2023.AugGPT: Leveraging ChatGPT for text data augmentation.arXiv preprint 10.48550/arXiv.2302.13007.
Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., King, J.R., 2023. Decoding speech from non-invasive brain recordings. Nat. Mach. Intell. 5, 1097–1107. https://doi.org/10.1038/s42256-023-00714-5.
van den Oord, A., Li, Y., Vinyals, O., 2019.Representation Learning with Contrastive Predictive Coding.10.48550/arXiv.1807.03748.
Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagen. Large-scale Hierarchical Image Database. CVPR09.
Denk, T.I., Takagi, Y., Matsuyama, T., Agostinelli, A., Nakai, T., Frank, C., Nishimoto, S., 2023.Brain2music: Reconstructing Music from Human Brain Activity.10.48550/arXiv.2307.11078.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019.BERT: pre-training of deep bidirectional transformers for language understanding, In: Proceedings of NAACL-HLT 2019, Association for Computational Linguistics.4171-4186.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020.An image is worth 16x16 words: Transformers for image recognition at scale.10.48550/arXiv.2010.11929.
Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X., 2023. Self-supervised learning for label- efficient sleep stage classification: a comprehensive evaluation. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 1333–1342. https://doi.org/10.1109/TNSRE.2023.3245285.
Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Lu, H., Song, R., Gao, X., Xiang, T., et al., 2022. Towards artificial general intelligence via a multimodal foundation model. Nat. Commun. 13, 3094. https://doi.org/10.1038/s41467-022-30761-2.
Ferrante, M., Boccato, T., Toschi, N., 2023a.Semantic Brain Decoding: from fMRI to Conceptually Similar Image Reconstruction of Visual Stimuli.10.48550/arXiv.2212.06726.
Ferrante, M., Ozcelik, F., Boccato, T., VanRullen, R., Toschi, N., 2023b.Brain Captioning: Decoding Human Brain Activity Into Images and Text.10.48550/arXiv.2305.11560.
Foerster, J.N., Assael, Y.M., de Freitas, N., Whiteson, S., 2016. Learning to Communicate with Deep Multi-agent Reinforcement Learning. Advances in Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, pp. 2145–2153.
Fu, K., Du, C., Wang, S., He, H., 2022.Multi-View Multi-label Fine-grained Emotion Decoding from Human Brain Activity.IEEE Transactions on Neural Networks and Learning Systems.
Geng, D., Alkhachroum, A., MeloBicchi, M., Cajigas, I., Chen, Z.S., 2021. Deep learning for robust detection of interictal epileptiform discharges. J. Neural Eng. 18, 056015 https://doi.org/10.1088/1741-2552/abf28e.
Geng, D., Chen, Z.S., 2021.Auxiliary classifier generalized adversarial network (AC-GAN) for interictal epileptiform discharge modeling and data augmentation, In: Proc. IEEE EMBS Neural Engineering Conference (NER), 5508-5518.
Gilja, V., Nuyujukian, P., Chestek, C.A., Cunningham, J.P., Yu, B.M., Fan, J.M., Churchland, M.M., Kaufman, M.T., Kao, J.C., Ryu, S.I., Shenoy, K.V., 2012. A high-performance neural prosthesis enabled by control algorithm design. Nat. Neurosci. 15, 1752–1757. https://doi.org/10.1038/nn.3265.
Goetschalckx, L., Andonian, A., Wagemans, J., 2021. Generative adversarial networks unlock new methods for cognitive science. Trends Neurosci. 25, 788–801. https://doi.org/10.1016/j.tics.2021.06.006.
Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S.A., Feder, A., Emanuel, D., Cohen, A., et al., 2022. Shared computational principles for language processing in humans and deep language models. Nat. Neurosci. 25, 369–380. https://doi.org/10.1038/s41593-022-01026-4.
Gong, C., Jing, C., Chen, X., Pun, C.M., Huang, G., Saha, A., Nieuwoudt, M., Li, H., Hu, Y., Wang, S., 2023. Generative AI for brain image computing and brain network computing: a review. Front. Neurosci. 17, 1203104 https://doi.org/10.3389/fnins.2023.1203104.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MA.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.
Groussard, M., Viader, F., Landeau, B., Desgranges, B., Eustache, F., Platel, H., 2009. Neural correlates underlying musical semantic memory. Ann. NY Acad. Sci. 1169, 278–281. https://doi.org/10.1111/j.1749-6632.2009.04784.x.
Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J., 2020.A review on generative adversarial networks: Algorithms, theory, and applications.10.48550/arXiv.2001.06937.
Habashi, A.G., Azab, A.M., Eldawlatly, S., Aly, G.M., 2023. Generative adversarial networks in EEG analysis: an overview. J. Neuroeng. Rehabil. 20, 40. https://doi.org/10.1186/s12984-023-01169-w.
Hassabis, D., Kumaran, D., Summerfield, C., Botvinick, M., 2017. Neuroscience-inspired artificial intelligence. Neuron 95, 245–258.
Hawkins, J., Ahmad, S., 2016. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. Front. Neural Circuits 10, 23. https://doi.org/10.3389/fncir.2016.00023.
He, H., Garcia, E., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284.
Ho, J., Jain, A., Abbeel, P., 2020.Denoising diffusion probabilistic models.arXiv preprint abs/2006.11239. 10.48550/arXiv.2006.11239.
Hodassman, S., Vardi, R., Tugendhaft, Y., Goldental, A., Kanter, I., 2022. Efficient dendritic learning as an alternative to synaptic plasticity hypothesis. Sci. Rep. 12, 6571. https://doi.org/10.1038/s41598-022-10466-8.
Hollenstein, N., Renggli, C., Glaus, B., Barrett, M., Troendle, M., Langer, N., Zhang, C., 2021. Decoding EEG brain activity for multi-modal natural language processing. Front. Hum. Neurosci. 15, 659410 https://doi.org/10.3389/fnhum.2021.659410.
Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y., 2013. Neural decoding of visual imagery during sleep. Science 340, 639–642. https://doi.org/10.1126/science.1234330.
Hsu, W.N., Bolte, B., Tsai, Y.H.H., Lakhotia, K., Salakhutdinov, R., Mohamed, A., 2021. HuBERT: Self-supervised Speech Represent. Learn. Masked Predict. Hidden Units arXiv:2106.07447.
Huang, W., Yan, H., Wang, C., Yang, X., Li, J., Zuo, Z., Zhang, J., Chen, H., 2021. Deep natural image reconstruction from human brain activity based on conditional progressively growing generative adversarial networks. Neurosci. Bull. 37, 369–379. https://doi.org/10.1007/s12264-020-00613-4.
Huth, A.G., De Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016a. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. https://doi.org/10.1038/nature17637.
Huth, A.G., Lee, T., Nishimoto1, S., Bilenko, N.Y., Vu, A.T., Gallant, J.L., 2016b. Decoding the semantic content of natural movies from human brain activity. Front. Syst. Neurosci. 10, 81. https://doi.org/10.3389/fnsys.2016.00081.
Huth, A.G., Nishimoto1, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014.
Jaegle, A., Borgeaud, S., Alayrac, J.B., Doersch, C., Ionescu, C., Ding, D., Koppula, S., Zoran, D., Brock, A., Shelhamer, E., Hénaff, O., Botvinick, M.M., Zisserman, A., Vinyals, O., Carreira, J., 2022.Perceiver io: A General Architecture for Structured Inputs & Outputs.10.48550/arXiv.2107.14795.
Joshi, M., Chen, D., Liu, Y., Weld, D.S., Zettlemoyer, L., Levy, O., 2020. SpanBERT: Improving pre-training by representing and predicting spans. Transactions of the Association for. Comput. Linguist. 8, 64–77. https://doi.org/10.1162/tacl_a_00300.
Kingma, D.P., Welling, M., 2013.Auto-Encoding Variational Bayes.arXiv:1312.6114.
Kostas, D., Aroca-Ouellette, S., Rudzicz, F., 2021. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. Front Hum. Neurosci. 15, 653659 https://doi.org/10.3389/fnhum.2021.653659.
Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., Poeppel, D., 2017. Neuroscience needs behavior: correcting a reductionist bias. Neuron 93, 480–490. https://doi.org/10.1016/j.neuron.2016.12.041.
Lakshminarasimhan, K.J., Avila, E., Pitkow, X., Angelaki, D.E., 2023. Dynamical latent state computation in the male macaque posterior parietal cortex. Nat. Commun. 14, 1832. https://doi.org/10.1038/s41467-023-37400-4.
Latimer, K.W., Yates, J.L., Meister, M.L.R., Huk, A.C., Pillow, J.W., 2015. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science 349, 184–187. https://doi.org/10.1126/science.aaa4056.
Lebedev, M., Nicolelis, M., 2017. Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation. Physiol. Rev. 97, 767–837. https://doi.org/10.1152/physrev.00027.2016.
Lee, W., Lee, J., Kim, Y., 2021. Contextual imputation with missing sequence of EEG signals using generative adversarial networks. IEEE Access 9, 151753.
Li, Y., Anumanchipalli, G.K., Mohamed, A., Lu, J., Wu, J., Chang, E.F., 2023b. Dissecting neural computations of the human auditory pathway using deep neural networks for speech. Nat. Neurosci. https://doi.org/10.1038/s41593-023-01468-4.
Li, W., Luo, H., Lin, Z., Zhang, C., Lu, Z., Ye, D., 2023a.A Survey on Transformers in Reinforcement Learning.10.48550/arXiv.2301.03044.
Liu, Y., Zhao, Z., Xu, M., Yu, H., Zhu, Y., Zhang, J., Bu, L., Zhang, X., Lu, J., Li, Y., Ming, D., Wu, J., 2023. Decoding and synthesizing tonal language speech from brain activity. Sci. Adv. 9, eadh0478 https://doi.org/10.1126/sciadv.adh0478.
Lu, H., Zhou, Q., Fei, N., Lu, Z., Ding, M., Wen, J., Du, C., Zhao, X., Sun, H., He, H., et al., 2022.Multimodal foundation models are better simulators of the human brain.arXiv preprint 10.48550/arXiv.2208.08263.
Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H., Hikida, T., 2021. Natural and artificial intelligence: a brief introduction to the interplay between ai and neuroscience research. Neural Netw. 144, 603–613. https://doi.org/10.1016/j.neunet.2021.09.018.
Malik, N., Bzdok, D., 2022. From YouTube to the brain: transfer learning can improve brain-imaging predictions with deep learning. Neural Netw. 153, 325–338. https://doi.org/10.1016/j.neunet.2022.06.014.
Metzger, S.L., Littlejohn, K.T., Silva, A.B., Moses, D.A., Seaton, M.P., Wang, R., Dougherty, M.E., Liu, J.R., Wu, P., Berger, M.A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G.K., Chang, E.F., 2023. A high-performance neuroprosthesis for speech decoding and avatar control. Nature 620, 1037–1046. https://doi.org/10.1038/s41586-023-06443-4.
Metzger, S.L., Liu, J.R., Moses, D.A., Dougherty, M.E., Seaton, M.P., Littlejohn, K.T., Chartier, J., Anumanchipalli, G.K., Tu-Chan, A., Ganguly, K., Chang, E.F., 2022. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. Nat. Commun. 13, 6510. https://doi.org/10.1038/s41467-022-33611-3.
Millet, J., Caucheteux, C., Orhan, P., Boubenec, Y., Gramfort, A., Dunbar, E., Pallier, C., King, J.R., 2023.toward A Realistic Model of Speech Processing in the Brain with Self-supervised Learning.10.48550/arXiv.2206.01685.

Moor, M., Banerjee, O., Abad, Z., Krumholz, H., Leskovec, J., Topol, E., Rajpurkar, P., 2023. Foundation models for generalist medical artificial intelligence. Nature 616, 259–265. https://doi.org/10.1038/s41586-023-05881-4.

Moses, D.A., Metzger, S.L., Liu, J.R., Anumanchipalli, G.K., Makin, J.G., Sun, P.F., Chartier, J., Dougherty, M.E., Liu, P.M., Abrams, G.M., Tu-Chan, A., Ganguly, K., Chang, E.F., 2021. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. N. Engl. J. Med. 385, 217–227. https://doi.org/10.1056/NEJMoa2027540.

Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., Mian, A., 2023.A Comprehensive Overview of Large Language Models.10.48550/arXiv.2307.06435.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031.

Ozcelik, F., VanRullen, R., 2023.Natural Scene Reconstruction from Fmri Signals Using Generative Latent Diffusion.10.48550/arXiv.2303.05334.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Pellicer, L.F.A.O., Ferreira, T.M., Costa, A.H.R., 2023. Data augmentation techniques in natural language processing. Appl. Soft Comput. 132, 109803 https://doi.org/10.1016/j.asoc.2022.109803.

Pereira, T., Shaevitz, J., Murthy, M., 2020. Quantifying behavior to understand the brain. Nat. Neurosci. 23, 1537–1549. https://doi.org/10.1038/s41593-020-00734-z.

Peterson, S.M., Steine-Hanson, Z., Davis, N., Rao, R.P.N., Brunton, B.W., 2021. Generalized neural decoders for transfer learning across participants and recording modalities. J. Neural Eng. 18, 026014 https://doi.org/10.1088/1741-2552/abda0b.

Poldrack, R., Mumford, J., Schonberg, T., Kalar, D., Barman, B., T, Y., 2012. Discovering relations between mind, brain, and mental disorders using topic mapping. PLoS Comput. Biol. 8, e1002707 https://doi.org/10.1371/journal.pcbi.1002707.

Qu, Y., Jian, X., Che, W., Du, P., Fu, K., Liu, Q., 2022.Transfer Learning to Decode Brain States Reflecting the Relationship between Cognitive Tasks.10.48550/arXiv.2206.03950.

Rafiei, M.H., Gauthier, L.V., Adeli, H., Takabi, D., 2022. Self-supervised learning for electroencephalography. IEEE Trans. Neural Netw. Learn. Syst. https://doi.org/10.1109/TNNLS.2022.3190448.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S.G., Novikov, A., Barth-maron, G., Giménez, M., Sulsky, Y., Kay, J., Springenberg, J.T., Eccles, T., Bruce, J., Razavi, A., Edwards, A., Heess, N., Chen, Y., Hadsell, R., Vinyals, O., Bordbar, M., de Freitas, N., 2022.A generalist agent.Transactions on Machine Learning Research 〈https://openreview.net/forum?id=1ikK0kHjvj〉.

Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., et al., 2019. A deep learning framework for neuroscience. Nat. Neurosci. 22, 1761–1770. https://doi.org/10.1038/s41593-019-0520-2.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022.High-Resolution Image Synthesis with Latent Diffusion Models, In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.10.1109/CVPR52688.2022.01042.

Rubin, T., Koyejo, O., Gorgolewski, K., Jones, M., Poldrack, R., T, Y., 2017. Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLoS Comput. Biol. 13, e1005649 https://doi.org/10.1371/journal.pcbi.1005649.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., Batista, A.P., 2014. Neural constraints on learning. Nature 512, 423–426. https://doi.org/10.1038/nature13665.

Saxe, A., Nelli, S., Summerfield, C., 2021. If deep learning is the answer, what is the question? Nat. Rev. Neurosci. 22, 55–67. https://doi.org/10.1038/s41583-020-00395-8.

Schneider, S., Baevski, A., Collobert, R., Auli, M., 2019.wav2vec: Unsupervised pre-training for speech recognition.arXiv:1904.05862.

Schneider, S., Lee, J., Mathis, M., 2023. Learnable latent embeddings for joint behavioural and neural analysis. Nature 617, 360–368. https://doi.org/10.1038/s41586-023-06031-6.

Schubert, P.J., Dorkenwald, S., Januszewski, M., Jain, V., Kornfeld, J., 2019. Learning cellular morphology with neural networks. Nat. Commun. 10, 2736. https://doi.org/10.1038/s41467-019-10836-3.

Seeliger, K., Güçlü, U., Ambrogioni, L., Güçlütürk, Y., van Gerven, M., 2018. Generative adversarial networks for reconstructing natural images from brain activity, 775–785–1790 Neuroimage 181. https://doi.org/10.1016/j.neuroimage.2018.07.043.

Shanechi, M., 2019. Brain-machine interfaces from motor to mood. Nat. Neurosci. 22, 1554–1564. https://doi.org/10.1038/s41593-019-0488-y.

Shen, G., Horikawa, T., Majima, K., Kamitani, Y., 2019. Deep image reconstruction from human brain activity. PLoS Comput. Biol. 15, e1006633 https://doi.org/10.1371/journal.pcbi.1006633.

Shi, J., Tripp, B., Shea-Brown, E., Mihalas, S., Buice, M.A., 2022. MouseNet: a biologically constrained convolutional neural network model for the mouse visual cortex. PLoS Comput. Biol. 18, e1010427 https://doi.org/10.1371/journal.pcbi.1010427.

Singhal, K., Azizi, S., Tu, T., et al., 2023. Large language models encode clinical knowledge. Nature 620, 172–180. https://doi.org/10.1038/s41586-023-06291-2.

Soingern, N., Sinsamersuk, A., Chatnuntawech, I., Silpasuwanchai, C., 2023. Data augmentation forEEG motor imagery classification using diffusion model. First International Conference on Data Science and Artificial Intelligence (DSAI) Proceeedings). Springer, Berlin, pp. 111–126.

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. Adv. Neural Inf. Process. Syst. 32.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2021.Score-based generative modeling through stochastic differential equations, In: International Conference on Learning Representations.〈https://openreview.net/forum?id=PxTIG12RRHS〉.

Speier, W., Arnold, C., Pouratian, N., 2016. Integrating language models into classifiers for BCI communication: a review. J. Neural Eng. 13, 031002 https://doi.org/10.1088/1741-2560/13/3/031002.

Sun, G., Zeng, F., McCartin, M., Zhang, Q., Xu, H., Liu, Y., Chen, Z.S., Wang, J., 2022. Closed-loop stimulation using a multi-region brain-machine interface has analgesic effects in rodents. Sci. Transl. Med. 14, eabm5868 https://doi.org/10.1126/scitranslmed.abm5868.

Takagi, Y., Nishimoto, S., 2023.High-Resolution Image Reconstruction with Latent Diffusion Models from Human Brain Activity, In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14453-14463.

Tang, S., Dunnmon, J., Saab, K.K., Zhang, X., Huang, Q., Dubost, F., Rubin, D., Lee-Messer, C., 2022.Self-supervised graph neural networks for improved electroencephalographic seizure analysis, In: International Conference on Learning Representations.〈https://openreview.net/forum?id=k9bx1EfHI_-〉.

Tang, J., LeBel, A., Jain, S., Huth, A.G., 2023. Semantic reconstruction of continuous language from non-invasive brain recordings. Nat. Neurosci. 26, 858–866. https://doi.org/10.1038/s41593-023-01304-9.

Thomas, A.W., Ré, C., Poldrack, R.A., 2022. Self-supervised learning of brain dynamics from broad neuroimaging data. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (Eds.), Advances in Neural Information Processing Systems. 〈https://openreview.net/forum?id=lMrpZ-ycIaT〉.

Tong, X., Xie, H., Fonzo, G., Zhao, K., Theodore, D., Satterthwaite, T.D., Carlisle, N., Zhang, Y., 2021.Dissecting symptom-linked dimensions of resting-state electroencephalographic functional connectivity in austim with contrastive learning. bioRxiv preprint 10.1101/2023.05.22.541841.

Vahdat, A., Kreis, K., Kautz, J., 2021. Score-based generative modeling in latent space. Adv. Neural Inf. Process. Syst. 34.

VanRullen, R., Reddy, L., 2019. Reconstructing faces from fMRI patterns using deep generative neural networks. Commun. Biol. 2, 193. https://doi.org/10.1038/s42003-019-0438-y.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., Attention is all you need, 2017.In: Advances in Neural Information Processing Systems.

Wagh, N., Wei, J., Rawal, S., Berry, B., Barnard, L., Brinkmann, B., Worrell, G., Jones, D., Varatharajah, Y., 2021. Domain-guided self-supervision of eeg data improves downstream classification performance and generalizability. In: Roy, S., Pfohl, S., Rocheteau, E., Tadesse, G.A., Oala, L., Falck, F., Zhou, Y., Shen, L., Zamzmi, G., Mugambi, P., Zirikly, A., McDermott, M.B.A., Alsentzer, E. (Eds.), Proceedings of Machine Learning for Health, pp. 130–142. In: 〈https://proceedings.mlr.press/v158/wagh21a.html〉 (PMLR).

Wang, R., Chen, X., Khalilian-Gourtani, A., Chen, Z., Yu, L., Flinker, A., Wang, Y., 2020. Stimulus speech decoding from human cortex with generative adversarial network transfer learning. Proc. IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 390–394, 10.1109/ISBI45749.2020.9098589.

Wang, R., Chen, X., Khalilian-Gourtani, A., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., Flinker, A., 2023. Distributed feedforward and feedback processing across perisylvian cortex supports human speech. Proc. Natl. Acad. Sci. USA 120, e2300255120. https://doi.org/10.1073/pnas.2300255120.

Wang, Z., Guet, C., 2022. Self-consistent learning of neural dynamical systems from noisy time series. IEEE Trans. Emerg. Top. Comput. Intell. 6, 1103–1112. https://doi.org/10.1109/TETCI.2022.3146332.

Willett, F.R., Avansino, D.T., Hochberg, L.R., Henderson, J.M., Shenoy, K.V., 2021. High-performance brain-to-text communication via handwriting. Nature 593, 249–254. https://doi.org/10.1038/s41586-021-03506-2.

Willett, F.R., Kunz, E.M., Fan, C., Avansino, D.T., Wilson, G.H., Choi, E.Y., Kamdar, F., Glasser, M.F., Hochberg, L.R., Druckmann, S., et al., 2023. A high-performance speech neuroprosthesis. Nature 1–6. https://doi.org/10.1038/s41586-023-06377-x.

Wolfram, S., 2023. What is ChatGPT doing… and Why Does it Work? Wolfram Media Inc.

Yamins, D., DiCarlo, J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356–465. https://doi.org/10.1038/nn.4244.

Yamins, D., Hong, H., F, C.C., Solomon, E.A., Seibert, D., DiCarlo, J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA 111, 8619–8624. https://doi.org/10.1073/pnas.1403112111.

Yang, C., Xiao, C., Westover, M.B., Sun, J., 2023. Self-supervised electroencephalogram representation learning for automatic sleep staging: model development and evaluation study. JMIR AI 2, e46769. https://doi.org/10.2196/46769.

Ye, J., Pandarinath, C., 2021.Representation Learning for Neural Population Activity with Neural Data Transformers.10.48550/arXiv.2108.01210.

Yoon, J., Jarrett, D., van der Schaar, M., 2019. Time-series generative adversarial networks. Adv. Neural Inf. Process. Syst. 5508–5518.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. 3320–3328.

Zhang, Q., Hu, S., Talay, R., Xiao, Z., Rosenberg, D., Liu, Y., Sun, G., Li, A., Caravan, B., Singh, A., Gould, J.D., Chen, Z.S., Wang, J., 2023. A prototype closed-loop brain-machine interface for the study and treatment of pain. Nat. Biomed. Eng. 7, 533–545. https://doi.org/10.1038/s41551-021-00736-7.

Zhang, K., Robinson, N., Lee, S.W., Guan, C., 2021. Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network. Neural Netw. 136, 1–10. https://doi.org/10.1016/j.neunet.2020.12.013.

Zhang, A., Xing, L., Zou, J., Wu, J.C., 2022. Shifting machine learning for healthcare from development to deployment and from models to data. Nat. Biomed. Eng. 6, 1330–1345. https://doi.org/10.1038/s41551-022-00898-y.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.Y., Wen, J.R., 2023.A survey of large language models.10.48550/arXiv.2303.18223.

Zhu, Q., Zhao, X., Zhang, J., Gu, Y., Weng, C., Hu, Y., 2023.EEG2vec: Self-Supervised Electroencephalographic Representation Learning.10.48550/arXiv.2305.13957.

Zhuang, P., Schwing, A.G., Koyejo, S., 2019.fMRI Data Augmentation Via Synthesis.arXiv preprint 10.48550/arXiv.1907.06134.